EUROPEAN COMMISSION
DIRECTORATE-GENERAL RESEARCH & INNOVATION

Open Innovation and Open Science
Research infrastructure

GRANT AGREEMENT

NUMBER 824091 — RISIS 2

This **Agreement** ('the Agreement') is **between** the following parties:

on the one part,

the **European Union** ('the EU'), represented by the European Commission ('the Commission'),

represented for the purposes of signature of this Agreement by Head of Unit, DIRECTORATE-GENERAL RESEARCH & INNOVATION, Open Innovation and Open Science, Administration and finance, Pascale CID,

and

on the other part,

1. 'the coordinator':

UNIVERSITE DE MARNE LA VALLEE (UPEM), established in BOULEVARD DESCARTES 5 CHAMPS SUR MARNE, MARNE LA VALLEE CEDEX 2 77454, France, VAT number: FR66199320565, represented for the purposes of signing the Agreement by President, Gilles ROUSSEL

and the following other beneficiaries, if they sign their 'Accession Form' (see Annex 3 and Article 56):

2. **AIT AUSTRIAN INSTITUTE OF TECHNOLOGY GMBH (AIT)**, established in GIEFINGGASSE 4, WIEN 1210, Austria, VAT number: ATU14703506,

3. **POLITECNICO DI MILANO (POLIMI)**, established in PIAZZA LEONARDO DA VINCI 32, MILANO 20133, Italy, VAT number: IT04376620151,

4. **CONSIGLIO NAZIONALE DELLE RICERCHE (CNR)**, established in PIAZZALE ALDO MORO 7, ROMA 00185, Italy, VAT number: IT02118311006,

5. **UNIVERSITEIT LEIDEN (LEIDEN)**, established in RAPENBURG 70, LEIDEN 2311 EZ, Netherlands, VAT number: NL809778282B01,

6. **THE UNIVERSITY OF SHEFFIELD (USFD)**, established in FIRTH COURT WESTERN BANK, SHEFFIELD S10 2TN, United Kingdom, VAT number: GB648238808,

7. **FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V. (FRAUNHOFER)**, established in HANSASTRASSE 27C, MUNCHEN 80686, Germany, VAT number: DE129515865,

8. **UNIVERSITA DELLA SVIZZERA ITALIANA (USI)**, established in VIA LAMBERTENGHI 10 A, LUGANO 6904, Switzerland, VAT number: CH507591,

9. **JOANNEUM RESEARCH FORSCHUNGSGESELLSCHAFT MBH (JOANNEUM)**, established in LEONHARDSTRASSE 59, GRAZ 8010, Austria, VAT number: ATU28781306,

10. **THE UNIVERSITY OF SUSSEX (UOS)**, established in SUSSEX HOUSE FALMER, BRIGHTON BN1 9RH, United Kingdom, VAT number: GB692712320,

11. **DEUTSCHES ZENTRUM FUR HOCHSCHUL- UND WISSENSCHAFTSFORSCHUNG GMBH (DZHW)**, established in LANGE LAUBE 12, HANNOVER 30159, Germany, VAT number: DE291239300,

12. **UNIVERSITY OF STRATHCLYDE (STRATHCLYDE)**, established in Richmond Street 16, GLASGOW G1 1XQ, United Kingdom, VAT number: GB261339762,

13. **UNIVERSITA DEGLI STUDI DI ROMA LA SAPIENZA (SAPIENZA)**, established in Piazzale Aldo Moro 5, ROMA 00185, Italy, VAT number: IT02133771002,

14. **NORDISK INSTITUTT FOR STUDIER AV INNOVASJON, FORSKNING OG UTDANNING (NIFU)**, established in ØKERNVEIEN 9, OSLO 0608, Norway, VAT number: NO976073169MVA,

15. **AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS (CSIC)**, established in CALLE SERRANO 117, MADRID 28006, Spain, VAT number: ESQ2818002D,

16. **THE SAMUEL NEAMAN INSTITUTE FOR ADVANCED STUDIES IN SCIENCE AND TECHNOLOGY OF THE TECHNION LTD (PSC) (SNI)**, established in TECHNION CITY, HAIFA 32000, Israel,

17. **TECHNOLOGICKE CENTRUM AKADEMIE VED CESKE REPUBLIKY (TC CAS)**, established in Ve Struhach 1076/27, PRAHA 160 00, Czech Republic,

18. **ATHINA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS (ATHENA RC)**, established in ARTEMIDOS 6 KAI EPIDAVROU, MAROUSSI 151 25, Greece, VAT number: EL999723442,

Unless otherwise specified, references to ‘beneficiary’ or ‘beneficiaries’ include the coordinator.

The parties referred to above have agreed to enter into the Agreement under the terms and conditions below.

By signing the Agreement or the Accession Form, the beneficiaries accept the grant and agree to implement it under their own responsibility and in accordance with the Agreement, with all the obligations and conditions it sets out.

EUROPEAN COMMISSION
DIRECTORATE-GENERAL RESEARCH & INNOVATION
Research infrastructure

ANNEX 1 (part A)

Research and Innovation action

NUMBER — 824091 — RISIS 2

Table of Contents

1.1. The project summary

Project Number [1]	824091	Project Acronym [2]	RISIS 2

One form per project

General information	
Project title [3]	European Research Infrastructure for Science, technology and Innovation policy Studies 2
Starting date [4]	01/01/2019
Duration in months [5]	48
Call (part) identifier [6]	H2020-INFRAIA-2018-1
Topic	INFRAIA-01-2018-2019 Integrating Activities for Advanced Communities
Fixed EC Keywords	
Free keywords	Science and technology, Innovation, STI indicators, Research and Innovation Policies, Knowledge Dynamics, Public Research, Firm Innovation Capabilities

Abstract [7]

The European Research infrastructure for science, technology and innovation policy studies (RISIS2) aims at building a data and services infrastructure supporting the development of a new generation of analyses and indicators. To develop a deeper understanding of knowledge dynamics and policy relevant evidence, the project goes beyond established quantitative indicators, developing positioning indicators, which take into account critical features of knowledge dynamics i.e. the importance of asymmetries in producers, in places and in themes. RISIS datasets are built keeping information on these three dimensions. To exploit them, new services dealing with actor identification, geographical information and thematic foci are developed, as well as semantic analytical capabilities. This project builds on RISIS1 (2014-18), which has demonstrated the relevance of such an approach and opened access to a first set of databases and services. RISIS2 gathers 19 partners aiming to transform the field of STI studies into an advanced research community. This step change is achieved by: (i) developing an e-infrastructure that supports full virtual transnational access by researchers, (ii) providing a vastly enlarged set of services tailored to field-specific needs (for problem-based integration of datasets, for exploring open data, and for supporting analytical capabilities of researchers), (iii) maintaining datasets dealing with firm innovation capacities, public research developments, R&I outputs and projects, and policy learning, (iv) developing new datasets on 4 key issues for research and policy (social innovation, non technological innovation, the role of PhDs in society, portfolios of public funding instruments). As reflected in the strong role of OpenAire in RISIS2, the infrastructure is fully inscribed into the open science movement. It is accompanied by a strong training, dissemination and communication effort to support the important widening of the community we aim at.

1.2. List of Beneficiaries

Project Number [1]	824091	Project Acronym [2]	RISIS 2

List of Beneficiaries

No	Name	Short name	Country	Project entry month[8]	Project exit month
1	UNIVERSITE DE MARNE LA VALLEE	UPEM	France	1	48
2	AIT AUSTRIAN INSTITUTE OF TECHNOLOGY GMBH	AIT	Austria	1	48
3	POLITECNICO DI MILANO	POLIMI	Italy	1	48
4	CONSIGLIO NAZIONALE DELLE RICERCHE	CNR	Italy	1	48
5	UNIVERSITEIT LEIDEN	LEIDEN	Netherlands	1	48
6	THE UNIVERSITY OF SHEFFIELD	USFD	United Kingdom	1	48
7	FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V.	FRAUNHOFER	Germany	1	48
8	UNIVERSITA DELLA SVIZZERA ITALIANA	USI	Switzerland	1	48
9	JOANNEUM RESEARCH FORSCHUNGSGESELLSCHAFT MBH	JOANNEUM	Austria	1	48
10	THE UNIVERSITY OF SUSSEX	UOS	United Kingdom	1	48
11	DEUTSCHES ZENTRUM FUR HOCHSCHUL- UND WISSENSCHAFTSFORSCHUNG GMBH	DZHW	Germany	1	48
12	UNIVERSITY OF STRATHCLYDE	STRATHCLYDE	United Kingdom	1	48
13	UNIVERSITA DEGLI STUDI DI ROMA LA SAPIENZA	SAPIENZA	Italy	1	48
14	NORDISK INSTITUTT FOR STUDIER AV INNOVASJON, FORSKNING OG UTDANNING	NIFU	Norway	1	48
15	AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS	CSIC	Spain	1	48
16	THE SAMUEL NEAMAN INSTITUTE FOR ADVANCED STUDIES IN SCIENCE AND TECHNOLOGY OF THE TECHNION LTD (PSC)	SNI	Israel	1	48
17	TECHNOLOGICKE CENTRUM AKADEMIE VED CESKE REPUBLIKY	TC CAS	Czech Republic	1	48
18	ATHINA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS	ATHENA RC	Greece	1	48

1.3. Workplan Tables - Detailed Implementation

1.3.1. WT1 List of work packages

WP Number[9]	WP Title	Lead beneficiary[10]	Person-months[11]	Start month[12]	End month[13]
WP1	Project management (NA)	1 - UPEM	42.00	1	48
WP2	Interaction with the Research community, Communication and Dissemination (NA)	4 - CNR	66.00	1	48
WP3	Preparation of long term sustainability of RISIS (NA)	1 - UPEM	13.00	1	48
WP4	RISIS Core Facility (JRA)	1 - UPEM	150.00	1	48
WP5	Core Datasets Maintenance (NA)	2 - AIT	92.00	1	48
WP6	Data Integration and Analysis tools (JRA)	6 - USFD	170.00	1	48
WP7	Advanced user communities and advanced analytical methods (NA)	8 - USI	56.00	1	48
WP8	Transnational access	2 - AIT	97.40	1	48
WP9	Deepening of Core datasets (JRA)	3 - POLIMI	140.00	1	48
WP10	RISIS Core datasets developments (JRA)	5 - LEIDEN	129.00	1	48
WP11	Ethics requirements	1 - UPEM	N/A	1	48
		Total	955.40		

1.3.2. WT2 list of deliverables

Deliverable Number[14]	Deliverable Title	WP number[9]	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D1.1	Preliminary Data Management Plan	WP1	1 - UPEM	Report	Public	6
D1.2	Data Management Plan 1st full version	WP1	1 - UPEM	ORDP: Open Research Data Pilot	Public	24
D1.3	Data Management Plan 2nd full version	WP1	1 - UPEM	ORDP: Open Research Data Pilot	Public	42
D2.1	Website and materials	WP2	4 - CNR	Websites, patents filling, etc.	Public	12
D2.2	First Report on activities developed for training, communication and interactions with users & stakeholders	WP2	4 - CNR	Report	Public	18
D2.3	Analysis of users' needs	WP2	4 - CNR	Report	Public	24
D2.4	Report on Press and media actions	WP2	4 - CNR	Report	Public	30
D2.5	Second Report on activities developed for training, communication and interactions with users & stakeholders	WP2	4 - CNR	Report	Public	36
D2.6	Policy Brief on achievements and lessons learned	WP2	4 - CNR	Report	Public	46
D3.1	Report on options for the future by the SWG	WP3	1 - UPEM	Report	Confidential, only for members of the consortium (including the Commission Services)	30
D3.2	Plan for the future	WP3	1 - UPEM	Report	Public	42
D3.3	Report on indirect contract approach	WP3	1 - UPEM	Report	Confidential, only for members of the consortium (including the Commission Services)	18

Deliverable Number[14]	Deliverable Title	WP number[9]	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D4.1	Joint-Opening of RCF beta version	WP4	1 - UPEM	Websites, patents filling, etc.	Public	18
D4.2	Opening of Open Data VRE	WP4	4 - CNR	Websites, patents filling, etc.	Public	18
D4.3	Frontend to access and navigate datasets	WP4	1 - UPEM	Websites, patents filling, etc.	Public	18
D4.4	Access to core Risis services	WP4	1 - UPEM	Websites, patents filling, etc.	Public	18
D4.5	Opening RCF data and services including external services linkage and VRE	WP4	1 - UPEM	Websites, patents filling, etc.	Public	36
D4.6	Usage Monitoring and infrastructure documentation	WP4	1 - UPEM	Report	Public	46
D5.1	Consolidated work plan on Maintenance	WP5	2 - AIT	Report	Public	6
D5.2	First interim report on Maintenance of RISIS core datasets	WP5	2 - AIT	Report	Public	18
D5.3	Second interim report on Maintenance of core RISIS datasets	WP5	2 - AIT	Report	Public	36
D5.4	Final report on Maintenance of core RISIS datasets	WP5	2 - AIT	Report	Public	46
D6.1	Consolidated report on the new services opened (based upon corresponding milestones)	WP6	6 - USFD	Other	Public	18
D6.2	First report on measuring RISIS impact through openscience tools	WP6	18 - ATHENA RC	Report	Public	24
D6.3	RCF methods & service integration	WP6	6 - USFD	Other	Public	36
D6.4	Second report on measuring RISIS impact through openscience tools	WP6	18 - ATHENA RC	Report	Public	46

Deliverable Number[14]	Deliverable Title	WP number[9]	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D6.5	RCF methods & service integration – final version	WP6	6 - USFD	Other	Public	46
D7.1	Data quality package report	WP7	13 - SAPIENZA	Report	Public	24
D7.2	Summary report on the first batch of usage cases	WP7	8 - USI	Report	Public	24
D7.3	Policy brief on advanced usage cases and their policy relevance	WP7	8 - USI	Report	Public	36
D7.4	Final report on usage cases and advanced communities	WP7	8 - USI	Report	Public	42
D8.1	Harmonized documentation and metadata for access	WP8	2 - AIT	Report	Public	6
D8.2	First interim report on Access	WP8	2 - AIT	Report	Public	18
D8.3	Updated documentation and metadata for access	WP8	2 - AIT	Report	Public	36
D8.4	Final report on Access	WP8	2 - AIT	Report	Public	48
D9.1	Consolidated work plan on datasets development	WP9	3 - POLIMI	Report	Public	12
D9.2	Interim documentation on new datasets development	WP9	3 - POLIMI	Report	Public	30
D9.3	Full documentation on developments made on RISIS core datasets	WP9	3 - POLIMI	Report	Public	42
D9.4	Policy briefs on new developments and their policy relevance	WP9	3 - POLIMI	Report	Public	42
D10.1	Opening of HR wiki	WP10	11 - DZHW	Websites, patents filling, etc.	Public	6
D10.2	Draft technical- method report of new datasets	WP10	5 - LEIDEN	Report	Confidential, only for members of the consortium (including the Commission Services)	12

Deliverable Number[14]	Deliverable Title	WP number[9]	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D10.3	Manual & Documentation of new datasets – first version	WP10	5 - LEIDEN	Report	Public	24
D10.4	Report of the handling of privacy aspects for the HR dataset	WP10	11 - DZHW	Report	Public	30
D10.5	Manual & Documentation of new datasets final version	WP10	5 - LEIDEN	Report	Public	36
D10.6	Policy brief on new datasets and their policy relevance	WP10	5 - LEIDEN	Report	Public	42
D11.1	POPD - Requirement No. 1	WP11	1 - UPEM	Ethics	Confidential, only for members of the consortium (including the Commission Services)	1
D11.2	GEN - Requirement No. 2	WP11	1 - UPEM	Ethics	Confidential, only for members of the consortium (including the Commission Services)	48

1.3.3. WT3 Work package descriptions

Work package number [9]	WP1	**Lead beneficiary** [10]	1 - UPEM
Work package title	Project management (NA)		
Start month	1	**End month**	48

Objectives

The overall objective is to insure a proper operational management of the project, both scientific and technical

Description of work and role of partners

WP1 - Project management (NA) [Months: 1-48]
UPEM, AIT, POLIMI, CNR, LEIDEN, USFD, FRAUNHOFER, USI, JOANNEUM, UOS, DZHW, STRATHCLYDE, SAPIENZA, NIFU, CSIC, SNI, TC CAS, ATHENA RC
Task 1 – Strategic governance of the project
This task is dedicated to the adequate functioning of the Governing Board. This includes the preparation of the annual review, the organisation of meetings of the Board, the organisation of the RISIS week that enables the Board to follow all activities before making strategic decisions for the year to come. This also includes the support and organisation of meetings of the Strategic Advisory Board (also once a year, during the RISIS week)

Task 2 – Operational management – Scientific aspects
The task supports the work of the Coordination Management Committee, that meets once a month and is in charge to insure the coordinated progression of the project. It also supports the work of the Project Review Board made of external reviewers for the selection of the projects for transnational access.

Task 3 – Administrative financial and legal management of RISIS
The UPEM management team is in charge of implementing the decisions taken by the Governing Board and the Coordination Management Committee.It has thus 3 tasks:
a) insure the financial operation and reporting of the project
b) develop a dedicated 'management website' that enables partners and the Commission to follow the activities of the project (it has four functions: keep all official documents of the project, gather all documents on 'activities' developed, keep all minutes and lists of decisions taken by the different boards and committees of the project, keep all milestones and deliverables of the project).
c) address all other issues (legal and administrative) raised by actual implementation.

Task 4 – Data Management Plan (DMP)
This task is in charge of developing and updating the data management plan. To support its development it includes the conduct of a collective legal expertise (on all aspects of IP and of data access and use).

Participation per Partner

Partner number and short name	WP1 effort
1 - UPEM	18.00
2 - AIT	2.00
3 - POLIMI	2.00
4 - CNR	2.00
5 - LEIDEN	2.00
6 - USFD	1.50
7 - FRAUNHOFER	2.00
8 - USI	2.00

Partner number and short name	WP1 effort
9 - JOANNEUM	1.00
10 - UOS	1.00
11 - DZHW	1.00
12 - STRATHCLYDE	1.00
13 - SAPIENZA	1.00
14 - NIFU	1.00
15 - CSIC	1.50
16 - SNI	1.00
17 - TC CAS	1.00
18 - ATHENA RC	1.00
Total	42.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D1.1	Preliminary Data Management Plan	1 - UPEM	Report	Public	6
D1.2	Data Management Plan 1st full version	1 - UPEM	ORDP: Open Research Data Pilot	Public	24
D1.3	Data Management Plan 2nd full version	1 - UPEM	ORDP: Open Research Data Pilot	Public	42

Description of deliverables

Data management plan
Annual reports

D1.1 : Preliminary Data Management Plan [6]

Preliminary Data Management Plan

D1.2 : Data Management Plan 1st full version [24]

Data Management Plan 1st full version

D1.3 : Data Management Plan 2nd full version [42]

Data Management Plan 2nd full version

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS4	Preliminary DMP	1 - UPEM	6	Report available

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS5	First RISIS week	1 - UPEM	12	All documentation available on line
MS16	Full legal expertise conducted	1 - UPEM	18	Report available
MS27	Second RISIS week	1 - UPEM	24	All documentation available on line
MS28	DMP first full version	1 - UPEM	24	Report available
MS50	Third RISIS week	1 - UPEM	36	All documentation available on line
MS61	DMP second full version	1 - UPEM	42	Report available
MS73	Fourth RISIS week	1 - UPEM	46	All documentation available on line

Work package number [9]	WP2	Lead beneficiary [10]	4 - CNR
Work package title	Interaction with the Research community, Communication and Dissemination (NA)		
Start month	1	End month	48

Objectives

The overall objective is twofold:
a) to increase the exposure to and the use of the RISIS infrastructure among academic researchers in the field of Science Technology and Innovation (STI) Studies in Europe and the integration of the STI scholars' community;
b) to develop an effective and continuous process of interaction between STI scholars, scholars from neighboring fields and stakeholders, in order to make RISIS data a key element of evidence-based policy-making for research and innovation.

Description of work and role of partners

WP2 - Interaction with the Research community, Communication and Dissemination (NA) [Months: 1-48]
CNR, UPEM, AIT, POLIMI, LEIDEN, USFD, FRAUNHOFER, USI, JOANNEUM, UOS, DZHW, STRATHCLYDE, SAPIENZA, NIFU, CSIC, SNI, TC CAS, ATHENA RC
The activities developed under this WP are directed to a broad audience, composed by:
i) scholars in the STI field,
ii) scholars from other neighboring R&D fields, and
iii) stakeholders organizations.
The latter include first Governments but also Regions, National Statistical Offices and European Statistical Units, organizations dealing with research and innovation (e.g. Universities and PROs), researchers and officers of relevant organizations, in particular those representing R&I Performers and Research Funding Organizations (e.g. Science Europe, EUA, LEU, EARTO, OECD) ans interested civil society organisations. WP3 is articulated in three tasks: Training, Dissemination and Communication.

Task 1 – Training
Training provides basic and advanced knowledge about the content of the RISIS infrastructure and how to use it, the methodologies to exploit the datasets for research aimed to supply evidence relevant for policy making. Activities build on the portfolio of training developed in RISIS1. RISIS 2 will continue to offer group training courses developing skills in using the RISIS datasets and platforms, and we will further develop online tools, tutorials and courses to support users and stakeholders remotely. Materials and documents associated with training will be freely available on the website at any time.
Training activities include the following:
- Tutorials: developing on-line tutorials to provide basic training for handling the datasets showing potentialities for analyses and research activities. This type of training is supposed to attract the attention of the different types of users toward RISIS;
- On-line training: provide periodical on-line training about how to use the datasets for research aims. In RISIS1 this type of training was done through short physical courses. We will adapt these events so that they can be accessible for a wide community (including those outside the users' community);
- On the job training: we have experienced during RISIS 1 on-site training for some datasets (e.g. ETER).
We will work to further develop this approach, with RISIS partners running training in places where there is a consistent community of users, which is developing research that would benefit from the use of the RISIS infrastructure but needs further assistance to learn how to use it for their specific research purposes.
We consider this kind of assistance also as a mean for further integrate the users STI community;
- Methodological courses: two-three days training courses dedicated to deepen specific methodologies mainly coming from activities developed within WP7. Methodological courses should enlarge the base of scholars using advanced quantitative methods;
- Schools: schools are devoted to train people using combination of different datasets to address key problems/analyses that are particularly relevant for policy purposes. Within Schools we can also realize spaces of 'creative training' as laboratories, where the trainers shall stimulate the trainees to reflect on their data needs and research questions, through direct participative techniques.
Training will also include dedicated courses and tutorials for integrating users of the STI community in the OpenAire environment, train them on how and why they take advantage from open access datasets with RISIS resources to

producing new means for R&D policy analysis. Training also addresses STI PhD courses developed in different countries in order to provide on-line training at the very beginning of the research career, and STI Master courses to introduce very early on these new capacities for the next generations. Courses in these cases will have contents properly designed for the different levels of their engagement. All participants to training courses (either on-line or physically) will provide their feedback after the event related to pros and cons from their experience through on-line questionnaires.

Task 2 – Dissemination

Dissemination will address targeted groups of users to improve the understanding of what RISIS can offer to their research effort (learning objective). Dissemination is also devoted to produce a change in the research practices of the existing STI community (transformative objective). Furthermore, dissemination shall develop productive interactions between researchers in the field, scholars from neighboring fields, policy makers, practitioners, and researchers working in stakeholders' organizations. Dissemination activities shall create occasions for discussing with a participative approach, relevant issues on STI policies, and the RISIS contribution to new developments and methodologies. All the Dissemination activities will benefit from the integration within OpenAire, which will further reinforce the possibility of reaching broad audiences of scholars and stakeholders.

We foresee four main group of activities:

- Attending international and national conferences in the fields of economics, management, finance, policy science, sociology, and indicators, as key mean for dissemination. WP3 either takes action or provides specific support to scholars for the organization of thematic tracks, conference sessions, mixed panels composed by scholars, practitioners and policy makers, on specific issues that can be addressed by using the RISIS data infrastructure. A set of core conferences, and 2 major conferences (EU-SPRI, STI-ENID) will be addressed, where RISIS must participate with dedicated events, presenting results and demonstrators of results realized throughout RISIS datasets. This set of activities shall increase the global visibility of the infrastructure, in all the areas of expected impact (See Part 2), especially transnational access and widening the users' base.
- 'Awareness rising' events with targeted audiences involving users from STI community, users from neighboring R&D communities and stakeholders organizations, to discuss items of interest, such as RISIS research outcomes, return experiences of users, planned developments of the research infrastructures, key research questions and ways to address them, role of the RISIS infrastructure for the goals of European integration. These events will have a strong participative orientation when they involve stakeholders; they can be working sessions within national and international conferences, dedicated workshops, or interactive learning seminars. Co-organization of participative events with stakeholders and representative of key research organizations (e.g. Science Europe and EUA) are also foreseen.
- Policy briefs: The main results coming from RISIS and their value to improve the use of data for evidence- policy making will be presented through short documents pointing out the main issues at stake, demonstrating the contribution provided by RISIS, and what new avenues for research are now open. Policy briefs will be produced at different time over the project duration by different WPs and will be coordinated by WP4. They can be not only paper documents but also take the form of videos or storytelling, providing evidences of how RISIS outcomes produce effects on research activities in the STI domain. Policy Briefs will be circulated around the broad community of scholars, stakeholders and policy makers through the RISIS communication channels, including the website for downloading.
- User oriented actions through a dedicated platform. RISIS provides a service where users can supply and find information linked to their experiences, results, problems deriving from the use of the infrastructure, free spaces for exchanges and interactions, and for building collaborative efforts; the platform should be integrated within the RCF. Access to OpenAire content and open science publishing (WP4) allows further integration of users from STI community and widening the existing community. WP3 will use the mentioned resources to build user-oriented actions to sustain the integration of the STI community.

Task 3 – Communication

Communication will run in tandem with the relevant core conferences engagement for dissemination purposes, policy briefs and awareness raising events.

The website is the most important mean to promote the dissemination of RISIS activities. It will be further improved during RISIS 2 becoming a tool for understanding what RISIS can offer, how you can join the community, what advantages you can have for your research work, and to access results and open resources produced; materials such as posters, printed promotion materials, templates, videos and story-telling, are also very important to this aim. A newsletter every six month is planned, to be distributed using a mailing list of users, stakeholders and policy makers, and accessible on the website. RISIS will also develop a blog for daily information and exchanges. The mentioned means will be combined for specific communication actions targeting researchers and analysts from stakeholder organisations.

Contacts with press at EU level and national level is foreseen -with the support of a communication specialist, at any time in order to improve the visibility of RISIS. Web-based exploration activities will be developed to be aware of sites where dissemination of RISIS can produce an impact attracting new users.

Using social media is a valuable way to keep in touch with people in the field, especially the young generation.

Actions are required to continue and improve the presence on Twitter and also for using other resources (e.g. YouTube) to share videos and promoting materials. Webinars will be used to present the results achieved. Online tools will be used that make it fairly straightforward.
Other strategic activities to fostering the impact plan, such as measures to promote project identity and branding will be developed as well, using also the possibilities provided by OpenAire, which will allow to monitor publications and acknowledgments to RISIS or related evidence of actual use (see WP6).

Organisation of the work package
CNR IRCRES is in charge of the WP central coordination. A Coordination Committee composed by other RISIS partners involved in WP3 supports the coordinated development of the different tasks, and the coordination with other WPs. It will meet physically once per year, during the RISIS week, in order to present and discuss the work done and prepare the future actions, according to the Plan for dissemination and exploitation of results.
We use a permanently open call for proposing the training activities (both on-line and physical courses) that will be analyzed by the Coordination Committee. A similar call will be opened for 'awareness raising events', and for organizing special tracks, panels or sessions at core conferences. The Committee will also develop top-down events when considered necessary.
WP3 will monitor users' views and needs through on-line feedbacks after activities developed (training, access, workshops, conferences special sessions, communication events)

Participation per Partner

Partner number and short name	WP2 effort
1 - UPEM	0.00
INRA	2.00
CCIP	2.00
2 - AIT	2.00
3 - POLIMI	3.00
4 - CNR	36.00
5 - LEIDEN	2.00
6 - USFD	2.00
7 - FRAUNHOFER	3.00
8 - USI	2.00
9 - JOANNEUM	2.00
10 - UOS	1.00
11 - DZHW	1.00
12 - STRATHCLYDE	1.00
13 - SAPIENZA	2.00
14 - NIFU	1.00
15 - CSIC	1.00
16 - SNI	1.00
17 - TC CAS	1.00
18 - ATHENA RC	1.00
Total	66.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D2.1	Website and materials	4 - CNR	Websites, patents filling, etc.	Public	12
D2.2	First Report on activities developed for training, communication and interactions with users & stakeholders	4 - CNR	Report	Public	18
D2.3	Analysis of users' needs	4 - CNR	Report	Public	24
D2.4	Report on Press and media actions	4 - CNR	Report	Public	30
D2.5	Second Report on activities developed for training, communication and interactions with users & stakeholders	4 - CNR	Report	Public	36
D2.6	Policy Brief on achievements and lessons learned	4 - CNR	Report	Public	46

Description of deliverables

Website, Media, Policy briefs, user's needs

D2.1 : Website and materials [12]

Website and materials

D2.2 : First Report on activities developed for training, communication and interactions with users & stakeholders [18]

First Report on activities developed for training, communication and interactions with users & stakeholders

D2.3 : Analysis of users' needs [24]

Analysis of users' needs

D2.4 : Report on Press and media actions [30]

Report on Press and media actions

D2.5 : Second Report on activities developed for training, communication and interactions with users & stakeholders [36]

Second Report on activities developed for training, communication and interactions with users & stakeholders

D2.6 : Policy Brief on achievements and lessons learned [46]

Policy Brief on achievements and lessons learned

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS6	Revised website, materials and video	4 - CNR	12	Altmetric indicators related to the website (& its elements) & of twitter
MS7	Training (first set of courses & tutorials)	4 - CNR	12	Full course documentation and tutorials on line
MS17	First batch of policy briefs & webinars	4 - CNR	18	Presence on line & downloading measures
MS29	Training (second set of courses & tutorials)	4 - CNR	24	Full course documentation and tutorials on line
MS30	User assessments & returns: first monitoring	4 - CNR	24	Blog altmetrics & report of return sessions
MS43	Second batch of policy briefs with participative events & webinars	4 - CNR	30	Presence on line & downloading measures
MS51	Training (third set of courses & tutorials)	4 - CNR	36	Full course documentation and tutorials on line
MS62	Third batch of policy briefs with participative events & webinars	4 - CNR	42	Presence on line & downloading measures
MS70	Training (last set of courses & tutorials)	4 - CNR	46	Results of the monitoring of users' assessments
MS71	User assessments & returns: final monitoring	4 - CNR	46	Blog altmetrics & report of return sessions

Work package number [9]	WP3	Lead beneficiary [10]	1 - UPEM
Work package title	Preparation of long term sustainability of RISIS (NA)		
Start month	1	End month	48

Objectives

The objective of this work package is to prepare the future long term sustainability of RISIS

Description of work and role of partners

WP3 - Preparation of long term sustainability of RISIS (NA) [Months: 1-48]
UPEM, AIT, POLIMI, CNR, LEIDEN, USFD, FRAUNHOFER, USI, CSIC
This work package deals with the preparation of the future of RISIS and its different components: the long term engagement of members, the development of a costing approach to shared elements of the infrastructure, the definition and deployment of the 'indirect contract' approach, the identification and analysis of options for the lasting operation of RISIS as a European infrastructure. This will be addressed through the constitution of a specific working group – the sustainability working group – that will be created from the start of the project.
The SWG presents its progress and issues requiring strategic decisions to the Governing Board every year on the different aspects identified. It will deliver a report on options by month 30. It will be use the different supports available for organizing wider discussions within the membership (via RISIS weeks) or by developing specific user and stakeholder meetings as planned in WP2.
It will issue a plan for the future well in advance of project termination.

Participation per Partner

Partner number and short name	WP3 effort
1 - UPEM	6.00
2 - AIT	1.00
3 - POLIMI	1.00
4 - CNR	1.00
5 - LEIDEN	1.00
6 - USFD	0.50
7 - FRAUNHOFER	1.00
8 - USI	1.00
15 - CSIC	0.50
Total	13.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D3.1	Report on options for the future by the SWG	1 - UPEM	Report	Confidential, only for members of the consortium (including	30

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
				the Commission Services)	
D3.2	Plan for the future	1 - UPEM	Report	Public	42
D3.3	Report on indirect contract approach	1 - UPEM	Report	Confidential, only for members of the consortium (including the Commission Services)	18

Description of deliverables

SWG reports about the future of the project

D3.1 : Report on options for the future by the SWG [30]

Report on options for the future by the SWG

D3.2 : Plan for the future [42]

Plan for the future

D3.3 : Report on indirect contract approach [18]

Report on indirect contract approach

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification

Work package number [9]	WP4	Lead beneficiary [10]	1 - UPEM
Work package title	RISIS Core Facility (JRA)		
Start month	1	End month	48

Objectives

The objective of this integrating Work Package is to provide a ground breaking technical infrastructure to support distant monitored and secured access to datasets and services through ergonomic user-oriented interfaces. The intended architecture is planned as a web-oriented architecture (WOA). This requires to share a common vision on the design, architecture, roadmap of IT development in order to develop and open the RCF in stages, with a first opening at M18. This infrastructure is based on a solid hardware and system architecture, including powerful processing and web servers, fast and secured network layer, dedicated application containers and monitoring, to ensure global security and a quality of service for the infrastructure. The IT Center of UPEM will be involved to empower the RISIS infrastructure.
This architecture will serve as a foundation to build full stack web services for 3 types of users:
- Users that will directly use the data API of the RCF to extract specific linked datasets for an external usage,
- Users of the frontend web applications of RCF that provide private and collaborative workspaces and access a set of services, processing internal and external datasets
- Users that benefit from the service API of the RCF to build their own application on top of the infrastructure, to be integrated into the RISIS 2 project after that.
The link from RCF toward Open Data communities will be held in coordination with D4Science. D4Science will design, deploy, and operate a VRE (Virtual Research Environment) to foster interaction between RISIS-RCF and OpenAire with the view to have the RISIS project to contribute to the OSaaS (Open Science as a Service), while linking a Community Based Infrastructure in science policy and innovation studies and the specific and relevant databases and contents imported from the OpenAIRE infrastructure.
This WP will be managed by UPEM (M. Barbier for coordination and P. Breucker for technological integration).

Description of work and role of partners

WP4 - RISIS Core Facility (JRA) [Months: 1-48]
UPEM, AIT, CNR, LEIDEN, USFD, JOANNEUM, CSIC
The WP will realize the design and the development of the modules of the RCF through time. Three milestones will organize the roadmap and delivery of developments.
1. The RCF will first start with the assemblage of existing resources:
- Enriched & harmonized datasets that are partially accessible on-line and all specified with metadata and condition of uses, and inter-operable
- A robust IT support, including server and software architecture, and human skills from existing platforms
- User-oriented procedures with 'virtual' transnational access to datasets
- Integration of existing services (enhanced and developed in WP6), for data mining and data analysis on heterogeneous datasets, including existing analytical and NLP platforms (Risis-Cortext, GATE, ORGREG, VOS-Viewer, D4Science-DataMiner)
2. The WP will then concentrate on the usage of bundles of services defined by advanced users and on the integration of heterogeneous datasets (notably Open Linked Data) in relation to numerous Open Science initiatives relevant for researchers of the RISIS Communities. This will be held by D4Science via the development of a specific VRE tailored to satisfy the RISIS needs and empowered by the RCF framework.
3. The WP will end with a short period dedicated to the full description and documentation of the RCF and its ability to inter-operate with other platforms of the European Open Science Cloud.

Task 1 – RCF Integration
6 subtasks organize the RCF integration.
ST1.1 Hardware and system architecture
Deliver a solid and secured hardware and system architecture that ensures the RCF's usability and quality of service. It will be implemented in coordination with UPEM's IT Resources Center (CRI). This task will deliver:
- servers and dedicated storage with their configurations, dedicated virtual machines and operating system configurations
- a layer of core network and system components (such as web server, containers, monitoring framework, system libraries)
- a backup system capable of restoring the full system and data in case of software or hardware malfunctions

- a monitoring system (based on tools like NAGIOS or Munin) that can alert administrators in case of malfunctions and keep the resource-usage history.
ST 1.2 Authentication & Authorization service
- Deliver a fully functional access control system, capable of handling the authentication of users from within the RCF, and from external partners applications.
- Deal with user-level, service-level and datasets-level authorizations.
The system provides a backend oAuth-based web service and a frontend administration tool to handle authorizations.
ST 1.3 User frontend
Development of web frontend enabling user to access datasets, services and visualization applications. Provides the main dashboard for the RCF to access workspaces, profile, projects, applications and so on.
ST 1.4 Service architecture
Implementation of an architecture for service access. The goal of this architecture is to ensure that internal services (such as dataset linking) can be integrated with external services such as those brought by Cortext (Text and Network analysis, Geographic), GATE (Text-Mining), and D4Science (DataMiner NLP Hub).
The RCF will access those external services via their APIs and process the results using its internal architecture and interchange format. All these external services will be developed and enhanced in WP5. The task here is about defining the common architecture to process results from a variety of services.
ST 1.5 Data integration
Provide the RCF with importing, linking and data storage capability. Implement an API handling data access from other modules, including security management. The first linking capabilities will be organizationoriented, using RISIS dataset OrgReg as a reference.
ST 1.6 Usage Monitoring
Development of a system describing global usages of RCF, from a systematical and dynamical point of view, based on underlying monitoring tools.
Task leader: UPEM (Philippe Breucker). Participants: AIT, CWTS, Joanneum, CNR, USFD, Ingenio

Task 2 – Open Data VRE
Deliver a tailored VRE operated and empowered by D4Science, in order to bridge RCF and OpenAIRE. This VRE will allow delivering to the RISIS community tailored and specific datasets collected by OpenAIRE and selected to satisfy the needs of the RISIS community. The VRE will therefore open a directional bridge from the open access and open science communities accessible through the OpenAIRE infrastructure.
ST 2.1 APIs design and documentation
To exploit this connection to OpenAIRE, shortly identified as publication and access, a full set of APIs will be properly designed and documented.
ST 2.2. VRE installation
The RISIS Open Data VRE will become a joint and interoperable facility of the RISIS infrastructure able to use the RISIS authentication and authorization, capable of exploiting the hardware resources operated and delivered by the UPEM's IT Resources Center (CRI) to ensure controlled access,
ST. 2.3. Exploitation of RCF services
Developing the environment of the access API to Open (Linked) Data with the objective of procuring the RISIS community with new datasets
Task leader: CNR. Participants: UPEM, CWTS, USFD

Task 3 – RCF Design and coordination
This Task aims at actively introduce the Users' needs and problem solving orientation in the design of the RCF in accordance with the feedbacks of advanced users and the networking and technical services in support of the D4Science initiative. It also ensures that implementation efforts are coherent with those needs. In order to achieve that, technical seminaries will be held during the project period.
ST 3.1 Implementation coordination
Coordination and management of the WP in relation to other WPs of the project: Interfacing with the other WPs of the project (Datasets WP and the assembly of IT persons of each Dataset Facility (in relation to Task 1 and 2 basically); the advanced users that specify their uses of resources & services and needs for creating
new indicators (in relation to Task-2 and 3 basically) ; the WP training that can open the RCF to CoP from various Scientific fields (in relation to Task-3 basically)
ST 3.2 RCF Progressive design
Represent needs and problems that are specific to certain research domains or research Community of Practices;
Use case of problem-solving design with: Nano; Career, Cancer, Gender, etc;
Specification and Design of the bridge with Open Access data empowered by D4Science, linking API of RCF and VRE to be developed in the WOA (SubTask 1.2 & 1.3).

ST 3.3 Organisation of Technical seminars
Technical seminar about e-infrastructure methodologies and state of the art. Reflecting about technical approach to the specific research problems existing in the RCF user and Open Science communities
Open Benchmarking Seminar about other experiences of Social Sciences and Humanities Infrastructure
Task leader: UPEM (Marc Barbier) Participants: AIT, CWTS, Joanneum, CNR, USFD, Ingenio

Participation per Partner

Partner number and short name	WP4 effort
1 - UPEM	20.00
INRA	48.00
2 - AIT	10.00
4 - CNR	24.00
5 - LEIDEN	16.00
6 - USFD	14.00
9 - JOANNEUM	10.00
15 - CSIC	8.00
Total	150.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D4.1	Joint-Opening of RCF beta version	1 - UPEM	Websites, patents filling, etc.	Public	18
D4.2	Opening of Open Data VRE	4 - CNR	Websites, patents filling, etc.	Public	18
D4.3	Frontend to access and navigate datasets	1 - UPEM	Websites, patents filling, etc.	Public	18
D4.4	Access to core Risis services	1 - UPEM	Websites, patents filling, etc.	Public	18
D4.5	Opening RCF data and services including external services linkage and VRE	1 - UPEM	Websites, patents filling, etc.	Public	36
D4.6	Usage Monitoring and infrastructure documentation	1 - UPEM	Report	Public	46

Description of deliverables

Usage, access and opening monitoring

D4.1 : Joint-Opening of RCF beta version [18]
Joint-Opening of RCF beta version

D4.2 : Opening of Open Data VRE [18]
Opening of Open Data VRE

D4.3 : Frontend to access and navigate datasets [18]
Frontend to access and navigate datasets

D4.4 : Access to core Risis services [18]
Access to core Risis services

D4.5 : Opening RCF data and services including external services linkage and VRE [36]
Opening RCF data and services including external services linkage and VRE

D4.6 : Usage Monitoring and infrastructure documentation [46]
Usage Monitoring and infrastructure documentation

Schedule of relevant Milestones

Milestone number[18]	**Milestone title**	**Lead beneficiary**	**Due Date (in months)**	**Means of verification**
MS8	Opening of VRE API	4 - CNR	12	RISIS VRE accessible
MS18	Opening of RCF beta version	1 - UPEM	18	RCF portal opened
MS31	Data services inserted in RCF	1 - UPEM	24	Services accessible directly through RCF
MS44	VRE fully enhanced with Open Data datasets	4 - CNR	30	Through operation of RISIS VRE
MS52	External data services (CORTEXT, GATE, D4Science) integrated	1 - UPEM	36	Access through RCF fully completed

Work package number [9]	WP5	Lead beneficiary [10]	2 - AIT
Work package title	Core Datasets Maintenance (NA)		
Start month	1	End month	48

Objectives

The overall objective of this WP is the integrated maintenance of the RISIS core datasets.
RISIS core datasets correspond to three criteria:
(a) they cover at least the EU and longer time periods (a critical dimension for indicator building);
(b) they represent core research, innovation and/or policy dimensions; and
(c) they have generated significant interest from users in RISIS1 and beyond.
Our core datasets have four main focuses:
a) Firm innovation capabilities (with our three firm datasets: CinnoB, VICO and CHEETAH);
b) European Public research and higher education organisations: RISIS-ETER (incl. OrgReg)
c) Output oriented datasets (with publications – CWTS publication database -, patents – IFRIS Patstat and R&D projects – EUPRO); and
d) Policy learning (with SIPER, the repository of policy measures and evaluations).
In addition, we will maintain a dataset on research careers, MORE, integrating newer MORE (MORE3) studies in the RISIS data nexus.

Description of work and role of partners

WP5 - Core Datasets Maintenance (NA) [Months: 1-48]
AIT, UPEM, POLIMI, LEIDEN, FRAUNHOFER, NIFU
RISIS core datasets require important maintenance efforts to keep their relevance for the field. Maintenance in this context mostly concerns the periodic updating of their information, i.e. collection of recent data, incorporate them in the databases, and integrate within the RISIS data nexus (see RISIS integrative dimensions, WP6). In this sense, maintenance is an issue of periodicity of updating, and incorporation of changes having happened compared during the period.
We face different situations depending upon the size of databases, the nature of the maintenance process and the extent of changes in past information. We have some datasets that will be updated annually, while other datasets will be updated bi-annually or only once during the project period. The most standardized form of maintenance, in form of clearly defined annual updates will be done for two of our output oriented datasets, the CWTS publication database that records information on publications given in the Web of Science and related bibliometric indicators (see WP8 for a detailed description of all datasets), and the EUPRO database, comprising systematic information on European R&D projects (and later also national, see WP9 on EUPRO development). While for the CWTS publication database periodicity concerns mostly the inclusion of new articles, EUPRO update basically refers to the integration of new projects (H2020, FP9, Eureka, JTIs, etc.), both feasible to be done at annual basis given the well developed update routines for both datasets. For the third output oriented dataset, IFRIS-PATSTAT recording standardized information on global patent applications, an annual update is not feasible as the incorporation of new patents requires reconsidering previous patents because of changing ownerships. Thus, maintenance for IFRIS-PATSTAT involves the advancement of the robustness of the dataset, and one full update is planned (in the third year of the project). It is important to note that maintenance for all datasets involves the ongoing cleaning, in particular of organisation names and harmonization with RISIS organisation registers, geolocalisation of organisation addresses, and topical annotation of data items based on semantic technologies (see WP6).

Turning to the datasets focusing on firm innovation capabilities, i.e. VICO, CHEETAH and CinnoB, the situation is more diverse and confronted by huge challenges related to the tracing of firm ownership relations. As for VICO, a dataset on European start-ups, and CHEETAH, a dataset on mid-sized firms showing fast growth rates, an appropriate strategy is to go for bi-annual updating in terms of the integration of new firms (from different sources), and the tracing of the survival dynamics of existing firms in the two datasets.
For CinnoB (CIB in RISIS 1), a dataset on the world's largest R&D investors (around 4000 in CIB), the issue of ownership relations is even more severe, given the extremely high number of subsidiaries of large firms, and the immense dynamics. Therefore, maintenance of CinnoB will be one major update that will include an updated list of the largest R&D performers and an update of the firms' patent portfolios up to 2020 using the updated version of IFRIS-PATSTAT.

Completely different situations in terms of updating are given, for the remaining core datasets, SIPER on policy evaluations, RISIS-ETER on public research organisations, and MORE on research careers. Maintenance for SIPER means to keep the database globally relevant, entailing searching de-facto continuously for new evaluations to include and to reach out to further countries and regions as appropriate. RISIS-ETER is maintained under a service contract of the European Commission and, therefore, does not require RISIS resources for maintenance. Two additional waves of data collection are foreseen during the project period. MORE will be subject to a smaller maintenance activity in WP5, integrating the newer MORE studies (MORE3) in relation to the already existing ones (MORE1 and MORE2).

Task 1 – Developing a consolidated concept and fully operational work plan for maintenance
Task 2 – Annual updating of EUPRO and CWTS Publication database
Task 3 – Bi-annual update of VICO and CHEETAH
Task 4 – New versions of IFRIS-PATSTAT and CinnoB
Task 5 – Ongoing integration of policy evaluations in SIPER
Task 6 – Integration of new RISIS-ETER and MORE datasets
Task 7 – Integration of updated datasets according to RISIS integrative dimensions

Participation per Partner

Partner number and short name	WP5 effort
1 - UPEM	12.00
CNRS	6.00
CCIP	6.00
2 - AIT	16.00
3 - POLIMI	24.00
5 - LEIDEN	12.00
7 - FRAUNHOFER	12.00
14 - NIFU	4.00
Total	92.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D5.1	Consolidated work plan on Maintenance	2 - AIT	Report	Public	6
D5.2	First interim report on Maintenance of RISIS core datasets	2 - AIT	Report	Public	18
D5.3	Second interim report on Maintenance of core RISIS datasets	2 - AIT	Report	Public	36
D5.4	Final report on Maintenance of core RISIS datasets	2 - AIT	Report	Public	46

Description of deliverables

Maintenance monitoring

D5.1 : Consolidated work plan on Maintenance [6]
Consolidated work plan on Maintenance
D5.2 : First interim report on Maintenance of RISIS core datasets [18]
First interim report on Maintenance of RISIS core datasets
D5.3 : Second interim report on Maintenance of core RISIS datasets [36]
Second interim report on Maintenance of core RISIS datasets
D5.4 : Final report on Maintenance of core RISIS datasets [46]
Final report on Maintenance of core RISIS datasets

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS1	13 datasets open: 9 core datasets, 4 datasets of interest	2 - AIT	1	Datasets accessible via RISIS datasets portal
MS9	Updated cheetah 1	3 - POLIMI	12	Updated database available for Access, reports on maintenance
MS10	Enriched IFRIS PATSTAT	1 - UPEM	12	Updated database available for Access, reports on maintenance
MS11	1st Annual update of CWTS Pub, EUPRO & SIPER	2 - AIT	12	Updated database available for Access, reports on maintenance
MS32	Updated VICO 1	3 - POLIMI	24	Updated database available for Access, reports on maintenance
MS33	2nd Annual update of CWTS Pub, EUPRO & SIPER	2 - AIT	24	Updated database available for Access, reports on maintenance
MS45	IFRIS Patstat new version	1 - UPEM	30	Updated database available for Access, reports on maintenance
MS53	Updated cheetah 2	3 - POLIMI	36	Updated database available for Access, reports on maintenance
MS54	3rd Annual update of CWTS Pub, EUPRO & SIPER	2 - AIT	36	Updated database available for Access, reports on maintenance
MS63	CIB2 complemented by new version of CinnoB	1 - UPEM	42	Updated database available for Access, reports on maintenance
MS64	Updated VICO 2	3 - POLIMI	42	Updated database available for Access, reports on maintenance

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS72	4th Annual update of CWTS Pub, EUPRO & SIPER	2 - AIT	46	Updated database available for Access, reports on maintenance

Work package number [9]	WP6	Lead beneficiary [10]	6 - USFD
Work package title	Data Integration and Analysis tools (JRA)		
Start month	1	End month	48

Objectives

This work package has the following key objectives:
a) Strengthen, interoperate services in consistent workflow and optimise them in light of integration within the RISIS Core Facility (RCF)
b) Widen significantly the type of research activities supported by RCF, through the integration of new data analytics tools and services, specifically by enabling flexible geographical analysis and multilingual content analytics.
c) Design and implement new visual analytics interfaces for RCF, to improve usability and harmoniseuser interactions, as well as enable researchers from the OpenAIRE infrastructure with the ability to access and use RISIS RI services and outputs.
The expected outcome is a much richer and more powerful RISIS research e-Infrastructure, that offers a wide, state-of-the-art range of theoretically-sound and practically applicable data analytics methods and algorithms, providing a broad statistical foundation for new research.

Description of work and role of partners

WP6 - Data Integration and Analysis tools (JRA) [Months: 1-48]
USFD, UPEM, AIT, POLIMI, CNR, LEIDEN, USI, JOANNEUM, UOS, CSIC, ATHENA RC
Task 1 – Integration of Existing and New Text Analysis and Semantic Annotation Tools and Services
This task will enhance the RISIS e-infrastructure in two ways - firstly, through the integration of scalable text analytics and semantic annotation tools and services, from two already established, complementary einfrastructures– CorTexT (https://www.cortext.net/- UPEMLV) and GATE Cloud (https://cloud.gate.ac.uk – USFD). Secondly – the RISIS RI will be enhanced through the development of new deep learning methods.
CorTexT provides text mining tools and socio-semantics analysis for corpus-level (inter-document) analysis, including data parser (txt, doc, json, ris and other proprietary format such as WoS, Scopus, Factiva,…), term extraction, generation of word2vec models, interaction networks (e.g. retweet networks), clustering, and graph-based community detection. A first integration of those services was completed in RISIS1.
Here UPEM will work on strengthening the already provided services and on complementing with new scripts (notably for geocoding and topic-modelling) and contribute to the portfolio of services (objective 1). GATE Cloud is complementary to CorTexT, as it provides numerous open-source multilingual tools and services for deep natural language processing analysis, on a document-by-document basis. This includes lexical analysis, Named Entity Recognition (NER), entity linking with Linked Open Data (LOD) resources), and ontology-based semantic annotation. USFD will integrate all these new multilingual text analysis tools and services in the RISIS research e-Infrastructure, thus widening significantly the kinds of research and data analysis supported (Objective 2). The output of the document-level linguistic and semantic annotations from GATE will also be made available to CorTexT, which can then use them in aggregate, at corpus level, for discovery of new networks, communities, and clusters. The integration will be based on well-defined APIs and standard data interchange formats (e.g. JSON, XML, CSV), to ensure interoperability and reusability.
Secondly, new deep learning methods will be developed to enable new research on the dynamics of science, technology, innovation and knowledge production, based on the datasets available in the RISIS Infrastructure. A specific new dimension will be the ability to use natural language processing tools and services in multiple languages. Addressing multilingual content analytics is essential, since it brings better understanding of multicultural events and topics and offers complementary information.
Task leader: USFD. Participants: UPEMLV, CWTS, CNR

Task 2 – Actors
A key outcome of the first RISIS project has been the establishment of two organizational registers, namely OrgReg for universities and public-sector organizations and FirmReg for firms. The register system allows to identify in an unambiguous way organizations and their linkages, by attributing stable identifiers that can be
used to match them with other datasets. This task aims to extend this register system. ORGREG is already online. It will be updated annually. Developments will deal with a better coverage of research hospitals, extending the university coverage worldwide (US coverage & 500 top ranking universities in the Shanghai ranking) and building a sub-register

for PRO (with characterisation & output variables). FIRMREG only exists as a prototype. The objective is to develop it and open it to the research community.
This requires the following steps: stabilise the perimeter (today made of the 3 RISIS firm datasets, by inclusion of firms present in other RISIS datasets); addition of basic data on firms (size, industry), inclusion of parent-subsidiary linkages and of R&D activities (publications, patents, trademarks, projects). After opening a procedure for its periodic extension and updating will be developed.
Task Leader: POLIMI, participants: AIT, CWTS, Joanneum, Sussex, USI

Task 3 – Tools for advanced geographical analysis
This task aims to reinforce the ability of the RISIS e-infrastructure to bring geographical analysis to the RISIS community. “70% of our text documents contain place name references” (Hill, 2006), but cannot be easily computed (ambiguities, spelling mistake, change of names). Moreover, when looking for statistics on geography, especially at the European level, finding a convenient aggregation level (administrative or based on economic data) that make able both to compare sub-regional dynamics across countries and to reflect the geographical continuity of socio-economic phenomena, is a difficult question.
This task will try to solve these issues by building services that will allow RISIS community to automatically add geographical information to their datasets in a very flexible way. The following group of services will be accessible:
- Data level: UPEM has developed a method that is able to massively find geographical coordinates from given addresses: a geocoding tool at a worldwide scale. One of its first steps is a Name Entity Recognition as accessible in T1, applied here on toponyms. This tool is already provided in CorText Manager, it will be accessible independently or as a prerequisite before using the two other analytical layers.
- Aggregated level: a first iteration of this service was implemented as a standalone application in RISIS.
This task will enlarge these approaches through two sets of services: one for clustering (based on geographical density and on core-periphery analysis) and a second one to calculate multimodal travel times (isochrones that show time needed to reach a given point);
- Visualisation: in coordination and close collaboration with task T4, T3 will develop a set of applications to simplify the creation of maps for non-geographers with various geographical elements: points (data level), shapes (aggregate level), networks, choropleths or heatmaps (for analysis).
Task leader: UPEM. Participants: UPEM, Ingenio, AIT, USI

Task 4 – Visual analytics
T4 will advance and integrate in the RISIS e-Infrastructure a range of scalable visual analytics methods, in order to provide interactive visual exploration over the outputs from the data analytics and text processing tools and services. A wide range of visualisations will be supported. The first half of this task will focus on the challenge of opening up and integrating the corpus-level visualisation existing tools from CorTexT (UPEM), VosViewer (CWTS) and the document-based GATE Cloud (USFD) visualisations, as they are all currently tightly coupled to their specific platform. The goal here is to componentise them and make them easily reusable and embeddable within the RISIS infrastructure, based on a common RISIS data interchange format. Additionally, challenges of using the mentioned tools to analyse and visualize the big data available within the RISIS infrastructure will be addressed. Secondly, UPEM and USFD will develop and integrate novel methods for context-aware visualisation of data, which will tie together corpus-level, document-level, and Linked Open Data semantic-based annotations and metadata. We will also aim to enable visual analytics across different RISIS datasets and domains, i.e. these methods will enable joint, interactive analysis. Thirdly CWTS will bring its competencies to align visualisation techniques on the data interchange format and have them presented and actionable in the RCF front end. More specifically, VosViewer standard formats for data input and output will be used to interchange data between the backend and frontend of RCF, and VOSviewer visualizations will form a major part of the RCF user-oriented front end.
Task leader: CWTS. Participants: USFD, UPEM, AIT, CNR

Task 5 – Processing of OpenAIRE content
This task will enable the RISIS2 tools and services to process OpenAIRE content, via APIs (http://api.openaire.eu). T5.5. will ensure the matching of the expected format and provide mapping onto the RISIS conceptual model. ARC will provide the expertise required to understand the data (completeness and provenance) and identify ways to enrich it to better match the expectations of RISIS RI. Specifically, T5 will enable RISIS2 services to process the following content:
- The Scholexplorer graph (scholexplorer.openaire.eu/), which provides access to a large collection of links between publications and datasets: 18 mi bi-directional links from 1000 publishers & data centers.
- The OpenAIRE scholarly works graph (api.openaire.eu), which offers access to a graph of 30M interlinked entities collected from 10,000 + data sources.
Task leader: ARC (OpenAIRE Legal Entity). Participants: CNR, USFD.

Task 6 – Virtual Research Environment (VRE) for Seamless Embedding of RISIS Outputs into OpenAIRE

The VRE is operated and empowered by the D4Science infrastructure and will support the publication RCF tools and services into the OpenAIRE infrastructure and in particular to the Community Dashboard that OpenAIRE will create for the RISIS community. The VRE delivers back to the OpenAIRE infrastructure the outputs of the RISIS infrastructure based on calibrated datasets analysis to explore the production of indicators over time, which have to be based on reliability, reproducibility and timely weighted over decades. The VRE will therefore:
- offer a virtual environment for collaborative analysis and discussion of RISIS contributions to OpenAIRE
- deliver a computational environment for exploration and validation of datasets dedicated to OpenAIRE communities.
Task leader: CNR Participants: UPEM, USFD, CWTS, Ingenio

Task 7 – Open Science Tools for Measuring the RISIS Impact
This task implements tools that monitor and quantify the outputs of RISIS and every open science publication, dataset, and experiment produced by using RISIS. This will enable the measurement of the research impact of the RISIS. In particular, this task will deploy, configure, and provide the technical support required to operate the OpenAIRE Research Impact Monitoring Dashboard (RID) for RISIS. The RID enables monitoring and statistics reporting of the artefacts (publications, datasets, software) generated by RISIS. Artefacts are associated (and reported) to RISIS by users, who can publish artefacts in a virtual space (deposition, DOI minting, and findability). All RI services developed in WP5 will also publish artefacts while scientists who are using these services will also produce new artefacts which will also be tracked automatically (e.g. experiments, methods, datasets). OpenAIRE itself will enrich this RISIS impact evidence automatically, by textmining publications to identify acknowledgments to RISIS (or related evidence).
Task leader: ARC. Participants: CNR, UPEM.

Participation per Partner

Partner number and short name	WP6 effort
1 - UPEM	12.00
INRA	6.00
CCIP	6.00
2 - AIT	6.00
3 - POLIMI	10.00
4 - CNR	24.00
5 - LEIDEN	10.00
6 - USFD	48.00
8 - USI	6.00
9 - JOANNEUM	12.00
10 - UOS	8.00
15 - CSIC	6.00
18 - ATHENA RC	16.00
Total	170.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D6.1	Consolidated report on the new services opened (based	6 - USFD	Other	Public	18

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
	upon corresponding milestones)				
D6.2	First report on measuring RISIS impact through openscience tools	18 - ATHENA RC	Report	Public	24
D6.3	RCF methods & service integration	6 - USFD	Other	Public	36
D6.4	Second report on measuring RISIS impact through openscience tools	18 - ATHENA RC	Report	Public	46
D6.5	RCF methods & service integration – final version	6 - USFD	Other	Public	46

Description of deliverables

Integration and impact monitoring

D6.1 : Consolidated report on the new services opened (based upon corresponding milestones) [18]
Consolidated report on the new services opened (based upon corresponding milestones)

D6.2 : First report on measuring RISIS impact through openscience tools [24]
First report on measuring RISIS impact through openscience tools

D6.3 : RCF methods & service integration [36]
RCF methods & service integration

D6.4 : Second report on measuring RISIS impact through openscience tools [46]
Second report on measuring RISIS impact through openscience tools

D6.5 : RCF methods & service integration – final version [46]
RCF methods & service integration – final version

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS2	CORTEXT, ORGREG & geoservices opened	1 - UPEM	1	CORTEXT, ORGREG & geoservices Accessible online
MS12	Opening of GATE resources	6 - USFD	12	Accessible online from RISIS portal
MS13	New release of ORGREG with worldwide universities and hospitals	3 - POLIMI	12	Accessible on CORTEXT Platform
MS19	Opening of tools enabling exploration and integration of Open- AIRE content	18 - ATHENA RC	18	Accessible online through RCF

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS20	Enriched geo services opened	1 - UPEM	18	Accessible online through RCF
MS21	Opening of first batch of new visualisation services	6 - USFD	18	Accessible online through RCF
MS22	First opening of firmreg	3 - POLIMI	18	Accessible online through RCF
MS34	Opening of new NLP and multilingual analysis capacity (V1)	6 - USFD	24	Accessible online through RCF
MS35	Opening of VRE for seamless embedding of RISIS outputs in OpenAire	4 - CNR	24	Accessible online through RCF
MS47	New release of ORGREG with PRO subregister	3 - POLIMI	30	Accessible online through RCF
MS55	FIRMREG first full version opened	9 - JOANNEUM	36	Accessible online through RCF
MS56	All services fully integrated within the RCF	6 - USFD	36	Accessible online through RCF

Work package number [9]	WP7	Lead beneficiary [10]	8 - USI
Work package title	Advanced user communities and advanced analytical methods (NA)		
Start month	1	End month	48

Objectives

a) To develop usage cases from RISIS datasets aimed at providing novel answers to policy questions through the use of advanced analytical methods.
b) To create advanced user communities around questions of interests and analytical methods.
c) To create a RISIS methodological support unit providing support on use of advanced methods and offering methodological training.

Description of work and role of partners

WP7 - Advanced user communities and advanced analytical methods (NA) [Months: 1-48]
USI, AIT, POLIMI, SAPIENZA
Task 1 – Developing a first batch of advanced usage cases (M1-M24)
Involved partners: USI, POLIMI, Sapienza, AIT.
The first batch of usage cases focuses on a number of relevant problems for which data integration in RISIS may provide novel answers, but where the adoption of more advanced methods is required in order to take into account the specificities of the data. The focus is on inferential methods, i.e. on methods that allow to test statistically hypotheses on the underlying phenomena, complementarily with the descriptive tools provided by the RCF. These usage cases have therefore both a substantive and methodological dimensions and will provide a first entry point in order to build user communities.
The first batch of usage cases deals with following relevant problems:
- How to take into account the multi-level nature of research systems with individuals nested within organizations and within countries and the multiple dependencies across these levels (for example organizational characteristics influencing individual productivity). Method: Multi-level models.
- How to account for spatial dependence in regression models. Spatial dependence refers to the correlation of observed values across geographical spaces (i.e. spatial autocorrelation), violating the independence assumption in regression models that have spatial entities (e.g. European regions) as units of observation, and thus, leading to biased estimates. Methods: spatial econometrics (allowing also for the estimation of spatial spillovers as explicitly policy relevant issue).
- How to take into account underlying (unobserved) heterogeneity of cases (particularly organizations) that affects the relationships between resources and activities, respectively to build sensible typologies of organizations. Methods: latent class models.
- How to treat longitudinal data taking into account the specificities of RISIS data, like short time-series presence of autoregressive variables and endogeneity. This use case
will specifically aim at providing guidance to users on the most suitable panel data methods to deal with their specific research question. Methods: (dynamic) panel data methods.
- How to compare in a sensible way the multidimensional performance of Decision Making Units at different levels of analyses by taking into account underlying differences in resourcing and environment. Methods: Advanced Benchmarking Models and Techniques (including basic productivity and efficiency analysis techniques such as Data Envelopment Analysis, Stochastic Frontier Models and more advanced models such as conditional efficiency measures including observed and unobserved heterogeneity).
Each usage case will provide an analytical report or journal paper and a full documentation of the methods, including links to relevant resources, documentation of the specific application to RISIS data and scripts required for that purposes.

Task 2 – Data quality and data preparation issues (M1-M48)
Involved partners: Sapienza, USI.
This transversal task for all cases studies will develop methods in order to analyse data quality and to lower barriers for users of RISIS datasets. This will involve:
- A basic data quality package to be implemented within the RISIS datasets, dealing with consistency issues, identification of deviant cases to detect data issues. The basic data quality package will be based on the experiences developed in RISIS 1 and in the ETER project.
- More advanced data quality and imputation techniques in order to address problems that are specific to each usage case, like missing cases for key variables or influential cases. This activity will be proposed by Sapienza within the

training on Data Quality in which participants will be asked to bring their own data and will experience the techniques proposed in the training on their own data. The training on Data Quality will be organized by Sapienza in Month 18. The Basic Data Quality Package will be made publicly available (including the feedbacks from the training activity in Data Quality) on the webpage of the project in Month 24.
- Data preparation for users. This task will involve detecting all data preparation issues required for statistical user of RISIS datasets and then implementing them in the RCF (for example through suitable scripts) to lower the barrier for data usage.

Task 3 – Implementing methodological user support within the D4Science infrastructure (M12-M48)
Involved partners: USI, POLIMI, Sapienza, AIT.
This task will involve the implementation of methodological support within the D4Science infrastructure by providing a mix of suitable mix of distant and presence support. Distant support will take the form of on-like services for data extraction, preparation and, possibly, also analysis, as well as also availability of tutorials and on-line materials, while presence support will include training activities (in presence and at distance) and dissemination activities like workshops and conference presentations.

Task 4 – Establishing a methodological support unit and developing additional case studies (M18-M48)
Involved partners: USI, POLIMI, Sapienza, AIT.
This activity will leverage on the experience of the first batch of case studies in order to create within RISIS a methodological support unit. The unit will be constituted by a people from the involved partners in this WP that own competences on the methods, on the RISIS data and on statistical software for usage cases to help
potential users for:
- Identifying the most suitable methods adapted to the characteristics of RISIS data and to their research question.
- Accessing and preparing the data for usage (in close connection with task 3.3).
- Developing new usage cases around additional methods and questions, including specifically also inferential models for social network analysis.
This activity will be organized through a first workshop at M18 in order to identify the interest of users and tentative usage cases, followed by a call for interest directed specifically to people outside RISIS. A second workshop will be organized at M24 in order to launch the new batch of case studies and to learn for the first batch how to implement them. Two further workshops will be organized at M36 for an interim assessment and at M46 for a final assessment of the usage cases and the user involvement. Participants for these cases will be taken care individually by a member of the unit for advice, support and implementation of analyses. Participants will be reimbursed for their expenses, while the implementation of the cases and related activities (training, data cleaning etc.) will be taken care by the methodological support
unit. In principle, the second batch of cases is expected to produce the same output as the first one.

Participation per Partner

Partner number and short name	WP7 effort
2 - AIT	6.00
3 - POLIMI	12.00
8 - USI	18.00
13 - SAPIENZA	20.00
Total	56.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D7.1	Data quality package report	13 - SAPIENZA	Report	Public	24

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D7.2	Summary report on the first batch of usage cases	8 - USI	Report	Public	24
D7.3	Policy brief on advanced usage cases and their policy relevance	8 - USI	Report	Public	36
D7.4	Final report on usage cases and advanced communities	8 - USI	Report	Public	42

Description of deliverables

Methodes to the users' community

D7.1 : Data quality package report [24]

Data quality package report

D7.2 : Summary report on the first batch of usage cases [24]

Summary report on the first batch of usage cases

D7.3 : Policy brief on advanced usage cases and their policy relevance [36]

Policy brief on advanced usage cases and their policy relevance

D7.4 : Final report on usage cases and advanced communities [42]

Final report on usage cases and advanced communities

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS23	First batch of usage cases successfully implemented	8 - USI	18	All documentation online
MS36	Methodological support unit operational and second batch of usage cases launched	8 - USI	24	List of cases and conditions of use of the unit online
MS37	First release of the D4Science methodological support infrastructure	4 - CNR	24	The advanced quantitative methods VRE operational
MS65	Second release of the D4Science methodological support infrastructure	4 - CNR	42	The VRE incorporates all documentation on second batch of methods

Work package number [9]	WP8	Lead beneficiary [10]	2 - AIT
Work package title	Transnational access		
Start month	1	End month	48

Objectives

WP8 will comprehensively deal with the development, organisation and implementation of Transnational Access of users to the different infrastructures (datasets) of RISIS. Access includes both conventional (onsite) and virtual transnational access to 16 distributed infrastructures that are described in the following:
Provision of access to the following infrastructure(s) (in alphabetical order):

1) Name of the infrastructure: CHEETAH
Location (town, country) of the infrastructure: Milan, Italy
Web site address: http://www.dig.polimi.it
Annual operating costs (excl. investment costs) of the infrastructure (€): ~ 100k
Description of the infrastructure: Cheetah contains geographical, industry and accounting information on three cohorts of mid-sized firms that experienced fast growth (i.e., average annual sales or employment growth greater than 20% for at least three years) during three growth periods: 2008-2011, 2009-2012 and 2010-2013. Mid-sized firms are defined by combining the EUROSTAT definition of medium-sized firm and the French definition of Entreprise de Taille Intermédiaire (ETI). The uniqueness of this database lies in the focus on this category of firms and its coverage in terms of countries. Cheetah currently includes 42,369 mid-sized firms, which are located in 30 European countries and Israel.
Services currently offered by the infrastructure: Cheetah is a brand new output of the first RISIS project. It is a unique source of data that offers to researchers new opportunities to study the growth dynamics of a relevant category of firms (i.e. mid-sized firms), which however has been almost neglected by the current academic literature on firm growth. The European dimension of Cheetah makes the data infrastructure well suited to address research questions concerning the institutional determinants of firm growth and geographical agglomeration effects.

2) Name of the infrastructure: CinnoB – Corporate Invention and Innovation Boards
Location (town, country) of the infrastructure: Champs sur Marne - France
Web site address: umr-lisis.fr
Annual operating costs (excl. investment costs) of the infrastructure (€): ~100k
Description of the infrastructure: CinnoB (former CIB) provides data and indicators at the firm level for the worldwide largest R&D corporate performers (about 4 000 parent companies and their 350 000 consolidated subsidiaries) that represent over 90% of world corporate industrial R&D. It includes a database of indicators related to R&D and innovation corporate activities at the firm level and, and a register to track back the links between the different sources used to produce the data. The indicators will originate from different resources available in the RISIS consortium: IFRIS-PASTAT, the CWTS Publication Database, EUPRO, the Industrial R&D Scoreboard and Orbis (aggregated data at the parent parent company level). This large set of resources will allow producing data and indicators related to: the corporate knowledge production (Patents/Trademarks/Publications/FP projects), R&D resources in firms and the general activities in firms (financial, sectoral data). Data and indicators will be produced for a period of time of 10 years (2005 -2015) (before and after the crisis). For patents the data will cover the period of time from 1995 to 2015 to provide data for panel analysis. In this sense, CinnoB is used to study many aspects of the dynamics of scientific and technological knowledge creation in large worldwide companies.
Services currently offered by the infrastructure: CinnoB is designed as a robust and user-friendly repository of indicators providing standardised indicators at the firm level (including the parent company and consolidated subsidiaries activities). It will cover many dimensions, i.e. industrial sector, technologies, scientific domains, geography, collaborations or financial information. It will benefit from specific recent developments such as the geocoding of addresses, the identification of key actors, and the tagging of thematic priorities. CinnoB will provide a diverse set of quantitative information to study the dynamics of knowledge production at fine scale in large firms in a given sector, technology or geographical space.

3) Name of the infrastructure: CWTS Publication Database
Location (town, country) of the infrastructure: Leiden, the Netherlands
Web site address: www.cwts.nl and www.leidenranking.com
Annual operating costs (excl. investment costs) of the infrastructure (€): 340k
Description of the infrastructure: The CWTS publication database is a full copy of Web of Science (WoS) dedicated to bibliometric analyses. The enhancements and improvement to the original version of WoS involve: harmonized main organisation names, matching cited references to source publications and some other data quality improvements. In this

database a permanent link is made between author affiliations and OrgReg identifiers. The database enables output and (scientific) impact analyses of any set of publications covered by WoS, using state of the art methods.
Services currently offered by the infrastructure: A public dataset to demonstrate the potential of the database, together with a complete documentation of data and methods used is available at leidenranking.com. In this dataset a set of around 800 universities worldwide is included with their research performance statistics. Through RISIS, a much larger list of organisations will be provided (including PRO's) with their research performance statistics. More details studies can be executed on-site at CWTS using the underlying database via a research visit.

4) Name of the infrastructure: EUPRO
Location (town, country) of the infrastructure: Vienna, Austria
Web site address: ait.ac.at
Annual operating costs (excl. investment costs) of the infrastructure (€): ~ 300k
Description of the infrastructure: EUPRO is a unique dataset providing systematic and standardized information on R&D projects, participants and resulting networks of the EU FP, starting from FP1, and recently integrating H2020 (until 2016), and other European funding instruments, such as EUREKA, COST and selected Joint Technology Initiatives (JTIs). It has been recently used intensively as a core facility in research activities that investigate structure, dynamics and impacts of project-based R&D collaboration, in particular to grasp and understand the development of the European Research Area (ERA). Basically, EUPRO covers cleaned and standardised information on R&D projects (such as project objectives and achievements, project costs, total funding, start and end date, contract type, information on the call), and their participants (standardized name of the participating organisation, organisation type, and geographical location).
Services currently offered by the infrastructure: EUPRO offers the access to high quality data on R&D projects funded by different European funders. The robustness of the dataset and its relevance for actual research issues of the field has been underlined by the number of transnational access requests and visits conducted in the first RISIS project. EUPRO offers substantial possibilities to address questions on structure and dynamics of knowledge creation, networking patterns of organisations in R&D projects, or impacts of publicly funded R&D, among others. Further, with its huge standardised set of organisational names it offers a great possibility to connect with other organisational level datasets, both within and external to RISIS.

5) Name of the infrastructure: IFRIS-PATSTAT
Location (town, country) of the infrastructure: Champs sur Marne, France
Web site address: umr-lisis.fr
Annual operating costs (excl. investment costs) of the infrastructure (€): ~100k
Description of the infrastructure: IFRIS-PATSTAT offers an enriched and cleaned version of the PATSTAT database, a database of all patent offices produced by the European patent office (EPO). In RISIS1, three major developments have been made to improve its relevance. The first one deals with geocoding addresses so as to follow the dynamics of invention. A second important development has been taking into account socalled artificial patents (13% of total priority patents). A third development – still at the pilot phase – deals with the recognition of individuals being legal persons in patent owners.
Services currently offered by the infrastructure: All these developments will be open and accessible. Users can query it along technological dimensions, geographical dimensions (at country and metropolitan levels), along actors and their agglomeration in sectors. Users can also use the sector-technology concordance map developed by Fraunhofer-ISI as a complementary resource to conduct sectoral studies. In addition, firms (and their subsidiaries) from CinnoB (see above) have been identified and agglomerated data at the firm level is integrated into the CinnoB DB. Identification of public actors (from ORGREG) and of firms from the other datasets (VICO and CHEETAH) has been initiated to insure a greater integration potential for RISIS users.

6) Name of the infrastructure: JOREP 2.0
Location (town, country) of the infrastructure: Rome, Italy
Web site address: http://www.ircres.cnr.it
Annual operating costs (excl. investment costs) of the infrastructure (€): 25k
Description of the infrastructure: JoREP is a unique database on European trans-national joint R&D programmes. JOREP stores raw panel data on joint R&D programmes and a basic set of descriptors of agencies participating to the programmes. The actual list of variables was subject to utility and feasibility controls. The current version 2.0, opened in June 2016 (on-site access), covers data for the period 2000-2014, with a specific focus on 2013 and 2014 - assumed as reference years. JOREP 2.0 has a geographical coverage of 32 countries for European-level programmes for the period 2010- 2014 and a table of indicators at the country level to allow spatial analysis.
Services currently offered by the infrastructure: The facility provides a quantitative basis for the monitoring of investments in joint R&D programmes, pointing out the policy rationales behind them and their impact. Each programme is described by a set of attributes (organisational, financial, etc.). The design of the database has been planned to store

data on the amount spent for joint R&D programmes and on their organisational characteristics (managing research funding organisations), according to the analyses of the modes of the ERA dynamics. Combined analyses with EUPRO are currently allowed for JTIs and EUREKA thank to the integration of the two datasets.

7) Name of the infrastructure: MORE
Location (town, country) of the infrastructure: Oslo, Norway.
Web site address: nifu.no
Annual operating costs (excl. investment costs) of the infrastructure (€): ~20k
Description of the infrastructure: MORE (Mobility Survey of the Higher Education Sector) is arguably the most comprehensive empirical study of researcher mobility available. The MORE dataset ('facility') is composed of two independent waves of extensive, Europe-wide surveys in this family: the pan-European MORE2 survey as well as its more limited predecessor, MORE1. MORE1 and MORE2 were carried out on behalf of the EU Commission among large samples of researchers at European universities in 2009 and 2012, respectively. MORE provides a unique lens on mobility patterns and career paths in Europe, including measures of flows of international (and sector) mobility, of factors that influence mobility, and of effects that can be linked to researcher mobility.
Services currently offered by the infrastructure: The MORE facility includes harmonization (e.g. of field of science, career stages) and links the data to the European Tertiary Education Register (RISIS-ETER). The linked information includes variables such as the distribution of staff by field of science, overall funding, and share of foreign students or staff, and provides the scope for in-depth analysis as it allows the research community to control for important institutional-level effects.

8) Name of the infrastructure: NANO ST DB
Location (town, country) of the infrastructure: Champs sur Marne, France
Web site address: umr-lisis.fr
Annual operating costs (excl. investment costs) of the infrastructure (€): -
Description of the infrastructure: The Nano S&T dynamics database (Nano) collects publications and patents between 1991 and 2011 about Nano S&T. One central characteristics of emerging S&T is that they do not correspond to pre-existing categorisations and require the elaboration of semantic based queries. IFRIS has developed a dynamic query gathering 1.18 million Nano-related publications and 735000 priority patents.
Services currently offered by the infrastructure: The facility offers harmonised and integrative analysis of Nano S&T dynamics globally, due to four types of enrichments that have been conducted: (i) categorisation and harmonisation of institutional affiliations, (ii) geolocalisation of all authors and inventors; (iii) geographical clustering of S&T activities; and (iv) thematic clustering of S&T activities.

9) Name of the infrastructure: PROFILE
Location (town, country) of the infrastructure: Berlin, Germany
Web site address: www.promovierendenpanel.de
Annual operating costs (excl. investment costs) of the infrastructure (€) : ~10k
Description of the infrastructure: ProFile is a longitudinal study focusing on the situation of doctoral candidates and their postdoctoral professional careers. The study started in 2009 and continued until 2017 when it was dissolved in the successor study Nacaps. Up to date, ProFile and its successor study Nacaps are the only longitudinal studies in Europe containing doctoral students from all scientific disciplines and regular information on the conditions of their doctoral training. ProFile is based on a systematic sample of doctoral candidates at universities and funding organisations in Germany. ProFile has seen visits during RISIS partly from researchers active in collecting data on careers themselves. ProFile and Nacaps thus serve as a beacon for similar data collection endeavours in other countries. The visits advance research on careers of PhDs and help build a network of data collectors on PhD careers in Europe.
Services currently offered by the infrastructure: The ProFile infrastructure gives access to 18 different cleaned and labelled panel surveys. These can be merged for multivariate analysis, event-history modelling etc. The data contain information on a number of topics including careers after conferral of the doctorate, supervision, mobility. A range of standard measures on personality traits, career intentions, satisfaction complete the high quality data. ProFile is the only study, which allows observing the transition from the PhD to a variety of post-doctoral employments including non-academic sectors directly. Being a multi-cohort study, it also enables analysis of changes in doctoral training conditions over time.

10) Name of the infrastructure: RISIS-ETER
Location (town, country) of the infrastructure: Graz, Austria
Web site address: https://risis-eter.orgreg.joanneum.at/search/filtered
Annual operating costs (excl. investment costs) of the infrastructure (€): ~50k
Description of the infrastructure: RISIS-ETER represents an extension of the Tertiary Education Register (ETER; https://www.eter-project.com/) database that is supported by the European Commission through a service contract.

RISIS-ETER provides an environment for enriching ETER with additional data in three respects: integrating additional variables from other RISIS datasets, particularly concerning research output (EU-FP participations from EUPRO, scientific publications from CWTS publications databse, patents from IFRIS-PATSTAT), extending the time coverage of ETER, and extending the geographical scope of ETER beyond Europe. RISIS-ETER is closely integrated and hosted by the same technical infrastructure as the register of public-sector research and higher education institutions OrgReg, but is a distinct database providing statistical data.
Services currently offered by the infrastructure: RISIS-ETER currently represents a unique data source on European HEIs that includes both information on inputs and educational activities (from ETER) and information on research outputs from RISIS datasets, including publications, European projects and patents. Data are available on distant access through the RISIS authentication service and can be downloaded for statistical analysis through a dedicated interface. RISIS-ETER also extends the time coverage of ETER (currently 2011-2014) by integrating the 2008 EUMIDA data.

11) Name of the infrastructure: SIPER
Location (town, country) of the infrastructure: Karlsruhe Germany
Web site address: http://si-per.eu/
Annual operating costs (excl. investment costs) of the infrastructure (€): ~50k
Description of the infrastructure: Science and Innovation Policy Evaluations Repository (SIPER) is a rich and unique database and knowledge source of science and innovation policy evaluations worldwide. Effects and efficiency of science, technology and innovation (STI) policies are typically assessed through a process of evaluation and illustrated in the evaluation reports. SIPER provides on-line access to those STI evaluation reports as well as data on the underlying policy measures at a single location. The database allows targeted searches and access to specific instruments and evaluation design for policy practitioners to learn from. It enables policy learning and academic research by providing an informed analysis of the database contents which allows users to pursue a range of research questions on evaluation approaches (methods, designs, objectives etc.), evaluation uses, impact, connection between characteristics of policy design and evaluation design, etc.
Services currently offered by the infrastructure: SIPER provides completely new repository of evaluations, with a unique characterisation system, offering an accessible database with a public portal to function as a service to the policy makers and academic communities. SIPER has developed a framework to characterise each evaluation report as well as the corresponding policy measure. SIPER focuses on evaluations after the year 2000. It has a global ambition, but starts with a focus on the Organisation for Economic Co-operation and Development (OECD) member countries/regions. SIPER is open to all interested parties, and is being widely used for policy learning and research purposes. Via RISIS, SIPER offers access to specific data no open in the public version.

12) Name of the infrastructure: VICO
Location (town, country) of the infrastructure: Milan, Italy
Web site address: http://www.dig.polimi.it
Annual operating costs (excl. investment costs) of the infrastructure (€): ~ 150k
Description of the infrastructure: VICO contains geographical, industry and accounting information on startups that received at least one venture capital investment in the period 1998-2014. Start-ups have been incorporated from 1988 onwards in seven European countries (Belgium, Finland, France, Germany, Italy, Spain, and the United Kingdom) and Israel. VICO also provides information on venture capital investment deals (e.g. deal date, total amount invested) and investors (e.g. investor type, investor experience, and its geographical location). Its uniqueness lies in the overall number of companies (17,863), the country coverage, and the extent of information gathered, thanks to the combination of multiple secondary data sources, namely Thompson One Private Equity, Zephyr, Crunchbase and Orbis.
Services currently offered by the infrastructure: VICO allows addressing a wide range of research questions concerning the characteristics, the evolution and the role of venture capital in Europe. It has been widely used by a large academic community to study the peculiarities of the European venture capital market and effectiveness of (different forms of) venture capital in supporting the performance of European start-ups. Results from academic research based on VICO data also attracted the attention of policy makers interested in assessing the impact of policy measures aimed at facilitating access to finance to start-ups. Data are currently available on site (at the Department of Management Engineering of POLIMI) for visits. During the first RISIS project, several scholars had the opportunity to access the data and to interact with researchers of POLIMI, which provided advice and customized support.

Infrastructures 1-12 already exist (from developments in RISIS 1); newly developed infrastructures (see WP10) will also be subject to transnational Access. They are listed below with some basic information on their orientation (in alphabetical order):
13) Name of the infrastructure: EFIL – European database of public funding
Location (town, country) of the infrastructure: Rome, Italy
Web site address: https://www.cnr.it/

Annual operating costs (excl. investment costs) of the infrastructure (€): -
Description of the infrastructure: The dataset will comprise systematic information on public R&D funding, shifting attention to research funding organisations (RFOs) and policy instruments in terms of their financial and other characteristics (e.g. type of allocation, competitive orientation). The units of observation will be countries but also relevant subnational spatial entities.
Services currently offered by the infrastructure: It will enable users to investigate public R&D funding in Europe at the level of RFOs and funding instruments in terms of a completely new information basis on the specific characteristics of these RFOs and funding instruments. Main emphasis is put on the financial aspects of funding (funding volume), but also the specific allocation modes, e.g. competitive orientation of the funding. With the integration with other RISIS infrastructures, e.g. at the organisational level of RFOs, specific questions cross-cutting datasets can be explored by users.

14) Name of the infrastructure: ESID – European Social Innovation Database
Location (town, country) of the infrastructure: Glasgow, UK
Web site address: https://esid.manchester.ac.uk/
Annual operating costs (excl. investment costs) of the infrastructure (€): -
Description of the infrastructure: ESID is a database on social innovation projects and actors in the European research area. ESID will collect and classify social innovation projects and actors by means of automated processes using text mining, natural language processing and machine learning, so the database can be sustainable and continue to collect data on projects and actors in future.
Services currently offered by the infrastructure: ESID will offer the following unique advantages: It will be thematically comprehensive and cover the full range of societal grand challenges and key enabling technologies. It will have significantly more recall than existing databases (magnitude of high thousands as opposed to the existing databases ranging between 50 and 1000 projects). With the flexible conceptual structure in terms of the social innovation definition it adopts, the user will also be able to narrow down by adjusting objectives, outputs, types of actors and their interactions. This will allow the user to search projects and actors according to various different elements of social innovation definition. Finally, as it is based on semiautomatic and automatic information retrieval and knowledge discovery techniques, it will be much more sustainable than the existing databases that rely on continued human coding.

15) Name of the infrastructure: PHD – Careers database
Location (town, country) of the infrastructure: Berlin, Germany
Web site address: http://www.dzhw.eu/
Annual operating costs (excl. investment costs) of the infrastructure (€): -
Description of the infrastructure: This dataset will for the first time collect comprehensive and systematic information on non-academic PhD careers, one of the highly debated issues in research and innovation studies. It will make use of different newly available sources (e.g. EUROPASS, EURAXESS, LinkedIn company websites, RPO websites) to gather information on such careers by scraping publicly available CVs, and bringing them to a systematic dataset.
Services currently offered by the infrastructure: It will allow users to investigate patterns of non-academic careers of PhDs (or hybrid careers) and sector mobility of PhDs from a cross-national perspective. With the inter-linking ambition to other RISIS infrastructures (e.g. OrgReg, RISIS-ETER, etc.), the investigation of a number of policy relevant question becomes possible.

16) Name of the infrastructure: Trademarks
Location (town, country) of the infrastructure: Karlsruhe, Germany
Web site address: https://www.isi.fraunhofer.de/
Annual operating costs (excl. investment costs) of the infrastructure (€): -
Description of the infrastructure: This infrastructure will comprise systematic information on European registered trademarks, viewed in the RISIS context as a marker for investigating the role of business models, organisational innovation and other 'non technological' forms of innovation. It will have systematic information on different types of trademarks, and most importantly the organisations (mainly firms) holding trademarks.
Services currently offered by the infrastructure: It will provide a novel resource as an important complementary view on non-technological innovation. With the planned set-up (see WP10 for details), it will also for the first time be possible for users to jointly investigate trademark application patterns in comparison to e.g. patenting behaviours of firms.

Description of work and role of partners

WP8 - Transnational access [Months: 1-48]
AIT, UPEM, POLIMI, CNR, LEIDEN, FRAUNHOFER, USI, DZHW, STRATHCLYDE, NIFU

This WP deals with the centralised and coordinated design and implementation of 'Transnational Access' in RISIS. Under 'Transnational Access' we understand the usage of RISIS infrastructures (datasets) by users located in other countries than the location of the respective accessed infrastructure, under the specific RISIS access rules (see below). We distinguish between :

• 'Physical' Transnational Access, conducted in form of a physical visit of the user to the location of the dataset of interest (see locations above)

• 'Virtual' Transnational Access, enabling the user to access RISIS datasets via online access channels (e.g. via the RISIS Core Facility, other data transmission tools, or even by email)

Transnational Access must be free of charge (which implies re-imbursement of visiting costs in case of physical Transnational Access; note that intra-national access is possible and also free of charge, but reimbursement of travels cannot be granted). Given the experiences from the first RISIS project, we roughly expect a share of about 20% of 'traditional' physical transnational access, and 80% of virtual transnational access, based on previous experiences in user requests.

Transnational Access is to be organised in a way that ensures all necessary logistical (e.g. offices), technological (e.g. computer environment) and supporting (e.g. scientific support, on-the-job training) requirements. Users entering RISIS must be selected, authenticated and monitored. Naturally, this requires stringent organisation and immense coordination efforts, both between RISIS and users, but also between dataset owners, the access coordinators. Therefore, Transnational Access is managed and coordinated centrally, both in technical and substantive terms, by a central RISIS Access coordinator (AIT) and dataset specific Access managers (all partners listed above). The Access coordinator will interact with the Access managers on the coordination and dataset specific access principles, and their proper integration in the RISIS datasets portal.

The latter will constitute the main online architecture for registration and authentication of users, and the evaluation and selection of research proposals in the first half of the project (later to be integrated in the RISIS Core Facility, see WP4). Note that RISIS will start with a refined version of the existing datasets portal in M1, in order to provide the possibility for accessing to existing datasets early on; a new portal incorporating the necessary advanced functionalities will be established until M12 of the project.

To successfully realise the objectives of WP8, the following subtasks have to be conducted:

Task 1 – Sustain and advance the RISIS datasets portal, constituting from M1 on our central web portal for users registering, accreditation, selection and monitoring

Task 2 – Harmonised documentation and metadata of all datasets available

Task 3 – Coordination of access across infrastructure owners (documentation, metadata, requirements of opening, etc.)

Task 4 – Coordination of access for users (selection of users incl. interaction with project review board, user guidelines, response to user problems/questions, etc.)

Task 5 – Monitoring and evaluation of projects conducted, and user feedback

Modality of access under this proposal:

The modality of access is equally organised across the different infrastructures listed above, and will be managed and coordinated by the central Access coordinator. We fully open all RISIS infrastructures listed above to all users that have a clear and significant research objective. The latter must be demonstrated in a short research proposal (up to 200 words) and a clarification which datasets are needed to address the objective.

This is referred to as 'controlled access'. The mechanisms of controlled access together with the RISIS Code of Conduct (already established in the first RISIS project) will provide the legal basis for access. Note that we offer a number of 'public resources' that researchers, once they register, access without further control.

This is the case today for parts of RISIS-ETER including the integrating services offered for actor delineation (ORGREG), and for SIPER. However, also for these two infrastructures, 'controlled' access will be organised when users want to e.g. draw on underlying metadata or additional, non-public parts.

Technically, access will be organized via the RISIS2 datasets portal, an online architecture for registration and authentication of users, and the evaluation and selection of research proposals (as an improved version of the existing RISIS1 dataset portal). It also comprises complete documentation, exemplary use cases, and metadata for each infrastructure. During the course of the project, the RISIS datasets portal will be integrated in the RISIS Core Facility (see WP4).

Access will have two modes, physical or virtual (see above). Both types require respective preparation activities by the respective team responsible for an infrastructure. Physical visits do not only require preparation, but also on-site support, and organisation of logistics (office, computer, etc.). Per unit of Access, we assume fixed costs, calculating 2.5 days for preparation per access, both for physical visits and virtual transnational visits; in case of physical visits, 3 days re-imbursement of travel costs for the user (max 750EUR per visit) are foreseen.

Our ambition is to largely increase the overall number of transnational visits compared to the first RISIS project, moving from 100 transnational visits in four years to around 400 visits. The unit of access offered per infrastructure, and resulting expected costs are as follows (in alphabetical order, distinguished between existing and new datasets):
Cheetah : Dataset Coordinator POLIMI, Units of access 40, Costs (PM)** 4,55, Physical visits Reimbursement** 5.600€
CinnoB : Dataset Coordinator UPEM, Units of access 30, Costs (PM)** 3,4, Physical visits Reimbursement** 4.200€
CWTS Publication : Dataset Coordinator LEIDEN, Units of access 40, Costs (PM)** 4,6, Physical visits Reimbursement** 5.600€
EUPRO : Dataset Coordinator AIT, Units of access 40, Costs (PM)** 4,6, Physical visits Reimbursement** 5.600€
IFRIS-Patstat : Dataset Coordinator UPEM, Units of access 30, Costs (PM)** 3,4, Physical visits Reimbursement** 4.200€
JOREP : Dataset Coordinator CNR, Units of access 10, Costs (PM)** 1,15, Physical visits Reimbursement** 1.400€
More : Dataset Coordinator NIFU, Units of access 20, Costs (PM)** 2,3, Physical visits Reimbursement** 2.800€
Nano : Dataset Coordinator UPEM, Units of access 10, Costs (PM)** 1,1, Physical visits Reimbursement** 1.400€
Profile : Dataset Coordinator DZHW, Units of access 20, Costs (PM)** 2,3, Physical visits Reimbursement** 2.800€
RISIS-ETER : Dataset Coordinator USI, Units of access 30, Costs (PM)** 3,4, Physical visits Reimbursement** 5.600€
SIPER : Dataset Coordinator Fraunhofer, Units of access 20, Costs (PM)** 2,3, Physical visits Reimbursement** 2.800€
VICO : Dataset Coordinator POLIMI, Units of access 40, Costs (PM)** 4,55, Physical visits Reimbursement** 5.600€
EFIL* : Dataset Coordinator CNR, Units of access 10, Costs (PM)** 1,15, Physical visits Reimbursement** 1.400€
ESID* : Dataset Coordinator Strathclyde, Units of access 30, Costs (PM)** 3,4, Physical visits Reimbursement** 4.200€
PHD* : Dataset Coordinator DZHW, Units of access 10, Costs (PM)** 1,1, Physical visits Reimbursement** 1.400€
Trademark* : Dataset Coordinator Fraunhofer, Units of access 20, Costs (PM)** 2,3, Physical visits Reimbursement** 2.800€
Total : Units of access 400, Costs (PM)** 45,6, Physical visits Reimbursement** 57.400€
*Datasets available from the third project year on (see WP10 for details on the development paths)
** based on 2.5 days preparation costs per unit of access, and additional 3 days for physical visits, with physical visits being estimated as 20% of total access.
*** in k€, based on up to 700EUR re-imbursement costs per physical access (estimated to be 20% of total units)

Support offered under this proposal
Thanks to the efforts for opening infrastructures in the first RISIS project, all necessary technical and logistic support in the form of respective computing and office environment is already given. Scientific support - both for virtual and on-site access – is provided by senior researchers in the field that have used their data infrastructures for different research purposes, but also by technical assistants.

Outreach to new users
This is essential und therefore in much more detail described in WP3 and WP7. Outreach to new users will be conducted by the main RISIS dissemination and communication channels, in particular the RISIS websites, the training courses offered to users, and intensive presentation at the core conferences of the field, but also of related disciplines.

Review procedure under this proposal
Transnational Access must be considered for users and/or user groups (lead by a ‘user group leader’) that articulate a request to access RISIS datasets and/or platforms in written form, stating the motivation and purposes of the work that is intended to be carried out within RISIS. Selection of such proposals will be done by the dataset owners and the Project Review Board, the selection panel of international experts set up by RISIS, based on criteria like scientific originality, policy relevance and clarity as well as feasibility of the proposal. This process is integrated in the RISIS datasets portal and is overseen and managed by the central Access coordinator (AIT).

A key of organizing Access is finally the monitoring and evaluation of visits. In this sense, the RISIS datasets portal (later integrated in RCF, see WP4) will include an automated monitoring system, keeping track of each researcher, and which datasets have been used in which way. Users will be contacted after 12 months and after 24 months of the access to provide information on dissemination of the research (mainly in form of publications but also other channels). The researcher confirms to cooperate on these matters when registering. We also expect users to engage into the user blog and forum, and to engage in our user return sessions and user participative events set up by WP3.

Participation per Partner

Partner number and short name	WP8 effort
1 - UPEM	0.00
CNRS	6.90
CCIP	7.00
2 - AIT	24.60
3 - POLIMI	17.10
4 - CNR	4.30
5 - LEIDEN	8.60
7 - FRAUNHOFER	6.40
8 - USI	5.40
11 - DZHW	6.40
12 - STRATHCLYDE	6.40
14 - NIFU	4.30
Total	97.40

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D8.1	Harmonized documentation and metadata for access	2 - AIT	Report	Public	6
D8.2	First interim report on Access	2 - AIT	Report	Public	18
D8.3	Updated documentation and metadata for access	2 - AIT	Report	Public	36
D8.4	Final report on Access	2 - AIT	Report	Public	48

Description of deliverables

Documentation on access

D8.1 : Harmonized documentation and metadata for access [6]
Harmonized documentation and metadata for access

D8.2 : First interim report on Access [18]
First interim report on Access

D8.3 : Updated documentation and metadata for access [36]
Updated documentation and metadata for access

D8.4 : Final report on Access [48]
Final report on Access

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS3	Revised Data Access portal for project selection fully operational	2 - AIT	1	Data portal accessible online
MS24	Data portal fully integrated in the RCF	2 - AIT	18	Accessible on the RISIS platform

Work package number [9]	WP9	Lead beneficiary [10]	3 - POLIMI
Work package title	Deepening of Core datasets (JRA)		
Start month	1	End month	48

Objectives

The overall objective of this WP is the development of RISIS core datasets. While maintenance activities of core datasets assure a periodic updating of existing information and indicators, development activities are needed to enlarge and deepen information contained in core datasets, allowing the production of new indicators on firm innovation capabilities, higher education & public research organisations, on S&T outputs and project based R&D collaborations, and on policy learning. This is expected to further increase the attractiveness of the RISIS infrastructure for users in different scholarly communities, as well its relevance for evidence-based policy making.

Description of work and role of partners

WP9 - Deepening of Core datasets (JRA) [Months: 1-48]
POLIMI, UPEM, AIT, CNR, LEIDEN, FRAUNHOFER, UOS, TC CAS
The work focuses on targeted developments concerning 7 core RISIS datasets. They are presented in turn while their development and management will be jointly conducted.

1) EUPRO (Involved partners: AIT, CNR, TC CAS)
The development of EUPRO deals with the inclusion of a complete new module, NATPRO, setting up a systematic information basis on nationally funded R&D projects (and potentially also extended to major regional or other programmes). Set-up as extension of EUPRO, NATPRO follows the same conceptual approach, i.e. its main units of observation will be projects and participating organisations, constituting the main elements to trace organisation-by-organisation networks. As in EUPRO, the main approach of NATPRO will be to scrap data from the publicly available data sources from the web. This will require to identify national research funding organizations (RFOs), to screen the public availability of project data, and to evaluate potential alternatives (e.g. data collection via national contact persons in RFOs or public authorities).
A specific effort will be done to cover new Member States from Eastern European countries. Additional information should cover the topic of the project, and – where applicable – project costs, funding instruments and funding calls.
Task 1.1 – Definition of a conceptual framework for EUPRO development
Task 1.2 – Analysis of the R&D policies in new Member States
Task 1.3 – Identification of main RFOs and investigating data availability country-by-country
Task 1.4 – Data collection of project-based information country-by-country (via web or directly via RFOs)
Task 1.5 – Cleaning, standardization and geolocalisation processing

2) CWTS Publication Database (Involved partners: CWTS)
The CWTS publication database will be enhanced by the following elements.
A first development deals with the classification at multiple levels of publications based on direct citation between them. This will provide a structure of all sciences (as far as covered by WoS) independent from journals in which papers are published, enabling a more in-depth analyses of output and a more accurate normalization of citation-based impact measurement. A second development is dedicated to the disambiguation of publications at author level. An algorithm will identify individuals in publication authors. The outcome of this process will provide proxies of individual researchers enabling analyses based on individuals. A third development deals with the addition of impact indicators based on altmetrics and citations. Publications will be tagged with regard to numbers of times they are mentioned in Twitter, Facebook, LinkedIn as well as in on-line news items, blogs, news and policy documents. A specific effort will be also done, using DOI linked to Crossref, to better capture the role of open science in academic research (through citations made). This is an important aspect to start understanding how open science changes knowledge dynamics. A forth development is the identification of funding organizations based on acknowledgements, in order to understanding the role of project-based funding in the dynamics of science. Data will be linked to the specific organizations in OrgReg. However, there will be an effort in identifying firms in coordination with FirmReg.
Task 2.1 – Definition of a conceptual framework for CWTS Publication Database development
Task 2.2 – Publication-based clustering
Task 2.3 – Disambiguation of authors
Task 2.4 – Building indicators on social media information and open science
Task 2.5 – Identification of funding organizations based on acknowledgements in publications

3) IFRIS-PATSTAT (Involved partner: Fraunhofer and UPEM)
Development of IFRIS-PATSTAT deals with the identification of European academic patents. This will allow drawing a more complete picture of the patent output of universities. The Identification of academic patents will be done by using a name-matching procedure to add university affiliations to the inventor information. Names of scientific authors (research-active university staff) from Scopus will be matched with inventor names in the IFRIS-PATSTAT dataset. In a subsequent phase, an effort will be done to identify patents whose investors are researchers in public research institutes to find out how these are different from universities regarding their commercialization activities.
For each academic patent, separation of the type of the applicant will be also provided (i.e. university, firm, individual). This informs about the ownership of academic patents and its trends across Europe.
Task 3.1 – Definition of a conceptual framework for IFRIS-PATSTAT development
Task 3.2 – Development of the name-matching procedure to add affiliations to inventors
Task 3.3 – Identification of the type of applicant

4) SIPER (Involved partner: Fraunhofer)
A first development effort deals with enlarging country coverage beyond Europe and North America, in order to include America and Asia. This requires to strengthen the already existing partnerships (e.g. with Brazil to cover South America), as well as to develop new partnerships to cover Asian countries. Finally, interactions
with OECD and other international and national bodies to link SIPER to other strategic intelligence databases are also part of the development activities.
A second development element will be the systematic analysis of performance of policy measures based on summary, conclusions, and recommendations of the evaluation reports. To this aim, the approach will be to develop specific ontologies for qualifying the performance associated to the evaluations already in SIPER. CORTEXT will help for the internal analysis of the corpus of evaluations and in the progressive building of a relevant ontology; and GATE will provide for systematic annotation once the ontology will have been stabilised.
Task 4.1 – Definition of a conceptual framework for SIPER development
Task 4.2 – Country coverage enlargement
Task 4.3 – Ontology development and definition of performance measures

5) CIB (Involved partner: UPEM)
The main development related to the large firms' R&D activity is the building of a repository CInnoB (Corporate Innovation Board) providing indicators on corporate R&D and innovation activities at the consolidated firm level for corporate R&D investors included in CIB2. CInnoB aims at being a user-friendly repository providing robust indicators on the R&D firms' activities based on a large set of data with time series from 2005 to 2015. It will rely on data provided in RISIS1 related the firms' activity and will be complemented with data related to their scientific publications, trademarks, participations to FPs and a set of more general information on the firms (sector, locations, financial data). It will provide indicators allowing the study of the firms' activities according to several dimensions: time, technologies, scientific domains, geographical location of the R&D activities, collaborations. CinnoB will also include a register to track back the links between the different sources used to produce the data.
CInnoB will aggregate at the firm's level corporate individual data (such as patents) already provided during RISIS 1 in order to calculate indicators to characterize the R&D activity of the firms and also extend its coverage on the corporate activities, including new data and indicators based on corporate publications, FP participations, trademarks, financial data on the firms. It will include methodological developments from RISIS1 such as the address geocoding, the organization type, the identification of the key actors (firms, universities…) and the tagging of specific thematic priorities (societal challenges and key enabling technologies) to calculate designed indicators according to relevant policy questions.
CInnoB aims at responding to interesting policy questions related to the dynamics of the R&D knowledge production and innovation capacities of the largest R&D corporate investors worldwide. It will allow to study in which environment is the corporate knowledge produced (geography and knowledge capacities of the surrounding area), the level and modes of collaboration the firms engage to produce scientific and technological knowledge.
A beta version of CInnoB (CInnoB1) will be produced at M24 based on the list of firms in CIB2. It will cover the period of time from 2005 to 2015. A limited set of indicators will be calculated for each set of data. A final version of CInnoB will be produced at M42 with a more complete set of indicators.
Task 5.1 – Definition of a conceptual framework for CIB development
Task 5.2 – Collecting the data from the different data sources
Task 5.3 – Setting a first list of indicators for CInnoB (version1)
Task 5.4 – Enlarging the list of indicators for CInnoB (version 2)

6) VICO (Involved partner: POLIMI)

Development of VICO is aimed at taking into account the peculiarities of the European supply of early stage finance, as well as the innovation patterns of start-ups.
The first development deals with investors. This will be conducted by collecting data on the ownership and governance of venture capital investors, with the aim of tracking their evolution across time. Specific attention will be devoted to corporate and governmental investors, by looking, e.g., at their level of autonomy from their parent organizations. This better characterization of investors will be finalized to the creation of an investor register as a section of FirmReg (see WP5). Beside venture capital investors, a specific effort will be devoted to a selective enlargement of investor types, including business angels and popular equity crowdfunding platforms (e.g., Seedrs and Crowdcube).
The second development deals with the characterization of start-ups' patterns of innovation in business models and organizational processes. This will require the development of appropriate ontologies related to business model and organizational process innovations of start-ups operating in selected specific fields (e.g. Internet start-ups, biotech start-ups, Industry 4.0 start-ups) and the test of the use of these ontologies on a sample of start-ups (in close connection with the work on RISIS integrative dimensions, WP5). These alternative innovation metrics will be compared to traditional innovation indicators, such as patents.
Task 6.1 – Definition of a conceptual framework for VICO development
Task 6.2 – Data collection on ownership and governance of venture capital investors
Task 6.3 – Selective enlargement to cover additional entrepreneurial finance channels
Task 6.4 – Characterization of start-ups' patterns of innovation in business models and organizational processes

7) CHEETAH (Involved partners: POLIMI, SPRU)
Development of Cheetah is aimed at extending our understanding on how mid-sized firms grow, as well as their innovation patterns.
The first development deals with fast-growth mid-sized firms' ownership structure and patterns of growth.
Through the collection of data on the ownership and governance of these firms, this extension will allow distinguishing specific categories (family firms, firms backed by private equity investors, and firms invested by large incumbent domestic or foreign firms). Furthermore, the collection of information on M&A activity will allow distinguishing organic and external growth patterns. From a policy perspective, a key issue is indeed to understand whether growth leads to employment creation, or whether it is driven by merge and acquisitions.
A second development focuses on the identification of changes in firms' patterns of innovation in business models and organizational processes. As in the case of VICO, the approach is based on the development of appropriate ontologies (in close connection with the work on RISIS integrative dimensions, WP5) related to business model and organizational process innovations. Careful consideration will be given to sectoral peculiarities of firms. Ontologies will be tested on a subset of fast-growth medium-sized firms, and alternative innovation metrics will be compared to traditional innovation indicators.
Task 7.1 – Definition of a conceptual framework for Cheetah development
Task 7.2 – Data collection on ownership and governance of fast-growth mid-sized firms
Task 7.3 – Data collection on fast-growth mid-sized firms' patterns of growth
Task 7.4 – Characterization of fast-growth mid-sized firms' patterns of innovation in business models and organizational processes

Participation per Partner

Partner number and short name	WP9 effort
1 - UPEM	12.00
CNRS	3.00
CCIP	3.00
2 - AIT	18.00
3 - POLIMI	27.00
4 - CNR	9.00
5 - LEIDEN	18.00
7 - FRAUNHOFER	28.00
10 - UOS	13.00

Partner number and short name	WP9 effort
17 - TC CAS	9.00
Total	140.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D9.1	Consolidated work plan on datasets development	3 - POLIMI	Report	Public	12
D9.2	Interim documentation on new datasets development	3 - POLIMI	Report	Public	30
D9.3	Full documentation on developments made on RISIS core datasets	3 - POLIMI	Report	Public	42
D9.4	Policy briefs on new developments and their policy relevance	3 - POLIMI	Report	Public	42

Description of deliverables

Documentation on datasets development

D9.1 : Consolidated work plan on datasets development [12]
Consolidated work plan on datasets development

D9.2 : Interim documentation on new datasets development [30]
Interim documentation on new datasets development

D9.3 : Full documentation on developments made on RISIS core datasets [42]
Full documentation on developments made on RISIS core datasets

D9.4 : Policy briefs on new developments and their policy relevance [42]
Policy briefs on new developments and their policy relevance

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS14	New dimension on open science of CWTS Publication Database opened	5 - LEIDEN	12	Information available in OrgReg
MS25	NATPRO with pilot countries part of EUPRO	17 - TC CAS	18	NATPRO (pilot) available for access
MS26	Investors data incorporated in VICO	3 - POLIMI	18	VICO investors data available for access

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS38	New dimension on acknowledgements of CWTS Publication Database opened	5 - LEIDEN	24	Information available in OrgReg
MS39	SIPER enlargement on policy performance (version 1)	7 - FRAUNHOFER	24	SIPER with policy performance measures (version 1) available for access
MS40	CinnoB opened (transformation of CIB)	1 - UPEM	24	CinnoB (version 1) available for access
MS48	VICO and Cheetah enlarged versions with innovation patterns (version 1)	3 - POLIMI	30	Data on innovation patterns (version 1) in VICO and Cheetah available for access
MS57	New dimension on social media of CWTS Publication Database opened	5 - LEIDEN	36	Information available in OrgReg
MS58	NATPRO fully opened	17 - TC CAS	36	NATPRO (version 1) available for access
MS66	VICO and Cheetah enlarged versions with innovation patterns (version 2)	3 - POLIMI	42	Data on innovation patterns (version 1) in VICO and Cheetah available for access
MS67	SIPER enlargement on policy performance (version 2)	7 - FRAUNHOFER	42	SIPER with policy performance measures (version 2) available for access
MS68	CinnoB opened (version 2)	1 - UPEM	42	CinnoB (version 2) available for access

Work package number [9]	WP10	Lead beneficiary [10]	5 - LEIDEN
Work package title	RISIS Core datasets developments (JRA)		
Start month	1	End month	48

Objectives

Objective(s)
This work package is dedicated to the development of new datasets. These new datasets provide an extended view on STI issues supporting the production of new policy relevant indicators and are naturally linked to datasets already in RISIS. Moreover they provide new nodes in the network of data to link to future new data and existing external (open) data. The new data regards social innovation projects, trademarks, a European database on public funding at the level of instruments and a dataset on the career information of PhD holders, particularly outside academia. In some more detail, the objectives per dataset are as follows.

ESID: European Social Innovation Dataset
ESID focuses on social innovation projects. A first version (ESID 1.0) already exists that relies on existing data sources. We use it as a seed data to train our machine learning model in understanding the semantic properties of the online presence of our entities. We will subsequently query a targeted number of indirectly relevant data sources that might include valuable information in identifying further entities, such as funding platforms (e.g. Kickstarter), legal entity registers (e.g. OpenCharities and OpenCorporates) and social media (e.g. Facebook) and open web (e.g. by utilising Bing search API). This objective of this data collection is to at least double the number of observations and, with it, improve our machine learning algorithms and develop an automatic process of updating. Furthermore, querying big data sources will enable ESID 2.0 to include data beyond Europe. ESID 1.0 is currently limited to a small number of features/ variables on the characteristics of entities. We will experiment with a number of variables that provide rich information including: organisational structures of the social innovation actors; types of involvement of actors in projects; public support for social innovation projects including crowdfunding and governmental support; and outputs, outreach and impacts of the projects.

R&D Database on public funding instruments
The objective of this new policy learning dataset is to characterize the effective portfolios of instruments of research funding organisations (RFO) in Europe, that allow developing evidence of the use of unstruments, on organisations profiles, and at different levels of policymaking (including regional level). There are two units of observations in the new dataset:
a) Funding instruments: the analysis of funding instruments shall figure out mission/objectives and subjects addressed. This allows for a much more reliable information about the orientation of the public R&D funding toward key objectives emerging at European/global level. Textual information can be generally retrieved from the websites, whilst data on funding are publicly available only in some cases;
b) Organizations: deepening the different functions performed by the RFOs and elements of their organizational forms (e.g. governance and autonomy), allow for a more fine-tuned typology of the intermediary policy level across European countries. This is crucial information to understand the balancing between policy actors at national and sub-national levels, and the relationships they build with performers.

TM Trademarks
The objective is to build a comprehensive dataset on trademarks. We will use the open trademark database by the EUIPO (formerly known as OHIM), where trademark registrations targeting the European jurisdiction are documented. Companies/applicants from any country can use this system to file trademarks in Europe. Absolute numbers have been increasing between 2000 and 2015 from about 60,000 to about 100,000 filings per year. The choice of EUIPO is justified by the limitations in access and content of the world level source (WIPO) and the focus of RISIS on Europe. To insure comparisons our choice is to focus on the other outstanding national trademark system, the USPTO trademark data that is available for bulk downloads and thereby offers a sufficiently large analytical potential.

HR Career Information of PhD holders
The objective is to develop a comprehensive and robust pan-European database on careers of PhDs based on CV information, publication and patent data. This database will start with two PhD cohorts from five participating countries (DE, ES, IL, NO, NL) but will be set up to allow easy expansion to other countries and cohorts.
The dataset will provide information on individual career trajectories and relate this to the PhD's individua characteristics (e.g. gender, knowledge fields of the dissertation, data on the PhD's publications and patents, characteristics of the doctorate granting universities and characteristics of dissertation supervisors (e. g. their publication experience). Validity

checks will be undertaken at different levels to verify the statistical power of the approach. The resulting dataset will be of high scientific value due to the unique combination of information from various sources to analyse individual career-trajectories with a focus on non-academic trajectories. We will design, collect and create a career dataset that can be useful to the research community as well as to advanced user groups to analyse CVs in a safe, effective, repeatable and comparative way. It will build on RISIS1 as well as other work pioneered in Europe to use new data-sources to understand (changing) patterns of knowledge production (e.g. ENMOB, 2004; Sisob, 2014). The effort will also liaise with ongoing standardization efforts within the European Committee for Standardization (eg CEN TC 353) to standardize both the content and the form of European CVs (see The European Learner Mobility project).

Description of work and role of partners

WP10 - RISIS Core datasets developments (JRA) [Months: 1-48]
LEIDEN, CNR, FRAUNHOFER, DZHW, STRATHCLYDE, NIFU, CSIC, SNI, TC CAS
The tasks below are organized by dataset :

Task 1.1 ESID Identification of Social Innovation Projects and Actors: develop the infrastructure to identify social innovation projects from online sources such as funding platforms (Kickstarter), legal entity registers (OpenCharities and OpenCorporates) and social media (Facebook) and open web (utilising Bing search API).
Task 1.2 ESID Characterising Social Innovation Projects and Actors: obtaining rich information on social innovation projects and actors identified (location; relationship to each other; social innovation types; legal entities and organisational structures of actors; types of involvement of actors in projects; funding supports (crowdfunding and governmental support); outputs, outreach and perceived impacts of projects.
Task 1.3 ESID Converting data to time series: convert ESID 1.0 from panel data to time series; create synthetic variables to track the evolution of these entities.
Task 1.4 ESID Integration of the dataset in the RISIS infrastructure: harmonise social innovation actors through OrgReg and FirmReg; utilise RISIS geo-coding tool for localizing projects; integrate the social innovation funding data to the R&D Funding Instruments and EUPRO datasets.

Task 2.1 R&D FIN Conceptual framework and methodology for data collection, based on the experience gained in the EC PREF experimental project and covering both national and regional levels.
Task 2.2 R&D FIN Perimeter: focused on RFOs operating in European countries including Central and Eastern countries. The first phase will focus on RFOS with 'important' funding capabilities, visible in the CWTS public DB in acknowledgement, also identified by the networks of national experts gathered through RISIS1 and PREF to complement information not available on the web. The perimeter will be enlarged in a second phase.
Task 2.3 R&D FIN Data collection: focused on publicly available resources using web-scraping techniques; with a limited survey on the quantitative side when financial data are not disclosed on the website. Text analysis (through RISIS services) will be mobilised to understand policy implementation and mission orientation of R&D funding.
Task 2.4 R&D FIN Data quality check and imputation of missing data: following processes proposed in WP7.
Task 2.5 R&D FIN Integration of the dataset in the RISIS infrastructure: Harmonisation with NATPRO (see WP9) and with ORGREG (see WP6) will be developed ex-ante. The possibility to integrate the CWTS publication data on acknowledgments on funding instruments in publication will be exploited and a further integration with SIPER will be developed as far as instruments went under evaluation processes.

Task 3.1 TM Extraction of trademark data and storage in SQL database including (a) bibliographic information: applicant/owner of the trademark (name as well as address information), date of application, registration, publication, expiry and opposition period, trademark number (ID), information on the representative, type of the trademark (word, figurative, sound, colour, 3D, etc.), language; (b) Classifications: Nice classification (42 classes), class(es) of the figurative elements of marks (Vienna Agreement) as well as sound marks (if possible). Within Nice classes, standardized text describes what the trademarks refer to (list of goods and services).
Task 3.2 TM Data preparation & name harmonization
Task 3.3 TM Matching to other RISIS datasets and integration in the RCF: link applicants to the RISISFirmReg Dataset, the CIB (Corporate Invention Board) data and the Cheetah Database on Fast Growing Medium Sized Firms for combined analyses.
Task 3.4 TM Updates: Annual updates will be done before the end of the project.

Task 4.1 HR Compile generic (country-level) data from official datasets that describe the doctoral population and relate the populate to current employment: National Education statistics, OECD's CDH light for 15 countries (2016), ILOSTAT Employment by Education, Eurostat labor market data. The resulting dataset builds a cross-country doctorate population (following CDH) and indicates employment outcomes for different cohorts.

Task 4.2 HR Data set enrichment of the CDH-plus dataset using more fine-grained data available at the level of five selected countries (DE, ES, IL, NL, NO), with the aim to allow for easy expansion to other countries and cohorts. This involves collecting information from PhD dissertation archives and registers for two cohorts (title of the dissertation, name of the degree granting institution and of the thesis supervisors); positioning of dissertations in the map of science using the CWTS algorithm and in the university patenting sub DB with the help from Fraunhofer (see WP9). This task also includes a privacy concept for data usage incl. the development of anonymization and aggregation routines to be applied to the final dataset.
Task 4.3 Dataset augmentation, exploring non-academic career trajectories: collect CV data for the two cohorts from the web at large (ResearchGate, Academia.edu, LinkedIn, Europass, OpenAire) to construct career trajectories. Combined with dissertation data from task 4.2 this will produce data regarding career pathways of academic, hybrid and non-academic careers for different fields of science.
Task 4.4 HR Further development of the wiki. Opening and maintening the semantic wiki from RISIS1, which is linking theory and measurement concepts to studies and available data in the field of research on research careers. The wiki provides an overview of data available in the field of research careers (by country, career stage, data type etc.) also for cross-sectional studies. The wiki will be opened for community editing under the moderation of DZHW. This will aid identifying gaps in the data landscape and point to potential for improvement. The wiki will assist in building a network of data users and producers in the field of research on research careers.

Participation per Partner

Partner number and short name	WP10 effort
4 - CNR	18.00
5 - LEIDEN	12.00
7 - FRAUNHOFER	16.00
11 - DZHW	16.00
12 - STRATHCLYDE	24.00
14 - NIFU	16.00
15 - CSIC	12.00
16 - SNI	12.00
17 - TC CAS	3.00
Total	129.00

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D10.1	Opening of HR wiki	11 - DZHW	Websites, patents filling, etc.	Public	6
D10.2	Draft technical- method report of new datasets	5 - LEIDEN	Report	Confidential, only for members of the consortium (including the Commission Services)	12

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D10.3	Manual & Documentation of new datasets – first version	5 - LEIDEN	Report	Public	24
D10.4	Report of the handling of privacy aspects for the HR dataset	11 - DZHW	Report	Public	30
D10.5	Manual & Documentation of new datasets final version	5 - LEIDEN	Report	Public	36
D10.6	Policy brief on new datasets and their policy relevance	5 - LEIDEN	Report	Public	42

Description of deliverables

HR, methodological report or technical report for all 4 datasets

D10.1 : Opening of HR wiki [6]
Opening of HR wiki

D10.2 : Draft technical- method report of new datasets [12]
Draft technical-method report of new datasets

D10.3 : Manual & Documentation of new datasets – first version [24]
Manual & Documentation of new datasets – first version

D10.4 : Report of the handling of privacy aspects for the HR dataset [30]
Report of the handling of privacy aspects for the HR dataset

D10.5 : Manual & Documentation of new datasets final version [36]
Manual & Documentation of new datasets final version

D10.6 : Policy brief on new datasets and their policy relevance [42]
Policy brief on new datasets and their policy relevance

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS15	First version of ESID DB on social innovation projects	12 - STRATHCLYDE	12	First version of ESID DB on social innovation projects
MS41	First version of the new dataset on Trademarks	7 - FRAUNHOFER	24	Opening
MS42	First version of the new dataset on R&D Funding instruments	12 - STRATHCLYDE	24	Opening

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS49	Pilot version of the new dataset on PHD non-academic careers	11 - DZHW	30	Opening (for testing)
MS59	ESID V2	12 - STRATHCLYDE	36	Opening
MS60	Trademarks DB V2	7 - FRAUNHOFER	36	Opening
MS69	PhD non academic career: first full version	11 - DZHW	42	Opening
MS74	R&D Funding instruments DB V2	4 - CNR	36	Opening

Work package number [9]	WP11	Lead beneficiary [10]	1 - UPEM
Work package title	Ethics requirements		
Start month	1	End month	48

Objectives

The objective is to ensure compliance with the 'ethics requirements' set out in this work package.

Description of work and role of partners

WP11 - Ethics requirements [Months: 1-48]
UPEM
This work package sets out the 'ethics requirements' that the project must comply with.

List of deliverables

Deliverable Number[14]	Deliverable Title	Lead beneficiary	Type[15]	Dissemination level[16]	Due Date (in months)[17]
D11.1	POPD - Requirement No. 1	1 - UPEM	Ethics	Confidential, only for members of the consortium (including the Commission Services)	1
D11.2	GEN - Requirement No. 2	1 - UPEM	Ethics	Confidential, only for members of the consortium (including the Commission Services)	48

Description of deliverables

The 'ethics requirements' that the project must comply with are included as deliverables in this work package.

D11.1 : POPD - Requirement No. 1 [1]

4.1. The applicant must check if a declaration on compliance and/or authorisation is required under national law for collecting and processing personal data as described in the proposal. If yes, the declaration on compliance and/ or authorisation must be kept on file. 4.2. If no declaration on compliance or authorisation is required under the applicable national law, a statement from the designated Data Protection Officer that all personal data collection and processing will be carried out according to EU and national legislation must be kept on file. 4.4. Detailed information on the procedures for data collection, storage, protection, retention, and destruction, and confirmation that they comply with national and EU legislation must kept on file.

D11.2 : GEN - Requirement No. 2 [48]

A report by the Ethical Committee describing the nature of ethical issues which arose from proposals together with measures taken to address these, and associated ethics approvals and other relevant documentation, must be submitted to the European Commission as a deliverable annually. The Ethics Committee must include an independent individual with a relevant expertise in the ethical requirements of H2020.

Schedule of relevant Milestones

Milestone number[18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification

1.3.4. WT4 List of milestones

Milestone number[18]	Milestone title	WP number[9]	Lead beneficiary	Due Date (in months)[17]	Means of verification
MS1	13 datasets open: 9 core datasets, 4 datasets of interest	WP5	2 - AIT	1	Datasets accessible via RISIS datasets portal
MS2	CORTEXT, ORGREG & geoservices opened	WP6	1 - UPEM	1	CORTEXT, ORGREG & geoservices Accessible online
MS3	Revised Data Access portal for project selection fully operational	WP8	2 - AIT	1	Data portal accessible online
MS4	Preliminary DMP	WP1	1 - UPEM	6	Report available
MS5	First RISIS week	WP1	1 - UPEM	12	All documentation available on line
MS6	Revised website, materials and video	WP2	4 - CNR	12	Altmetric indicators related to the website (& its elements) & of twitter
MS7	Training (first set of courses & tutorials)	WP2	4 - CNR	12	Full course documentation and tutorials on line
MS8	Opening of VRE API	WP4	4 - CNR	12	RISIS VRE accessible
MS9	Updated cheetah 1	WP5	3 - POLIMI	12	Updated database available for Access, reports on maintenance
MS10	Enriched IFRIS PATSTAT	WP5	1 - UPEM	12	Updated database available for Access, reports on maintenance
MS11	1st Annual update of CWTS Pub, EUPRO & SIPER	WP5	2 - AIT	12	Updated database available for Access, reports on maintenance
MS12	Opening of GATE resources	WP6	6 - USFD	12	Accessible online from RISIS portal
MS13	New release of ORGREG with worldwide universities and hospitals	WP6	3 - POLIMI	12	Accessible on CORTEXT Platform
MS14	New dimension on open science of CWTS Publication Database opened	WP9	5 - LEIDEN	12	Information available in OrgReg
MS15	First version of ESID DB on social innovation projects	WP10	12 - STRATHCLYDE	12	First version of ESID DB on social innovation projects
MS16	Full legal expertise conducted	WP1	1 - UPEM	18	Report available

Milestone number[18]	Milestone title	WP number[9]	Lead beneficiary	Due Date (in months)[17]	Means of verification
MS17	First batch of policy briefs & webinars	WP2	4 - CNR	18	Presence on line & downloading measures
MS18	Opening of RCF beta version	WP4	1 - UPEM	18	RCF portal opened
MS19	Opening of tools enabling exploration and integration of Open- AIRE content	WP6	18 - ATHENA RC	18	Accessible online through RCF
MS20	Enriched geo services opened	WP6	1 - UPEM	18	Accessible online through RCF
MS21	Opening of first batch of new visualisation services	WP6	6 - USFD	18	Accessible online through RCF
MS22	First opening of firmreg	WP6	3 - POLIMI	18	Accessible online through RCF
MS23	First batch of usage cases successfully implemented	WP7	8 - USI	18	All documentation online
MS24	Data portal fully integrated in the RCF	WP8	2 - AIT	18	Accessible on the RISIS platform
MS25	NATPRO with pilot countries part of EUPRO	WP9	17 - TC CAS	18	NATPRO (pilot) available for access
MS26	Investors data incorporated in VICO	WP9	3 - POLIMI	18	VICO investors data available for access
MS27	Second RISIS week	WP1	1 - UPEM	24	All documentation available on line
MS28	DMP first full version	WP1	1 - UPEM	24	Report available
MS29	Training (second set of courses & tutorials)	WP2	4 - CNR	24	Full course documentation and tutorials on line
MS30	User assessments & returns: first monitoring	WP2	4 - CNR	24	Blog altmetrics & report of return sessions
MS31	Data services inserted in RCF	WP4	1 - UPEM	24	Services accessible directly through RCF
MS32	Updated VICO 1	WP5	3 - POLIMI	24	Updated database available for Access, reports on maintenance
MS33	2nd Annual update of CWTS Pub, EUPRO & SIPER	WP5	2 - AIT	24	Updated database available for Access, reports on maintenance
MS34	Opening of new NLP and multilingual analysis capacity (V1)	WP6	6 - USFD	24	Accessible online through RCF

Milestone number[18]	Milestone title	WP number[9]	Lead beneficiary	Due Date (in months)[17]	Means of verification
MS35	Opening of VRE for seamless embedding of RISIS outputs in OpenAire	WP6	4 - CNR	24	Accessible online through RCF
MS36	Methodological support unit operational and second batch of usage cases launched	WP7	8 - USI	24	List of cases and conditions of use of the unit online
MS37	First release of the D4Science methodological support infrastructure	WP7	4 - CNR	24	The advanced quantitative methods VRE operational
MS38	New dimension on acknowledgements of CWTS Publication Database opened	WP9	5 - LEIDEN	24	Information available in OrgReg
MS39	SIPER enlargement on policy performance (version 1)	WP9	7 - FRAUNHOFER	24	SIPER with policy performance measures (version 1) available for access
MS40	CinnoB opened (transformation of CIB)	WP9	1 - UPEM	24	CinnoB (version 1) available for access
MS41	First version of the new dataset on Trademarks	WP10	7 - FRAUNHOFER	24	Opening
MS42	First version of the new dataset on R&D Funding instruments	WP10	12 - STRATHCLYDE	24	Opening
MS43	Second batch of policy briefs with participative events & webinars	WP2	4 - CNR	30	Presence on line & downloading measures
MS44	VRE fully enhanced with Open Data datasets	WP4	4 - CNR	30	Through operation of RISIS VRE
MS45	IFRIS Patstat new version	WP5	1 - UPEM	30	Updated database available for Access, reports on maintenance
MS46	Extended topic based annotation services		6 - USFD	30	Accessible online through RCF
MS47	New release of ORGREG with PRO subregister	WP6	3 - POLIMI	30	Accessible online through RCF
MS48	VICO and Cheetah enlarged versions with innovation patterns (version 1)	WP9	3 - POLIMI	30	Data on innovation patterns (version 1) in VICO and Cheetah available for access

Milestone number[18]	Milestone title	WP number[9]	Lead beneficiary	Due Date (in months)[17]	Means of verification
MS49	Pilot version of the new dataset on PHD non-academic careers	WP10	11 - DZHW	30	Opening (for testing)
MS50	Third RISIS week	WP1	1 - UPEM	36	All documentation available on line
MS51	Training (third set of courses & tutorials)	WP2	4 - CNR	36	Full course documentation and tutorials on line
MS52	External data services (CORTEXT, GATE, D4Science) integrated	WP4	1 - UPEM	36	Access through RCF fully completed
MS53	Updated cheetah 2	WP5	3 - POLIMI	36	Updated database available for Access, reports on maintenance
MS54	3rd Annual update of CWTS Pub, EUPRO & SIPER	WP5	2 - AIT	36	Updated database available for Access, reports on maintenance
MS55	FIRMREG first full version opened	WP6	9 - JOANNEUM	36	Accessible online through RCF
MS56	All services fully integrated within the RCF	WP6	6 - USFD	36	Accessible online through RCF
MS57	New dimension on social media of CWTS Publication Database opened	WP9	5 - LEIDEN	36	Information available in OrgReg
MS58	NATPRO fully opened	WP9	17 - TC CAS	36	NATPRO (version 1) available for access
MS59	ESID V2	WP10	12 - STRATHCLYDE	36	Opening
MS60	Trademarks DB V2	WP10	7 - FRAUNHOFER	36	Opening
MS61	DMP second full version	WP1	1 - UPEM	42	Report available
MS62	Third batch of policy briefs with participative events & webinars	WP2	4 - CNR	42	Presence on line & downloading measures
MS63	CIB2 complemented by new version of CinnoB	WP5	1 - UPEM	42	Updated database available for Access, reports on maintenance
MS64	Updated VICO 2	WP5	3 - POLIMI	42	Updated database available for Access, reports on maintenance
MS65	Second release of the D4Science methodological support infrastructure	WP7	4 - CNR	42	The VRE incorporates all documentation on second batch of methods

Milestone number[18]	Milestone title	WP number[9]	Lead beneficiary	Due Date (in months)[17]	Means of verification
MS66	VICO and Cheetah enlarged versions with innovation patterns (version 2)	WP9	3 - POLIMI	42	Data on innovation patterns (version 1) in VICO and Cheetah available for access
MS67	SIPER enlargement on policy performance (version 2)	WP9	7 - FRAUNHOFER	42	SIPER with policy performance measures (version 2) available for access
MS68	CinnoB opened (version 2)	WP9	1 - UPEM	42	CinnoB (version 2) available for access
MS69	PhD non academic career: first full version	WP10	11 - DZHW	42	Opening
MS70	Training (last set of courses & tutorials)	WP2	4 - CNR	46	Results of the monitoring of users' assessments
MS71	User assessments & returns: final monitoring	WP2	4 - CNR	46	Blog altmetrics & report of return sessions
MS72	4th Annual update of CWTS Pub, EUPRO & SIPER	WP5	2 - AIT	46	Updated database available for Access, reports on maintenance
MS73	Fourth RISIS week	WP1	1 - UPEM	46	All documentation available on line
MS74	R&D Funding instruments DB V2	WP10	4 - CNR	36	Opening

1.3.5. WT5 Critical Implementation risks and mitigation actions

Risk number	Description of risk	WP Number	Proposed risk-mitigation measures
1	RCF architecture does not match userdriven objectives (limited)	WP4	A specific coordination task with documentation website, online development forge, project roadmap and task management application should insure direction and limit risks to delays in delivering the RCF
2	RCF services do not open at due date (M18) (possible)	WP8	Mitigation is focused on users that need access to multiple datasets for their projects. The transitory solution developed until RCF opening will be continued.
3	NATPRO extension of EUPRO faces strong data availability issues	WP9	Mitigation has been anticipated for central and Eastern European countries with a dedicated effort with classical approaches.
4	SIPER has difficulty to operationalize the ontology on the performance of policy measures	WP9	Mitigation is about reducing the scope to a subset of policy measures using classical means based on expert analysis
5	VICO and Cheetah face difficulty in gathering automatically data on ‘non technological’ innovation	WP9	The contingency plan is to limit the scope of the activity to a corset of firms and use more traditional inquiry methods
6	The PhD career dataset is not able to retrieve enough CVs dealing with PhD non academic careers	WP10	The proposed approach is to first build an enriched PhD dataset (which should not face any major risk). It will thus offer at least a partial and new dataset to the research community. The second step is following PhD holders on the web: here the risk is that we follow ‘easily’ those with a research career while our core objective is the others. The only contingency plan will be to build a prototype only in those countries and/or fields where we have enough information.

1.3.6. WT6 Summary of project effort in person-months

	WP1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	WP9	WP10	WP11	Total Person/Months per Participant
1 - UPEM	18	0	6	20	12	12	0	0	12	0		80
· CNRS	0	0	0	0	6	0	0	6.90	3	0	0	15.90
· INRA	0	2	0	48	0	6	0	0	0	0	0	56
· CCIP	0	2	0	0	6	6	0	7	3	0	0	24
2 - AIT	2	2	1	10	16	6	6	24.60	18	0		85.60
3 - POLIMI	2	3	1	0	24	10	12	17.10	27	0		96.10
4 - CNR	2	36	1	24	0	24	0	4.30	9	18		118.30
5 - LEIDEN	2	2	1	16	12	10	0	8.60	18	12		81.60
6 - USFD	1.50	2	0.50	14	0	48	0	0	0	0		66
7 - FRAUNHOFER	2	3	1	0	12	0	0	6.40	28	16		68.40
8 - USI	2	2	1	0	0	6	18	5.40	0	0		34.40
9 - JOANNEUM	1	2	0	10	0	12	0	0	0	0		25
10 - UOS	1	1	0	0	0	8	0	0	13	0		23
11 - DZHW	1	1	0	0	0	0	0	6.40	0	16		24.40
12 - STRATHCLYDE	1	1	0	0	0	0	0	6.40	0	24		32.40
13 - SAPIENZA	1	2	0	0	0	0	20	0	0	0		23
14 - NIFU	1	1	0	0	4	0	0	4.30	0	16		26.30
15 - CSIC	1.50	1	0.50	8	0	6	0	0	0	12		29
16 - SNI	1	1	0	0	0	0	0	0	0	12		14
17 - TC CAS	1	1	0	0	0	0	0	0	9	3		14
18 - ATHENA RC	1	1	0	0	0	16	0	0	0	0		18
Total Person/Months	42	66	13	150	92	170	56	97.40	140	129		955.40

1.3.7. WT7 Tentative schedule of project reviews

Review number [19]	Tentative timing	Planned venue of review	Comments, if any
RV1	21	Brussels	

1.3.8. WT8 Summary of transnational / virtual access provision per installation

Access provider short name	Short name of infrastructure	Installation		Installation country code[21]	Type of access[22]	Unit of access	Unit cost (€)	Min. quantity of access to be provided	Access costs[23]		Estimated number of users	Estimated number of projects
		number[20]	Short name						On the basis of UC	As actual costs		
1 - UPEM	Nano	1	3. Nano	FR	TA-ac	Visits		10.0		8250	10	10
1 - UPEM	Patstat	1	2. Patstat	FR	TA-ac	Visits		30.0		25500	30	30
1 - UPEM	CinnoB	1	1. CinnoB	FR	TA-ac	Visits		30.0		25500	30	30
2 - AIT	Eupro	1	4. Eupro	AT	TA-ac	Visits		40.0		40250	40	40
3 - POLIMI	Vico	1	5. Vico	IT	TA-ac	Visits		40.0		34125	40	40
3 - POLIMI	Cheetah	1	6. Cheetah	IT	TA-ac	Visits		40.0		34125	40	40
4 - CNR	Efil	1	8. Efil	IT	TA-ac	Visits		10.0		7906.25	10	10
4 - CNR	Jorep	1	7. Jorep	IT	TA-ac	Visits		10.0		7906.25	10	10
5 - LEIDEN	CWTS	1	9. CWTS	NL	TA-ac	Visits		40.0		34500	40	40
7 - FRAUNHOFER	Trademark	1	10. Trademark	DE	TA-ac	Visits		20.0		19250	20	20
7 - FRAUNHOFER	Siper	1	11. Siper	DE	TA-ac	Visits		20.0		19250	20	20
8 - USI	Eter	1	12. Eter	CH	TA-ac	Visits		30.0		29750	30	30
11 - DZHW	PhD	1	14. PhD	DE	TA-ac	Visits		10.0		6875	10	10
11 - DZHW	Profile	1	13. Profile	DE	TA-ac	Visits		20.0		14375	20	20
12 - STRATHCLYDE	Esid	1	15. Esid	UK	TA-ac	Visits		30.0		23375	30	30
14 - NIFU	More	1	16. More	NO	TA-ac	Visits		20.0		20125	20	20

1. Project number

The project number has been assigned by the Commission as the unique identifier for your project. It cannot be changed. The project number **should appear on each page of the grant agreement preparation documents (part A and part B)** to prevent errors during its handling.

2. Project acronym

Use the project acronym as given in the submitted proposal. It can generally not be changed. The same acronym **should appear on each page of the grant agreement preparation documents (part A and part B)** to prevent errors during its handling.

3. Project title

Use the title (preferably no longer than 200 characters) as indicated in the submitted proposal. Minor corrections are possible if agreed during the preparation of the grant agreement.

4. Starting date

Unless a specific (fixed) starting date is duly justified and agreed upon during the preparation of the Grant Agreement, the project will start on the first day of the month following the entry into force of the Grant Agreement (NB : entry into force = signature by the Commission). Please note that if a fixed starting date is used, you will be required to provide a written justification.

5. Duration

Insert the duration of the project in full months.

6. Call (part) identifier

The Call (part) identifier is the reference number given in the call or part of the call you were addressing, as indicated in the publication of the call in the Official Journal of the European Union. You have to use the identifier given by the Commission in the letter inviting to prepare the grant agreement.

7. Abstract

8. Project Entry Month

The month at which the participant joined the consortium, month 1 marking the start date of the project, and all other start dates being relative to this start date.

9. Work Package number

Work package number: WP1, WP2, WP3, ..., WPn

10. Lead beneficiary

This must be one of the beneficiaries in the grant (not a third party) - Number of the beneficiary leading the work in this work package

11. Person-months per work package

The total number of person-months allocated to each work package.

12. Start month

Relative start date for the work in the specific work packages, month 1 marking the start date of the project, and all other start dates being relative to this start date.

13. End month

Relative end date, month 1 marking the start date of the project, and all end dates being relative to this start date.

14. Deliverable number

Deliverable numbers: D1 - Dn

15. Type

Please indicate the type of the deliverable using one of the following codes:

R	Document, report
DEM	Demonstrator, pilot, prototype
DEC	Websites, patent fillings, videos, etc.
OTHER	
ETHICS	Ethics requirement
ORDP	Open Research Data Pilot

16. Dissemination level

Please indicate the dissemination level using one of the following codes:

PU	Public
CO	Confidential, only for members of the consortium (including the Commission Services)
EU-RES	Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)
EU-CON	Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)
EU-SEC	Classified Information: SECRET UE (Commission Decision 2005/444/EC)

17. Delivery date for Deliverable

Month in which the deliverables will be available, month 1 marking the start date of the project, and all delivery dates being relative to this start date.

18. Milestone number

Milestone number:MS1, MS2, ..., MSn

19. Review number

Review number: RV1, RV2, ..., RVn

20. Installation Number

Number progressively the installations of a same infrastructure. An installation is a part of an infrastructure that could be used independently from the rest.

21. Installation country

Code of the country where the installation is located or IO if the access provider (the beneficiary or linked third party) is an international organization, an ERIC or a similar legal entity.

22. Type of access

VA	if virtual access,
TA-uc	if trans-national access with access costs declared on the basis of unit cost,
TA-ac	if trans-national access with access costs declared as actual costs, and
TA-cb	if trans-national access with access costs declared as a combination of actual costs and costs on the basis of unit cost.

23. Access costs

Cost of the access provided under the project. For virtual access fill only the second column. For trans-national access fill one of the two columns or both according to the way access costs are declared. Trans-national access costs on the basis of unit cost will result from the unit cost by the quantity of access to be provided.

Research Infrastructure for Science and Innovation Studies 2

History of changes

Part A	1.2 List of Beneficiaries: University of Manchester (Partner 8) removed. Number of partner changed.	p.4
Part A	1.3.1 WT1 List of work packages: WP11 Ethic requirements added. Changes in the name of WP1 and WP2 WP3 Preparation of long term sustainability of RISIS added. Number of work packages changed. Person/Month changed in WP1, WP8, WP3	p.5
Part A	1.3.2 WT2 List of deliverables: Number of deliverables changed. D2.1 and D2.4 led by CNR. D1.3, D1.6 and D1.8 go to WP3 Preparation of long term sustainability of RISIS D1.2, D1.4, D1.5, D1.7 deleted	p.6-7
Part A	1.3.3. WT3 Work package descriptions: WP11 Ethic requirements added. D11.1 added WP11 Ethic requirements added. D3.1, D3.2 and D3.3 added	p.10
Part A	1.3.3. WT3 Work package descriptions: WP1 Project management: Person/Month changed for Fraunhofer and UNIMAN removed. 13 Person/Month go to WP3 Preparation of long term sustainability of RISIS. Changes in the description of work D1.3, D1.6 and D1.8 go to WP3 Preparation of long term sustainability of RISIS	p.13
Part A	1.3.3. WT3 Work package descriptions: WP2 Interaction with the Research community, Communication and Dissemination: Person/Month changed for CNR and UNIMAN removed. Changes in the description of work and role of partners. D2.1 and D2.4 led by CNR	p.16-18
Part A	1.3.3. WT3 Work package descriptions: WP4 RISIS Core Facility: Person/Month changed for Fraunhofer and UNIMAN removed.	p.23
Part A	1.3.3. WT3 Work package descriptions: WP5 Core Datasets Maintenance: Person/Month changed for Fraunhofer and UNIMAN removed.	p.26
Part A	1.3.3. WT3 Work package descriptions: WP8 Transnational access: Person/Month changed for Fraunhofer and UNIMAN removed. Changes in the description of work and role of partners (SIPER).	p.37
Part A	1.3.3. WT3 Work package descriptions: WP9 Deepening of Core datasets: Person/Month changed for Fraunhofer and UNIMAN removed. Changes in the description of work and role of partners (SIPER).	p.47-48
Part A	1.3.6. WT6 Summary of project effort in person-months: Person/Month changed for Fraunhofer and CNR. UNIMAN removed. 13 Person/Month from WP1 to WP3	p.62
Part B	3.1.2. The project's 11 work packages section removed as included in Part A	p.42
Part B	3.1.4. RISIS Deliverables section removed as included in Part A	p.45
Part B	3.2.2. Key milestones section removed as included in Part A Table 3.2b: Critical risks for implementation removed as included in Part A	p.46
Part B	Table 3.2c: Summary of trans-national/virtual access provision deleted	p.52
Part B	3.4. Resources to be committed : changes as UNIMAN has been removed Changes in the subcontracting and addition of Table 3.4.b for AIT transnational access	p. 54-56
Part B	4. Members of the consortium: University of Manchester (Partner 8) removed. Changes in Fraunhofer partner. Changes in Third parties sections	p.88-91
Part B	5. Changes in the Ethics sessions	p. 129-132

Table of Contents

1. Excellence

This part is organised in four sections as required. We start with the objectives and relations with the topic of work-programme (1.1). As we ask for an advanced community, we devote the second section to demonstrate our achievements (through the RISIS 1[1] project) that justify we candidate to an advanced community (1.2). We then explain our ambition (1.3), before presenting our concept and methodology (1.4).

1.1. Objectives

Science, Technology and Innovation (STI) policies have been faced with numerous requirements for change since the end of the years 2000. Issues of sustainable development have triggered a debate about the need for transformative policies (e.g. Schot & Steinmueller, 2016). In a rather similar vein, a discussion about the need to renew the role of the state re-emerged recently (Mazzucatto, 2013) and to rethink our approaches to societal challenges (L. Georghiou and the ERA Rationales Report, 2008 and more recently the RISE report, 2018). Other colleagues also insist on the wide-ranging shifts that take place around innovation linked to three major phenomena that deeply question the locus and content of research and innovation policies: the over-dominance of the service economy (and other non-technological forms of innovation, e.g. business models), globalization, the shifting national production structures, and new user-based developments that embed new values into innovation (e.g. sharing economy, crowd sourcing, political consumption, social responsibility) (Laredo, 2017). These new phenomena question our accumulated knowledge base, and first of all, the STI indicators that provide the backbone of most if not all quantitative resource for 'evidence based' policies.

The central developments since the beginning of the 21st century have been to complement statistically based indicators with 'positioning' indicators that keep track of actors, spaces and topics. This has driven to multiple costly developments of targeted datasets that faced difficulty in being maintained and remained *de facto* poorly accessible. This explains why the European Commission 7th Framework Programme funded a first Infrastructure project, RISIS 1 (EC-GA n°313082, 2014-2018), as an experiment to open, integrate and complement existing datasets. This is the success of this first stage that makes us ready to turn into an 'advanced community'.

The central aim of RISIS is to support the production of new policy relevant indicators. It addresses a small but fast growing community, where Europe plays a world pivotal role and which is organised around two learned associations: The European association of indicators designers (ENID) that organises the world central academic conference on indicators; and the European forum for research and innovation policy studies (EUSPRI forum). It addresses more widely researchers in the constitutive disciplines of the community that deal with science and innovation studies (in management, in economics, in sociology and in geography mostly). **The project thus addresses directly and entirely the theme "Research infrastructures for the assessment of science, technology and innovation policies"** by sharing the six central objectives set out by the call.

[1] *We use the term RISIS 1 to speak of the previous project (EC-GA n°313082, 2014-2018), RISIS 2 to speak of specific features of the new project we candidate for, and RISIS at large when we discuss generic features of the activity and of the community.*

1.1.1. Facilitating "transnational access"

The call and the project focus on "Transnational Access" to highly relevant core datasets of the field, and on services that enable their integration for new indicator production. Even though we shall rely more and more on open data, we still depend (and for quite some time) on organised sources that are privately owned (especially for firm and output information). We have thus developed a so-called "hybrid model", where RISIS gives access to 'enriched' versions of datasets only for 'publishable research'. This model requires that we organise transnational access based on peer-reviewed projects proposed by researchers (we call it "controlled access"). RISIS 1 has opened 14 datasets between 2015 and 2017. This opening is based on physical transnational access and thus, *de facto*, covers projects mobilising only one dataset. But during RISIS 1, we have demonstrated that the power of RISIS lies on "problem-based" integration and simultaneous access to multiple datasets. Our central objective, and we consider it as a step change, is to move to "virtual transnational access". This requires the development of a "RISIS Core Facility" (RCF). Though it uses multiple existing bricks, it remains a challenging issue if only to integrate "controlled" access. We plan to have a first version operational within 18 months. This will enable to achieve our ambitious goal of moving from 100 projects/visits over four years (RISIS 1) to 100 projects/visits per year. This is only feasible linked to the combination of the next three objectives.

1.1.2. Enlarge and strengthen the services offered

A first step in building the infrastructure was enabling the access to individual datasets that were not accessible before (or not in the cleaned, enriched and harmonised format proposed). We just mentioned that the power of RISIS as an infrastructure lies in the combined mobilisation of datasets. This is why we have and will spend important efforts developing data integration services. The asymmetric distribution of knowledge production explains two critical data integration services based on "actors" (for us actors are not individuals but organisations, e.g. universities or firms) and "geography" (with a high concentration in a few metropolitan areas worldwide). In RISIS 1 we have developed a first generation that we want to deepen in RISIS 2, especially we wish to extend our actor based resources from Europe to the world level. The growing debates about the directionality of policies (societal challenges and now "missions") push for a third radically new topic-focused data integration service based on ontologies and semantic annotation techniques. Our plan is to open it during the first year and to develop demonstrations to provide incentives for users. In terms of data analysis services, we provide to our users a unique on-line platform for semantic analysis, CORTEXT. This is a clear success of RISIS 1 and a major lever for the type of new approaches and indicators the field asks for and RISIS 2 will support. It has gained recognition well beyond the borders of our research community, both from researchers of neighbouring communities and from policymakers. CORTEXT treats today approximately 200 projects per month, most users undertaking multiple projects. The ambition is to enlarge progressively the user base, and arrive to 1000 different users per year at the end of the project. This will be a major achievement demonstrating RISIS ability to enlarge its user base (see 1.1.4).

1.1.3. "Further integrate and open research data infrastructures in the field of science, technology and innovation (including social innovation)"

We have now a clear view of core RISIS datasets that are focused on firm innovation capabilities, higher education & public research organisations, on S&T outputs and project based R&D collaborations, and on policy learning. An important issue is to maintain them (since they loose relevance within 2-3 years if they are not updated), and we plan targeted deepening (the most important one dealing with project-based collaborations). We also plan to develop new datasets around a limited set of key issues focusing on new innovation processes, on human resources (HR) and on policy learning. First, we focus on a highly debated research and also rising policy issue in HR: patterns and drivers of "non-research careers" of PhD graduates (recent data tend to show that 70% of PhD holders non longer work in research after 7 years). Second, deal-

ing with policy learning, there are numerous debates on funding channels, but we know very little about the variety of individual instruments and their combination, an important issue to rethink "policy mixes". Thirdly and fourthly, we shall develop two new datasets to deepen our knowledge of new forms of innovation: "non-technological" forms (using as a marker to identify them, trademarks) and "social innovation" (as explicitly mentioned by the call, integrating and deploying a pilot database on social innovation projects and their characteristics under development). We expect all to have first experimental opening within two years and we also plan rapid integration of the two datasets on innovation to enlarge actor and topic-based analyses.

1.1.4. Widen the user base

There is a specialised core of our community focused on quantitative approaches, models and indicators that is well aware of RISIS and the possibilities it offers. This base is central for the development of new indicators from our integrated datasets. Our databases have specific features that hinder the application of state of the art quantitative methods, in particular when moving from descriptive to explanatory analysis. This is why we plan important efforts on supporting advanced quantitative treatment capacities, developing a methodological support service accompanied by use cases, dedicated use workshops. A major step forward in promoting advanced user communities is the integration within RISIS of the "virtual research environment" proposed by D4Science, which we expect will have a strong impact on the wider community. It will also provide new exploration capabilities by linking to the world of OpenAIRE resources. To widen our user base, we shall undertake a specific set of awareness raising activities (in most conferences of the field, by developing use cases and tutorials) and we shall reinforce the already very important training effort we have successfully initiated in RISIS1, adding to classical physical courses, online training. We shall also do two specific efforts targeted to young researchers: supporting collective training projects at master level, and entering into specific relations with established PhD programmes. An important indirect aspect of widening our user base is to give a far wider visibility and access to RISIS "products" (articles, presentations, but also teaching material, documentation, use cases, etc.). RISIS is clearly inscribed into the open science movement (we only support "publishable" research): inscribing RISIS 2 in OpenAIRE translates this dual objective, and will no doubt participate to widening our user base.

1.1.5. Foster the innovation role of the infrastructure

Our core stakeholder community is made of policymakers at different levels: European, national and regional, but also at city level and at organisational level, in universities, public research organisations (PROs), civil society organisations and even at firm or profession level. Most of them mobilise publicly available indicators and it is thus an important activity for RISIS 2 to produce policy briefs and to develop interaction channels with stakeholders. The variety of stakeholders and of their interests drives us to develop a specific policy of targeted interactions privileging their collective associations (like EUA for universities). Whenever there are specific demands, we are restricted by our hybrid model, and have to ask the operators of the specific datasets (within their agreements of use with the original private source) to handle them. But more and more, while doing these specific treatments, the different teams mobilise data, tools and services from the RISIS infrastructure. **We thus have agreed on one important principle for the future: RISIS 2 as a consortium should receive a percentage of these contracts to support the overall infrastructure. We call this "indirect contracting".** Framing this has strong legal implications we have to address, but we expect to be able to see it at work in the second half of the project.

1.1.6. Ensure the long term sustainability of the infrastructure

The infrastructure depends mostly on the lasting mobilisation of human resources for the maintenance of our datasets, services and RISIS Core Facility (RCF). One central issue for long-term sustainability

thus lies in the fact that institutes, centres or groups participating to the project are officially tasked by their institutions for doing so (both for datasets and services). During RISIS1, the number of groups with such official tasking has widely increased letting us quite confident about the long-term future of the infrastructure. We also expect that our "indirect contracting" approach will cover in the future a significant part of the infrastructure's "shared running costs". However, our key objective is to go one step further and have this infrastructure established as a European infrastructure of interest in the ESFRI roadmap.

1.2. Project history and dynamics

1.2.1. A field in need of a distributed infrastructure

The field covered is "science and innovation studies". This field has two original characteristics:

a) It is born "interdisciplinary" from its beginning in the years 1960, mixing economics, political science, sociology and management. This is well illustrated by its core journal "Research Policy" which is a rare case of a journal ranked among the best in three different disciplines. At the same time, it explains still today why the field has often difficulties to be "located" in the established disciplinary framework, research groups or centres being often classified either in sociology, in economics or in management.

b) It has had a quantitative arm from the beginning. In fact the "moment" that drove to the development of the field is linked with the emergence of "Science policies" at OECD (with the famous Frascati manual). Similarly the structuration of databases around outputs (publications and patents) has conditioned the emergence of scientometric approaches in the years 1970. Data and indicators have thus been one critical dimension of the research field. It has driven to the development of specific data collection and databases mostly handled by statistical offices (for the input indicators) and through international databases (for the scientific and technological output).

The beginning of the years 2000 drove to questioning both the framework which underlies the construction of S&T indicators – the input-output model as process – and the type and level of data constructed – at the national level and with four types of performing actors (higher education institutions, private companies, public research organizations, government). New questions have boomed: is the delineation of actors at the national level still relevant when multinational firms dominate the economic landscape and when 200 of them represent half of world industrial R&D? Can we consider universities as similar when 200 of them in Europe represent 80% of total university publications in Europe? How to reflect the ways states allocate money and what balance does (or should) exist between core and competitive funding? How to monitor the development of new sciences when frontier research becomes a policy priority and transcends disciplinary borders? Etc. These questions have driven the community to speak of a new type of indicators, positioning indicators (Lepori et al., 2008) focusing on actors and their networks in innovation systems. They have driven public authorities (in particular the EC) to fund on an ad-hoc basis targeted databases, which have remained most of the times experimental. This movement has been reinforced by both the development of open access (multiplying the sources of "publicly available" data) and of increasing computer power (enabling to mobilise software on large corpuses). These elements explain why we applied to the EC-FP7 and developed the RISIS 1 project. We took RISIS 1 as an experiment with two concomitant dimensions: would we be able to open and/or produce datasets that interest the community? Would we be able collectively to make a step change in what is offered and in the ways we act as a 'producing community'? We address these in turn to explain why we consider that we are ready for a step change and move to an advanced community.

1.2.2. RISIS1: Major advances in datasets opened and services offered

We organise our achievements (and limitations) along the four objectives of RISIS1:

a) The first objective of RISIS1 was to professionalise, maintain and open a set of existing datasets along six major themes: firm innovation dynamics, European integration, public research dynamics, technological dynamics, PhD & researchers careers, R&I policy support tools. This was an important task that raised cognitive, technical and legal issues. By the end of 2017, fourteen datasets have been opened, the first since mid-2015 and the last just recently. Most datasets are accessible through physical transnational visits (only one dataset, SIPER, is a "service"). This has enabled us to exceed our objective of 100 projects during RISIS1, with an interesting marker (at the last conference of ENID in Paris 34 out of 136 presentations directly used RISIS1 datasets and/or services). The fourteen datasets are grouped as follows: 4 datasets on firm innovation capabilities, including FIRMREG, the firm register; 3 datasets on scientific and technological outputs, including one dataset to analyse emerging technologies; 2 datasets on European integration; 2 datasets on universities and research organisations; 2 datasets on PhD and researchers; 1 dataset on policy learning dealing with policy evaluations (see box 1 below).

Firm Innovation capabilities	
CIB	Corporate invention board / measures inventive activities of large firms & their degree of internationalisation / covers 2000 world largest firms until the 2008 crisis (CIB1) / enlargement to 4000 firms & post crisis (CIB2) planned for 2018)
VICO	Dataset on high tech start-up supported by venture capital / offers longitudinal data on their key characteristics
Cheetah	Fast growing Midsize firms / new dataset with a first version opened in 2017
FIRMREG	Reference database on private actors, combining the firms from the 3 firm datasets (CIB, VICO and Cheetah) with their linkages, enabling actor-level harmonisation at European level / A pilot version will be opened for test in 2018
Scientific and technological capabilities	
CWTS publication database	Offers access (under condition) to harmonised WOS data underlying the Leiden Ranking (with actor harmonisation and geocoding)
Patstat IFRIS	Enriched Patstat dataset (especially with addresses enrichment & geocoding, treatment of artificial patents, focusing on organisation-based patents) (periodic updating) / accessed presently through CIB
Nano S&T database	Demonstrator dataset on dynamics of emerging technologies / based on publications and patents / focused on organisations & geoclusters
European Integration	
EUPRO	Offers harmonised data of all EC FP projects enabling longitudinal analyses / progressively enlarged to key 'joint programmes' (in particular Eureka and JTI)
JOREP	Provides characteristics of 'joint and open' programmes between EU member states / will be progressively integrated in enlarged EUPRO
Universities and research organisations	
RISIS ETER	Based on ETER Dataset on European universities, updated through direct EC projects / enlarged by integrating previous research versions (EUMIDA) for longitudinal analysis / being integrated in a wider register (see ORGREG below)
ORGREG	Reference database on European public actors: includes ETER with enlarged information on publications, European projects and patents; integrates a new register on European public research organisations (PRO) and university hospitals / fully accessible through CORTEXT (see below)

Research careers	
MORE	Integrates the successive MORE studies (directly funded by the EC) on the mobility of EU researchers / offers anonymised microdata
Profile	Panel based data on German early career researchers / serves as a starting point to develop a PhD career platform (that will integrate data from the numerous national surveys)
Policy learning	
SIPER	Repository of European and other countries (mostly OECD) evaluations of science and innovation policies (in conjunction with the IPP platform of OECD and world bank) / opened as a free access service for researchers and policymakers in November 2016

Box 1: Datasets opened during RISIS1

b) The second objective of RISIS 1 was to develop the ability to integrate datasets to address specific problems. This has driven us to develop collective processes of harmonisation and to develop two modalities of integration centred on actors and spaces. The first focuses on actors, and we have developed two registers on firms and on European public actors. ORGREG has been opened in 2017 as a "common good" accessible to all researchers, and has been implemented in all RISIS datasets, enabling actor-based integration. The second modality focuses on "geography", and in particular metropolitan areas. This has required developing a new CORTEXT geocoding service (focused on large corpuses with millions of addresses to geocode, and based on open source resources) that has been implemented in all relevant RISIS datasets and is now opened as a service associated to CORTEXT (see below). The figure below gives an idea of the intensity of interlinking after harmonisation and implementation of integrating activities.

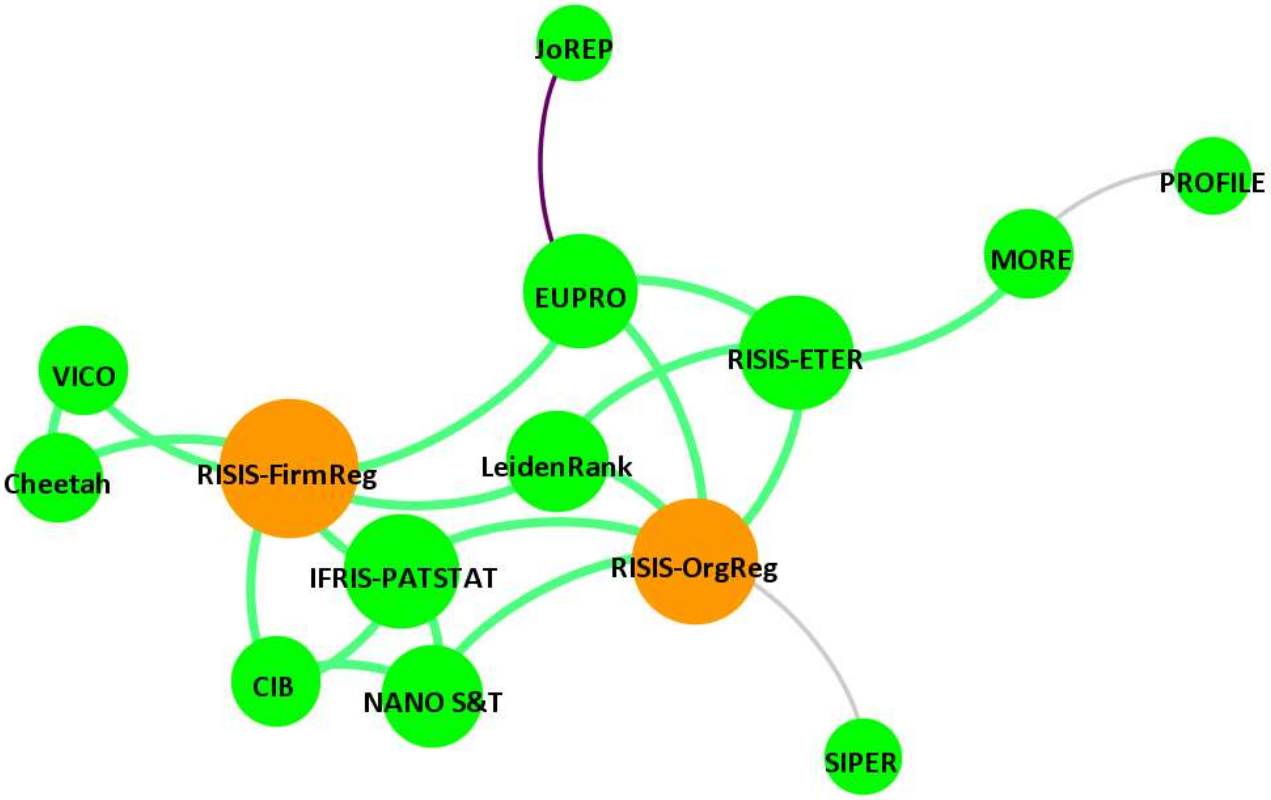


Figure 1: Inter-linking of RISIS datasets after implementation of integration activities (links based on harmonization of actors and/or geography; node size corresponds to the number of links)

In order to show the power of such integration approaches, we have developed demonstrators that have been presented in different conferences and to policymakers (in particular in an internal workshop for EC services) (see box 2)

Box 2 - Demonstrators and their roles: extract of the agenda of the EC-RISIS internal workshop (26/10/2017)

(1) Open, harmonise, enrich and integrate existing RISIS datasets
For this aspect we shall present 2 demonstrators:
- Demonstrator 1 is about how large European firms explore and exploit new knowledge at world level. This is enabled by the combination of 3 datasets dealing with large firms (CIB), with patents (IFRIS Patstat) and with publications (CWTS publication DB) combined with new tools enabling combination.
- Demonstrator 2 deals with the role of European programmes in supporting nano sciences and technologies. This is enabled by the combination of 3 datasets (on publications, patents and European projects, the EUPRO database) supported by the development of a specific nano query and by specific 'geolocation' developments to analyse the situation and networks at the metropolitan level.
(2) Present first results on new datasets enabling to bring new understanding to hot issues
- Demonstrator 3 presents a first exploration of the new RISIS dataset on European fast growing mid-sized firms, opened in 2017.
(3) Demonstrate the new possibilities of connecting RISIS datasets to outside datasets
- Demonstrator 4 compares European and US universities using the university component of the new RISIS dataset on European research organisations ORGREG (an enriched version of ETER including publications, patents and European projects) and the US data set on universities (IPEDS).

c) The third objective of RISIS 1 was to develop platforms and tools to exploit new heterogeneous datasets and explore capacities to integrate open-access datasets. In RISIS 1 the focus has been on semantic analysis. Most corpuses built from the web are textual and require semantic analyses. The CORTEXT platform was developed to fit with the type of semantic analysis needed by field-specific questions. It was opened in 2014 in a beta version and an enlarged more user-friendly version was opened in RISIS in October 2015. The last analytics show that over 200 projects are conducted monthly by different users, half of them French, 25% from other EU countries and the last 25% from all over the world. This is far beyond all expectations (when created the long-term objective was of 100 users per year)[2].

CORTEXT	On-line platform for semantic treatment of large textual corpuses / opened in beta format in 2014, upgraded into RISIS_CORTEXT in 2016 / on average 200 research projects per month in 2016 worldwide / offers new functionalities since the end of 2017 with access to European actors (ORGREG) and to CORTEXT geocoding service (focusing on large datasets)

[2] *We invested in another semantic platform, SMS, dedicated to articulate RISIS datasets to external open-access databases, using RDF techniques. There were two other ambitions: provide RDF based enrichment services, and organise a seamless flow between SMS and CORTEXT. It was a very ambitious objective and we have not been able to insure distant access to SMS nor to connect it to CORTEXT. There was also a gradual shift between the developers of SMS (VU Amsterdam) and the other members about the existence of multiple integrating mechanisms against a unique technical choice based upon RDF (which may be a long term solution but does not apply in present conditions if we wish to develop a new generation of policy relevant indicators) and about how to consider the future RCF between developing it de novo based on SMS or mobilising existing ones (and our present choice of D4Science). This has driven VU Amsterdam to decide to leave the consortium at the end of RISIS1.*

d) The last objective of RISIS 1 was to prepare for further deployment. This latter objective has undergone massive changes since the beginning of the project. We had conceived uses as mostly dealing with one dataset and to be accessed via physical transnational access. We now consider that many uses will require a combined use of our datasets, with aggregations and extractions enabled by our data integration tools. This entails the systematic possibility of access at a distance, in a client-server architecture enabling users to have a workspace within which they can undertake all the data extraction, reconfiguration and analysis they require. This has driven us, in anticipation of RISIS 2, to design a computer architecture of the "RISIS Core Facility" (RCF) and to look for external software sources. This has enabled to identify critical aspects that require tailoring whatever the system we choose. This is in particular the case about the structuration of "controlled access" associated to virtual transnational access, where it is not only an issue of authentication and authorisation, but an issue of automatic "parametrisation" of the elements of datasets required by researcher projects. We are conducting tests about how to proceed under different architectures. This will put us in a position to operationalize the RCF rapidly if we get further support. The inclusion of our field in the call for "advanced communities" is thus a unique opportunity for us to make a step change in the services we offer to our community.

1.2.3. RISIS1 The building of a "producing community"

RISIS 1 started as a collection of datasets that had been developed for research and were both experimental and internal to one given group. The first ambition was thus to turn them into robust, reliable and documented datasets, so that researchers in the field could access them. We chose to do it collectively, meaning that the processes were both defined collectively and implementation was collectively 'controlled'. Doing so took us more time than expected (one and a half year instead of one), but had three important effects: (i) it developed strong links between participants that had seldom shared before on the ways they were building and maintaining datasets; (ii) it produced a first layer of harmonisation of data, and at the same time enabled all to measure the importance of further harmonisation work to be done; and (iii) it developed an organised capacity to integrate more datasets, enabling thus to consider a progressive and regular enlargement of the coverage offered by RISIS as a resource for the community. We have thus opened progressively 14 datasets for the community, fulfilling the central objective of the first project.

This first collective efforts enabled to completely transform the two 'networking activities' we had planned on the harmonisation of actors and geography within all datasets, as a privileged entry to integration. There are important reasons for putting these two aspects as central for all future work on datasets, and future datasets themselves. The asymmetric distribution of knowledge production between actors and between geographical spaces (mostly metropolitan areas) is by far, in our view, the central policy issue in knowledge production. We have developed important efforts to produce new 'data integration services' that are operational and address the constraints of indicator building (in particular reliable longitudinal coverage).

For spatial analysis, we have progressed in steps. The problem is that we have very large datasets to geocode (e.g. over 100 million addresses for patents, evolving at a rate of 5 million a year). We thus considered that we needed a tool tailored for such quantities, with as central objective geocoding at the city or "cluster" level. A first tool was tested, and shared within the project through a dedicated "summer school". This enabled to conceive a full service, built during 2017 as a "common good" and based on open source resources (the software and the documentation are on CORTEXT-Github). The CORTEXT geocoding service is now fully accessible to the community for their own purposes through the RISIS-CORTEXT platform. To convince the community of the power of such integration, we have decided to produce demonstrators on given problems (see above the Nano demonstrator).

For actors, where the asymmetry is very strong (200 large firms represent half of world industrial R&D; the 200 largest European Universities ranked by Leiden represent 80% of total university publications in Eu-

rope), we opted for classical methods for identifying them. We thus engaged in the production of two reference databases, for public research (ORGREG) and for private research (FIRMREG), two developments that were not anticipated in the initial project and thus illustrate very well the emergence of a "producing community". ORGREG is a reference database on universities, public research organisations and research hospitals that enable to identify them in all our datasets (and thus connect different resources: for instance we have added publications, patents and participation to European programmes for all universities). The production process of ORGREG is an illustration of the building of RISIS as a community: we have used the work on ETER to learn how to retrieve universities in our datasets (many of them are present under more than 50 written forms). We have used our ability to identify public research organisations in our datasets to build an initial list of PROs and then ask each of us to check and complement their national situation (as we do not cover the 28 member states, we also have subcontracted colleagues in the other member states as a path for progressive involvement and potential future integration). In order to "industrialize" ORGREG and have a robust database, we reorganised our partnership and resource to include a new partner in charge of this work, Joanneum. Today, as for the RISIS geocoding tool, ORGREG is fully accessible by the community through the RISIS-CORTEXT platform. Again we have used one demonstrator to show the power of this type of integration focusing on universities (see above).

FIRMREG, which is still under development, translates the evolution of one of the core priorities of RISIS, offering renewed views on the innovation dynamics of firms. RISIS started in helping two "experimental" datasets, on world large firms (CIB) and on venture capital backed start-up firms (VICO), to update and enlarge. CIB has been developed to follow the inventive activities of the world largest firms (1000 in a preliminary version, over 2000 in the version opened for users in RISIS 1, 4000 in the new version to open in 2018) while VICO was updated and systematically enlarged to fully cover in Europe the financial developments of most venture capital investors and the firms they support. And we considered it crucial to add a new dataset on fast growing mid-sized firms in Europe (CHEETAH, built in cooperation and opened in 2017). The joint work has progressively driven to a strong evolution in the construction of the three datasets: each has a different entry point but they all evolve now in the same direction: a reference database for the identification of firms at corporate level (on average large firms have over 500 subsidiaries, the largest ones being above 40000), a firm "characterization" database (gathering the financial, HR and market aspects) and a specific database on "innovation" aspects (patents, publications, European projects for the time being, trademarks and innovation types and models in the future round). The on-going development of FIRMREG organises the consolidation of the identification of firms and the linkages between the three datasets. This progressive evolution, not anticipated at all at the start of the project, is for us another sign of the deepening of the community at work to produce datasets useful for the community at large interested by science and innovation studies. Again we have developed one demonstrator to show the power of such type of integration dealing with the exploration and exploitation patterns of large European pharmaceutical and chemical firms (see above).

A third central dimension of RISIS 1 project – beyond opening of datasets and "tools fostering integration", has been associated with the development of a computer-based platform. It translates the hybrid model RISIS is promoting in the ways to change relationships with data in our field. Open access progressively transforms deeply the ways researchers in our field access "sources". The field is borne with the idea that we do not need exhaustive information about science, technology and innovation, but that we need indicators that enable to follow evolutions, changes and transformations. This requires stable longitudinal data, which is an important constraint on the ability to adapt to new questions. For a long time, the only quantitative sources have been produced by surveys, mostly surveys on financial data produced by statistical offices and aggregated at European and international levels (with Eurostat and OECD). Even output data (around publications and patents) are still being treated through simple (but often conflict-laden) measures only. They are however, as most open source data, very rich in text, so that progressively new approaches associated with semantic

analysis have been developed. For a long time, these analyses have been limited by computing power, requiring extensive manual pre-treatments, limiting de facto their scope. RISIS1 decided to support this movement, and has chosen to focus on semantic analysis. CORTEXT is a platform for on-line treatment of large heterogeneous textual corpuses that existed at an experimental stage at the start of RISIS 1. We have pushed its "industrialization" with a real success, well beyond what was anticipated: we now have more projects by researchers from the world every month that what had been planned as a yearly objective (around 200 different projects per month). This has driven us to articulate complementary resources for enriching the types of analysis enabled, adding possibilities for harmonizing actors and geography in semantic treatments. CORTEXT thus operates today as a platform offering researchers different ways to enrich and analyse their corpuses.

Visits, integration capabilities and the success of the CORTEXT approach drove to reconsider our whole approach to opening and services to the research community. More and more we will face researchers willing to extract data from different datasets and combine them (as in our demonstrations). More and more we will face researchers knowledgeable in manipulating data and software, even more so following our intense training efforts. And more and more our "integrating" tools will provide researchers with multiple possibilities of analysis. This means that we have to change our understanding of access associated with visits to move to "generalised distant access". This requires considering a user oriented computer infrastructure which will provide users with a workspace, where they will find all the dataset extracts they require for conducting their project, as well as the shared resources available for enriching and treating. It will also keep the services offered by CORTEXT that enables researchers to bring their own complementary datasets in their working space to combine with RISIS datasets. We thus asked for an amendment of the contract to start reflecting and testing about the future infrastructure. Three aspects have been explored to enable faster deployment in a second round: define with each dataset the options of access, so as to anticipate work about the necessary interfaces; work on user pathways and corresponding dashboards; and define key elements of the future computer infrastructure and organise tests (in particular about 'controlled access' through authentication and project-based authorisations). In our view, all these elements contribute demonstrate that we are ready to move towards an "advanced community".

1.3. RISIS 2 ambition

We consider that we are at a turning point in the field dealing with the articulation between qualitative and quantitative approaches. Before, the classical sequence was to explore qualitatively an emerging issue before being in a position to develop surveys that enable to measure the extent and forms of the phenomenon under study. Now we see a more complex interaction between the two: new conceptual frameworks can be explored and "positioned" through new types of quantitative exploration. This enables to better delineate on what, on/with whom and where we should conduct more in depth analyses, before eventually to conduct targeted surveys. This change is not specific to our field and might turn central for the social sciences in the coming years. But it requires that we develop the adequate core datasets and the adequate data integration and analysis tools. And it also requires that we facilitate "exploration" of the fast growing open source data as a source for new indicators. Our experience on integrating tools shows that their generic forms require active tailoring both to be adapted to field questions and practices, and to be user-friendly for non-specialists. This is why we consider that the idea of generic platforms for the whole of social sciences might not prove that easy to develop and to embed in social sciences, and that we propose a "prototype" of what could be future platforms dedicated to a given field taking into account the central reference datasets of the field as well as its research practices (and thus the specific tailoring of generic tools for them to become both relevant and user-friendly).

We think we are now in a position – mobilising the fast growing generic IT solutions available – to build such a platform for our field. This requires significant investment and builds one core dimension of the project we propose. By adding the topic based data integration tool, we shall offer the core range of integration services needed in our field. We consider that we have already offered the community a new central analysis resource through our CORTEXT semantic platform. We also think that by developing a "methodological support unit" (and its back-grounding resource through documented use cases and advanced workshops), we will strongly improve the capacity of the community to generate new and advanced indicators for STI policies. This finally requires that we are far more proactive during the next years in raising awareness, in building demonstrators and show-cases, in training our colleagues and in deepening our efforts for the next generation (at master and PhD programmes).

If we succeed in adequately deploying these dimensions, the field at large will significantly change at the end and will be on the way to turn into a true advanced research community.

1.4. RISIS 2 "concept" and "methodology"

This section is dedicated to the ways for moving toward our core ambition and achieving our goals. It is organised along the three lines of infrastructure proposals: networking, access and research activities. All activities dedicated to interactions with users and stakeholders (in particular the very important training activities) are dealt with in this section, and they are also an intrinsic component of how we intend to maximise impact (section 2.2).

We start with networking activities that we take as all the activities that deepen and enlarge our interactions with our user community and help us be more integrated as a consortium. The core of networking activities is dedicated to the collective introduction of enhanced and more user-friendly services. They deal first and foremost with the development of an infrastructure supporting generalised virtual transnational access, that we call the RISIS Core Facility (RCF). It will bring a transformative change in our relations with users in two dimensions: one is easier and less time-consuming access, and the other is the possibility to access simultaneously and combine the different datasets of interest for the researcher's project. It is thus transformative both for our relations with researchers, but more widely for the evolution of our community. This is a necessary condition. But it will be successful only if we accompany it with five complementary "services":

a) "Up-to date" and "relevant" core datasets that call for core dataset maintenance and deepening. The former relates for us with networking aspects while the latter is clearly a research dimension.

b) "Data integration services": the three key dimensions for new indicator development lie at actor, geography and topic levels. We want to extend the first two ones already developed in RISIS 1 and develop the third one, ontology building, which is central when "societal issues" and "missions" increasingly direct the framing of policies.

c) "Data analysis services": Here we have three objectives. One is to broaden the range of services offered by reinforcing our successful semantic platform, CORTEXT and integrating the GATE Cloud. The second is to enlarge our visualisation capabilities. The third one is of a different nature and aims at increasing the capacity of use by researchers in our field, of advanced quantitative methods. This is an issue because of the nature of our datasets (strong asymmetries, role of outliers…). The former require mostly IT and software developments, while the latter is mostly an issue of better delineating conditions of use through usage cases and advanced training workshops.

d) Integrating in our services the "virtual research environment" offered by D4Science to promote community building and increasing the capacity of colleagues to explore more widely open source data (in particular OpenAIRE resources) in view of testing new indicators.

e) Pro-active policy on interaction with users and the community at large (in particular through extensive training and participation).

Our approach to access is focused on transnational access, based upon peer-reviewed projects. We have explained the reasons that justify this choice. We also have devised a quite efficient selection process in RISIS 1. Our goals for RISIS 2 are very high compared to the overall size of the specialised community. They are explained by the new targets we follow: in attracting more researchers from the field beyond those already interested (thanks to our communication and extensive training efforts), in relation with master and PhD training, and in attracting researchers from neighbouring fields interested in "science and innovation studies" (in great part because of the added value of our enriched core datasets). The development of the RCF combined with an expected growth of concomitant access to multiple datasets drives us to adopt a centralised management of access.

Finally, we focus research activities on the further development of our core datasets (see above) and on the development of four new targeted datasets that have been carefully selected to be both academically interesting and policy relevant, addressing major transformations in innovation and knowledge dynamics that are critical for the shaping of policies.

1.4.1. RISIS Core facility

To define the nature of the transformation expected, it is useful to start from the present situation. We have a layered access: (i) Present data integration services (for actors and space) and data analysis services (CORTEXT semantic platform) are "open access" once researchers engage in following ethical conditions of use (RISIS Code of conduct); (ii) To access datasets, researchers need to propose projects that are peer reviewed (all this process is handled on-line via the RISIS datasets portal, with a permanent call and an efficient process, usually less than two weeks), and access is organised in the form of visits, dataset per dataset.

We aim at a completely different situation at the end of stage 1 (month 18) of RISIS 2: transnational access will also be "virtual" and enable the simultaneous use of multiple datasets, while we keep the possibility of physical visits for advanced issues requiring direct collaborations. This has three implications that are the objectives of this first phase: (i) Users will have on-line workspaces where they access the elements of the datasets that are needed for their projects. This means the implementation of user interfaces and the implementation of what we call "controlled access" (i.e. access based upon the peer-reviewed selection of researchers' projects). (ii) For this access to work, IT interfaces need to be built between the RCF and the datasets and include the different levels of access that can be granted to researchers (we use the term "parametrization" to qualify this activity). (iii) Interfaces must be developed so that users can mobilise the data integration services (in a first round: the CORTEXT geocoding service, ORGREG and a first round of GATE tools focused on ontology building) and connections have to be developed for linking to CORTEXT semantic analysis.

This will be a beta version, which will be made available on-line to the RISIS community for use and testing, enabling progressive optimisation over the second stage of the project, while incorporating progressive enlargements (in services, in particular visualisation tools and in datasets, especially the opening of new datasets).

The central activity during this second stage will however be focused on the other major new access service we wish to provide our community, access to "Virtual Research Environments" (VRE) enabled by the incorporation of D4Science resources. This incorporation has also another important dimension that is developing interoperability with OpenAIRE infrastructure and datasets.

1.4.2. RISIS core datasets – maintenance and targeted developments

RISIS core datasets correspond to three criteria: (a) They cover at least the EU (not only one country) and long time periods (a critical dimension for indicator building); (b) They represent core research, innovation and/or policy dimensions; and (c) They have generated significant interest from users in RISIS 1 and beyond. Our core datasets have four main focuses:

a) Firm innovation capabilities with our three firm datasets: CIB, VICO and CHEETAH;

b) European Public research and higher education organisations: RISIS-ETER & PRO databases included into ORGREG.

c) Output oriented datasets (with publications – CWTS publication database -, patents – IFRIS Patstat – and projects – EUPRO); and

d) Policy learning (with SIPER, the repository of policy measures and evaluations).

RISIS role is to maintain them and to insure their targeted advancement that will enhance their use and thus provide new services to users. These two dimensions are examined in turn.

RISIS core datasets also include the two actor-based registers of "research performers": ORGREG for public research (higher education institutions, public research organisations, research hospitals, research performing NGOs and Research funding organisations); and FIRMREG. They are key supports for our integration services and their maintenance and development are presented in "data integration services".

RISIS core datasets are complemented by datasets of interest for which we offer access, once they have fulfilled our quality insurance and harmonisation efforts. But we do not update them. This is the case for the German PhD career database (Profile), for the database integrating the MORE surveys on researcher mobility, for the JOREP database on trans-border funding programmes in Europe, and for the Nano S&T database, which has been developed as a demonstrator for approaches and tools to analyse emerging trends in S&T. It is part of the role of RISIS 2 to continue incorporating such datasets and providing access to them. This requires that we facilitate their 'problem based' integration with other datasets, aligning them with our data integration services. This is presented in the third section.

Finally an important aspect deals with the integration of the datasets in the RISIS Core Facility (see point 1.4.1 above). Together they build what we call the RISIS datastore. This requires important work to insure that we can handle smoothly "controlled access" for the virtual transnational visits, enabling our users to find in their workspace all the extracts they require from the different RISIS datasets for undertaking their projects. This relationship between the RCF and each dataset is handled both at RCF level, but it also requires that each dataset has an 'IT integration engineer' that is able to interact with the RCF team and do the necessary developments at the dataset end.

Maintenance of existing core datasets: All datasets in social science require important maintenance efforts to keep their relevance, mostly in terms of periodic updating of their information. Maintenance is thus an issue of periodicity of updating, and incorporation of changes having happened compared during the period (one being mergers and acquisitions, which, for instance, are far more important in the public sphere that what we anticipated at the beginning of RISIS 1). We face different situations depending upon the size of

databases and the extent of changes in past information. For SIPER updates concern the inclusion of new evaluations: this is time consuming because of the need to create the relevant metadata for each evaluation. It is de facto a continuous process:

- For the CWTS publication database, periodicity concerns mostly the inclusion of new articles. Changes due to mergers and acquisitions mostly concern universities and can be dealt with separately. This explains why a yearly periodicity is favoured for indicator-based analyses. The same applies for EUPRO where the incorporation of new projects is done on an annual basis, while, as with CWTS publication database, changes in organisations included, can be tracked separately.
- On the contrary every update of the IFRIS Patstat database requires reconsidering previous patents (in particular because applicants can stop maintaining them, and because changes of ownership play an important role in the life of patents, and enable to follow mergers and acquisitions, that otherwise are not feasible to follow on their own). So the choice has been to do it only periodically (since it takes near to 1 year to reconstruct the enriched dataset). We introduce in 2018 a new version and plan only another one in 2021.
- It is even more complex for CIB, the database on large firms, due both to the importance of mergers and acquisitions, but also of sales / acquisitions of business lines between large firms. This is why CIB has been built on a "retrospective" approach (building a profile at a given date and looking at the progressive construction of firm knowledge base, see Laurens et al. 2015 in Research Policy for a full explanation). Each update of CIB is thus a full reconstruction of the dataset. In RISIS 1 we opened the first version in 2015 and decided of a major update only at the end of RISIS 1, when the full potential of use of the dataset was demonstrated: it took more than one year to conduct it and it will de facto open at the very end of RISIS 1. We thus intend to only do one update during RISIS 2, in the third year, while conducting a large transformation of CIB into the Corporate Innovation Board (see below in developments).
- VICO and CHEETAH face intermediary situations, where updates of information for firms incorporated in the dataset can be done yearly, as can be done the incorporation of new firms, whether through screening the different sources (VICO) or introducing new cohorts (CHEETAH). What is more complex is to follow survival and dynamics of existing firms (e.g. IPOs or being bought by other firms): such treatments will be done with a different periodicity (every 2 years).

New Developments within the existing core datasets: New services require that beyond maintenance we develop new dimensions within our datasets. This was the case in RISIS 1 by augmenting data with geographical coordinates to enable flexible comparisons at the level of metropolitan areas (or even creative cities) and for incorporating our organisations registers (ORGREG for public research and FIRMREG for firms). We still have important efforts to make in order to deploy FIRMREG in all our datasets; but the core of transversal efforts will focus on "topic" focused integration and the articulation with "ontology" tools (see section on data integration services). In the following paragraphs, we explain the 'targeted' developments we plan and the reasons that justify them. We anticipate:

- A major effort to enlarge our core dataset on project funding from European Commission level to national level making of EUPRO a coordinated set of project-based datasets
- Targeted extensions for our two output oriented datasets: CWTS publication database and IFRIS-Patstat
- Adding a new functionality to SIPER on the performance of policy instruments, and developing specific alliances (in America and in Asia) to make it global
- Structuration efforts turning CIB into the Corporate Innovation board offering multiple indicators organised at the firm level and structured into panel data
- Enlargements of VICO and CHEETAH focused on patterns of innovation and governance of firms (ownership and investors).

EUPRO: The EUPRO dataset was before RISIS 1 only focused on projects funded by the Framework Programme (offering an actor-harmonised database of participants enabling robust network-based analyses of European integration dynamics (FP-PRO). RISIS 1 has enabled to enlarge it to other European-level efforts, based on the JOREP database (in particular EUREKA and JTIs). This provides a more encompassing view of European integration dynamics at organisational level. To better position such dynamics and study complementarities between project-based funding mechanisms, we consider it central to enlarge it to national project-based funding. This is made feasible by the efforts done by an increasing number of research funding agencies in Europe to provide online information. This will require an important effort, and as we wish to cover from the start new Member States, we have added a new partner from the Czech Republic with a deep knowledge about R&I policies of new member states. We also aim at progressively integrating major regional schemes and major NGO funding research (such as the Welcome Trust or the Axa foundation). At the end of RISIS 2 we should thus have an advanced EUPRO composed of three components: ECPRO (with the different initiatives at European level: FP, Eureka, JTI), NATPRO (with national research funding agencies) and OTHERPRO (still at a limited stage with major regional funding agencies and NGOs).

CWTS publication Database: The CWTS publication database offers access to Web of Science publications with a refined thematic decomposition (4000 themes). This thematic decomposition is a unique resource to link publications to societal issues, directly through ontology building (see corresponding WP) and indirectly through citations in social media (see below). It also has deployed an important effort in harmonising universities, PROs and clinical research organisations. And there will be a shared effort in identifying firms in coordination with FIRMREG and other firm datasets. The effort on acknowledgements as a source for understanding the role of project-based funding in the dynamics of science will be fostered in coordination with ORGREG, and a specific effort will be done, using DOI, to better capture the role of open science in academic research (through citations made). This is an important aspect to start understanding how open science changes knowledge dynamics. Finally, there is a growing interest in disambiguation at author level to promote analyses about knowledge circulation through social media or involvement of researchers in social media. This last effort will require that we are very cautious for privacy reasons, about access and use (only at adequate aggregate levels) of such developments.

Patstat-IFRIS: Major efforts have been conducted in RISIS 1 to enrich the Patstat database, dealing with priority patents: reincorporating so-called artificial patents (13% of the database), vastly improving the coverage of addresses and geocoding them systematically, delineating individuals from firms to better analyse co-patenting, and systematising the semantic use of IPC classes (patents are unique by the richness of their classifications – around 75.000 – but numbers need to be translated into written language, a non-trivial development). This is why we shall focus on only one further development. There is a strong policy interest in university patenting as an engagement into the application of knowledge. However recent work done in Germany (Dornbusch et al 2013; Dornbusch, Neuhäusler 2015) has shown that patents held by universities represent probably just half of total university based patenting activity (mostly because of patents filed by firms or public research organisations with university inventors). A quite powerful method developed by Fraunhofer, based on inventor-author delineation (see also CWTS publication database) has proven performative to address this issue, and we wish to generalise it to European countries. This will remain an internal process and users will only access aggregated information at organisation level.

SIPER: A first effort deals with enlarging country coverage. We need to go beyond Europe and North America. Within RISIS 1 a first effort has been made with Brazilian colleagues to cover South America (as such it is already an achievement as our Brazilian colleagues are funded by national sources). We intend to develop new partnership to cover key Asian countries (this partnership is essential if only for language reasons). A second development will be the inclusion of data on policy performance. The new repository demonstrates the value of metadata about policy evaluations. In this first version, metadata concerns two

dimensions: characteristics of the related policy, characterisation of the evaluation done. We miss a third dimension to make it a broader policy-learning tool: the analysis of performance of policy measures (based on summary, conclusions, and recommendations of the evaluation reports) will require developing a new approach to coding performance, probably by a systematic use of semantic analysis (which is made easier using RISIS-CORTEXT facility). This builds the second enlargement to be done to be even more relevant for our users. Finally we shall strengthen the integration with other international initiatives, and especially the OECD STI outlook, which is now an online tool, enabling links to SIPER at the level of policy measures.

CIB: CIB is focused on the dynamics of the worldwide largest firms. In RISIS 1 we have made an important effort: (a) To enlarge in a systematic way the coverage (we move from 2.000 to 4.000 firms complementing our use of the EU scoreboard (2.000 firms per year, 3.000 over the period 2008-2017) by the top patenting firms worldwide (based on a systematic analysis of top patenting actors made by WIPO); and (b) To construct their consolidated perimeter (more than 330000 subsidiaries identified). CIB was only concerned by "invention activities", using patents to measure the degree of internationalisation of firms. The integration of CIB within RISIS is changing its nature by enabling articulation with other sources: publications (experiments on pharma, biotech & chemistry firms have shown how productive this is), European Commission's projects, R&D expenditures (at least for the 3000 firms contained in the European Commission scoreboard), plus financial data (following the same rules as in VICO and CHEETAH) and we hope thanks to RISIS 2, trademarks. This integration of new data entails a radical change in the nature and scope of CIB, that will enable all types of panel based treatments. This explains why that we have decided to change the meaning of CIB turning it into the Corporate Innovation Board. The core of the effort is thus to structure CIB along these lines to offer a really enlarged service to users.

VICO: The extensive use of VICO by users in RISIS 1 demonstrates the interest of giving access to a longitudinal, aggregative and validated dataset of start-up firms in Europe. This explains why we plan two new developments to better serve the community. The first one deals with investors: the information of funding sources is for the time being limited to venture funds, while we witness a significant enlargement of other sources (in particular equity crowd funding platforms). Structuring funding actors in an investors register will be the first enlargement of VICO. Secondly, results show that a growing number of start-up firms are marginally linked to new S&T developments. We face here the issue of innovation patterns and the role of new business models. The articulation with the work done on trademarks (see below) is important here, as well as more specific developments based on semantic analysis and machine learning techniques.

CHEETAH: This latter issue of innovations patterns and new business models is even more important for fast growing mid-sized firms. Thus the same approach as for VICO will be used to address this issue. We also are in a very early development of Cheetah with very limited knowledge about ownership (domestic vs. global firms; private vs. public; role of private equity firms) and about patterns of growth (organic vs. acquisitions). This will build a second development so that we can go beyond identifying these firms, and characterise their dynamics and their patterns of innovation.

Integration of datasets of interest: RISIS already integrates four datasets of interest (Profile, More, Nano S&T and Jorep). We need to enlarge this role, and without engaging in the development & maintenance of such datasets, we want to assure users of their robustness and reliability. We also want to insure that they are prepared for our data integration services, especially for actor and space-based analyses. We intend to use the procedure developed in RISIS 1 to insure both aspects, and will engage with the operators of potential datasets of interest to conduct the necessary operations. We also need a clear process for identifying such datasets: this will be the role of the work package team to propose such datasets, and we shall use the normal collective procedures at project level (see Governance structures) for final decision.

1.4.3. RISIS integration services: actors, spaces and topics

One central reason why a new generation of datasets (but also a set of rankings), distinct from statistical sources, has emerged at the turn of the 21st century, links to the very dynamics of knowledge production, its strong asymmetric distribution. 200 large groups represent half of world industrial R&D; in Europe the 200 largest universities (out of more than 2.500) produce 80% of articles published by European universities. Similarly studies dealing with nanotechnologies have shown that 80% of publications and an even greater percentage of patents were concentrated in just over 200 clusters worldwide. These asymmetric distributions call for a new type of dataset that keeps the identity and address of actors, requiring thus to be built on open data. Following actors in different sources (and in particular outputs and collaborations) and geocoding addresses of organisations and their authors and inventors have become two critical resources for providing new policy evidence. This is why RISIS 1 made important efforts to develop a geocoding service that can apply to very large batches and to develop registers that enable to retrieve European public actors and firms throughout our output and project datasets. On both aspects we have gone further in RISIS 1 than what was expected: both the CORTEXT geocoding service and the ORGREG register of universities, PROs and research hospitals covering Europe are on-line. Both are fully open and are free resources for researchers, accessible through the RISIS-CORTEXT platform (see below). Both have also been implemented in our firm, output and project datasets, offering a new unique service to users for actor and/or geography-based analyses.

This provides the basis on which we wish to extend RISIS integration services. In the following paragraph we first describe the developments we intend to do for actors and geography. Then we will explain why and how we wish to offer users a completely new service within RISIS, focused on topics.

Complementing and extending actor registers and geocoding services: The next step in the geocoding services is to develop extensions to enable users to aggregate data bottom-up or at the required level (with specific developments focused on the different approaches for metropolitan areas). These aspects have given rise to initial developments (see CORTEXT Github for the bottom-up approach developed), but require further industrialisation and integration within CORTEXT and the RCF. ORGREG will require periodic maintenance (especially about university mergers). There is still significant improvement to make to the present version about the coverage of research hospitals and the ways to take into consideration university hospitals. But we wish to focus on two major aspects. First, we have identified Public research organisations (PRO) and Research funding organisations (RFO) but we miss the basic characterisation features (as the ones we have for universities, thanks to the incorporation of ETER); And we need to improve the way in which we treat large multi-thematic and multi-located PRO. Second, our ambition for ORGREG (at least for universities) is to extend beyond Europe through a progressive internal effort but also and more and more through coordination with interested partners internationally (as we are starting with Carnegie for US universities). However the core of the work will be to develop the firm register (FIRMREG) that remains at a pilot stage. It requires important methodological improvements and legal clarifications before we can integrate it in our output and project datasets (for the time being only patents are available for firm datasets).

Toward topic based integration services: The directionality of innovation policies and their ability to be problem solving has become a central policy issue that is asking for renewed characterisation of scientific, technological and innovation activities. Within RISIS 1 we did a full experiment of what can be learnt when being able to disentangle an emerging field, using Nano S&T as an example. This requires both relevant queries (the Nano one is a good example of such complexity, which drove to its publication in a high level academic journal, Research Policy) but also "tagging" in the source databases (we discovered that it is far from being trivial). This "local" exercise (which took altogether around 5 years) told that it was necessary to find other resources if we wanted to generalise possibilities for other emerging technologies and even more for societal challenges and "missions" to cope with the issue of directionality. This is why we have integrat-

ed a new partner, the University of Sheffield, which is one of the few European specialists in methods for automatic ontology learning for tagging data items (part of the GATE open-source infrastructure for text analysis and natural language processing, https://gate.ac.uk). They are already participating to a companion project, KNOWMAK, which has made first steps towards tailoring GATE tools and services to the needs of our field, developing ontologies for both key enabling technologies and Societal Challenges (as defined by the EU). This first experiment has been very positive, but it has also demonstrated that research and innovation are a difficult "field" for ontology building that will require both further developments (The Glove environment does not fit well with S&T and we have had to build a dedicated learning one) and a number of experiments before we can deploy a generic "ready to use" service. Our objective in RISIS 2 is thus to very fast integrate GATE tools and services as they currently stand as new services for "advanced" users and, in collaboration with other RISIS 2 partners, to continue with adapting and extending the GATE algorithms and tools to better address key societal challenges, to offer our users more user friendly services and, for some "hot" policy issues, "ready-made" tagging in our core datasets.

Integrating "integrating services" in the RCF: One important work is that all these services become resources that are also available by users in their workspaces. Today we use the CORTEXT platform to provide a virtual access both to ORGREG and to our geocoding services. These need to be included into the overall RCF as well as the connection with and use of GATE. This requires significant integration effort, and our experience from RISIS 1 has demonstrated that it should not be underestimated. There is also an important reason why such RCF integration is needed: even if all these services are openly available online, we consider it critical that our users benefit from tailored virtual research environments, access to aggregated datasets, follow our research methodology, engage with our values, subscribe to our charter, and participate actively in RISIS community in an identifiable way.

1.4.4. Data analysis services

In RISIS 1 we have focused on the key types of data analysis required by this new generation of datasets that are mostly textual. We have thus integrated a unique platform for semantic analysis. CORTEXT is one major success of RISIS 1 offering semantic preparation, analysis and visualisation of large heterogeneous textual corpuses. We anticipated 100 projects a year by 2017, we have had over 200 projects every month, from a far wider remit than the core STI community, for both research and more and more for training, and also for researchers and analysts in stakeholder organisations (we have just organised a master class at OECD for both internal users and for national representatives). Our ambitions for RISIS2 are to deepen this success, firstly by developing further CORTEXT's corpus analysis capabilities, and secondly by widening the types of analysis offered through integration of GATE's document-level natural-language processing tools and services. Ultimately, this will help them jointly to address all aspects of *multi-lingual analysis* and development, a critical issue when dealing with societal issues. One important aspect is linked with the *visualisation* of semantic analyses and we expect a step change in this dimension during RISIS2 by integrating multiple open source visual analytics tools (starting with VosViewer).

A last important aspect of the much needed data analysis services lies in the ability of the RISIS community to use the most advanced quantitative and statistical analysis tools to build a new generation of indicators, since the characteristics of our datasets pose specific and difficult problems to their use. This is addressed in a specific section based upon the development of exemplary cases of use, upon research workshops and online tools to create targeted advanced user communities and the creation of a specific methodological support unit.

1.4.5. Promoting advanced user communities

We offer our community a set of datasets (projects, patents, publications, firms…) with integrative dimensions (actors, spaces and soon topics). But their mobilisation for analysis is not easy in methodological terms. Researchers face three issues that make the application of advanced quantitative methods complex: heterogeneity in the different dimensions considered, dependence on soft characteristics of context, specific difficulties linked to the datasets themselves: missing data, different types of data (e.g. categorical vs continuous), and unusual distribution (lognormal, role of outliers…).

These difficulties were already visible in the first round of user experiences we have had in RISIS 1. We thus think that we need to provide additional support to users that wish to use advanced methods. The ambition is thus to provide researchers with a set of complementary resources. This requires that we build usage cases of the methods selected with resources, documentation and tutorials. These cases will also help us to organise data pre-treatment needed for using these methods (Integration into the RCF lies less in the integration of software as they are mostly generic and distributed, but in the transformation of datasets for being used). These usage cases will support the mobilisation of interested users through "hands-on" workshops that combine questions and methods, and participate in building targeted communities interested in mobilising the new method and infrastructure. The deployment of the D4Science Virtual Research Environment (VRE) will deepen such mobilisation and help the deployment of advanced user communities that accelerate the pace of research and also participate actively in enlarging the user base (here we take it as a major lever for extension in science and innovation studies at large, gathering colleagues from neighbouring fields). This will be complemented by a RISIS methodological service to provide support and guidance to users of these methods (the fact that we rely on transnational access becomes here an advantage to be pro-active with our users).

The plan is to organise two batches (of 24 months each) that address selected methods. We will start with a defined subset of methods, and then organise a consultation for the definition of the two other batches (with the possibility that a method raises such interest that we continue it in follow-up batches). The first batch is focused on organisations, discussing typologies, multi-level embedding and benchmarking issues. This drives us to focus on 5 methods: latent class models, multilevel models, dynamic panel data methods, spatial econometrics (allowing for the estimation of spatial spillovers) and advanced non-parametric models (DEA & similar approaches applied to efficiency and benchmarking). For each an important aspect is to consider specifically issues of data quality and imputation techniques (applied to the case or specific to the datasets used). Doing so we aim at providing a completely new service to our users with, and we anticipate, a strong impact on the robustness and relevance of indicators supporting research and innovations policies.

1.4.6. Training and Interactions with our user communities and key stakeholders

Training: Reinforcing and enlarging our community, i.e. the STI policy field at large, is a critical objective of the project. We target those in our community that are eager to work with advanced new data and with our new data analysis services. We also target and educate those in our community that are still reluctant to use these new quantitative approaches. Thirdly and more generally, we have a specific role in training the young generation of researchers very early on in using quantitative data more creatively. Training in RISIS 1 achieved impressive results in terms of number of scholars attracted and positive assessment received from attendees. These results encourage us to be proactive, and push us to complement short courses and summer schools by multiplying tutorials and on-line courses (especially for the use of datasets or integrating tools). We also expect to put a stronger focus on new methodologies and on the combined use of datasets to address new research questions dealing with issues of policy relevance. Finally we shall more strongly target the young generation entering into relations with specialised masters and PhD programmes to foster the development of collective and/or individual projects mobilising RISIS resources.

Interacting with our research user community: The very good results in training also tell us that we have been too shy in promoting the infrastructure and the training courses. We thus intend to be more present at conferences, and articulate awareness and training events with the major conferences in the field, in particular the STI conferences, as well as reach out to related fields. RISIS 1 also demonstrated that we need to engage users more in the dynamic process of defining, building and using the infrastructure. We thus intend to learn more about our users, not only the projects they develop and the type of use they make of datasets, but also the insertion of such approaches in their overall research efforts, the lessons they derive from use, their needs for improvement in RISIS services, and their willingness to engage. For the latter, we will develop improved interaction channels (a global user forum, specific user blog, specific user events at conferences, dedicated 'programming' workshops, social media presence). We have also seen through our training and visits, the important presence of colleagues from neighbouring fields. Science and Innovation studies at large sit at the encounter of four major disciplines (management, political science, sociology and economics, with more and more connections with geography). In all these fields there are targeted interests for our datasets (especially firm and output datasets). We intend to be more present in their major events so as to raise awareness and use, in particular advanced uses of our data and analytics services. We also intend to develop advanced methodological workshops to foster this use (see above).

A pro-active communication strategy: These activities need to be supported by a pro-active communication strategy. We started late in RISIS 1, avoiding to promote the infrastructure before we were sure that we would deliver what we expected. But catching up has been quite rapid since we have adopted (in 2017) a specific communication strategy and created a communication team that will go on promoting RISIS in the next project. We have now a lively website. We have developed material (in particular videos) supporting it. And we are developing a targeted communication strategy toward stakeholders through policy briefs. We anticipate a more regular production in RISIS 2 and we shall accompany colleagues through specific training in writing policy briefs. But we are still far from our expectations. We thus will reinforce the communication efforts along two lines. First, we need a dedicated effort in producing relevant material to demonstrate the interest and power of using the infrastructure, in a more adapted format (e.g. storytelling, webinars). Second, we need to multiply the channels for communication. The website is a key starting point. We see that the diffusion of a periodic newsletter remains an important awareness-raising instrument, provided it is complemented by a well-elaborated mailing list; and we shall reinforce our presence on social media.

Interacting with stakeholders: Our main stakeholders are policy and strategy makers in government (national, regional and of course European), universities and research organisations, firms and professional associations, and more and more civil society organisations. One important lesson from RISIS 1 is that we need to be aware of and engage with researchers and analysts within stakeholder organisations. These "stakeholder researchers" are often critical to the development of "productive interactions" between the academic community and stakeholders, and to the mobilisation of our resources for developing evidence for policies. We thus need to engage specifically with them. Beyond contractual modes, we have selected three interaction modes: opening all our training activities; engaging them with our users in "programming events", and developing "participative events" targeted to different stakeholders, if possible in connection with their representative associations (EUA, Science Europe, …) which play an important role in fostering new practices. Finally, RISIS as an infrastructure positions itself as complementing the important data gathering and indicator building role of statistical offices. They have played very supportive roles in promoting quite a few of our activities (e.g. Eurostat and national statistical offices for ETER; OECD with SIPER or the use of semantic analysis for policy analysis). It is thus important that we engage in a new step pushing for more in-depth exchanges and regular structured interactions discussing issues of data quality, evolution of existing datasets, development of new datasets and implications of new data gathering methodologies.

1.4.7. Transnational access to RISIS datasets, tools and services

The overall approach selected by RISIS is to favour transnational access, meaning that we give access to researchers on the basis of the projects they propose. On top of this, most of our services (for data integration and analysis) are "free" resources that researchers, once they register, access without further control. This is already the case for the integrating services offered by RISIS for actor delineation with ORGREG, for geocoding and of course for semantic analysis with CORTEXT. And it will be the case within one year for the GATE service for topic-based annotation.

Our ambition is to increase significantly the overall number of transnational visits compared to RISIS 1 moving from 100 visits in 3 years to around 400 visits during RISIS 2. We also expect to shift to "virtual" transnational visits within 18 months of RISIS 2 activity, with an overall balance of 75-80% of virtual visits the last year of the project. As we have said, this enlargement is ambitious, and it takes into account, all the efforts we intend to do (see section 1.4.6) to deepen our penetration of the community, to act upon the young generation, and to extend to neighbouring fields.

The way we organise access will also radically change, not for users, but for the consortium. For users we keep the central access through the RISIS dataset portal that presents datasets, with complete documentation, exemplary use cases, and metadata. We shall progressively add exemplary cases of "actor-based", "geography-based" and "topic-based" combinations of datasets in order to push for simultaneous use of datasets. The process for proposing a project and for getting it evaluated is fully on line and, rather than periodic calls, is open permanently. Thanks to the dedication of dataset teams and thanks to our very committed "Project Review Board", our normal decision time takes on average just under two weeks.

For the consortium we wish to radically change our approach to the management of visits, if only because more and more visits will concern more than one dataset. We thus propose a central management of visits operated by AIT. We expect all teams in charge of one dataset to nominate an "access manager" that will be the contact point of AIT both for the selection of projects and the organisation of visits, taking in charge all logistical aspects for physical visits and taking in charge the "parametrisation" of access for users. This will enable a simpler articulation of Access with the RCF and a more robust integration of datasets in the datastore. It will also help in devising a better follow-up of our users and, even more important, to push for user engagement in our activities (see section 1.4.6).

Our estimates are based on opening from Month 1, 12 datasets opened during RISIS 1; the 3 firm datasets (CHEETAH, CIB and VICO), the 3 datasets on outputs and collaborative projects (CWTS publication database, Patstat IFRIS, EUPRO), the dataset on universities and public research organisations (ORGREG), the repository on policy evaluations (SIPER) and the 4 datasets of interest (Profile, More, Nano S&T and JoREP). It also anticipates a progressive opening of the new datasets (even in their provisional versions): end of year 1 for social innovation; end of year 2 for trademarks, for the NATPRO extension of EUPRO, and for instrument-based financial analysis; end of year 3 for "non-academic careers" (taking into account the adaptation to privacy issues).

1.4.8. New dataset development

For RISIS, research is first and foremost developing new services that help heterogeneous integration of the datasets included. The development of new datasets is only justified when facing important gaps in our knowledge supporting evidence-based policies. Thus the arguments to do so combine a research and a policy argument.

We proved in RISIS 1 that this could be a productive and useful approach if we look at the new datasets we have produced to fill in gaps, recognised for a long time but not yet addressed. This is well illustrated by the

new dataset on "fast growing mid-sized firms" (Cheetah) or by the new repository rendering accessible to all (researchers and policymakers) existing policy evaluations (SIPER), which plays a central role in learning about policies. Similarly reconfiguring the ETER database into a far wider and encompassing register of organisations within the public sector in Europe (ORGREG) and integrating it to "project" and "output" databases has opened completely new alleys for analysing public sector research.

In RISIS 2 the same analysis has driven us to propose a focused effort on 4 key policy issues that have not yet found any robust quantitative answers. The call itself is explicit on one, social innovation, which we propose to approach through developing an aggregative database of social innovation projects. However this covers only partially one of the key changes academics and policymakers discuss about innovation that is the role of business models, organisational innovation and other "non-technological" innovation. We consider the latter a crucial issue for a better and more comprehensive framing of innovation policy. Following exploratory work done, we consider that trademarks could build an important and relevant marker of "non-technological" innovation. As we have done for publications and patents, we wish to build an enriched and researcher friendly trademark dataset. This is the second important new development we propose. The third development addresses our policy learning dimensions. We know the importance of competitive public funding in research and innovation at large. But our knowledge of effective channels for its allocation remains limited, in particular the effective use and combinations of instruments that build the "operational" policy mixes: this question is highly debated within the academic community and at the same time critical for evidence based policy-making if we wish to understand why we have so much difficulty in shifting funding for addressing both ends of the spectrum: the ability to support breakthrough innovation and the ability to redesign action for coping with societal challenges. The objective is to build a European database of public funding disaggregated at the instrument level. The fourth development lies in our will to offer new insights to the academic and policy debates on Human Resources and the role of PhD holders in society. There has been numerous works on academic PhD careers and our review in RISIS1 has concluded that it was useful to help their integration through maintaining a wiki helping the conceptual and targeted integration of the numerous local databases. But the core result is that we miss the vast majority of PhD holders since it is widely considered that half never engage in a research career after their PhD and that 70% are no longer in research after 7 years. But we know little about the jobs they occupy, and the role of the PhD in their careers, a critical issue for the shaping of higher education policies. Our fourth development is to nurture this academic and policy debate by developing an operational dataset targeted on the role of PhD diploma in society.

These 4 datasets share in common their ability to complement with our core datasets, and they will be developed in a way as to implement from the start all our integrative dimensions (about actors, spaces and/or topics). The following paragraphs present the efforts to be done on these four new datasets.

Focusing on non-technological innovation: Most work dealing with innovation highlights the fact that we can no longer reduce innovation to technological innovation, that we have to care for other types of innovations that do not translate into the classical markers of technology-led innovations (in particular patents, but also involvement into science or technology led collaborative projects). This is a very visible result of innovation surveys, but these remain at best at a meso level forbidding the possibility to analyse bottom-up new processes and new key factors beyond those observed for the last twenty years. This is why we are in search of new markers. Quite a few studies (e.g. Schmoch 2003, Sandner & Block 2011, Gotsch & Hipp 2012) have demonstrated the potential of trademarks as "closeness to market" and "diffusion" indicators as well as being able to grasp non-technical dimensions of innovations, e.g. service innovations, but also software and IT-related parts, which are hardly patentable. But this remains experimental mainly because of the work to be done to stabilise a relevant dataset. Further, to prove useful it also has to be articulated to the work done in the other output datasets and with our firm datasets. Matching trademark data with other RISIS data sources will considerably enrich their analytical potential and go beyond what has been done in most of the existing

studies so far. This effort explains why we have asked the best specialist of the issue, and furthermore one of the most competent actors in indicators building to join RISIS, which is Fraunhofer ISI institute.

Grasping the realm of social innovation: Within the broad term of "non-technological innovation", an important dimension deals with "social innovation". We cannot avoid the fuzziness of the term, the multiplicity of definitions, both at political and at research levels. An important exploration is being done in a companion project to RISIS 1, KNOWMAK (EC-GA n°726992), which has enabled to propose an open ended approach around the four dimensions that are constitutive of most definitions: answering societal needs; the driving role of social actors (mostly under the form of civil society organisations) in conducting them and/or the intensive interactions with social actors; the clearly defined social effects or impacts, and the nature/extent of the transformations that they introduce to address their objectives. A first pilot database is being realised that mobilise novel machine learning techniques. This exploratory database will be integrated in RISIS 2 and opened to researchers in the first year of RISIS 2 to favour interactions and returns before the main database is produced. Again this entails the introduction of a new partner that is specialised in social entrepreneurship, Strathclyde University.

Deepening our knowledge of public funding at the instrument level: Financial data about public funding of research and innovation exist since the early stages of science policies. However, they have been built with the objective of nurturing analyses at the national level in an aggregate way. Core policy efforts since the beginning of the years 2000 have been to differentiate between types of funding (core vs. competitive) and between the different instruments mobilised (whether speaking of policy mixes or portfolios). However these ambitions have proven very difficult to operationalize in the traditional frame of "Frascati": the developments made in the PRIME network of excellence (see 2007 SPP special issue) are still not integrated in the financial data produced by statistical offices. This explains why the European Commission funded an exploratory project (PREF), developed by participants in RISIS 1, to propose analyses of public funding at a more disaggregated level. Based on this experience we now think that we can develop, based on public information, a robust European level database of public funding at the instrument level. While we have a view of the variety of available instruments, we still lack knowledge of their effective use and about the effective policy portfolios. This disaggregation will enable to renew comparative analyses both between countries and at regional level. As such it will provide a unique complement to the SIPER database, offering researchers and policymakers two unique entries to the analysis of research and innovation policy and policy measures.

The societal relevance of non-academic careers of PhD holders: The funding of doctoral studies represents a significant proportion of higher education funding (including grants to doctoral candidates). It is thus important to know the social returns of such investments. It was not a central issue when the core of PhD holders used to go in public (or private) research professions. But it becomes a different issue when the majority of PhD holders enter other activities whether in the public, private or civic society worlds: do positions occupied and roles played in society justify the public investment? The efforts in RISIS 1 have helped us develop conceptual frames to address this issue and they have helped build a semantic wiki enabling to share efforts widely. But they have shown us that traditional survey-based or panel-based approaches will not provide suitable answers in a reasonable time horizon (around 3 to 5 years) or at a reasonable cost. This is why we wish to test new approaches based on open access and the very large number of CVs that are publicly available. We know the difficulty of treating CVs, but focusing only on one central issue (CVs mentioning PhD in studies and with a last position outside academia) will increase feasibility. The teams in charge, especially DZHW, which manages the PROFILE database, are well aware of security and ethical rules presiding to the development of such a dataset.

2. Impact

In this section we examine first the six dimensions of expected impact as is mentioned in the call; and detail the key performance indicators that should manifest the degree of achievement of RISIS2 objectives. We then discuss the measures taken for maximising impact. These include: (a) The RISIS2 plan for the dissemination and exploitation of results; (b) Communication activities; (c) A two-tier approach to data generated incorporated in the RISIS data management plan: open access to research results and "controlled access" to primary datasets included in the RISIS infrastructure; (d) the Approach developed to the sustainability of RISIS infrastructure, and corresponding plan.

The reader will see when reading this section that the notion of 'impact' as well as the processes through which such impact is obtained are not only an object of academic debate; but play an in-depth role in the way we define and deploy activities. The classical view is that policies should be evidence-based, meaning that the central role of indicator producers is to produce relevant background information to policy debates. However it is not through simply producing new evidence that impact is assured. Spaapen and van Drooge in their very influential 2011 article have shown that impact requires that researchers and stakeholders enter into 'productive interactions' for this new evidence to become part of stakeholders framing of the situation. This insight is what drives the activities we have developed for "maximising" the chances of a potential impact, both vis-à-vis our community and vis-à-vis our stakeholders.

2.1. Expected Impact

The work programme identifies six major expected impacts. Section 1 has shown that we have organised the project *de facto* along these six major dimensions. We shall here recall how in practical terms we have organised the project to generate impact and how we translate it into measurable effects. In order to avoid too much repetition, two of the impacts have been grouped together. A table at the end of the section recapitulates the impact expected and the criteria for measuring their success.

2.1.1. Impact 1: Further integrating and opening research data infrastructures in the field of science, technology and innovation (including social innovation)

RISIS 2 addresses this question through 4 major angles:

a) ***The further integration of RISIS core datasets*** thanks to our data integration services (see impact 4) on organisations, geography and topics will offer researchers multiple new alleys for research.
b) ***The integration of four new datasets*** dealing with important but lasting research and policy issues will provide resources for real new developments in the community on new forms of innovation and the role of societal actors play in innovation, on the role of the PhD in society, on the effective role of different policy instruments and the de facto mixes they build for deepening policy learning. We expect to open first versions end of year 2, and start see effects in term of visits and presentations at conferences before the end of the project)
c) Our ***new data integration services*** will enable the community to address the fast growing issue of societal challenges and the ways in which research and innovation can contribute to them. This transversal ability to read RISIS databases of outputs and projects (enriched by the new datasets) provides researchers with a unique new capacity, without having to operate massive simplifications at aggregate level. A first version will be offered at the end of year 1, fully integrated in the RCF and available through it at month 18, and become a tailored service by month 30. In order to help the community to make this move, we shall have topic-based demonstrators (we expect 5 of them over the project).

d) We will *connect the community to more and more open data* enabling far wider exploration (via the opening to the growing number of open access datasets thanks to the D4Science capabilities articulated with the RCF). We speak of exploration because few of these datasets can be used as such to build new indicators (which require reliable longitudinal data). However, this exploration is important for the dynamics of the field and of RISIS itself, as it will help decide which datasets should be transformed to become a reliable source for new indicators. Here the RCF-D4Science virtual research environment and search tools for external open access datasets will play a central role. It means that this third component will de facto not play significantly before the end of the second year.

Indicators for measuring success:

- Share of projects mobilising multiple datasets thanks to the use of one of our data integration services (Success target: 33% the 4th year of the project)
- Effective opening of the new datasets in time, and attractiveness (i.e. number of visits by the end of the project – Success target: 70)
- Effective opening of ontology building and topic-based annotation (GATE services), systematic tagging of RISIS datasets along 5 "societal challenges" by the end of year 3, and attractiveness (i.e. number of projects developed based on or mobilising GATE, expected number at end of project: 100).
- Number of projects combining RISIS Core datasets with open access datasets to explore new directions for indicator building (Success target: 50)

2.1.2. Impacts 2&3: Facilitating transnational access and widening the user base

RISIS2 will have two core activities with a strong impact on this objective

a) We propose a step change in moving from 'physical' to *'virtual' transnational access* by month 18 of the project with the opening of the RISIS Core Facility (RCF). This change will not only simplify access for users independently of their geographical location. It will also radically enlarge the 'user base' as the research topics that researchers can address will be multiplied through the simultaneous access to multiple datasets and the possibilities of combination offered by our data integration services. This is why we can plan this very important increase in use.
b) We propose a set of measures to *widen the user base through a specific plan.* The plan will deal will the following aspects: (i) *training*: we expect 31 training courses over 4 years, with a growing use of 'blended' formats that combine training activities requiring physical presence with on-line (both synchronous as asynchronous) training activities. We also expect to widely develop tutorials. This will be complemented by a special activity toward *STI master courses and PhD programmes* (to introduce very early on these new capacities for the next generations); (ii) *'awareness raising events'* at the 2 major annual conferences; (iii) enlarging *the use of advanced quantitative methods for modelling and indicator building*, through the development of specific material and *advanced research workshops*.
c) One aspect we consider crucial to widen usage is to develop *interactions with research users through a variety of means*: user return sessions at conferences, user blog and forum, user programming sessions about future needs of data and services.

Indicators for measuring success:

- Overall number of transnational visits over the 4 years of the project (Success target: 400)
- Expected acknowledgements to RISIS in articles and open-access documents (success target: respectively 75 and 100)[3]
- Effective opening of the RCF in due time
- Share of projects using "virtual transnational access" (Success target: 80% the fourth year of the project)
- Share of projects mobilising multiple datasets thanks to the use of one of our data integration services (Success target: 33% the fourth year of the project)
- Number of researchers trained through our courses (Success target: 500),
- Number of interactions with specialised master and PhD programmes (20 the fourth year of the project)
- Number of downloading of tutorials, use cases and other relevant documentation & material (Success target: 1000)
- Participants in RISIS user blog, user return sessions and user programming sessions (Success target: 800)

2.1.3. Impact 4: Enlarging and strengthening the offered services

An important part of RISIS 2 is to offer the research community with new data integration services and new data analysis services. Most of the times their effect is combined with other aspects to further integration of the community (Impact 1b) and widening the user base (Impact 2a and 2b). Many of the KPI to measure impact go through these main impacts, beyond the monitoring of use and interactions developed with users about conditions of use and future developments. Four main aspects are to be highlighted:

a) Together our three data integration services provide a new resource for combining datasets both RISIS datasets and external datasets. All are open online for all 'accredited' researchers for their own corpuses as well as when using RISIS core datasets. We thus expect a very important use of the services as such by researchers in search of adequately shaping their study corpus (success target: 300 cases of use in fourth year).
b) RISIS 2 will offer a completely new service based on *topic identification* (though the ontology tagging service of GATE). This is a major addition because of the rising research interest in analysis at the level of targeted topics not recognisable through existing classifications, for instance in the case of possible future policy missions pursued by European framework programmes (see impact 1.3).
c) We extend the realm of *semantic analysis tools* we offer through CORTEXT (based on corpuses) with the GATE services (based on documents). We also extend the realm of visualisation tools (in particular VosViewer) helping to mobilise results. What is aimed through these targeted developments is to maintain the high rate of growth in use we have been witnessing the last 3 years. This analysis service is a major channel, which supports the widening of our user base.
d) Finally, we expect a strong enlargement of research users from addressing the difficult issue of *use of advanced quantitative methods* in our field (due to the specific features of knowledge dynamics and corresponding datasets). In order to foster the uptake and diffusions of these advances methods, we plan a set of measures aiming to make these methods more easily accessible, in particular through well-documented use cases and research workshops. A dedicated methodological support unit will promote their use, in particular through the organisation of special tracks at major conferences.

[3] We expect that one third of visits will generate articles with a two years delay between visits and papers. And we expect a complementary similar return for other documents than articles (conference presentations, DOI reports mainly) but with one year of delay. These figures should double when looking 2 years after the end of the project.

Indicators for measuring success:

- Number of cases of mobilisation of our data integration services (the latter can be mobilised in projects but are also available on line for all accredited researchers) (Success target for the 3 services: 300 cases the fourth year of the project)
- Number of different users for CORTEXT (on average one user conducts multiple projects): 1000 the fourth year of the project.
- Downloading of use cases and documentation for advanced quantitative methods (Success target: 200 the fourth year of the project)
- Number of researchers mobilising the VRE to build advanced user communities (Success target: 200)

2.1.4. Impact 5: Fostering the innovation role of such infrastructures

We see two channels through which foster the innovation role of RISIS taking into account the nature of our stakeholders and the channels though which we can impact them. Our impact in society lies in the production of new quantitative evidence that helps in the setting of new policy agendas, in nurturing policy debates and supporting evidenced-based policymaking and the widespread dissemination of this evidence in society. But RISIS does not produce this quantitative evidence directly, but provides the means to produce, while offering access to data and support to analysis. In practice this evidence is mostly produced by the researchers that mobilise our infrastructure to produce new knowledge. RISIS infrastructure also helps researchers identify new trends in the STI system, formulate new research questions and produce new knowledge and indicators. Of course most of RISIS members (especially those producing datasets) specialise in the field and do produce indicators and policy relevant analysis, using RISIS datasets.

a) Potential Impacts are thus associated with the efforts we do to *disseminate results* obtained through the use of our datasets and services. To ensure impact we will develop measures supporting wide dissemination of these results produced by our research users. We will do so globally through the development of *policy briefs* and though the extensive use of RISIS communication instruments: the *website*, a widely circulated *newsletter*, webinars, blog, social media (see section 2.2.2).
b) This will be complemented by *targeted interaction with stakeholders*. Our stakeholders are not only Governments[4], but also regions, and research and innovation organisations. We will privilege their European representation such as Science Europe for PROs, EARTO for RTOS, TAFTIE for RFOs or EUA for universities) and engage with them to jointly organise workshops and/or working sessions to present results and discuss their needs (in particular for new indicators) with a view of learning from these interactions about the future directions of infrastructure development we may prioritise.
c) We expect that a number of different stakeholders also have specific requirements that RISIS datasets and services can address. We are constrained here by the 'private' nature of the main sources we mobilize. But de facto RISIS partners responsible for opening these sources have specific agreements to do consultancy. We have thus agreed of a principle that we call *'indirect contracting'*, whereby the corresponding RISIS partners will contract directly with stakeholders but will return to the RISIS consortium a percentage of the contract for the use of RISIS services[5]. This will be part of the RISIS 2 consortium agreement and, with support from legal services (see under management) we expect to develop a robust

[4] In RISIS1 we have developed a practice of internal seminars with the Commission to present our results that we intend to continue and enlarge if asked for by other Governments. See section 1.2.
[5] The experiment done in RISIS1 with research infrastructures in Marine Biology (via the EMBRIC project) has demonstrated the value of such an approach. It has demonstrated that the spectrum of stakeholder potential use is quite large, in particular touching upon other research communities for their international positioning and the building of their strategies.

and practical solution for end of year 2, and thus complement our global impact on stakeholders by more direct services to specific users.

d) This will be complemented by a *specific action towards researchers and analysts from our field that stakeholder organisations employ* (the ones mentioned above, but also OECD, EIB, JRC or NESTA). They play an important role in the circulation of results and tailoring to the policy issues of the stakeholders that employ them. This is an important audience that we wish to engage deeply in the activities of the infrastructure: beyond our transversal activities, we shall give them access to all our training efforts; once accredited, they will access our open services; we shall include them in all our user-oriented actions especially for discussing future developments.

Indicators for measuring success:

- Statistics following the use of the website and effective circulation of the newsletter
- Level of downloading of policy briefs produced (Success target: 1000 during the project)
- Number of targeted events and of participants (Success target: 8 events, 240 stakeholders mobilised at these events)
- Monitoring the returns of our 'indirect contracting approach' (success target over the 2 last years of the project: 10 indirect contracts)
- Number of 'stakeholder researchers' accredited (Success target: 100) and number involved in our training and user programming sessions (Success target: 100).

2.1.5. Impact 6: Ensuring long term sustainability to their integration

Datasets supporting indicators are long-term endeavours (if only because of the need of longitudinal analyses that are at the essence of indicators). They thus require that teams, centres, groups or institutes have an *official mission* (or responsibility and goal) to develop them, translating the long-term engagement of the institution they belong to or are affiliated to. RISIS1, as a project, has been successful in pushing more institutions to engage and/or national long-term finance to be granted.

a) Our datasets are costly first and foremost in Human Resources and this, in our view, is the most important aspect dealing with the long-term sustainability of RISIS from the supply side. This also applies for most of our services. This builds a first critical criterion to measure the degree of sustainability: the number of core partners (i.e. in charge of core datasets and of our services), which have been given this mission by their institution. Our objective is that the number is above 8 by 2020 so that we can enter in an ESFRI recognition process.

b) With 'virtual' access and the construction of the RCF, RISIS will have a significant 'shared infrastructure' to cover and maintain. While the project will help us to develop it, it is still difficult to measure its running, maintenance and improvement costs. It is a clear objective included in the consortium agreement to assess these needs (which are hard at this stage to calculate) during the third year of the project. We also at this time will be able to estimate the returns from our 'indirect contracting' approach. This will help us to develop a distributed approach of the funding of the fully shared part of the infrastructure and build our long-term sustainability plan (see section 2.2.4).

2.1.6. What are the main barriers to impact?

If we take out critical risks in project development (see part 3), we are quite confident that 'virtual' transnational access complemented by our data integration and analysis services will in fact enlarge the user community (the present process only based on physical visits has enabled us to obtain globally our RISIS 1 objectives). So we do not foresee any major barrier to the fulfilment of our objectives in term of research usage.

The situation differs widely when discussing impact is society. It is not because we offer new resources for indicator building and production of new quantitative evidence, and because we heavily support these activities and the dissemination of the results, that the latter will be taken up in STI policy debates and STI policies. Readers have just to think about the lasting prevalence of the figure of 3% of GDP that numerous academic studies have shown not to be a relevant target for most countries. We do not help building new products, *we help policy debate to be more rooted in analyses, more evidence based, and less driven by stereotypes or anecdotal evidence*. We do not replace policy debates and the values (and ways of 'judging' dynamics) that come out of such public debates. Very often it is through a long and indirect process of circulation, that results percolate and influence policymaking. Traceability is thus a tricky issue, even for specialists like some of the members of this project, in evaluating impacts of research in society (e.g. Joly at al., 2015). In this respect we cannot take for granted the involvement of stakeholders in productive interactions, this presupposes a long-term view that is often difficult to maintain. This is an intrinsic barrier that justifies important 'dissemination and exploitation' efforts, and the plan we propose.

The *other important barrier lies in the 'innovation dimension' and is legal*. The central endeavour of RISIS is to foster open science. But we depend very much on privately owned public databases, out of which we build enriched datasets that can be integrated. All the members responsible of such datasets have developed agreements for use for 'publishable research', and most of them have individual rights for exploitation in paid studies. But this does not apply to collective use and even less to public opening to the community. This is why we have developed the principle of 'indirect contracting'. We need to assess its robustness and clarify the situation, in term of IP, use and open access. This will be conducted very early on in the project, so that we can mitigate that legal risk.

Table 2.1 A recapitulation of expected impacts, of the indicators of success and of the corresponding dimensions of the plan for dissemination and exploitation of results and its main indicators

Area of impact	Type of impact	Who is impacted	Indicators for measuring success	Dissemination & exploitation plan	Main indicators for the D&E plan
Impact 1 further integration & opening of datasets	1.1 - Further integration of core RISIS datasets	Users from STI community Users from Neighbouring R&D communities Stakeholders Organizations	Share of projects mobilising multiple datasets (target: 33% in the 4th year)	Special track/ sessions at core conferences around results Awareness raising events at major conferences	At least 2 conferences per year 2 webinars (30-40 people) 3 or 4 per year
	1.2 - Integration of the four new datasets	Users from STI community Users from Neighbouring R&D communities Stakeholders Organizations	Effective Opening of the first version of the new datasets Attractiveness (number of visits by end of project:70)	Training & tutorials Webinar Participative events with stakeholder organisations Policy Briefs User forum	6 training (online or physical) 4 webinars (30-40 people) 3 workshops and events (120 people involved in total) 6 policy briefs circulated
	1.3 - New data integration services	Users from STI community Users from Neighbouring R&D communities Stakeholders Organisations	First version of topic-based service on-line at M12 Number of projects mobilising the service (100 at end of project) 5 demonstrators of tagging, generalised throughout RISIS datasets	Training & tutorials Demonstrators at core Conferences User & stakeholder programming workshop (to discuss interest)	4 training (online or physical)/80 people involved 2 Conferences per year/60 people involved per year 5 policy briefs on the 5 topics (‘missions’) selected 3 workshops (years 2, 3&4) / 90 people
	1.4 - Exploration of open access datasets	Users from STI community Users from Neighbouring R&D communities Stakeholders Organizations	Search tool for open data M24 Number of projects at end of project incorporating open access data (target: 50)	Training & tutorials Webinars Demonstrators at core Conferences User & stakeholder programming workshop (to discuss interest)	4 training (online or physical)/90 people involved 2 webinars (30-40 people) 2 conferences per year/60 participants per year 2 workshops (years 3&4) / 60 people

Impacts 2&3 Transnational access and widening the user base	2.1 - Virtual transnational access	Users from STI community Users from Neighbouring R&D communities	Virtual access available M18 Number of TA in 4 years (target: 400) Share of virtual TA (target: 80% at the 4th year) Share of projects mobilising multiple datasets (target: 33% in the 4th year)	Special track/ sessions at core conferences around results Awareness raising events at major conferences User return sessions at core conferences User programming participative workshops	At least 2 conferences per year 3 or 4 per year At least 2 per year / 50 participants per year At least 1 per year (overall 200 users involved)
	2.2 - Training	Users from STI community Users from Neighbouring R&D communities Researchers belonging to stakeholder organisations	Number of courses (target 34) & number of trainees (target 500) Structured exchanges with specialised master & PhD programmes Use of Tutorials (target: 1000)	Organising training Presentations at PhD programmes, support to the development of student projects (collective & individual) Dissemination of tutorials	Number of courses (target 34) & number of trainees (target 500) Relations established with 20 programmes in the 4th year Tutorials: number of downloads over the project (1000)
	2.3 - Interactions with research users	Users from STI community Users from Neighbouring R&D communities	800 research users engaged in the different formats proposed	User return sessions at conferences User blog and forum User programming sessions	2 sessions per year/60 participants per year 1 session per year / 50 participants
Impact 4 Enlarging and strengthening offered services	4.1- Data integration	Users from STI community Users from Neighbouring R&D communities Researchers belonging to stakeholder organisations	Number of cases of mobilisation of our data integration services (target: 300 in the 4th year)	Training & tutorials Demonstrators at core Conferences (and in awareness raising activities – see 21) User forum	6 training (online or physical) Tutorials: number of downloads over the project (600) 2 conferences per year/60 participants per year
	4.2 - New service on topic identification (already presented in 12)	idem	See 12	See 12	See 12

	4.3 - Semantic analysis	idem	Enlargement of CORTEXT user base (1000 different users, with over 3000 projects)	Training Tutorials Webinar Stakeholder oriented 'master classes'	At least 6 training (online or physical) (120 people involved) Target: 1000 downloads 2 webinars (120 people involved) target: 2 per year
	4.4 - Enlarging the use of advanced quantitative methods	Users from STI community Users from Neighbouring R&D communities	10 methods covered in 2 batches Research workshops about advanced uses (1 per case, 200 researchers mobilised) Progressive mobilisation of RISIS VRE to build advanced research communities	On line use cases & documentation Participative events at core conferences	Download target: 500 over the project Specific sessions in core conferences (5 sessions and 150 participants) 200 researchers mobilised by end of project
Impact 5 Fostering the innovation role of the infrastructure	5.1- Policy briefs	Users from stakeholder organisations	Policy briefs: number of downloads (target 1000 over the project-	Organising & supporting the production of policy briefs	Main indicator: number of downloads (target 1000 over the project)
	5.2 - Targeted interactions with the different types of stakeholders	Users from stakeholder organisations	Number of events organised (target 8) and number of individuals from stakeholder organisations mobilised (target: 240)	Organising the targeted events Wide dissemination of their results (in particular new needs identified) through communication channels (see section 222).	Number of individuals from stakeholder organisations mobilised (target: 240) Main indicator: number of results downloads (target 500 over the project)
	5.3 - Indirect contracting approach	Users from stakeholder organisations	Process operational at M24 Active monitoring of deployment (success target: 10 over the last 2 years of the project)		
	5.4 - Specific actions towards researchers in stakeholder organisations	Researchers & analysts from stakeholder organisations (SO R&A)		Active communication activities (see section 222) targeting this subgroup	Main indicators of success: 50 SO R&A involved in training 100 SO R&A involved in our user programming sessions (see 21)

2.2. Measures to maximise impact

Measures to maximise impact are organised around four central aspects:

a) A plan for the dissemination and exploitation of results
b) A two-tier approach to data generated: open access to research results and 'controlled access ' to primary datasets included in the RISIS infrastructure.
c) Approach to the sustainability of RISIS infrastructure and its components, and corresponding plan.
d) Communication activities

For easing the reading of the section we have decided to put communication activities just after the plan for the dissemination and exploitation of results. The table above recapitulates per impact our measures for measuring success, and presents the main dimensions of our plan for dissemination and exploitation of results with their main indicators.

2.2.1. Plan for the dissemination and exploitation of results

We have taken very seriously the requirement by Horizon 2020 that projects need to develop a dissemination and exploitation plan (D&E plan) both to develop a precise strategy, and to avoid *ad-hoc* solutions. And we have integrated the main objectives of the plan in the methodology section of Part 1 (section 1.4.6.) where objectives are decomposed in training objectives, interaction with our research community and interactions with stakeholders. In the following paragraphs we first detail the audiences of the plan before examining the main instruments of the plan. The ways in which the plan is mobilised to maximise the expected impact is presented in Table 2.1.

The implementation of the plan will require further elaboration, in particular to sequence the activities and expected achievements. We expect to work in stages starting with a first 18-month activity plan prepared at the outset of the project. The monitoring of this first phase will help the Governance of the project (see Part. 3) decide of the required adjustments for the second phase.

The audiences of the plan:

As a research infrastructure, our key target is made of researchers that use our datasets to produce new knowledge and new STI policy indicators. As we have said, the STI field lies at the encounter of four main disciplines. So that it is first made of a core of researchers that recognize themselves as STI researchers by participating to the conferences of the field and publish in the journals of the field. A raw estimate drives to consider that it is made of around 2.000 researchers in Europe only. But it gathers a very large crown of researchers who in their field do consider research and innovation as one of the dimensions they consider. Our plan for dissemination and exploitation, first addresses both types of researchers.

We target those in our community that are eager to work with advanced new data and with our new data analysis services. We also target and educate those in our community that are less aware about the use of quantitative approaches beyond the traditional surveys. Thirdly and more generally, we have a specific role in training the young generation of researchers very early on in using quantitative data more creatively.

We produce and support the production of indicators to support public debates on STI policies and policy-making. As we have said above, it represents a wide audience not only because it serves governments and members of parliament at different levels (European, National or regional), but also because it serves the stakeholders involved in those debates, universities and research organisations, firms and their professional associations, and more and more civil society organisations. It also serves, as we have demonstrated in RISIS 1, different research communities that organise transversally at the European level (such as the marine biologists regrouped in EC research infrastructures). Quite a number of stakeholder organisations incorporate researchers and analysts that play a key role in mediating between our results and policy recommendations. For all those we propose in the plan complementary approaches: from the start a generic approach based on

dissemination of results, and a targeted approach focused on different professions or types of actors; a specific approach targeting researchers in stakeholder organisations; and when we will have set our "indirect contract" approach, possibilities to serve the needs of individual stakeholders.

The main instruments incorporated in the plan:

We have regrouped them around 3 main dimensions: training, policy briefs and other materials synthesizing results, interaction activities:

a) Training is a critical dimension as was demonstrated by the high success encountered during RISIS 1 project. There were two types of training courses: one to introduce to new datasets and their use, the other focused on data analysis methodologies. We wish to maintain them, complement dataset-focused courses with data integration focused courses. And we wish, using an existing infrastructure of one of our partners, to develop more and more courses online. We also intend to increase the number of summer or winter schools that enable to deepen thematic aspects and the combined use of different datasets. All in all we expect to deploy 31 training courses over 4 years. An important complementary dimension lies in tutorials, and we have high expectations about their use (see downloading objectives in table 4). Finally the experience made in RISIS 1 to organise a direct contact with given PhD and/or master programmes is proving very productive and this appears no doubt as a potentially rich tool to promote these new approaches in the new generation. But it requires time and progressivity.
b) Classically in our field, the way results are circulated to stakeholders take the form of policy briefs (a term that covers multiple forms of synthesis). The plan will invest in supporting researchers to develop policy briefs from their results, whether in a classical written form of through other means (e.g. video and storytelling). Their circulation (see communication activities) is important, but a clear measure of success lies in their online access and in the number of downloading they generate.
c) As we mentioned in the introduction of this part, direct interactions and exchanges are critical to the circulation of knowledge; this is why we give a great importance to all forms of interactions. The presence in the international conferences of our field as in the key international conferences of neighbouring fields is a pre-requisite. There we wish to de-multiply "awareness raising events" (to enlarge the number of effective users), to support academic sessions and tracks that help presenting and discussing key results (in particular using advanced quantitative methodologies). We also wish to take advantage of conferences to organise user participative events (about their return of experience and about potential evolutions in services and in relevant datasets to incorporate). These will complement online supported interactions (user blog, social media…). A second form of direct interactions lies in the development of specific participative events with our stakeholders. We expect to involve many in our user programming events. But we also wish to develop "targeted" workshops and interactive learning events with specific audiences (most of the times we expect co-organised with their European associations, such as EUA for universities…).
d) To this, should be added our open science policy for all results produced with the use of RISIS datasets and services see section 2.2.3.). The link constructed at the RCF level with D4Science (see section 1.4.1.) provides a direct and systematic link with OpenAIRE, thus being part of our dissemination efforts of all results produced by RISIS 2.

The mobilisation of instruments for maximising impact:

We have chosen here to put a clear table that connects the instruments with the expected impacts, so that it gives a clear picture on how the plan is going the help maximising impact. Table 2.1 above, is organised by impact, recalling our indicators for measuring success, identifying relevant D&E activities and identifying the main indicators enabling to assess the quality of deployment of the D&E plan.

2.2.2. Communication activities

To be successful, the dissemination and exploitation plan required to be supported by a pro-active communication strategy to promote knowledge and awareness of RISIS toward the relevant users, and to stimulate the willingness of stakeholders to engage in two-way exchanges about the value and significance of results for evidence-based research policies.

Our communication activities will build on the initial developments in RISIS 1. We started late in RISIS 1, avoiding to promote the infrastructure before we were sure that we would deliver what we expected. But catching up has been quite rapid since we have adopted (in 2017) a specific communication strategy and created a professional communication team that will go on promoting RISIS in the next project. We have now a lively website. We have developed material (in particular videos) supporting it. During the first year, the website will be further reinforced and adjusted in order to deal with the new objectives and audiences. And we are developing a targeted communication strategy toward stakeholders through policy briefs (a critical dimension of the D&E plan). We anticipate a more regular production in RISIS 2 and we shall accompany colleagues through specific training in writing policy briefs. But we are still far from our expectations. We thus will reinforce the communication efforts along two lines:

a) First, we need a dedicated effort in producing relevant materials to demonstrate the interest and power of using the infrastructure, in more adapted formats (e.g. storytelling, webinars). This means that we need to engage users in discussing their experience (in particular through a specific blog, and the users' platform); and that we need more diverse material targeting our stakeholders (videos, leaflets, brochures). We expect to operate a step change in this direction within the first 18 months of the project.
b) Second, we need to multiply the channels for communication. The website is a key starting point. We see that the diffusion of a periodic newsletter remains an important awareness-raising instrument, provided it is completed by a well-elaborated mailing list; and we shall reinforce our presence on social media. Both activities shall start from M6 and become regular at M12.

RISIS will use the contribution of a communication specialist for improving the capability to be in contact with specialized magazines, and to enlarge the linkages with sites where there is room for further exploitation of RISIS results. Also, the professional competence in communication will help to develop measures in line with the scope of the project and targeted to different audiences. What we foresee is a feedback loop of communication-dissemination and exploitation to rise the visibility of RISIS in Europe and globally.

The implementation of communication activities will be monitored by following the now classical analytics for website, social media and blogs (in particular visits and downloads) and number of articles in specialized press and press release produced.

2.2.3. Research results and data generated

The project is based on the opening of datasets to researchers. Discussing data generated requires differentiating between the datasets themselves and their use (alone or in combination) in the research projects selected for transnational access. We thus make a critical difference between both. We have an open access systematic policy for the research results obtained through the use of RISIS datasets and their underlying aggregated data. And we have a data management plan that covers both our datasets and the software mobilised and/or developed within RISIS.

Research results and underlying data: a fully open access policy

All research results are publishable (this is a condition for accessing RISIS datasets). They are listed in the project website. The principle (which will be inscribed in the new version of the code of conduct and opera-

tionalized in our project selection process) is that authors archive the document, report, presentation, article or chapter following the rules of their institutions, in their self-archiving place, and authorise RISIS to upload them in OpenAire (see below). For articles, we follow the 'green' model, and ask that a preliminary version is accessible in open archives.

We decompose underlying data in two dimensions: The aggregated data that supports the document archived. This data is usually a specific extraction and combination of aggregated data from RISIS datasets. This underlying data will follow the same rules as documents. The primary data is linked to the RISIS datasets that are mobilised by users. These are dealt in the data management plan.

One critical development of RISIS2 is linked to the incorporation in the RCF of D4Science elements, and in particular those that organise a systematic connection with OpenAIRE. This will insure systematic opening of all RISIS documents, and we expect both a greater visibility and greater circulation and use. This will also help us follow the use of RISIS products.

RISIS2 Data Management Plan (DMP):

This data management plan covers the datasets RISIS integrates, maintains and develops. It is a clear objective from what we have developed in section 1 that the core effort we develop in RISIS is to make data findable, accessible, interoperable and reusable (FAIR). An initial DMP was built within RISIS 1 using the Horizon 2020 FAIR Data Management Plan Template.

RISIS as we have explained is made of a quite large number of individual datasets. One central work of RISIS1 has been to organise a collective approach to the technical and cognitive harmonisation of datasets, to share and apply the same principles for quality assurance, to provide users with complete metadata and with full documentation of the datasets. All this has been fully documented on reports that are available online. RISIS 1 has also developed data integration services that aim at enabling 'problem-based' data interoperability. It required all datasets to evolve for enabling integration (e.g. all datasets have been geocoded; the harmonisation of actors pushed by our 2 registers has been implemented into the different datasets). These are for us critical elements for insuring that data are inter-operable. All this, we consider, provides a solid starting base to deploy more systematically a DMP within the first 6 months of the project.

Curation and preservation are critical aspects of the sustainability of the infrastructure. Here we maintain a decentralised approach, whereby both are organised at the level of each dataset, following the rules of the institution that deploys them. We simply insure that it is fully done. For the time being, re-accessibility of older versions has never been an issue, as we maintain continuous access to the updated datasets. But it will have to be considered, taking into account legal aspects associated to the private origin of many datasets. Another aspect that we have to systematise is the versioning of datasets (especially when they are periodic updates). Many datasets do it already but we have not yet developed a systematic and co-ordinated approach to the issue.

There are clearly legal aspects that impact on data sharing. Presently, these are distributed and managed by each dataset. But the central principle that has been transversally arrived at, is that data operators are authorised to give access to public researchers provided they produce 'publishable' research (what the community calls the 'hybrid model'). This legal base has been robust in a fully distributed approach complemented by the use of physical transnational access. But it has to be further checked with the creation of the RCF and the development of virtual transnational access. This is why we plan a full legal review from the start of the project, and we have 18 months to make our 'hybrid model' robust and lasting.

At present, we have limited ethical aspects as our datasets are built on public data (even privately owned), and the analyses made and indicators built consider aggregated levels (at the organisational or geographical

level). The only situations where we face privacy issues are with the work on Human Resources. There our decentralised approach helps. These issues are handled by the dataset operator, DZHW in Germany, who is one of the largest holders of such databases and has an extended practice of anonymization and a developed policy for controlled access and use. These aspects are further developed in section 5.

These elements drive us to consider the DMP as an intrinsic part of the management of the project. The DMP will thus be constructed as part of WP1.

Software and services developed within or mobilised by RISIS

We face two complementary situations in RISIS:

- Numerous targeted developments done in RISIS (for instance those enabling the enriching of IFRIS-PATSTAT) are completely open-source and are archived and documented in GitHub (in this case, CORTEXT Github).
- RISIS mobilises "value added services" and supports their tailoring and further development. It does not impact upon the intellectual property rights of the developers (for instance INRA for the CORTEXT platform). But RISIS involvement is linked to the fact that the owner is dedicated to open science, and guarantees access and use for research activities ("strictly for research, teaching or training purposes"). This means that the service has developed clear "general terms of use" that are part of the mutual engagement between the service and users (though the charter they sign for being accredited). This includes in particular the presence of acknowledgements in documents based on the use of the service, and keeping mentions of the service when mobilising graphs, tables and any other visualisation produced by the service (e.g. "Powered by www.cortext.net").

It will be part of the legal review mentioned to make a systematic review of the status of software and value added services mobilised by RISIS 2, and to insure full harmonisation of rules applied (keeping in mind national legal requirements that operators of services and datasets face).

2.2.4. Approach to the sustainability of RISIS infrastructure and its components

We have explained at length that RISIS 1 was an experiment to demonstrate the interest and feasibility of a research infrastructure for our field, science, technology and innovations policy studies, covering also the broader remit of "science and innovation studies". RISIS 2 is there to fully materialise it, in particular developing a platform that enables virtual transnational access to community researchers and the use of a wide range of 'tailored' services.

It is thus critical that we prepare its lasting operational capacity and its lasting sustainability. We consider this as part of the strategic management of RISIS 2, and it will be handled through a specific working group and task from the start of RISIS 2. This objective and the working group will be an integral part of the project consortium agreement. Five aspects will be addressed:

a) The inscription of the work done in RISIS (whether the maintenance of a dataset and/or the deployment of value added services) by a group / centre / institute as an "official mission" by its institution. We have shown how intensive in Human Resources is the infrastructure. Thus our first objective for insuring the sustainability of the infrastructure is to safeguard this HR investment over time. Multiple procedures can be thought of, but as a first step we wish to develop a memorandum of agreement between members about such HR engagements.
b) The evaluation of the shared costs of the infrastructure: The costing of research infrastructures is still an open issue and debate. We wish here to focus only on the shared dimensions. From RISIS 1 come the website, the dataset portal, the communication activities, the harmonisation and quality control activities.

RISIS 2 will build a major extension with the RISIS Core Facility and all its connected components. It will be the task of the working group to monitor these activities, and evaluate what are the costs that need to be shared between members. We expect a clear picture to start emerging at Year 3 of the project.

c) The development and monitoring of RISIS "indirect contracting" approach: We have explained above that RISIS as such is constrained by the "private" nature of the main sources we mobilize. The "hybrid model" that has been built enables opening to colleagues for publishable research, but not to address specific stakeholders' needs. The latter can only be satisfied by the operators of the datasets we have opened, and for which the operators have the right to sale studies and consultancy work. The more the datasets are enriched, the more integrated they are, the more dataset operators mobilise RISIS resources while making these contracts. This is why we have agreed on a principle, that will be part of the consortium agreement we call "indirect contracting": it agrees that RISIS is entitled to receive a percentage of the contract for the use of RISIS services. The working group will be in charge to operationalize this principle, both in term of defining what is a relevant percentage, and in term of modalities: where to locate it, how to use it, and what for. It will also monitor the deployment of this new resource, as it is an important part of the long-term sustainability of the infrastructure.

d) The identification of options to organise lasting operation of the infrastructure. Long-term sustainability requires that we find adequate legal solutions for the infrastructure. Do we need a specific legal structure (and under which format: ERIC, NGO…)? Are there conventional solutions that can prove as efficient? What types of governance mechanism, especially when there are very different investments by partners? The working group will be in charge to study the pros and cons of each solution, and to organise the discussion with member institutions about the preferred approach.

e) The recognition as a European infrastructure of interest. The working group will be in charge to study whether the long-term sustainability of RISIS requires an official recognition by the European Strategy Forum on Research Infrastructures (ESFRI), and to propose directions to the Governance board about the type of relationships that should be looked for.

3. Implementation

3.1.Work Plan, work packages, deliverables

3.1.1. Project structure

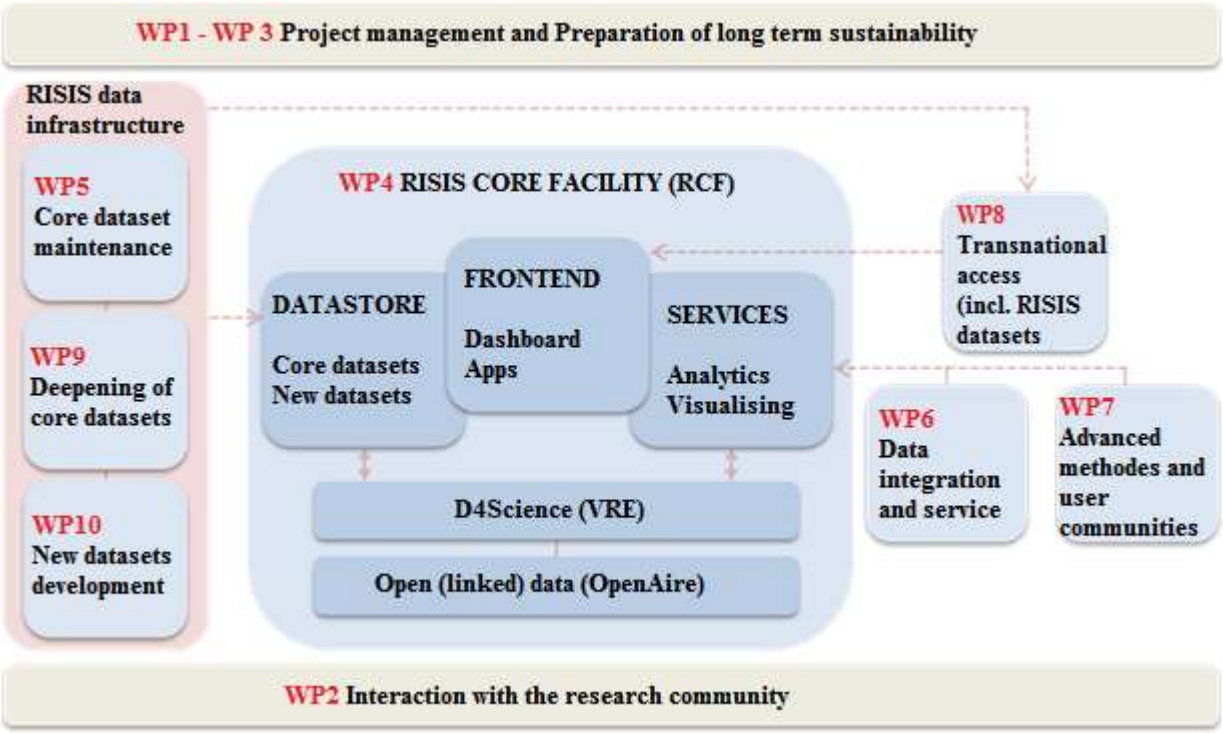
The project, like the RCF, is organised around 3 major dimensions and activities:

a) A front end, focusing on users, the ways they access RISIS, work within RISIS and build RISIS user communities. At the core is the RISIS Core facility (**WP4**). The core facility supports virtual transnational access (**WP8**) and is accompanied by all the efforts we do to raise awareness, train researchers and interact with them (WP2) and to help them build active user communities (mobilising D4Science VRE, **WP7**).

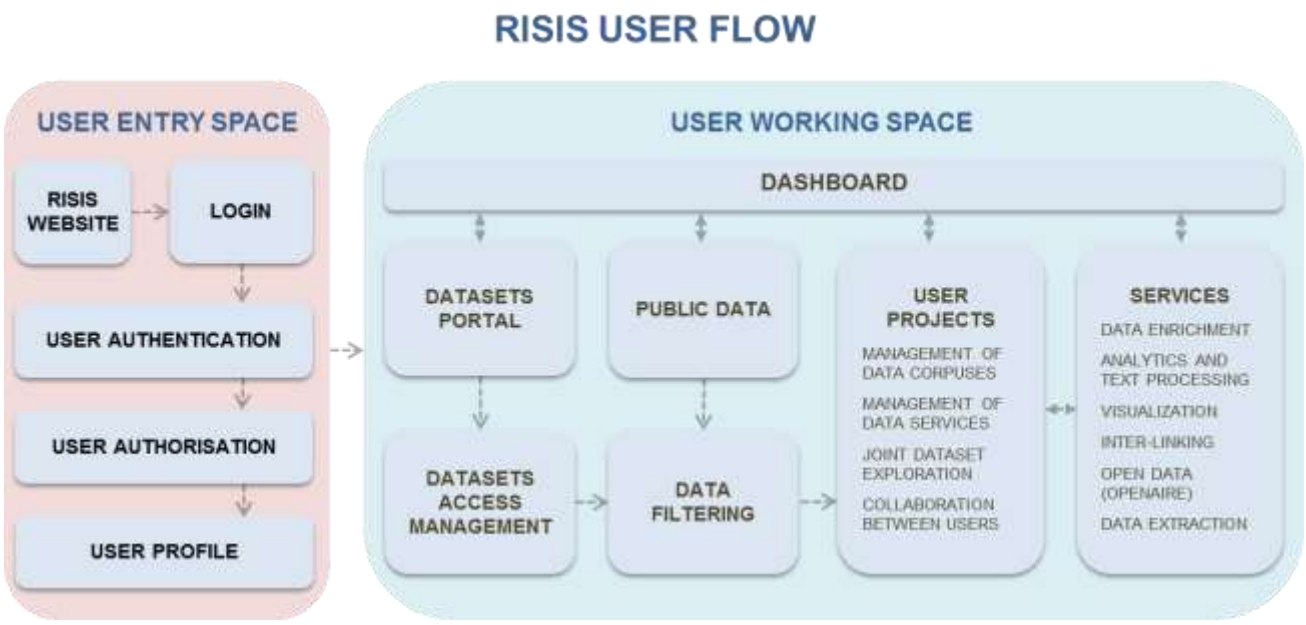
b) A service layer that helps users organise problem based integration of RISIS datasets (with possibilities to complement with their own datasets) – this entails the data integration and analysis services (WP5) and methodological support for advanced quantitative methods (WP6).

c) A data layer that gathers the core RISIS datasets that we maintain (**WP5**) and enlarge (**WP9**), the datasets of interest for which we insure reliability and harmonisation for integration (**WP4**), and the new datasets that we develop and will progressively open (**WP10**).

This is complemented by transversal activities – management both operational (WP1) and strategic (WP3) supported by ethical reviews (WP11), communication and dissemination & exploitation activities (WP2). Table 3.1.a lists the work packages while graph 3.1.a and 3.1.b provide an organisational and functional overview of the project.



Graph 3.1.a Organisational view of the project



Graph 3.1b A functional view (from the side of users)

The different work packages in term of networking, access and joint research, are organised as follows

- We have chosen to gather all access activities in one work package (**WP8**).
- We have organised networking activities to support our community (WP2), to favour their integration and analytical capabilities (**WP6 and WP7**) and to insure continuity in relevance through the maintenance of our datasets (**WP5**).
- We have clear joint research in order to offer new reliable data sources, either enlarging the coverage of existing datasets (**WP9**) or creating new datasets (**WP10**).
- And we have **WP4**, which for the project internally is a major development (Joint Research), but seen from users it is simplifying and widening access.

3.1.2. Deployment of the project and RISIS timing

The project is structured over a four-year period. We have one central critical deliverable and milestone at Month 18, when the RCF opens. We wish to see most of our developments implemented at the end of year 3 (M36). We have a last year with full deployment, which is critical to have a real measure of enlargement. Our first key central date is thus M18. At that time we shall have the following achievements:

- Opening of the RCF and with it of virtual transnational access (**WP4 & WP8**)
- Opening of RISIS VRE will be conducted (**WP4**)
- In term of data services (**WP6**), the new service for topic-based annotation (GATE) will have been opened (from M12) with a first demonstration online; the geo services will provide new resources for ad-hoc choice of the level of geographical agglomeration,
- In term of actor based data integration services (**WP6**) ORGREG will have witnessed its first enlargement (US universities, European hospitals) and FIRMREG will open a first beta version (at M18).
- Our support to the use of advanced quantitative methods (**WP7**) will have produced a first round of 5 methods with documented use cases and research workshops.
- All core RISIS datasets will have been opened from M1 and most will have witnessed one round of updating (**WP5**)
- A first version of the dataset of social innovation projects (ESID) will have been opened (at M12) (**WP10**)

Our second central date will be M36 by which we expect most developments to have produced their effects so that users can have one full year of full benefits from the fully-fledged RISIS2 infrastructure:

- All RISIS data integration and analysis services will be accessible by users through the RCF in their workspaces (**WP4**)
- The VRE and all linkages to external resources will be fully operational. (**WP4**)
- All improvements to our data integration services (in particular the full tailoring of the topic-based GATE services) and to our analysis and visualisation services will have been conducted. The core improvements to our two actor registers (ORGREG and FIRMREG) will have been conducted (**WP6**)
- The full realm of documented case of advanced quantitative methods will be online and collaborative research promoted by the use of the VRE (**WP7**)
- By this date all core RISIS datasets will have been fully updated (**WP5**), except CinnoB (M42); and all targeted developments for core RISIS datasets (**WP9**) will have been conducted (opening of CinnoB, first SIPER enlargement on the performance of policy measures, the addition of NATPRO to EUPRO, first enlargements of VICO and Cheetah on firm innovation patterns & business models, all CWTS publication database additions on acknowledgements, the role of open science & the articulation with social media)
- A first version of the 4 new datasets (**WP10**) (on social innovation projects, on trademarks, on funding instruments and on PHD non-academic careers) will be open for research usage (in some cases final versions will only be available at month 42).

Graph 3.1 below proposed a synthetic view of the achievements planned. As the project is mostly about giving access to improved and new datasets, and to improved and new services, we consider that milestones telling about dates of opening for researchers, and activities developed to support them provides a better view of the dynamics and achievements of the project. This is why we have positioned here Graph 3.1. It is a type of simplified Gantt chart that shows the dates at which progressive new services are offered to our user community. It clearly highlights the two main turning points of the project: at month 18 when we open the RCF and at month 36 when most of our anticipated developments are online and accessible by researchers.

Graph 3.1: Synthetic view of the achievements planned

	M1	M12	M18	M24	M30	M36	M42	M46
FRONT END								
WP3 Interaction with research community	Website operational	First newsletter + 6 Training courses	First batch of policy briefs	8 Training courses	First batch of policy briefs / participative events	10 Training courses	Second batch of policy briefs / participative events	10 Training courses
WP4 RISIS Core Facility (RCF)			Opening and operation of RCF					
WP8 Transnational Access	Dataset portal operational for Transnational Access		Access procedures integrated in RCF					
SERVICE LAYER								
WP6 Data integration & analysis services	Cortext, OrgReg and geoservices opened	Opening of GATE + VosViewer + OrgReg enlarged I	Enriched geoservices opened + FirmReg V.1		Extended annotation services + OrgReg enlarged II	New visualisation tools + FirmReg V.2		
WP7 Advanced quantitative methods		2 use cases on line & first workshops		5 use cases, workshops finalized		5 use cases debated on the VRE		
DATA LAYER								
WP5 Maintenance of RISIS Core datasets	13 Datasets open: 9 core datasets + 4 Datasets of interest	Updates: Cheetah, CWTS Pub, EUPRO, SIPER Enriched IFRIS Patstat		Updates: VICO, CWTS Pub, EUPRO, SIPER	IFRIS Patstat new version	Updates: Cheetah, CWTS Pub, EUPRO & SIPER	New version CIB2 / CinnoB	Updates: VICO, CWTS Pub, EUPRO & SIPER
WP9 Deepening of RISIS core datasets		New dimension on open science of CWTS Pub DB opened in ORGREG	NATPRO pilot countries part of EUPRO + investors data incorporated in VICO	Acknowledgements in CWTS pub + SIPER enlargement V1	V.1 of CinnoB + VICO & CHEETAH enlarged V.1	NATPRO part of EUPRO + Social media in CWTS pub	SIPER enlargement V.2 VICO & CHEETAH enlarged V.2	
WP10 New datasets		ESID V.1		Trademarks V.1, EFIL V.1	PhD V.1	ESID V.2	Trademarks V.2, EFIL V.2	PhD V.2

3.2.Management structures, milestones and procedures

3.2.1. Management structures and their working

The central strategic decision making body will be the **Governing Board** of project members. Our choice is one voice for each member, and the need of a two-thirds majority for each decision. It will meet at least once a year physically and this will be at the end of the annual **RISIS week** which is an important feature as it enables not only all participants in the project to meet but also members to follow in detail what is taking place in the project. This will be complemented by an electronic voting procedure in case of urgent problems, new events requiring RISIS to take a position or in case of lasting disagreement in the Coordination Management Committee. The coordinator is responsible of the use of this electronic procedure.

The Governing board is advised by a **Strategic Advisory Board (SAB)**. Its role is to discuss long-term choices, including major shifts compared to the strategy in place. This high level body will be made of 6 persons, coming both from the academic and the policy side and knowledgeable about quantitative studies and indicators. We expect members to participate to RISIS weeks so as to have a deep knowledge of the project and to give their views and suggestions in a specific exchange with the Governing Board, before its meeting.

We have chosen a simple operational management structure, with **one Coordinator** (that also coordinates the WP management) and the **8 Work Package Leaders**. Each WP leader is responsible of the activities of his/her work package. Together they build the **Coordination Management Committee (CMC)** that meets once a month with a specific role of insuring all the connections between WP (which play a central role in the effective delivery of the project). Every 6 months the CMC does a complete review of activities, based upon WP leaders reports. All necessary adjustments require the agreement of the CMC (with a 2/3rds majority rule if needed, but working in PRIME and in RISIS1 told us it very seldom happened. In case of absence of relevant majority, decisions are sent back to the Governing Board through the electronic procedure, see above).

The coordinator prepares, with the support of the CMC, once a year an overall report on the development of the project and submits to the Governing Board the necessary adjustments and/or new developments that are deemed necessary. The report is sent at least 15 days in advance to the members of the Governing Board and of the Strategic Advisory Board.

The scientific management of the project is supported by the **Project Review Board** in charge of the external evaluation of projects asking for access (on similar grounds of the very successful RISIS 1 Project Review Board) and by the **RISIS Ethical Committee (REC)** in charge of all ethical issues and in particular of privacy issues (both for dataset construction and dataset uses). The committee will be mobilised both ex-ante (when the institutions of proposers do not have their own ethical committees) and ex-post when seized by project members about potentially un-ethical behaviours. Researchers selected can also call on the REC in case of disagreement about the “privacy” constraints put on them.

The PRB is made of one president and of a group of selected peers that the president mobilises depending upon projects. The president of the PRB is part of the Strategic Advisory Board. He can select within members of the PRB, “thematic” vice-presidents if judged necessary. We wish a limited but operational Ethical Committee chaired by a specialist of these issues in our domain.

In order to prepare for the future, a specific **Sustainability Working Group** will be created from the start, made of representatives of members, to implement our sustainability plan (see section 224). It will report to the CMC every 6 months and to the Governing Board at each meeting.

The **management team** assists the coordinator and is in charge of:

- All administrative and financial matters

- The organisation of RISIS weeks and Governing Board meetings
- Supporting the activities of the SAB
- Coordinating the sustainability working group
- Coordinating the data management plan, and in particular organising the full legal review at the start of the project (see section 2.2.3.).

3.2.2. Critical Risks analysis

Few of our milestones face 'critical' risks. Many of them may witness delays without endangering the overall deployment of the project, nor requiring changes in the operational choices made. They will most of the times drive users to use pre-existing versions for a longer time than expected (either of datasets or services). This is why we have chosen to present a detailed analysis of the risks we face work package by work package. It clearly shows, we hope, that our approach has been to minimize risks in the development processes adopted. The reader will have thus a view of the many risks we face locally and how we have prepared even for these. This analysis thus explains why we have only selected few 'critical' risks gathered in the critical risk table.

The sequence below explains them and closes up with **Work Package 4** and the RCF, which is the transformative dimension of the overall project. We also consider management risks, not on the day to day management but on the three critical aspects not for the project itself but for after the project for its sustainable future.

WP2 risks: WP2 is about communication and interaction activities. We inherit from RISIS1 an established team made of one specialist in our field (from CNR) and one communication specialist (from UNIMAN). The team has been working for now one year, has established robust connections with most partners, has developed a stream of videos, has organised very successful events (at the Commission, within the STI conference, not withstanding the final RISIS event end of June 2018 in Brussels with the Commission) and has established a full process for policy briefs writing (with training and support to teams). The team is further complemented by the dedicated training management team in CNR that has organised the full process for the very successful training activity in RISIS1 (both in term of pushing colleagues to organise training, in the success of these training events and on the wide availability of the training material developed). So we do not see any significant risk in meeting the different milestones considered.

Risks associated to RISIS Core datasets (WP5, WP8 & WP9): While in RISIS 1 preparing for opening was a challenge entailing a set of risks which we had to address by evolving in our approaches to technical and cognitive harmonisation, we no longer face this risk in RISIS2. They are all opened, all well documented with complete metadata, and, for many, with a set of specific tables that simplify the use of our data integration services. All datasets have also gone through at least one round of updating, so that procedures are now well established, and even in cases when updating is really complex (in particular IFRIS Patstat and CIB), it is now fully mastered and documented (with all the enrichment scripts published in CORTEXT Github). So we should not encounter the delays we have faced in RISIS1 in term of anticipated updating. This has also driven us to a more pragmatic approach for datasets where updating is not only made of recent additions but requires extensive longitudinal transformations (this is in particular the case of firm datasets). Furthermore certain datasets (like RISIS-ETER, MORE or PROLIFE) have their updates funded as part of specific projects, which reduces risks hugely.

A specific analysis needs however to be made in term of targeted additions. We have taken three major risks: in the enlargement of EUPRO to integrate NATPRO as additional EUPRO module on nationally funded projects; in the analysis on 'innovation modes' for VICO and CHEETAH, in the inclusion of a new dimension in SIPER about the performance of policy measures. The latter are examined in Box 3.2a. It shows that we might face delays, but we should not face issues of reconsidering the working process.

Box3.2a - The analysis of risks taken for the development of RISIS Core datasets and measures to address them

For EUPRO we have made tests about availability of data online with respect to the NATPRO extension. While date seem to be publicly available via the web for many large research funders of EU 15 member states, data availability for the newer member states seems critical. This has driven us to devise a specific process (and add a specific partner to RISIS) in order to engage in data collection for central and eastern European countries (also tackling respective language issues) so that they are present from the start. We might face delays, but we do not expect to have to reconsider our ways for gathering information.

For SIPER, we need to build an ontology for qualifying the performance of policy measures. This applies on all the evaluations already in SIPER. Even though, there are external resources, in particular at OECD and World Bank that can help in this process, this is a really new development. In a way, it is made possible because of the existence of RISIS services: CORTEXT will help for the internal analysis of the corpus of evaluations and in the progressive building of a relevant ontology; and GATE will provide for systematic annotation once the ontology will have been stabilised. What we do not master at this stage is the level of manual input that will be required to arrive to an 'operationalisable' ontology. There is thus a risk of limited deployment, which we have to face, even though we have tried with UNIMAN to seriously quantify the level of investment required.

For VICO and CHEETAH, analysing innovation practices and business models of firms (which are not based on traditional technology based innovations) requires extensive data mining and scraping, based upon the identification of the core vocabulary corresponding to 'non technological' and/ or business model innovation. The risk we face is of two natures: one lies in the information we can retrieve (and thus on a partial coverage of firms); the other is in the analysis and the identification of a set of core patterns of innovation. The team being specialised in both aspects (innovation patterns and data mining techniques), we may face delays, but do not foresee any need in considering alternative approaches. A contingency plan, if needed, will be to limit the scope of the activity to a core set of firms.

As far as access is concerned, we have already tested ways in which we give access to multiple datasets without having to have physical visits to all, mostly by gathering at one place the extracts needed so that one visit is enough. We have also tested de facto 'virtual' access by transferring the 'elements' of datasets needed

to the researcher for conducting his/her project. This is time intensive for dataset owners, but it enables 'virtual' transnational access to de facto operate from month 1 of RISIS 2. This is not a satisfactory situation, but we operate it as a transitory solution not to penalise our users. This mode of operation is also an answer to the risk of delays in the opening of the RCF.

Risks associated with new datasets (WP10): The risks differ widely between the 4 new datasets planned, and we consider that there are major risks in only one case.

The social innovation projects dataset (ESID) is presently being piloted in a companion project (KNOWMAK) and thus it guarantees a first opening at the end of year 1 (still already a few thousands projects included). The main risk lies in the extension and the power of the machine learning process that has already been developed and calibrated to move from existing online datasets to 'search in the wild' individual social innovation projects. The risk is thus less in the operation of the approach than in its ability to 'cover' well the set of projects 'out there'. Apart from big numbers the only way to know better lies in interactions with social actors in given places or in given domains so that they estimate the degree of coverage of the dataset.

Similarly the European dataset on trademarks is built upon pilot experiments already done by the most specialised team on this topic in Europe, ISI Fraunhofer. The central issue lies less in the dataset itself than in its articulation with our registers (of firms and organisations) which raise complex matching issues that both Fraunhofer and UPEM have in great part already faced with patents. We are thus confident to find adequate technical solutions, but the time and effort required for this might drive to potential delays. This is why we have proposed one intermediary milestone (for a pilot version) so as to better master when we shall be able to fully open this new important dataset to study 'non technological innovation'.

The dataset on funding instruments is based on a large scale pilot exercise conducted in another EC project (PREF). This has enabled to already develop the conceptual framework and test the operational difficulties to face. Like NATPRO above, the main problems lie in online access to data. As we wish to cover Europe, like EUPRO, we have introduced a new partner in charge of covering Central and Eastern European countries. So we may face issues of partial coverage, but these should be solved progressively, so that we are rather confident that our research users will be able to use the new dataset at least for one full year.

The main risk we take however is about the fourth dataset on Human Resources. The topic is so important (the non-academic careers of PhDs) that we consider it critical to devote important efforts to it. We face one major challenge: our ability to identify in the multiple CV datasets, individuals with a PhD and being in a non 'research' position after a certain number of years. In order to limit the risks the choice is to build first a calibrated PhD dataset from the multiple sources available today. Using a first set of countries where there is a long tradition of information about PhDs and complementing it by resources from other RISIS datasets, will help to define a robust methodology for extending it at EU level, insuring also an adequate coverage of fields. We shall open this first dataset that should enable to address our two central research questions: identifying a limited set of profiles and critical factors enabling them; addressing the role of the PhD in such profiles. But this still remains a challenge as the way of building the dataset remains fully open. We thus intend to work step by step to be in a position to rethink the process if needed.

Risks associated to RISIS services (WP6 and WP7): RISIS 2 starts with an online platform CORTEXT that is user friendly, widely used and offers access to our presently developed other services on actors (access to the ORGREG platform) and on geography (CORTEXT geo-services). Our users thus already benefit from a first and significant layer of data integration and analysis services. We have set targeted developments for the 3 services, most of which do already exist at the pilot stage and/or have tested the methodologies to use.

We have said that we wanted to introduce a third service based upon the ability to read transversally our datasets (as well as outside datasets) along specific topics. This is a crucial entry with the rise of policy objectives associated to emerging technologies, societal challenges and now so-called 'missions'. None of them correspond to a given classification, thus the importance of such semantic based ontology driven services. Our choice has been to integrate one of the most used platforms in Europe GATE based on document analysis and the word embedding approach. This has already been tested in a companion EC project (KNOWMAK) that had enabled to organise a tagging process for millions of items, and to build a dedicated learning base for field relevant ontology building. We have no doubt that the GATE platform will be open for RISIS users at latest at the end of year 1. But we anticipate with Sheffield colleagues who operate GATE that to tailor it fully to our field and to make it user friendly will require time, and a set of demonstrator cases. This is why we have taken a conservative approach to its full opening, that is month 30. Again there is work to do, but no real critical risk faced.

Our last service aims at empowering our user base for the use of advanced quantitative methods. This is linked to the very specific nature of our datasets that makes it difficult to mobilise the new quantitative methods. We have identified 5 major ones for a first round. The approach has been to select teams in the consortium that already master one of these methods to apply it to a relevant case for the field mobilising if possible more than one dataset, and to document it fully (in term of data preparation, of handling missing values, of treating outliers, of mobilising them in the existing software and of analysing results). We know that it is a central condition to enlarge the use of RISIS datasets, and we have very dedicated partners pushing for this development. The organisation of research workshops where interested users come with their questions and their combined dataset is not either a source of risk (all our training sessions have always had a queuing line!). What is less measurable is how these interested users will seize the virtual research environment that we shall propose for going further in organising the take-up of such methods and in building advanced communities.

So all in all, we do not anticipate any major risk coming from our data integration and analysis services. This has been a clear choice made: what we are looking for is the use of services and how they enlarge and change the community. To focus on this, we need to minimise developments and focus on tailoring services to the needs of our community, and making them user-friendly for this community (taking into consideration its specific characteristics). Our progressive approach (initiated in RISIS1) enables us in RISIS2 to very fast offer the full set of services we consider critical for the future of the community.

Risks associated to RISIS Core Facility (WP4): The RISIS core facility will give access to researchers to all RISIS datasets in a linked structure, and to a very extensive set of services and web applications to explore, connect and analyse those data (T.1). It will also connect users to the Open Data communities by constructing a bridge between the core infrastructure of RCF and a specifically designed D4Science VRE (T.2), which provides a direct access to Open Datasets platforms and first to OpenAire. What are the risks we face for such a 'breakthrough' development for our community? For the opening of the RCF we have identified 2 major risks on the effective design of the architecture and on software development.

Concerning the design of the RCF, the risk taken are related to the capacities of the architecture to be aligned with the goals it is supposed to achieve, and thus that the tool does not match our user-driven expectations. To address this we have chosen two complementary approaches: one is not to get trapped in the preferred solution of one actor. This is why we have chosen a distributed development based on the IT competences already demonstrated by partners with their own platforms (UPEM, Sheffield, Joanneum, ISTI) and on data handling (Leiden and AIT). This goes along, our second choice, with an important coordination task (T.3.3), which puts management and coordination tools in the WP, to keep it in coherence with research problems. This will be built on a solid documentation website describing the conception and development parts (building one of the key deliverable of the project, D3.5), an on-line development forge (code versioning, issue

tracking, technical documentation), and a project roadmap and task management on-line application that will handle the tasks discussion and decision-making stages. On top of that, developing frameworks will be used for the development of the architecture, to ensure good practice and the continuation of the code.

As for the software development part, the risk taken is on the production of the main frontend to open it for users at month 18. This has been a long-standing preoccupation well recognized in RISIS1 and with an amendment in the project so that we could prepare for it. A first front-end has thus been already prototyped and tested. And extensive work has been done with the core datasets to already integrate them in the future infrastructure. This has enabled to already develop a proof of concept of the future RCF. For us it is the insurance that the kind of frontend we are building will be suited for the problems we are trying to solve and that it is within our reach.

Contrary to our failed attempt in RISIS 1, to connect our users to the Open Data world, we have chosen to integrate a well-established project with already numerous applications online, D4Science. Involving them reduces risk dramatically as they already have a full-featured framework. Furthermore we are organizing a process, where in a first phase access will be provided directly (as is the case today for our other services, CORTEXT in particular) before in a second phase to be fully integrated in the work processes of our users.

This is why we really expect to be on time for a first release and opening at month 18. We do not expect with this approach to face strong design problems (that would drive us to reconsider the approach selected), but rather delays. The consequences of delays for users are of two natures: (a) their capacity to access at a distance multiple datasets simultaneously, and (b) the conditions of use of our integration services. We have explained in **WP8** risks that we have organized a 'manual' process for users to access multiple datasets early on in the project (either by providing access in one place or in organizing the transfer of the elements needed). This is time consuming for data providers and for the overall organizer of access, but it already works. So the risk that we face is that this 'transition' period lasts longer, but the delay in opening will not endanger point (a). This is the same for point (b): today users using our data integration services have to bring their data to the service platforms to get them enriched and thus prepared for integration. Our ambition with the RCF is that these services are accessible by users from their workspace in a continuous flow. If we are delayed this will not forbid their use, but drive to use longer the present process.

We are also taking risks related to hardware and data storage. The risk is of a natural hazard or a security breach that could lead to the destruction of all or part of the infrastructure. We answer to that risk in two ways:

- First, we involve the UPEM IT Resource centre in the conception and management of the hardware infrastructure (they handle the whole UPEM university servers and network and have more than 20 years of experience; the RISIS-1 project was handled by the same team, and we survived a flood and several major server malfunctions like the loss of disks and power supplies, without any data losses or even an interruption of service of more than a few hours).
- Second, by developing our own monitoring and backup system (T.3.1.1 & T.3.1.6) capable of alerting administrators in case of any problem (being hardware or software related), and restoring the whole infrastructure in case of a partial or total destruction of the data and software.

Risks associated to the preparation for the future (WP3): Should there be any 'critical' risk in management? Not in the traditional sense. We have learnt from past experience and are quite sure that the structures, mechanism and processes will enable to operate in a fluid way the overall project. But within WP1 we have a number of 'strategic activities' that might have strong consequences upon the future of the infrastructure (beyond project time):

a) We need to assess very early on the legal robustness of our hybrid model (see section 2.1 on barriers and section 223). The present approach is based upon bilateral links between private providers and RISIS data operators, and, though initial reviews tell that it should not face major problems, we consider it crucial to have a full legal review that will stabilise the situation. This analysis might drive us to inflexions in this approach. We however do not anticipate major shifts. Starting early will give us time to address the few situations that might require adaptations.
b) The second critical step lies in our ability to jointly develop the operational conditions of the principle established for 'indirect contracting'. We have given ourselves 2 years to operationalize it.
c) The third critical point lies in the basis on which sustainability can be arrived at, that is the fact that our member groups, units or institutes have a lasting mission given by their legal institution, to commit resources to maintain the datasets and/or services they bring to RISIS. This is already the case of the most involved groups, but we need both to enlarge it and arrive at a collective view.

These are 3 critical risks for the future, but not for the project as such. The 3 dimensions are part of the consortium agreement as principles and elements for joint working. Our answer is to give time to maturation, and thus to start very early in the project and organise a focused way to address these issues. We do not see alternative ways to face these risks.

3.3.The consortium as a whole

RISIS combines unique datasets and well established data services. The consortium is based on the developers and operators of both datasets and services.

It gathers first the developers and operators of the 14 datasets opened in RISIS1 (UPEM, AIT, University of Leiden, Politecnico di Milano, CNR-IRCRES, **Fraunhofer (former** University of Manchester), USI, Joanneum, DZHW, NIFU, Sussex University). New members specialised in the field of the new datasets to be developed complement them (Fraunhofer, Strathclyde university, CSIC-INGENIO, Czech Academy of Sciences, Technion). It gathers second the operators of services (UPEM, University of Sheffield, CNR-ISTI, University of Leiden, Athena Research Centre, USI and Sapienza). This is this combined group that builds the overall relevance of this consortium.

What is furthermore critical to the quality of the consortium is that participants are used to work together thanks mostly to RISIS1, but also to companion projects (this is in particular the case with the University of Sheffield or the University of Strathclyde) or to long traditions of bilateral exchanges on data handling (in particular between UPEM and Fraunhofer). All have worked in European projects already. Nearly all have been coordinators of such projects. This is in particular the case of all our WP leaders (coming from UPEM, CNR-IRCRES, AIT, Politecnico di Milano, Univ of Sheffield, and Univ of Leiden), an important aspect as our WP are quite large and complex. Finally, the coordinator has a long experience of complex projects, having been in particular the coordinator of the PRIME Network of excellence (2004-2009) and of RISIS1 project (2014-2018).

All partners do not play the same role:

- A first group is made of the RISIS Core dataset holders: The University of Leiden for the CWTS publication database, AIT for EUPRO, Polimi for two firm datasets (VICO and CHEETAH), **Fraunhofer** for SIPER, UPEM for the third firm dataset (CIB on large firms) and for the IFRIS_PATSTAT DB, USI and Joanneum for ETER and ORGREG. They have extensively worked together in RISIS1 to arrive at a technical and cognitive harmonisation, for testing and deploying internally the integration services, and for working on the future RCF. To these should be added CNR-IRCRES for its central role in organising all the training activities, and the interactions with the research community and our stakeholders.
- RISIS2 is making a clear enlargement of the second group of partners, dealing with data services and with the move toward an e-infrastructure and the open science world. This has driven to widely enlarge the IT-based competences of this social science project, adding to UPEM (its online semantic CORTEXT platform and its geoservices), the University of Sheffield and its GATE platform, CNR-ISTI with D4Science and ARC for integrating VRE in RISIS and organising a lasting link with OpenAire. We are quite confident about this enlargement as all partners have a long experience in working with social sciences. This is in particular the case for UPEM where IT developers and STI researchers are part of the same lab, or of Gate colleagues from Sheffield that work extensively with STI researchers in Leiden and Paris, at AIT and USI to tailor GATE to the issues of our field.
- A third group is made of partners that are fully specialised in one of our critical developments. They are real experts in the developments we are making. This is the case of Fraunhofer-ISI for patents and trademarks, of the University of Strathclyde for social entrepreneurship and social innovation, for the University of Sussex SPRU on firm innovation, and for the partners we have gathered to address the difficult and risky issue of non-academic PhD careers (with DZHW, NIFU, Technion-SNI and CSIC-Ingenio). In term of services, this is the case of Sapienza on data quality issues.

3.4.Resources to be committed

The project effort and support asked for amounts to 8.492M€:

- Management costs represent 7% including the DMP, and preparation for the future.

- Access (transnational, mostly virtual TA) represents 10%.
- Training, interactions with users and stakeholders and communication represent 12%.
- Other networking activities (dealing with data maintenance, data integration services and data analysis services) represent 32%
- Joint research activities dealing with the RCF and with new data sources represent 40%.

In term of type of expenditures:

- 68% of the budget is for Human Resources
- 9% for other direct costs
- 19% for overheads
- 3% for subcontracting.

Work Package number	Work Package title	Person months	Personnel costs	Other direct costs	Sub-contracting	Total
1	**Project Management (NA)**	**42**	**255.250**	**64.000**	**104.312,5**	**503.375 €**
2	**Interaction with the Research community, Communication and Dissemination (NA)**	66	384.000	298.000	160.000	**1.012.500 €**
3	**Preparation of long terme sustainability of RISIS (NA)**	**13**	**80.250**	**13.000**		**116.562,50€**
4	RISIS Core Facility **(JRA)**	150	904.000	58.000		**1.202.500 €**
5	Core Dataset Maintenance **(NA)**	92	566.000	32.000		**747.500 €**
6	Data Integration & Analysis services **(JRA)**	170	1.025.000	63.000		**1.360.000 €**
7	Advanced Methods and User communities **(NA)**	56	350.000	82.000	18.000	**558.000 €**
8	Transnational Access	97.4	613.350	57.400		**838.437,50 €**
9	Deepening the Core Datasets **(JRA)**	140	852.000	62.000		**1.142.500 €**
10	New Dataset Development **(JRA)**	129	760.000	49.000		**1.011.250 €**
11	**Ethics requirements**					

Human resources committed:

As we have just mentioned, the core expense for this infrastructure is in Human Resources (68% of the total budget). The total investment is calculated at 955 person-months with a balanced effort between the IT based infrastructure and services, and more STI based maintenance and development of datasets (respectively 39% and 38%).

Other direct Costs:

The other direct costs represent 9% of the total budget. The core of other direct costs is linked to training sessions and research workshops for advanced quantitative methods (46%).

Access cost has been estimated at 10% It covers only the costs associated to physical transnational access (while virtual transnational access is estimated at 80% of all transnational access planned).

We have limited equipment needs (just for complementing UPEM present capacity for the RCF, **WP4**, 3%).

The collective approach to all our developments, even when using all distance facilities, has a cost visible in the amount of training and subsistence for teams working on the project (34%).

Other direct costs remain in control with only 9%.

We provide one table detailing costs per partner, with specific explanations when costs are important.

Participant	TA[6]	Training T&S & organisation costs[7]	Other T&S (for team)	Equipment	Other goods & services	Total	Justification
1.UPEM		36.000	**60.000**	25.000	15.000	**136.000**	Equipment for the RCF (**WP4**)
2.AIT	57.400	18.000	**23.000**		12.000	**110.400**	AIT manages all T&S associated with physical access
3. POLIMI		54.000	**19.000**		7.000	**80.000**	3 training sessions (WP2) and 3 research workshops (**WP7**)
4.CNR		18.000	**19.000**		**13.000**	**50.000**	
5. LEIDEN		18.000	**19.000**		11.000	**48.000**	
6.UFSD		18.000	**12.500**		4.500	**35.000**	
7.FRAUNHOFER		**27.000**	**18.000**		**7.000**	**52.000**	
8.USI		63.000	**10.000**			**73.000**	2 training sessions (WP2) and 5 research workshops (**WP7**)
9.JOANNEUM		18.000	**9.000**		2.000	**29.000**	
10.UOS		9.000	**8.000**		2.000	**19.000**	
11.DZHW		9.000	**7.000**			**16.000**	
12.STRATHCLYDE		9.000	**7.000**			**16.000**	
13.SAPIENZA		18.000	**7.000**			**25.000**	
14.NIFU		9.000	**8.000**			**17.000**	
15.CSIC		9.000	**11.500**		2.500	**23.000**	
16.SNI		9.000	**11.000**			**20.000**	
17.TC CAS		9.000	**6.000**			**15.000**	
18. ATHENA RC		9.000	**5.000**			**14.000**	
TOTAL	**57.400**	**360.000**	**260.000**	**25.000**	**76.000**	**778.400**	
%	**7%**	**46%**	**34%**	**3%**	**9%**	**100%**	

(according to the budget table in section 3 of the project administrative forms).

[6] All T&S costs for physical access are managed by the coordinator of Access

[7] We consider the cost of 1 training session including travel & subsistence of participants, materials and specific organisational costs at 9000€. When training takes place at a distance, T&A are replaced by costs for the online system we shall use. We apply the same costing for our research workshops on advanced quantitative methods (WP6).

Table 3.4b: 'Other direct cost' items (travel, equipment, other goods and services)

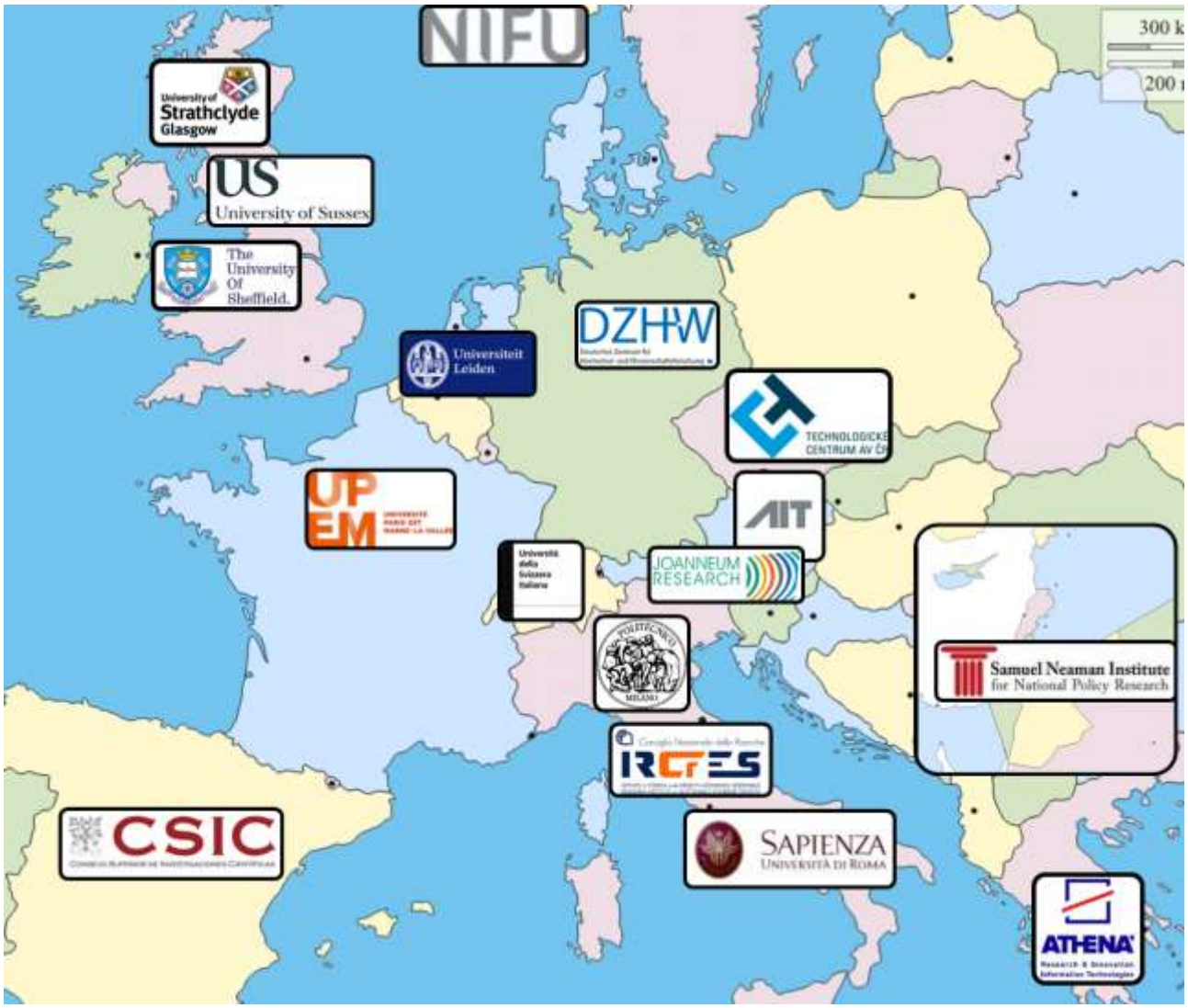
2. AIT	Cost (€)	Justification
Travel & subsistence for trans-national access (if applicable)	57.400 €	Centralisation of all the physical visits' reimbursements (20% of the overall visits so 80 visits)
Other Travel		WP1 : 3.000€ for RISIS WEEK, Governing board, Steering committee WP3 : 1.000€ for the SWG meetings WP4 : 4.000€ for Travel and subsistance WP5 : 4.000€ for Travel and subsistance WP6 : 5.000€ for Travel and subsistance WP7 : 4.000€ for Travel and subsistance WP9 : 14.000€ for Travel and subsistence and consummables
Equipment		
Other goods and services		WP2 : 18.000€ for the organisation of 2 trainings
Total	110.400	

Appendix: list of references mobilised in the project

- Dornbusch, F., Schmoch, U., Schulze, N., Bethke, N., 2013, Identification of university-based patents: A new large-scale approach, *Research Evaluation*, 22, 1, 52-63
- Dornbusch, F., Neuhäusler, P., 2015, Academic Patents in Germany, Studien zum deutschen Innovationssystem Nr. 6-2015, Berlin: Expertenkommission Forschung und Innovation (EFI)
- Georghiou, L. et al., 2008, Challenging Europe's research: Rationales for the European research area (ERA), Report of the ERA expert group, European Commission, Luxembourg
- Georghiou L et al., 2018, Mission oriented research and innovation policy, a RISE perspective, European Commission, doi:10.2777/426921
- Gotsch, M., Hipp, C., 2012, Measurement of innovation activities in the knowledge-intensive services industry: a trademark approach, *Service Industries Journal*, 32, 13, 2167-2184
- Joly P.B., Gaunand A., Colinet L., Laredo P., Lemarie S., Matt M., 2015, ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization, Research Evaluation, 1-14, doi:10.1093/reseval/rvv015
- Larédo P., 2017, The next generation in STI studies, keynote speech for the 50th anniversary of science and innovation studies at the University of Manchester, Manchester, November 4, 2017, available on IFRIS website
- Laurens P., Le Bas C., Schoen A., Villard L., Larédo P., 2015, the rate and motives of the internationalisation of large firm R&D (1994-2005): toward a turning point, *Research Policy*, 44, 765-776. http://dx.doi.org/10.1016/j.respol.2014.11.001
- Laurens P., Villard L., Schoen A., Laredo P., 2017, The artificial patents in the PATSTAT database: how much do they matter when computing indicators of internationalisation based on worldwide priority patents? *Scientometrics*, DOI 10/1007/s11192-017-2578-5
- Laurens P., Schoen A., Yegros A., Laredo P., 2017, Exploration of knowledge in European large firms in the Chemicals and Pharma/biotech sectors: level and mode of collaboration in the corporate scientific publications and patents, available at *https://sti2017.paris*
- Lepori B. et al., 2007, Indicators for comparative analysis of public project funding: concepts, implementation and evaluation, *Research Evaluation*, 16, 4, 243-255,
- Lepori B., Barré R., Filliatreau G., 2008, New perspectives and challenges for the design and production of S&T indicators, *Research Evaluation*, 17(1), 33–44
- Schmoch, U., 2003, Service marks as novel innovation indicator, *Research Evaluation*, 12, 2, 149-156
- Mazzucato M., 2013, The Entrepreneurial State: debunking public vs. private sector myths, London: Anthem Press
- Mazzucato M., 2018, Mission-oriented research & innovation in the European Union, European Commission, doi:10.2777/360325
- OECD, 2015, Guidelines for collecting and reporting data on research and experimental development, Frascati Manual, Paris: OECD.
- Sandner, P.G.; Block, J. (2011): The market value of R&D, patents, and trademarks. In: Research Policy, 40, S. 969-985;
- Spaapen J., Van Drooge L., 2011, Introducing productive interactions in social impact assessment, *Research Evaluation*, 20, 3, 211-218, https://doi.org/10.3152/095820211X12941371876742

4. Members of the consortium

4.1.Participants (applicants)



4.1.1. Partner 1: Université Paris-Est Marne-la-Vallée (UPEM)

Partner profile
The official partner coordinating the project in the name of LISIS will be the Paris-Est Marne-la-Vallée University (UPEM), currently coordinating the European Infrastructure project RISIS1 and Co-creation project KNOWMAK. UPEM is an interdisciplinary institution hosting more than 15 laboratories with around 300 researchers and 11. 000 students. UPEM carries out the missions of public services with two primary objectives: to develop employability of students (10% of the total number of apprentices studying in French universities are enrolled at the UPEM, 94% of employment rate in science and technology) and outstanding research (participation in 4 French Laboratories of Excellence, 1 ERC grant). LISIS - Laboratory in Interdisciplinary Sciences for Innovations and Society - is a Joint Research Unit (JRU) gathering 37 researchers and 33 doctoral students and post-docs in sociology, management and economics from the University Paris-Est Marne-la-Vallée, the Engineering School "ESIEE", the CNRS and the INRA. LISIS is organised around three major stakes for social sciences on innovation. The first is to revisit concepts and frameworks for analysing innovation in society dealing with collective dimensions of innovation processes, the inclusion of controversies and public debates, and shifting values and valuation processes. One important aspect is to analyse organisational and institutional implications of such transformations in innovation processes. The second is methodological with a clear engagement in combining extensive fieldwork with revised quantitative approaches: the unit has a strong component dedicated to new ways to analyse open data through new tools and software tailored to our field (with the CORTEXT platform, geolocation and geoclustering software). Working with societal actors, firms as well as Civil Society Actors is the third major characteristics of LISIS working practices. This has driven to organise activities around 4 'axes': transitions, emergence & transformations; socialisation and 'reappropriations' of innovation; valuation and evaluation of research & innovation; analysis of digital traces of sciences in society. In term of policies it focuses on research and innovation policies and the on-going transformations of higher education. In the latter, LISIS has been engaged into a large project developing and implement a new approach to identifying and measuring societal impacts of public research. CRI (Informatic Resources Center) who will manage the hardware of the RISIS infrastructure is the Centre of IT Ressources of the Paris-Est Université in Marne la Vallée (more than 12 000 students, 6 Gradutate schools, 20 research units). CRI is a resource centre of 31 persons, 2/3 of them being IT Engineers. The CRI is divided in two teams: one dedicated to servers (storage and computing servers) and networks notably to fulfil specific needs of research units, the other one is dedicated to the engineering and management of web application frameworks and IT services for university departments, professors and students. The infrastructure of servers of UPEM is managed and sustained by the CRI. This large infrastructure entails 60 servers mainly operating under Linux (debian), but also HPUX, BSD and Windows. For the last 20 Years UPEM has always supported the existence of an up-to-date infrastructure of computing servers notably to answer the needs of research units. The cluster of computing servers is accessible only to research laboratories. At the moment, the Hardware of processing capacities is based on a cluster of 40 nodes of processing servers (bi-Xeon à 2,8GHZ, 2 G de RAM, 32 bits).

Staff profile			
Name	Philippe LAREDO	**Gender**	M
Short CV / description of work experience relevant to the project	Prof. Philippe Larédo is a invited professor at UPEM, at the University of Manchester (MBS, Manchester Institute of Innovation Research) and he the coordinator of the first RISIS project. He is a specialist of public policies for research and innovation. He has conducted numerous evaluations and been member on a number of EC high-level groups. He has a 20-year experience in strategy-making at the Regional level. From 2003 to 2009 he was the coordinator of the PRIME network of excellence (www.prime-noe. org) and is one of the board members of the EU-SPRI Forum (www.euspri-forum.org) that continues, through direct support from member universities, the main activities. His present interests are on the emergence of new technologies, and on the evolving policy mixes in the field of		

	research and innovation.		
Name	Antoine SCHOEN	**Gender**	M
Short CV / description of work experience relevant to the project	Prof. Antoine Schoen is Engineer in Technology management (Université de Technologie de Compiègne) and holds a PhD in economics (Université Paris-Sud). He has long lasting interests in indicators with work at the French observatory for Science and technics (OST, 2003-2005) and at IPTS (2005-2008). During this period he developed a new corpus on genomics inventions, he managed the PRIME project on the characterisation of universities (OEU) and conducted a transversal analysis of the involvement of universities in the EU Framework Programmes. Since 2008 he is a professor at ESIEE and a researcher at LISIS. This is where he has developed the Corporate Invention Board.		
Name	Patricia LAURENS	**Gender**	F
Short CV / description of work experience relevant to the project	Dr. Patricia Laurens is a researcher from CNRS working at LISIS. Until 2004 she worked at CNRS Laboratory of Applications of High Power Laser as Researcher and was the project manager on Laser-Matter Interaction and Laser Surface Treatment. From 2004 to 2010, she worked at OST as project manager in bibliometrics, R&D indicators, and the exploitation of R&D databases. Since joining IFRIS-LISIS she has specialised in the analysis of internationalisation patterns of multinational firms and has become a specialist in the use of patent data.		
Name	Marc BARBIER	**Gender**	M
Short CV / description of work experience relevant to the project	Marc Barbier is Researcher at INRA, head of the INRA unit at Cité Descartes (INRASenS) and director since 2009 of the CorTexT Platform for Textual Analysis and Corpus of IFRIS. He specialises in the longitudinal analysis of socio-economic and socio-political transformations of the agricultural word. This has driven him progressively focus on quantitative approaches and on the semantic analysis of large textual corpuses.		
Name	Lionel VILLARD	**Gender**	M
Short CV / description of work experience relevant to the project	Lionel Villard (M) is a senior lecturer at ESIEE and a researcher at LISIS. His research focuses on data mining, sci-entometrics and visualisations, dealing in particular with geographical agglomeration and knowledge dynamics in the development of high-tech fields (e.g. nanotechnology, biotechnology). He specialises (1) in processing large volumes of data (publications, patents, press articles and financial data) collected from heterogeneous sources in structured databases; (2) in their analysis, in particular the spatial projection of this data (including mapping through GIS); (3) in producing quantitative indicators. These tools and methods built as part of his research pro-vide necessary foundations to the analysis of scientific and technological fields. Currently he is an active member of the European Infrastructure project RISIS.		
Name	Philippe BREUKER	**Gender**	M
Short CV / description of work experience relevant to the project	Philippe Breuker (M) is an INRA senior IT engineer working at LISIS. He is specialized in web development and data mining. He has a background in science and a degree in engineering. He has 15 years of experience in web development, server architecture and maintenance. He has been and is the developer of the Cortext platform since 2008. There, he conceived and developed the architecture of the platform (backend and frontend) as well as its server computing infrastructure. He was also in charge of maintaining its information system. Within INRA he has been coordinator of an IT engineering structure on Textual Analysis (www6.inra.fr/cati-icat), and part of a work group on Open Data (www6.inra.fr/datapartage). He has led the IT team preparing the RISIS Core facility during the RISIS1 project.		

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
The rate and motives of the internationalisation of large firm R&D (1994-2005): toward a turning point?	Research Policy, 44, 765-776	2015	Laurens P., Le Bas C., Schoen A., Villard L., Larédo P.
Logics of integration and actors' strategies in European joint programs	Research Policy, 43, 2, 391-402	2013	Lepori B., Reale E., Larédo P.
Policy-making in science policy: the 'OECD model'unveiled	Research Policy, 42, 3, 801-81	2013	Henriques L. and Larédo P.
ASIRPA: A comprehensive theory-based approach to assessing the societal impacts of a research organization	Research Evaluation	2015	Joly P.B., Gaunand A., Colinet L., Laredo P., Lemarie S., Matt M.
The artificial patents in the PATSTAT database: how much do they matter when computing indicators of internationalisation based on worldwide priority patents?	Scientometrics, 114, 1, 91-112	2017	Laurens P., Villard L., Schoen A., Laredo P.

Relevant previous projects or activities			
Project name	RISIS - Research infrastructure for research and innovation policy studies (EC-GA N°313082) http://risis.eu/ coordinated by UPEM	**Dates (start/end)**	2014-2018
Short description	The RISIS project aims at creating a distributed research infrastructure to support and advance science and innovation studies. This will give the field a strong scientific push forward, and at the same time, provide a radically improved evidence base for research and innovation policies, for research evaluation, and for the quality of policy relevant indicators. The field of science and innovation studies is interdisciplinary, and is related to political sciences, sociology, management and economics. It has a strong quantitative core - with specialties such as scientometrics, technometrics and more widely indicators design - but for many important questions data were lacking or small scale only. This has made the field too much dependent on a few pre-existing datasets. However, during the last decade important efforts have been undertaken to develop new datasets on burning issues such as industrial R&D globalisation, patenting activities of firms, university performance, Europeanisation through joint programming, or the dynamics of nano S&T. Another new characteristic of the field is the development– together with computer scientists – of software platforms for collecting, integrating and analysing ever more data.		
Project name	KNOWMAK - Knowledge in the making in the European society (EC-GA N° 726992) http://www.knowmak.eu/ coordinated by UPEM	**Dates (start/end)**	2017-2019
Short description	KNOWMAK project aims at developing a web-based tool, which provides interactive visualisations and state-of-the-art indicators on knowledge co-creation in the European Research Area (ERA). It is structured around three integrative elements: Research topics (by developing ontologies around Societal Grand Challenges and Key Enabling Technologies), Actors (with a focus on the quadruple helix and the involvement of societal actors in knowledge co-creation) and Geographical spaces (with a focus on multiple level metropolitan, regional, national and European spaces and their interconnectedness). The tool combines three main data sources: established indicators of scientific and technological knowledge production based on scientific publications and patents; information on		

	knowledge in the making derived from research projects' descriptions; information on social innovation projects and user attention to knowledge production derived from the Internet and from social media. The integrative elements (topics, actors, space) allow for the interlinking of data items, to produce a characterisation of different dimensions of knowledge in the making.		
Project name	VERA - Forward Visions on the European Research Area (EC-GA N° 290705) http://eravisions.eu/	**Dates (start/end)**	2012-2015
Short description	The VERA project aims to provide relevant strategic intelligence for the future governance and priority-setting of the research, technology, development and innovation (RTDI) system in Europe and for better adapting science, technology and innovation policy to the shifting global environment and upcoming socio-economic challenges. For this purpose VERA carries out an in-depth stocktaking of RTDI related forward looking activities in Europe and internationally and a thorough review of trends and drivers of long-term change of European RTDI governance. On the base of these insights VERA develops scenarios on the evolution of the European Research Area, assesses the critical issues for the ERA's future capabilities emerging from these scenarios, explores subsequent strategic options and ultimately generates a set of policy recommendations for responsive and future oriented multi-level, multi-domain RTDI policy strategies.		

Relevant infrastructure	
Name	**Description**
CIB (Corporate Invention Board)	The CIB (Corporate Invention Board) dataset is a database characterising the patent portfolios of the largest industrial firms worldwide. The CIB combines information extracted from the Industrial R&D Investment Scoreboard (EU Commission), the ORBIS financial database and the PATSTAT-IFRIS patent database - an enriched version of the Patstat EPO database.
NANO	The Nano S&T dynamics database (Nano) developed by LISIS collects publications and patents between 1991 and 2011 about nano S&T. One central characteristics of emerging S&T is that they do not correspond to pre-existing categorisations and require the elaboration of semantic based queries. IFRIS has developed a dynamic query gathering 1.18 million publications and 735000 priority patents. Four types of enrichments have been organised dealing with: (i) categorisation and harmonisation of institutional affiliations, (ii) geolocalisation of all authors and inventors; (iii) geographical clustering of S&T activities; and (iv) thematic clustering of S&T activities. It offers for each publication, 14 'main units of observation' and 11 for patents.
CORTEXT	CorText Manager provides an online platform interfacing users with a range of analysis tools developed by CorText corpus.This application enables to upload data sets from disparate sources, and initiate treatments (scripts) to perform remote analyses/maps of primary data. A private space allows users to launch and test on their own chains of treatments before possibly making results public.

Contribution to the project
In RISIS2, UPEM will bring its specific competences concerning the design and operation of computer platforms for overall RISIS integration (the RISIS CORE Facility) and for the management of data analysis tools (the CORTEXT platform) complemented by the UPEM IT service for the implementation of the overall IT structure. UPEM will also bring its specific competences on geographical analysis (wit the CORTEXT geoservices). It will finally bring its long lasting competences on two specific datasets linked to firm innovation capabilities (the CIB dataset for large firms) and on patent outputs (the IFRIS Patstat DB). UPEM has extensive experience in coordinating large-scale networks in the field of S&T policy and indica-

tors, as witnessed by the coordination of the PRIME Network of Excellence, the KNOWMAK project and of the Infrastructure action on Research infrastructures for the assessment of science, technology and innovation policy (RISIS1). Work undertaken within KNOWMAK and the first RISIS project to integrate research infrastructure in the domain of research policy and S&T indicators will be key element for the implementation of RISIS2.

4.1.2. Partner 2: Austrian Institute of Technology GmbH, (AIT)

Partner Profile
With about 50 employees, the Center for Innovation Systems & Policy at AIT Austrian Institute of Technology is Austria's largest and leading research institute in the area of research, technology and innovation policy. It has been active in the field since the late 1980ies and contributing to a broad range of research and policy initiatives in the field since then. The Department is using regularly a broad range of national and international statistical data sources and STI indicators, with in-house access to key patent and publication data bases, data on R&D collaborations, finance, various types of survey data, etc. With the EUPRO database AIT disposes of the most comprehensive and best standardized database of EU framework programmes, covering FP 1 to FP 7 and allowing a vast range of econometric, spatial econometric and social network analysis of R&D collaboration patterns in Europe. The Department has strong in-house expertise in quantitative methods such as econometrics, social network analysis, survey techniques, text mining, simulation and modeling, but also in qualitative methods of enquiry like interviews, workshops and focus groups. Over the past years, it has also built up a strong track record in forward-looking methods such as scenario development, Delphi-techniques, technology roadmapping, and identification of weak signals and emerging technologies. The Innovation Systems Department has long-standing experience in both RTI policy research and policy advice to regional national governments, European institutions and international organisations. It is member of national and international networks for policy advice (platform evaluation, European Techno-Economic Policy Support network ETEPS, ERAWATCH network) and scientific research (e.g. EU-SPRI European Science Policy Research Initiative).

Staff profile			
Name	Thomas SCHERNGELL	**Gender**	M
Short CV / description of work experience relevant to the project	Thomas Scherngell (acting as AIT internal project coordinator) joined the Innovation Systems Department of AIT in December 2007. He holds a venia docendi (habilitation) in Economic Geography and Regional Science, received from the Vienna University of Economics and Business in 2012. He received his M.A. in Economic Geography (2003) from the University of Vienna (Mag. phil.), and his PhD (2006) from the same university (Dr. rer. nat). Prior to working at AIT, Thomas Scherngell was a full time university assistant at the Department of Economic Geography and GIScience at the Vienna University of Economics and Business Administration (2002-2007). He is an expert in economics of innovation and technological change as well as in regional science and spatial analysis. His research on R&D networks in Europe have been published and cited in major international journals. Further he has been working extensively on spatial analysis methods, with a special emphasis on spatial econometric methods and spatial interaction modelling, and on network analysis methods, including Social Network Analysis (SNA) and methods from statistical physics.		
Name	Matthias WEBER	**Gender**	M
Short CV / description of work experience relevant to the project	Matthias Weber has been Head of Research, Technology and Innovation (RTI) Policy Unit at Innovation Systems Department since 2000. He has been working for almost twenty years on the monitoring and analysis of innovation systems, RTI-policy, and long-term strategies for transforming research and innovation systems. His expertise covers a broad range of thematic areas (e.g. ICT, transport, energy, security, environment) as well as structural matters of RTI-policy (e.g. R&D collaboration networks, human resources, priority setting). Matthias Weber is regularly advising national and European institutions on matters of RTI policy, and he was member of the EC high-level groups on Key Technologies (2005), and on ERA Rationales (2007/08), and recently appointed member of EFFLA, the European Forum on Forward-Looking Activities. Matthias Weber is also President of the European Techno-Economic Policy Support Network ETEPS, lecturer and member of the steering		

	board of the IEV Innovation Economics Vienna joint Master- and PhD-programme of AIT and Vienna University of Economics.		
Name	Michael BARBER	**Gender**	M
Short CV / description of work experience relevant to the project	Michael Barber has a B.S. in physics from Michigan Technological University and a Ph.D. in physics from Washington University (Dissertation: Studies in Neural Networks: Neural Belief Networks and Synapse Elimination). In 2006, he began his present position as Scientist at the AIT innovation Systems Department. Previously, he was a visiting scientist at the Center for Mathematical Sciences of the University of Madeira and a postdoctoral researcher at the Institute for Theoretical Physics of the University of Cologne. His research interests include complex systems and networks, machine learning, structure and function of R&D networks, geography of collaborations, neural networks, and statistical physics. He is highly skilled in the application of analytical and computation methods to the investigation of complex systems and networks.		
Name	Xheneta BILALLI-SHKODRA	**Gender**	F
Short CV / description of work experience relevant to the project	Xheneta Bilalli-Shkodra is a Junior Expert Advisor in the Center for Innovation Systems & Policy at the Austrian Institute of Technology GmbH. She graduated in computer science at the South East European University in Tetovo/Macedonia. Following her bachelor studies, she studies computer science (master program: Business Informatics) at the Technical University of Vienna. She is experienced in programming, data gathering and databases. Her focus at AIT is on (web) data extraction (programming in Ruby) and data processing, in particular with respect to the maintenance and update of the AIT EUPRO database on European Framework Programs.		
Name	Martina DÜNSER	**Gender**	F
Short CV / description of work experience relevant to the project	Martina Dünser is a Junior Scientist in the Innovation Systems Department at the AIT Austrian Institute of Technology GmbH, where she is currently working on simulating knowledge production of regional innovation systems, especially in the life sciences sector, using agent-based modelling. Martina Dünser holds a master's degree in economics from the Vienna University of Business and Economics (WU) and studies statistics at the University of Vienna. Her special interests lie in the field of quantitative data analysis, data visualization and econometrics.		
Name	Barbara HELLER-SCHUH	**Gender**	F
Short CV / description of work experience relevant to the project	Barbara Heller-Schuh joined AIT in the year 2002. Here she works as a researcher in the Innovation Systems Department. She holds a Master's degree in History, German Language and Literature studies from the Karl Franzens University in Graz, Austria (Mag. phil.) and in Knowledge Management (Master of Advanced Studies, MAS and Master of Science, MSc) from the Danube University in Krems, Austria. Her main research interests cover RTI policy, analysis of R&D collaboration networks as well as governance and funding systems for universities. She is involved in research and consulting projects dealing with the exploration of collaborative R&D projects on different policy levels in order to analyse network structure and dynamics of collaboration patterns. She carried out studies analyzing the governance and funding systems of universities based on different indicators. Furthermore, she is responsible for the maintenance of the EUPRO database, a comprehensive relational database containing information about all accessible projects and their participants of the EU-Framework Programmes (at the moment more than 60.000 projects and participants).		
Name	Georg ZAHRADNIK	**Gender**	M
Short CV / description of work experience relevant to	Georg Zahradnik is Junior Scientist at the Innovation Systems Department. He graduated in economics at the Vienna University of Economics and Business. He is experienced in data gathering and databases, econometrics, and has worked on the internationalisation of research activities, the European ICT sector and on sectoral innovation systems in Europe. In course of a recent project on the internationalisation of business R&D he has been coordi-		

the project	nating the data gathering process resulting in an unique dataset on R&D expenditure of foreign-owned firms collected from Eurostat, the OECD, and national statistical offices. He has been involved in a number of consultancy and research projects for Austrian as well as international clients.

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
Bridging centrality as an indicator to measure the 'bridging role' of actors in networks: An application to the European Nanotechnology co-publication network	Journal of Informetrics	2017	Berge, L., Scherngell, T. and Wanzenböck, I.
Integration Processes in European R&D: A comparative spatial interaction approach using project based R&D networks, co-patent networks and co-publication networks	Geographical Analysis (47, 349-375)	2015	Lata, R., Scherngell, T. and Brenner, T.
Embeddedness of regions in European knowledge networks. A comparative analysis of inter-regional R&D collaborations, co-patents and co-publications	The Annals of Regional Science (53, 337-368)	2014	Wanzenböck, I., Scherngell, T. and Brenner, T.
The Geography of Networks and R&D Collaborations	Springer, Book Series: Advances in Spatial Science	2013	Scherngell, T. (Ed.)
Towards an integrated European Research Area? Findings from Eigenvector spatially filtered spatial interaction models using European Framework Programme data	Papers in Regional Science (92, 555-577)	2013	Scherngell, T. and Lata, R.

Relevant previous projects or activities			
Project name	RISIS - Research infrastructure for research and innovation policy studies (EC-GA N°313082) http://risis.eu/	**Dates (start/end)**	2014-2018
Project name	KNOWMAK - Knowledge in the making in the European society (EC-GA N° 726992) http://www.knowmak.eu/ coordinated by UPEM	**Dates (start/end)**	2017-2019
Project name	PREF - Analysis of national public research funding, by theme and by type of allocation (JRC Contract N° 154321)	**Dates (start/end)**	2015-2016
Short description	The PREF project aims to collect information on and provide an analysis of national public research funding, by theme and by type of allocation (competitive project based funding versus institutional funding) including an overview of the evolution and current state of public research funding in European and selected non-European Countries		
Project name	Impact of R&D Networks - The impact of R&D Networks across European regions. Evidence from a spatial econometric perspective (Austrian National Bank)	**Dates (start/end)**	2015-2016
Short description	This project focuses on assessing the significance of policy induced R&D network structures for the creation of new knowledge at the regional level. The overall objective is to		

	measure the impact of EU funded R&D networks on knowledge production in European regions.		
Project name	SD-Net - The spatial and temporal evolution of R&D Networks (Austrian Science Fund)	**Dates (start/end)**	2009-2012
Short description	This project focuses on the spatial and temporal evolution of R&D collaborations in Europe as captured by collaborative research projects of organizations that are funded by the European Framework Programmes (FPs)		

Relevant infrastructure	
Name	**Description**
EUPRO	AIT owns and maintains the EUPRO database that will not only be one of the RISIS datasets made available in the infrastructure, but also functions as central input for other work packages (e.g. those working on organisation registers). EUPRO presently comprises information on over 90,000 research projects funded by the European Framework Programmes (EU FP), and other European programmes such as EUREKA and selected Joint Technology Initiatives. It records harmonized information on projects and participating organizations, such as project objectives and achievements, project costs, total funding, start and end date, contract type, a standardized subject index, a freely specified index, information on the call, the participating organisation, contact person with contact details, organization type, and geographical location. EUPRO raw data are based on publicly accessible data (e.g. CORDIS). AIT has undertaken a substantial effort to significantly improve quality and the level of standardization of the data (e.g. correction of heterogeneous spellings of organization names, different languages, inconsistency of organization types and organizational levels, etc.), and to retrieve and add missing data. The EUPRO database is regularly updated by AIT, usually each year. In its current form, it contains homogenized, standardized, and regionalized data from 1984 to date. It thus offers coverage, data quality and reliability that is superior to any other data base of its kind.

Contribution to the project

AIT will contributes in an number of workpackages in theoretical, methodological and operational terms. The main role of AIT will be the coordination of **WP5** and **WP8**:

- In **WP8**, AIT will coordinate the development, organisation and implementation of what is to be considered as the very core of a research infrastructure from a user perspective, namely the Trans-national Access of researchers and potentially other users to the RISIS datasets, tools and services. Given the move to an advanced community, the organisation and Access mechanisms must be advanced profoundly in comparison to the first RISIS project, most importantly moving from exclusively physical trans-national access to distant trans-national access.
- In **WP5**, AIT will coordinate the work carried to update, deepening and maintaining of existing RISIS datasets, with a special focus on the extension of the EUPRO dataset to integrate beneficiary information from national funding programmes.

In addition, AIT contributes to a number of other WPs building on its long-standing experience in policy oriented analysis and research in different policy areas. AIT brings in competences in economics of innovation, national and sectoral innovation systems, the internationalisation of innovation, R&D and innovation strategies of firms, EU R&D policy as well as knowledge on indicators for measuring innovation and R&D.

4.1.3. Partner 3: Politecnico di Milano (POLIMI)

Partner Profile
Politecnico di Milano (POLIMI) was established in 1863 and it is currently ranked as one of the most outstanding European universities in engineering, architecture, and design (http://www.polimi.it/en/university/rankings/). Among its most eminent professors over the years there are the mathematician Francesco Brioschi (its first Director) and Giulio Natta who won the Nobel Prize in Chemistry in 1963. POLIMI has always focused on the quality and innovation of its teaching and research, developing a fruitful relationship with the business and productive world by means of experimental research and technological transfer. POLIMI is organized in 12 Departments. The partner Department is the Department of Management, Economics and Industrial Engineering (DIG) (http://www.dig.polimi.it). DIG is one of the largest departments in the University. Its staff includes around ninety tenured professors. The DIG faculty is responsible for several graduate and undergraduate courses in the fields of Management, Economics and Industrial Engineering within master and bachelor programs offered by Politecnico. Moreover, DIG has a three-year PhD program in Management Engineering. The research team of POLIMI within this project will be coordinated by Massimo G. Colombo, Full Professor of Innovation Economics, Entrepreneurship and Entrepreneurial Finance, and Deputy Director for Research and Rankings at DIG, and will include an additional Full Professor, an Associate Professor and two Assistant Professors from DIG. The team has developed considerable scientific expertise over the past decade in the fields of economics of innovation, entrepreneurship and entrepreneurial finance, attracting funds and sponsorships from several sources, including the European Commission, governmental institutions, industry associations, and private companies. The team has also developed key competences in quantitative econometric methods, big data analytics, and in the construction of large data infrastructures aimed at supporting policy-relevant research activity in the above-mentioned research fields: namely, the RITA database, VICO database and the RISIS infrastructure. The RITA database has been created in 1999 and since then it has launched several research projects on Italian young high-tech firms. The dataset has been updated and extended in 2002, 2004, 2007 and 2009. Around 2,000 firms are currently included in the RITA dataset. The VICO project has been financed by the FP7 between 2008 and 2011. The aim of the project was to understand the impact of VC financing on the economic performance of European innovative entrepreneurial ventures as reflected by their innovation, growth and competitiveness. The project created a unique hand-collected, large-scale longitudinal database on European high-tech companies and venture capital investments. The RISIS FP7 project (started in 2014 and currently ongoing) created a distributed research infrastructure to support and advance science and innovation studies. In this latter project, POLIMI is responsible for the work-package dedicated to the development and updating of data infrastructures at the company-level (the above-mentioned VICO database on venture capital and high-tech companies, the CIB on large scale innovators and the Cheetah database on fast-growing medium sized companies). Finally, the team is currently involved in the H2020 KNOWMAK project, aimed at creating a web-based tool, which provides interactive visualisations and indicators on knowledge co-creation in the European research area.

Staff profile			
Name	Massimo G. COLOMBO	**Gender**	M
Short CV / description of work experience relevant to the project	Massimo G. Colombo is Full Professor of Innovation Economics, Entrepreneurship and Entrepreneurial Finance, and Deputy Director for Research and Rankings at DIG – Politecnico di Milano. The scientific activity of Massimo G. Colombo is mainly in entrepreneurship, entrepreneurial finance, and economics of innovation (with a focus on open innovation), with a special focus on high-tech entrepreneurship. Massimo G. Colombo is author (or co-author) of 9 books and more than 90 international publications in highly reputed international journals. These latter include Science, Strategic Management Journal, Journal of Business Venturing, Entrepreneurship Theory and Practice, Research Policy, Journal of		

	Industrial Economics, Industrial and Corporate Change, International Journal of Industrial Organization, Journal of Economics and Management Strategy, Journal of Economic Behavior and Organization, Journal of Banking and Finance, Small Business Economics, Journal of Small Business Management, among the many others. He edited special issues on Venture Capital, Journal of Technology Transfer, Industry & Innovation, Journal of Small Business Management, European Management Review, Small Business Economics, Journal of Product Innovation Management, and Strategic Entrepreneurship Journal. He is Editor in Chief of Economia e Politica Industriale - Journal of Industrial and Business Economics and Associated Editor of Journal of Small Business Management. He was the Scientific Coordinator of the aforementioned VICO FP7 project. VICO has officially recognized by the EU Commission as a success case. Currently, he is the coordinator of the research unit of Politecnico di Milano of the RISIS project (FP7).		
Name	Cristina ROSSI LAMASTRA	**Gender**	F
Short CV / description of work experience relevant to the project	Cristina Rossi Lamastra is Full Professor of Business and Industrial Economics at DIG – Politecnico di Milano. She holds a PhD in Economics of Innovation from Sant'Anna School of Advanced Studies in Pisa. Her research interests are in the area of entrepreneurship, organizational economics, and user innovation. Cristina has published on these topics in Science, Management Science, Entrepreneurship Theory and Practice, Research Policy, Economic Letters, Industry & Innovation, R&D Management, Long Range Planning, Small Business Economics, Managerial and Decision Economics, among others. She is Associate Editor of Journal of Small Business Management and of Journal of Industrial and Business Economics. She edited special issues on Strategic Entrepreneurship Journal, Journal of Product Innovation Management, Industry & Innovation, Journal of Small Business Management. At the Politecnico di Milano School of Management, Cristina served as director of the Executive MBA Part Time and as member of the Steering Committee of the PhD program in Management Engineering. She was Representative at Large of the Division of Technology and Innovation Management of the Academy of Management and General Representative at Large of the Interest Group of Entrepreneurship and Strategy of the Strategic Management Society.		
Name	Evila PIVA	**Gender**	F
Short CV / description of work experience relevant to the project	Evila Piva is Associate Professor at DIG – Politecnico di Milano. Her main competences are in the fields of entrepreneurship. Her research interests deal mainly with peculiarities and performance of startups founded by academic personnel and university students; economic and managerial aspects of collaborations established by entrepreneurial ventures, including strategic alliances, collaborations with communities of users and developers, relationships between start-ups and public research organizations, acquisitions of high-tech start-ups; organizational design of start-ups. On these themes Evila Piva has publishes in highly reputed international journals including Research Policy, Strategic Entrepreneurship Journal, Small Business Economics, Industry & Innovation, Journal of Small Business Management, Long Range Planning, among the others. Evila has participated in several research projects founded by the European Commission, including the aforementioned VICO project.		
Name	Massimiliano GUERINI	**Gender**	M
Short CV / description of work experience relevant to the project	Massimiliano Guerini is Assistant Professor at DIG – Politecnico di Milano. He holds a PhD in Management Engineering from the Politecnico di Milano and he was Visiting Fellow at the Harvard Business School. His research interests are in the area of entrepreneurship, entrepreneurial finance, and technology transfer. Massimiliano has published on these topics in Journal of Business Venturing, California Management Review, Journal of Corporate Finance, Research Policy, and Small Business Economics, among others. Massimiliano has participated in several research projects founded by the European Commission, including the aforementioned VICO, RISIS and KNOWMAK projects. Starting from January 2017, he is coordinating the research unit of Politecnico di Milano of the KNOWMAK project.		

Name	Vincenzo BUTTICÈ	**Gender**	M
Short CV / description of work experience relevant to the project	Vincenzo Butticè is Assistant Professor at DIG-Politecnico di Milano. Vincenzo has been a visiting fellow at the Copenhagen Business School and at the Indiana University-Purdue University of Indiana. In 2017, he obtained the PhD in Management Engineering from the Politecnico di Milano. His research is about crowdfunding and entrepreneurial performance. Vincenzo ha published on these topics on Entrepreneurship Theory and Practice.		

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
University specialization and new firm creation across industries	Small Business Economics, 41(4), 837-863.	2013	Bonaccorsi, A., Colombo, M. G., Guerini, M., & Rossi-Lamastra, C.
Serial crowdfunding, social capital, and project success	Entrepreneurship Theory and Practice, 41(2), 183-207	2017	Butticè, V., Colombo, M. G., & Wright, M.
Cash from the crowd	Science, 348(6240), 1201-1202	2015	Colombo, M. G., Franzoni, C., & Rossi-Lamastra, C.
Firms' genetic characteristics and competence-enlarging strategies: A comparison between academic and non-academic high-tech start-ups	Research Policy, 41(1), 79-92	2012	Colombo, M. G., & Piva, E.
Governmental venture capital in Europe: Screening and certification	Journal of Business Venturing, 31(2), 175-195	2016	Guerini, M., & Quas, A.

Relevant previous projects or activities			
Project name	RISIS - Research infrastructure for research and innovation policy studies (EC-GA N°313082) http://risis.eu/	**Dates (start/end)**	2014-2018
Project name	KNOWMAK - Knowledge in the making in the European society (EC-GA N° 726992) http://www.knowmak.eu/ coordinated by UPEM	**Dates (start/end)**	2017-2019
Project name	PICO – Academic entrepreneurship from knowledge creation to knowledge diffusion (EC-FP6)	**Dates (start/end)**	2006-2009
Short description	This project focuses on Research-Based Spin-Offs (RBSOs), that is to say firms that are created around scientific knowledge generated in public research organizations (PROs). RBSOs play a key role in transforming scientific knowledge into new knowledge that is embodied in new technologies, products, services and processes. Hence, they play an important role as an engine of innovation and growth for Europe. In this project we seek to understand the conditions under which RBSOs effectively contribute to the generation of new knowledge and its dissemination in the environment in which they are embedded.		
Project name	PRIME – Policies for research and innovation in the move towards the European research area (EC-FP5)	**Dates (start/end)**	2002-2004
Short de-	The network brings together 11 groups from 7 countries. It is interdisciplinary covering eco-		

scription	nomics, sociology, political science, history and management. It explore the conditions of creation of a network of excellence on the shaping of research and innovation policies in the move towards the ERA.		
Project name	VICO – "Financing Entrepreneurial Ventures in Europe : Impact on innovation, employment growth, and competitiveness (EC-FP7 GA N° 217485)	**Dates (start/end)**	2008-2011
Short description	The aim of the VICO project is to assess the impact of Venture Capital and Private Equity (VC/PE) financing on the economic performance of European innovative entrepreneurial ventures as they are reflected by their innovation output, employment, growth, and competitiveness, and the role VC/PE investors play in helping these firms bridge their resource and competence gap beyond the provision of financing. Particular attention will be devoted to the institutional variety of VC/PE investors in Europe. We will consider factors affecting both the demand for and the supply of VC/PE financing, at micro and macro level. In particular, we will analyze how the effects of VC/PE financing on the economic performance of portfolio companies are moderated by i) the characteristics of the investor, ii) the characteristics of invested firms, and iii) the characteristics of the business and institutional environment in which firms operate.		

Relevant infrastructure	
Name	**Description**
VICO	The VICO 2.0 database contains geographical, industry and accounting information on startups that received at least one venture capital or early stage equity investment in the period 1998-2014, and operating in seven European countries (Belgium, Finland, France, Germany, Italy, Spain, and the United Kingdom) and Israel. Its uniqueness lies in the overall number of companies (17,863), the country coverage, and the extent of information gathered (thanks to the combination of data provided by different proprietary datasets, i.e. Thompson One Private Equity, Zephyr, Crunchbase and Orbis). The database is widely used in the academic and policy making community.
CHEETAH	The Cheetah database contains geographical, industry and accounting information on mid-sized firms that experienced fast growth during three growth periods (cohorts) starting at 2008, 2009 and 2010 in 30 European countries plus Israel. The uniqueness of this database lies in its coverage in terms of countries and number of firms included in the database. The database presently includes 42,369 firms.

Contribution to the project

POLIMI is involved in a number of work packages of the present project. It will mainly contribute by:

- Coordinating the work package on the development of RISIS core datasets. Development activities are aimed at deepening and enlarging information contained in core datasets, allowing the production of new indicators and opening new opportunities for research;
- Maintaining and expanding two RISIS core datasets, namely the VICO dataset on European venture capital backed start-ups and the Cheetah dataset on European fast-growing mid-sized firms, as well as managing the access process in order to allow researchers to use them effectively;
- Coordinating the design, development and consolidation activities of register of firms (FirmReg), which represents a key facility in the harmonization of organizations;
- Providing methodological support for the use of advanced econometric techniques for analyzing panel data.

In addition, the POLIMI team will leverage its considerable scientific expertise in the fields of economics of innovation, entrepreneurship and entrepreneurial finance, for the dissemination of project results in the respective communities of scholars as well as among national and supranational policymakers.

4.1.4. Partner 4: Consiglio Nazionale delle Ricerche (CNR)

Partner Profile
National Research Council (CNR) is the largest public research institution in Italy, the only one under the Research Ministry performing multidisciplinary activities. The Scientific network of CNR is organized in Research Institutes; two Institutes participate in RISIS, IRCRES (leading institute), and ISTI. IRCRES-Research Institute on Sustainable Economic Growth (www.ircres.cnr.it) is a public entity, one of the CNR Institutes. It is finalized to the study of sustainable economic growth, innovation, research indicators, public policies and research institutions. The headquarter is located in Moncalieri (Turin); other units are in Milan, Genova and Rome. The Unit of Rome has a long-lasting experience on research policy and innovation studies, with a special focus on R&D funding and evaluation, governance of R&D systems, STI indicators. IRCRES was built up along 40 years of activities by a group of researchers characterized by a high level of interdisciplinarity between economic, political and social sciences, and by strict linkages with external environments as academia, policy makers and economic actors. From 2005 the Institute enlarged the scientific range towards social and political sciences and new economic directions (as higher education policy, research and impact evaluation, science and society relationships, labor market and environmental impact). ISTI-Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" is one of the largest CNR Institutes (www.isti.cnr.it) and its main mission is to producing scientific excellence and playing an active role in technology transfer. CNR-ISTI is organised in 16 technology centres or labs, each of them pursuing a well-defined set of scientific objectives. State-of-the-art technology development and training are other activities performed by the centres. The team participating in this project belongs to the "Multimedia Networked Information System Laboratory", which consists of 60 researchers and technicians conducting research and development of (i) software architectures and systems supporting data infrastructures and large distributed multimedia information systems and (ii) technologies for the management, distribution and fruition of multimedia and multi-type information and data. This team has a long experience in participating to many EU-funded projects relevant to the topics addressed in the RISIS2 project. In particular, the team has coordinated from the technical point of view the stream of projects DRIVER and OpenAIRE and from the scientific point of view the stream of projects DILIGENT-D4Science-D4Science-II-iMarine-BlueBRIDGE that have played a central role in the definition and implementation of the D4Science Hybrid Data Infrastructure and of a set of enabling services tailored to meet basic data management and processing requirement of different scientific communities. Other H2020 related projects the team is involved in are: ENVRIPlus, EGI_ENGAGE, SoBigData, eInfraCentral, PARTHENOS and EOSCpilot. It also participates in RDA-Europe project.

Staff profile			
Name	Emanuela REALE	**Gender**	F
Short CV / description of work experience relevant to the project	Dr. Emanuela Reale, social scientist, is Senior Researcher at the Research Institute on Sustainable Economic Growth IRCRES, National Research Council CNR of Italy. Her main areas of interest are: higher education policy, governance of universities and public research organizations, R&D funding mechanisms, research evaluation, and STI indicators. She was principal investigator in several national and European projects dealing with higher education, research evaluation and research policy; she was coordinator of tenders of the European Commission on transnational interdisciplinary research and government R&D funding. She was Vice President of the Italian Evaluation Association-AIV in 2009-2013, and Vice President of the European Forum for Studies on Policies for Research and Innovation-EU-SPRI (2009-2015). Currently, she is Member of the Board of the Consortium of Higher Education Researchers CHER. She publishes in several international journals and books including Research Policy, Science and Public Policy, Research Evaluation, Higher Education.		

Name	Antonio ZINILLI	**Gender**	M
Short CV / description of work experience relevant to the project	Dr. Antonio Zinilli, is a PhD in Applied Social Sciences. He is a postdoctoral fellow at IRCrES-CNR, National Research Council of Italy, Research Institute on Sustainable Economic Growth. Furthermore, he is a lecturer in Sociology at University of Rome "La Sapienza". He has scientific research interests in: research evaluation, performance-based research funding systems and their impact on the academic context, network modeling, network analysis methods and spatial statistics. He won a scholarship by Eu-SPRI Phd Circulation Award and he was visiting researcher at the Austrian Institute of Technology in Wien (Austria) for three months. He has attended many conferences in Italy and abroad with paper presentations and publications in peer-review journal.		
Name	Serena FABRIZIO	**Gender**	F
Short CV / description of work experience relevant to the project	Dr. Serena Fabrizio, finished her PhD in Communication Science at Sapienza University of Rome, defending a thesis titled "Youth, media and civic engagement. Internet as sociality place or civic activation soace?" on topics related to new technologies as a tool for training new collective subjects and access to knowledge. From 2008 to 2015 she collaborated on Research University Projects within the Department of Communication and Social Research of Sapienza University on topics related to the representation and governance of social problems. Since June 2015 she has a research fellowship at IRCRES-CNR collaborating on several European and national research projects. Main research interests concern policies for higher education and governance systems, analysis of public funding on R&D at national and international level, evaluation of impact of Social Science and Humanities (SSH) research.		
Name	Lucio MORETTINI	**Gender**	M
Short CV / description of work experience relevant to the project	Dr. Lucio Morettini, is Researcher at the National Research Council of Italy - IRCRES. He has taken his PhD in Economics at the University “La Sapienza” of Rome with a thesis on Economics of Education in 2010. He is interested in research policy and research evaluation, effects of university design on job market and influence of social background on education choices. Lucio has been Research Fellow at National Research Council of Italy – Institute for the Study of Regionalism, Federalism and Self Government (ISSIRFA – CNR) where he has developed a dataset of public and private R&D expenditure at local level. This dataset has been used to develop analysis on the use of European R&I structural funds and his effects on local system of innovation.		
Name	Andrea ORAZIO SPINELLO	**Gender**	M
Short CV / description of work experience relevant to the project	Dr. Andrea Orazio Spinello, sociologist, is a Research Assistant at CNR-IRCRES. After undertaking a 2nd-level Professional Master’s Programme on evaluation-related topics, he received his Ph.D. in “Communication, Research, Innovation” (curriculum “Methodology of Social Sciences”) at the Department of Communication and Social Research of Sapienza University of Rome in 2018. His doctoral thesis focused on the effects of the exploitation of research infrastructures on the work of social scientists. His research interests are mainly dealing with the analysis of research funding systems and the dynamics related to transnational R&D programmes.		
Name	Donatella CASTELLI	**Gender**	F
Short CV / description of work experience relevant to the project	Dr. Donatella Castelli is Research Director at CNR-ISTI and leader of the InfraScience research group. Her research interests include digital libraries and data infrastructure content modeling, interoperability and architectures. She is author of several research papers in such fields. Under her supervision, the InfraScience team coordinated and participated in several EU and nationally funded projects on Digital Libraries and Research Data Infrastructures. In particular, she has been the technological coordinator of the DRIVER, DRIVER-II, OpenAIRE, and OpenAIREplus projects. In this role she has contributed to the development of the D-Net software toolkit. Dr. Castelli is currently technical manager of the OpenAIRE2020 project and the scientific director of the BlueBRIDGE project and		

	of the D4Science infrastructure. She is also a member of the RDA Europe Expert Group that promotes research and cross-infrastructure coordination at global level and member of the COAR Strategic Committee		
Name	Pasquale PAGANO	**Gender**	M
Short CV / description of work experience relevant to the project	Dr. Pasquale Pagano is Senior Researcher at CNR-ISTI. He has a strong background and experience on models, methodologies and techniques for the design and development of distributed virtual research environments (VREs) which require the handling of heterogeneous computational and storage resources, provided by Grid and Cloud based e-Infrastructures, and management of heterogeneous data sources. He participated in the design of the most relevant distributed systems and e-Infrastructure enabling middleware developed by ISTI-CNR. He is currently the Technical Director of the D4Science Data Infrastructure, Technical Director of H2020 BlueBRIDGE project and CNR lead person for the EGI-ENGAGE one. In the past, he has been involved in the iMarine, EUBrazilOpenBio, ENVRI, Venus-C, GRDI2020, D4Science-II, D4Science, Diligent, DRIVER, DRIVER II, BELIEF, BELIEF II, Scholnet, Cyclades, and ARCA European projects		
Name	Leonardo CANDELA	**Gender**	M
Short CV / description of work experience relevant to the project	Dr. Leonardo Candela is a researcher at CNR-ISTI. He has relevant expertise in the area of Virtual Research Environments development. He was involved in various EU-funded projects including CYCLADES, Open Archives Forum, DILIGENT, DRIVER, DELOS, D4Science, D4Science-II, DL.org, EUBrazilOpenBio, iMarine, BlueBRIDGE, ENVRI and ENVRI PLUS. He was a member of the DELOS Reference Model Technical Committee and of the OAI-ORE Liaison Group.His research interests include Data Infrastructures, Virtual Research Environments, Data Publication, Open Science, Digital Library [Management] Systems and Architectures, Digital Libraries Models, Distributed Information Retrieval, and Grid and Cloud Computing.		
Name	Paolo MANGHI	**Gender**	M
Short CV / description of work experience relevant to the project	Dr. Paolo Manghi is Researcher at CNR-ISTI since 2006, member of the InfraScience research group. His research focuses on data models for Digital Library Management Systems, data curation in digital libraries, and autonomic service-oriented data infrastructures, with special care to the realization and technological transfer of open source software. He is author of several research papers in such fields. Moreover, he has participated and coordinated the research and technical activities of several EC projects, such as DRIVER, DRIVER-II, OpenAIRE, OpenAIREplus, HOPE, EFG, and several others whose focus was on the realization of aggregative data infrastructures for scientific communication and science. He is currently coordinating research and development activities supporting the production data infrastructures for the OpenAIRE2020 project, SoBigData.eu project, and the EAGLE project. Finally, he is conceiver and designer of the D-NET Software Toolkit (http://www.d-net.research-infrastructures.eu), a framework for the construction of autonomic and sustainable aggregative data infrastructures.		
Name	Alessia BARDI	**Gender**	F
Short CV / description of work experience relevant to the project	Dr. Alessia Bardi received her MSc in Information Technologies in 2009 at the University of Pisa, then a PhD in Information Engineering at the Engineering Ph.D. School "Leonardo da Vinci" of the University of Pisa and works as researcher at CNR-ISTI. Today she is a member of the InfraScience research group, part of the Multimedia Networked Information System Laboratory (NeMIS). She is involved in EC funded projects for the realisation of aggregative data infrastructures. Her research interests include service-oriented architectures and data infrastructures for e-science and scholarly communication.		

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
Logics of integration and actors' strategies in European Joint Programs	Research Policy, 43(2) 391-402	2014	Lepori B., Reale E., Laredo P.
Challenges in higher education research: the use of quantitative tools in comparative analyses	Higher Education, 67, 409-422	2014	Reale E.
Analysis of National Public Research Funding-PREF – Final Report	Luxembourg: Publications Office of the European Union	2017	Reale, E.
Repositories for Open Science: the SciRepo reference model	MTSR 2015 - Metadata and Semantics Research. 9th Research Conference (Manchester, UK, 9–11 September 2015). Proceedings, pp. 298 - 311. Emmanouel Garoufallou, Richard J. Hartley, Panorea Gaitanou (eds.). (Communications in Computer and Information Science, vol. 544). Springer	2015	Assante M., Candela L., Castelli D., Manghi P., Pagano P.
A social networking research environment for scientific data sharing: the D4Science offering	GL15 - Fifteenth International Conference on Grey Literature. The Grey Audit: a Field Assessment in Grey Literature (Bratislava, Slovakia, 2-3 December 2013). Proceedings, vol. 15 pp. 151 - 158. D.J. Farace, J. Frantzen, GreyNet InternationalGrey Literature Network Service (eds.). (GL Conference Series, ISSN 1385 2316, vol. 15). TextRelease, Amsterdam,	2014	Assante M., Candela L., Castelli D., Gioia A., Mangiacrapa F., Pagano P.
The D-NET software toolkit: a framework for the realization, maintenance, and operation of aggregative infrastructures	Program, vol. 48 (4) pp. 322 - 354. Emerald	2014	Manghi P., Artini M., Atzori C., Bardi A., Mannocci A., La Bruzzo S., Candela L., Castelli D., Pagano P.

Relevant previous projects or activities			
Project name	RISIS - Research infrastructure for research and innovation policy studies (EC-GA N°313082) http://risis.eu/	**Dates (start/end)**	2014-2018
Project name	PREF - Analysis of national public research funding, by theme and by type of allocation (JRC Contract N° 154321) Coordinated by CNR	**Dates (start/end)**	2015-2016
Project name	JOREP - Investments in joint and open R&D programs and analysis of their economic impact (EC Tender 2009/S 158 229-748) Coordinated by CNR	**Dates (start/end)**	2009-2013
Short description	JOREP aim was to build a dataset on characteristics of transnational research programmes and develop case studies on the impact they had with respect to the planned objectives, including the development of interdisciplinary research		

Project name	OpenAIRE, OpenAIREPlus, OpenAIRE2020 and OpenAIRE-Advance (https://www.openaire.eu)	**Dates (start/end)**	2009-2020
Short description	These are flagship projects on open access for information and data, supported by technology, services and policy building		
Project name	BlueBRIDGE - Building Research environments for fostering Innovation, Decision making, Governance and Education to support Blue growth (www.blue.bridge-vres.eu)	**Dates (start/end)**	2015-2018
Short description	Aims at building research environments for fostering innovation, decision making, governance and education in the context of the Blue Growth Societal challenge		
Project name	EOSCpilot - The European Open Science Cloud for Research Pilot Project (https://eoscpilot.eu/)	**Dates (start/end)**	2017-2018
Short description	The EOSCpilot project supports the first phase in the development of the European Open Science Cloud (EOSC) as described in the EC Communication on European Cloud Initiatives [2016].		

Relevant infrastructure	
Name	**Description**
JOREP	JoREP is a database on joint R&D programmes. It provides a quantitative basis for the monitoring of investments in joint R&D programmes in the countries belonging to the European Research Area (ERA), pointing out the policy rationales behind them and their impact. The set of data aims at describing when, how and serving what purposes European-level initiatives and bilateral/multilateral joint R&D programmes are combined. JoREP 2.0 also supports the analysis of important ERA dynamics and Europeanization processes through the study of the behaviour of main national actors (i.e. funding agencies).
D4Science (www.d4science.org)	D4Science is a Hybrid Data Infrastructure serving a number of Virtual Research Environments exploited in the context of several European projects and international partnerships. It delivers its services to more than 5,000 users spread in more than 50 countries worldwide.
OpenAIRE (https://www.openaire.eu)	Is the European Data Infrastructure for Scientific Open Access. CNR-ISTI is responsible of the OpenAIRE development and pre-production infrastructure.

Contribution to the project

CNR-IRCRES team has consolidated individual experiences and institutional expertise on delivering research and research services within international programs including the activities developed in the RISIS Project. On this base, CNR-IRCRES is able to contribute to the project for the following tasks:

- Theoretical background on funding policies for research and characterization of research funding organizations and funding portfolio,
- Development of conceptual framework and creation of a new dataset on financial instrument data on the base of experiences developed within other research projects;
- Support training of new generation of scholar in STI studies;
- Dissemination of results achieved and supporting actions toward integration of the STI research community;
- Development of awareness rising events devoted to produce social and political impact through stakeholders' productive interactions.

CNR ISTI team has a sound experience in research and development of (i) software architectures and systems supporting data infrastructures and large distributed multimedia information systems and (ii) technologies for the management, distribution and fruition of multimedia and multi-type information and data. On

this base, CNR-ISTI will contribute to the project with the following expertise:

- Design and implementation of hybrid data infrastructures supporting Virtual Research Environments;
- Design and implementation of aggregative data infrastructures for scientific and scholarly communities;
- Aggregation and curation of Open Access data for the implementation of the Open Science Paradigm.

4.1.5. Partner 5: Leiden Universiteit (LEIDEN)

Partner Profile
The Centre for Science and Technology Studies (CWTS) is a research institute of the Faculty of Social Sciences from Leiden University. It is a globally leading centre in science and technology studies with a focus on the study of the dynamics of knowledge communication processes in science and technology, as well as on directly related themes such as evaluation of research performance, knowledge diffusion, and innovation. CWTS's mission is to combine and integrate a diversity of theoretical frameworks (e.g. actor network theory, communication and citation theories, neo-institutional theory) and methodological approaches, including interviews and ethnography, focus groups, surveys, scientometrics, and computer simulation. CWTS has more than 20 years extensive experience in the construction and the application of advanced bibliometric methods, and associated science technology indicators. Bibliometric analysis is used in basic research for the design and construction of quantitative indicators to study the development of science, technology and science-based innovations. The CWTS applies this methodology for evaluation and assessment purposes, often in contract work, services and consultancy, as well as for monitoring and mapping developments within science and innovation systems. Our Leiden Ranking is recognised as one of the most reliable citation based rankings of large universities. A key asset of the centre is its high quality processing of citation databases and the data expertise in managing large citation databases of its staff. As a result, the citation analyses produced by CWTS are based on corrected and verified data of high quality, different from the public citation indexes commonly available on the web. CWTS is one of the oldest, most experienced, and largest scientometric centres in the world it has the responsibility to take the initiative in the required methodological and theoretical advances in the field.

Staff profile			
Name	Ed NOYONS	**Gender**	M
Short CV / description of work experience relevant to the project	Ed Noyons (PhD) is a senior researcher at the Leiden University based Centre for Science & Technology Studies (CWTS). As an academic institute, CWTS serves many clients with their contract research supplying bibliometric data for research assessment procedures. Ed has, as member of the CWTS project board, the function of the supervisor and coordinator of almost all (contracted) studies, and is attached to CWTS for over 20 years. He has contributed to over 200 of bibliometric studies, which range from application on national level to research group level. In parallel, he has been involved in and coordinated many projects internally to improve CWTS procedures and processes. Ed has worked for many clients on different levels and from different backgrounds. Examples are the European Commission, Dutch, and foreign (mostly European) national government agencies, universities, and research managers. He is the (co-) author of more than 40 journal publications, and thus contributes on a regular basis to the international literature in the field of quantitative science studies. Furthermore, he has been involved many times in the two main International conferences (ISSI and S&T Indicators) as program chair. Since 2011 he is the assistant director of CWTS and member of the Management Team and board. Presently Ed's main research interests involve structures, mapping of science and their use in science policy and research management, field delineation, and multi-dimensional, in particular societal impact of science.		
Name	Thomas FRANSSEN	**Gender**	M
Short CV / description of work experience relevant to the project	Thomas Franssen (PhD) leads a research line on the effects of research governance on epistemic cultures in the humanities and social sciences within the Science and Evaluation Studies (SES) research group. His work focusses on changes in funding arrangements, performance evaluation and research infrastructure on research practices and content of research in different disciplines within the humanities and social sciences. He has a PhD (2015) in cultural sociology from the University of Amsterdam. Methodologically he specializes in comparative empirical analysis using a mixed-method approach combining bibliometric and qualitative		

	techniques.		
Name	Nees Jan VAN ECK	**Gender**	M
Short CV / description of work experience relevant to the project	Nees Jan van Eck (PhD) started working as a researcher at Centre for Science and Technology Studies (CWTS) of Leiden University in 2008. He is involved in various science mapping projects and works on the development of science mapping software (VOSviewer). His main research interests are science mapping and bibliometric indicators. In addition, Nees Jan is managing the ICT department of CWTS. Nees Jan has published in all major bibliometric journals, and he is a member of the editorial boards of Journal of Informetrics and Scientometrics.		
Name	Alfredo YEGROS-YEGROS	**Gender**	M
Short CV / description of work experience relevant to the project	Alfredo Yegros-Yegros (PhD) has been working at CWTS since 2012. His current research revolves around quantitative studies of science and technology. More specifically, the analysis of public-private research interactions and knowledge flows, the study of science-technology linkages, and the study of methods potentially able to capture societal impact of scientific research are some of his research interests. At CWTS, Alfredo also conducts contract research for several clients from different countries. Alfredo Yegros has an MSc in Library and Information Science (2002) – Universitat de València and Universitat Politècnica de València and he holds a PhD in Documentation, Universitat Politècnica de València (2012). Within the framework of his PhD he visited the Science and Policy Research Unit (SPRU) at the University of Sussex.		
Name	Clara Maria CALERO MEDINA	**Gender**	F
Short CV / description of work experience relevant to the project	Clara Maria Calero Medina (PhD) is a researcher at the Leiden University based Centre for Science and Technology Studies (CWTS). As an academic institute, CWTS serves many clients with their contract research supplying bibliometric data for research assessment procedures. Clara is attached to CWTS for over 14 years. Clara has, as member of the CWTS project board, the function of the supervisor and coordinator of contracted studies. She has contributed to hundreds of bibliometric studies, which range from application on national level to research group level. In parallel, she has been involved in and coordinated many projects internally to improve CWTS procedures and processes. Clara has worked for many clients on different levels and from different backgrounds. Examples are the European Commission, Dutch, and foreign (mostly European) national government agencies, universities, and research managers. She is the (co-) author of more than 15 journal publications, and thus contributes on a regular basis to the international literature in the field of quantitative science studies.		
Name	Rodrigo COSTAS COMESANA	**Gender**	M
Short CV / description of work experience relevant to the project	Rodrigo Costas Comesana (PhD) is one of CWTS's most experienced researchers, and joined Centre for Science and Technology Studies (CWTS) of Leiden University in 2009 after having worked as a researcher at the Spanish National Council for Scientific Research (CSIC). He has a strong background in library and information science and more than 15 years of experience as a researcher in the field scientometrics, combining this with ample knowledge on information management, databases design, bibliometric analysis, statistics and programming. His specialism is the development of new bibliometric tools and indicators and the development of analytical tools for the study of research activities based on quantitative data (combining different sources). The main focus of his research activities include the development of individual-level bibliometrics, the use of funding acknowledgments for scientometric research, and more recently, the development of 'altmetric' studies. The latter is a field in which CWTS has become one of the international leading research centers. Rodrigo has published extensively on all these topics.		
Name	Bijan RANJBARSAHRAEI	**Gender**	M
Short CV / description of work	Bijan Ranjbarsahraei (PhD) is a data scientist at the Centre for Science and Technology Studies (CWTS) of Leiden University. He has a PhD in Social Networks Analysis from Maastricht University, and has carried out his postdoctoral research within the AIDA project – focused		

experience relevant to the project	on Research Positioning and Research Trend Identification - at Delft University of Technology. Bijan is responsible for maintaining the CWTS database environment for which he continuously explores the possibility to use various data sources for development of new databases. He also works on novel visualization methods to facilitate the analysis and interpretation of bibliometric data.		
Name	Inge VAN DER WEIJDEN	**Gender**	F
Short CV / description of work experience relevant to the project	Inge van der Weijden (PhD) is a senior researcher at CWTS. Her main research interests are situated in the area of career paths of academics, academic leadership and management and research evaluation. Special attention is given to the academic gender balance. She has sixteen years of research experience in these research themes. She has specific experience with qualitative studies (interviews, focus group discussions) and quantitative studies (questionnaires, bibliometrics). Currently, she is a board member of the European Sociology of Science and Technology Network. She has served several times as an advisory expert to both the Diversity Office and the Human Resource Management department at Leiden University. Between 2006 and 2011 she worked in research policy for the Rathenau Institute (The Netherlands) and was involved in various studies concerning the organization and development of science systems. She holds a PhD in organizational sciences from the VU Amsterdam (2007), with a background in policy and management in medical biology (MSc 1999).		
Name	Vincent TRAAG	**Gender**	M
Short CV / description of work experience relevant to the project	Vincent Traag (PhD) leads the research line on modelling the research system within the Quantitative Science Studies (QSS) research group of the CWTS. His main interests are mathematical models in the social sciences with a focus on (social) networks. In addition to his scientific research, Traag also acts as a bibliometric consultant at the CWTS. Vincent obtained his Master in sociology (cum laude) from the University of Amsterdam (2008). Coming from a computer science background, and taking up mathematics during his studies in sociology, he went on to obtain a PhD in applied mathematics in Louvain-la-Neuve, Belgium (2013). During his PhD he studied methods for detecting communities in complex networks, resulting in a Python software package. In addition, he applied this methodology in several fields across the (social) sciences, ranging from citation networks to international relations. He joined the CWTS in 2015.		
Name	Josephine BERGMANS	**Gender**	F
Short CV / description of work experience relevant to the project	Josephine Bergmans started as a project coordinator of a couple of institutional projects at CWTS in September 2017. Her aim is to organize and manage these institutional projects properly and to find a future working mode for the institutional projects CWTS is participating in. In 2016 Josephine graduated from the master Sociology and Social Research of Tilburg University and Università degli Studi di Trento. Immediately after, she started working in the two-year graduate programme of Leiden University, TU Delft and Erasmus University.		

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
Characterization, description, and considerations for the use of funding acknowledgement data in Web of Science	Scientometrics, 108(1), 167-182	2016	Paul-Hus, A., Desrochers, N., & Costas, R.
The Leiden Manifesto for research metrics	Nature, 520 (7548), 429-431	2015	Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I.
Scientists have most impact when they're free to move	Nature, 550(7674), 29–31	2017	Sugimoto C.R., Robinson-Garcia N., Murray D.S., Yegros-Yegros

			A., Costas R. & Lariviere V.
Software survey: VOSviewer, a computer program for bibliometric mapping	Scientometrics, 84(2), 523-538	2010	Van Eck, N. J., & Waltman, L.
The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation	Journal of the American Society for Information Science and Technology 63(12): 2419-2432	2012	Waltman L.R., Calero-Medina C.M., Kosten J., Noyons E.C.M., Tijssen R.J.W., Van Eck N.J.P., Leeuwen T.N. van, Raan A.F.J. van, Visser M.S. & Wouters P.
A new methodology for constructing a publication-level classification system of science	Journal of the American Society for Information Science and Technology 63(12): 2378-2392	2012	Waltman L.R. & Van Eck N.J.P.
Virtual knowledge: experimenting in the humanities and the social sciences	MIT Press	2012	Wouters, P., Beaulieu, A., Scharnhorst, A., & Wyatt, S.

Relevant previous projects or activities			
Project name	RISIS - Research infrastructure for research and innovation policy studies (EC-GA N°313082) http://risis.eu/	**Dates (start/end)**	2014-2018
Project name	KNOWMAK - Knowledge in the making in the European society (EC-GA N° 726992) http://www.knowmak.eu/	**Dates (start/end)**	2017-2019
Project name	PRINTEGER – Promoting Integrity as an Integral Dimension of Excellence in Research (EC-GA N° 665926) https://printeger.eu/	**Dates (start/end)**	2015-2018
Short description	The mission of Printeger is to enhance research integrity by promoting a research culture in which integrity is part and parcel of what it means to do excellent research, and not just an external and restrictive control system. To promote such a culture, an improved governance of integrity and responsible research has to be informed by practice: the daily operation of researchers and the tensions of a complex research system. PRINTEGER will contribute to improve adherence to high standards of integrity in research warranting high levels of public support for the sciences. In the short term, it will do so by improving integrity policies of national and international research organisations, but also by providing better tools for research leaders and managers. In the longer term, PRINTEGER will contribute to improve ethical awareness and reflection through the education of new generations of scientists with next generation educational tools. Immediate contributions of PRINTEGER will include raised attention for realistic and effective integrity measures through dissemination, including a large conference, and immediate trial and use of much improved educational resources for teaching research ethics to future and young scientists.		
Project name	BE-OPEN – Boosting engagement of Serbian universities in open science (EC-ERASMUS+) http://www.beopen.uns.ac.rs/	**Dates (start/end)**	2016-2019
Short descrip-	Open science, open innovation, and openness to the world are set as one of the main re-		

tion	search and innovation strategies in the European Research Area. Open science provides an effective framework for the enhancement of research transparency, dissemination, collaborations, and accountability.BE-OPEN is a structural project funded under the Erasmus+ Capacity Building in Higher Education action. The project is coordinated by the University of Novi Sad, Serbia. The main objective of the project is to develop conditions for the implementation of the core principles of Open Science at universities in Serbia.

Relevant infrastructure	
Name	**Description**
LEIDEN RANKING	The CWTS Leiden Ranking 2017 offers key insights into the scientific performance of over 900 major universities worldwide based on bibliometric data. A sophisticated set of bibliometric indicators provides statistics on the scientific output, impact and collaboration of universities. The dataset included in RISIS provides these statistics for a larger list of organizations (including PROs). Moreover, the underlying data can be accessed via a local visit at CWTS.

Contribution to the project
With our bibliometric tools and evaluation expertise the CWTS will contribute to this project in several work packages. With the CWTS publication database, which is an enhanced version of Web of Science, we will contribute to the core dataset work package by cleaning and harmonizing the author affiliations in a consistent manner on an annual basis. The CWTS will lead the work package of new dataset developments and will contribute to the RCF infrastructure as well.

4.1.6. Partner 6: The University of Sheffield (USFD)

Partner Profile
The Natural Language Processing (NLP) group at the University of Sheffield is one of the largest and most successful research groups in text mining and language processing in the EU. The group is based in the Department of Computer Science, which also includes world-class teams in the areas of speech, knowledge and information processing, biotechnology, and machine learning for medical informatics. The NLP group has world-leading research record in the fields of social media analytics, multilingual information extraction, machine translation, NLP infrastructures (GATE), standardisation, machine learning methods for NLP, terminology extraction, NLP methods for semantic annotation, Linked Open Data and the Semantic Web. We will build on the results of previous EC and national projects, including KNOWMAK (developing indicators for European knowledge co-creation), SoBigData (research infrastructure for scientific discovery), OpenMinted (text mining infrastructure for scientific publications), AnnoMarket (cloud-based NLP services), CLARIN (language resources infrastructure), and NeOn (automatic creation of ontologies from text). USFD also has extensive experience in innovation and knowledge transfer activities, in particular collaborations and consulting for companies (both SMEs and large corporates), government bodies, and other organisations.

Staff profile			
Name	Kalina BONTCHEVA	**Gender**	F
Short CV / description of work experience relevant to the project	Professor Kalina Bontcheva is the holder of a prestigious EPSRC career acceleration fellowship, working on text analysis and summarisation of social media. Prof. Bontcheva received her PhD on the topic of adaptive hypertext generation from the University of Sheffield in 2001. She has been a leading developer of the GATE text analytics infrastructure since 1999. Her main interests are social media mining and summarisation, information extraction, natural language generation, and text summarization. Kalina Bontcheva is currently deputy-coordinator of the SoBigData research infrastructure project, and recently coordinated the PHEME FP7 project on computing veracity of social media content; as well as leading the Sheffield team in COMRADES and OpenMinted. Previously she coordinated the EC-funded TAO STREP project on transitioning applications to ontologies and contributed to the TrendMiner, uComp, MUSING, SEKT, and MI-AKT projects.		
Name	Diana MAYNARD	**Gender**	F
Short CV / description of work experience relevant to the project	Dr. Diana Maynard is a Senior Research Fellow, who has been involved in research in NLP and text mining since 1994. She received a PhD in Automatic Term Recognition from Manchester Metropolitan University in 2000. Her main research interests are in information extraction, opinion mining, social media analysis, web science, term extraction, ontology development, and the semantic web. Since 2000 she has led the development of USFD's open-source multilingual IE tools, and is currently PI on the EU KNOWMAK project where she leads the work on ontologies and annotation. She has led research on a number of UK and EU projects including COMRADES, DecarboNet, Arcomem, KnowledgeWeb and NeOn. She regularly provides consultancy and training on NLP and GATE in the public and private sector, is advisor to two start-up companies, and is on the advisory board for the OECD project DSIP (Digital Science and Innovation Policy and Governance) and the joint EC-OECD project REITER.		
Name	Mark A. GREENWOOD	**Gender**	M
Short CV / description of work experi-	Dr Mark A. Greenwood received his PhD from the University of Sheffield in 2006 for work on Open-Domain Question Answering from large text collections. Since 2005 he has been a research associate on a number of NLP projects including RESuLT, X-Media,		

ence relevant to the project	LarKC, Khresmoi, ForgetIT, and most recently the EU funded OpenMinTeD project. Since 2009 he has been a member of the GATE team making contributions to the development of both the core framework and plugins. Mark has been involved in the planning and running of two international workshops on Information Retrieval for Question Answering (IR4QA) as well as reviewing for numerous conferences, workshops and journals and examining MSc and PhD theses, and has published widely.

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
Real-time Social Media Analytics through Semantic Annotation and Linked Open Data	Proceedings of WebSci 2015, Oxford, UK	2015	D. Maynard, M. A. Greenwood, I. Roberts, G. Windsor, K. Bontcheva
TwitIE: An Open-Source Information Extraction Pipeline for Microblog Text	Proceedings of the International Conference on Recent Advances in Natural Language Processing	2013	K. Bontcheva, L. Derczynski, A. Funk, M.A. Greenwood, D. Maynard, N. Aswani
Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics	PLoS Comput Biol 9 (2)	2013	H. Cunningham, V. Tablan, A. Roberts, K. Bontcheva
NLP Techniques for Term Extraction and Ontology Population. Bridging the Gap between Text and Knowledge - Selected Contributions to Ontology Learning and Population from Text	P. Buitelaar and P. Cimiano (editors). IOS Press	2007	D. Maynard, Y. Li and W. Peters
Large-scale, Parallel Automatic Patent Annotation	Proceedings of 1st International CIKM Workshop on Patent Information Retrieval	2008	M. Agatonovic, N. Aswani, K. Bontcheva, H. Cunningham, T. Heitz, Y. Li, I. Roberts and V. Tablan

Relevant previous projects or activities			
Project name	KNOWMAK - Knowledge in the making in the European society (EC-GA N° 726992) http://www.knowmak.eu/	**Dates (start/end)**	2017-2019
Project name	OpenMinted – Open Mining Infrastructure for Text & Data (EC-GA N°654021) http://openminted.eu/	**Dates (start/end)**	2015-2018
Short description	OpenMinTeD aspires to enable the creation of an infrastructure that fosters and facilitates the use of text mining technologies in the scientific publications world, builds on existing text mining tools and platforms, and renders them discoverable and interoperablethrough appropriate registriesand a standards-based interoperability layer, respectively. It supports training of text mining users and developers alike and demonstrates the merits of the approach through several use cases identified by scholars and experts from different scientific areas, ranging from generic scholarly communication to literature related tolife sciences, food and agriculture, and social sciences and humanities.		

Project name	SoBigData – Social Mining & Big Data Ecosystem Research Infrastructure (EC-GA N°654024) http://www.sobigdata.eu/	**Dates (start/end)**	2015-2018
Short description	SoBigData proposes to create the Social Mining & Big Data Ecosystem: a research infrastructure (RI) is providing an integrated ecosystem for ethic-sensitive scientific discoveries and advanced applications of social data mining on the various dimensions of social life, as recorded by "big data". SoBigData will open up new research avenues in multiple research fields, including mathematics, ICT, and human, social and economic sciences, by enabling easy comparison, re-use and integration of state-of-the-art big social data, methods, and services, into new research. It will not only strengthen the existing clusters of excellence in social data mining research, but also create a pan-European, inter-disciplinary community of social data scientists, fostered by extensive training, networking, and innovation activities.		
Project name	AnnoMarket – Annotation Resource Marketplace in the Cloud (EC- FP7) Coordinated by USFD	**Dates (start/end)**	2012-2014
Short description	ANNOMARKET aims to revolutionise the text annotation market, by delivering an affordable, open market place for pay-as-you-go, cloud-based extraction resources and services, in multiple languages. This project is is driven by a commercially-dominated consortium, from 3 EU countries and with 43% of the budget assigned to SMEs. The key differentiating feature of ANNOMARKET is its open marketplace concept. In addition, the Software-as-a-Service (SaaS) model reduces the complexity of deployment, maintenance, customisation, and sharing of text processing resources and services, making them affordable to SMEs – both users and resource providers.		
Project name	NeOn – Social Mining & Big Data Ecosystem Research Infrastructure (EC-FP6) http://www.neon-project.eu	**Dates (start/end)**	2006-2009
Short description	The growing availability of information has shifted the focus from closed, relatively data-poor applications, to mechanisms and applications for searching, integrating and making use of the vast amounts of information that are now available. Ontologies provide the semantic underpinning enabling intelligent access, integration, sharing and use of data. As ontologies are produced in larger numbers and exhibit greater complexity and scale, we now have an opportunity to build a new generation of complex systems, which can make the most of the unprecedented availability of both large volumes of data and large, reusable semantic resources. At the same time, we face a challenge: current methodologies and technologies, inherited from the days of closed, data-poor systems, are simply not adequate to support the whole application development lifecycle for this new class of semantic applications. NeOn has developed both tool support and an associated methodology, to make possible the development of such a new generation of semantic applications, with the overall goal of producing economically viable solutions handling the whole life-cycle of these applications.		

Relevant infrastructure	
Name	**Description**
GATE	The Sheffield team has been developing the world-leading open-source GATE text mining infrastructure (http://gate.ac.uk) for the last 20 years. GATE has a vibrant user community (41,000 software downloads in the past 12 months and 265,000 in the past 9 years). GATE has a repository of over 150 text mining and NLP models and algorithms, including Infor-

	mation Extraction (IE), biomedical text mining, ontology-based semantic annotation, machine earning for IE, and NLP evaluation tools, as well as many text mining plugins provided by 3rd parties. GATE has been used for a wide variety of European projects in the past, and forms the core infrastructure for the text mining work to be carried out in the project by USFD.

Contribution to the project
The Sheffield team is one of the leading research groups in the EU in analysing both traditional forms of text and social media through text mining and semantic annotation. In the project, Sheffield will lead the work on integrating tools and services in the RCF, in particular integrating the GATE Cloud Natural Language Processing platform and the development of new tools for scalable multilingual text analysis and search services, building on work on annotation, ontology development, and topic classification in the KNOWMAK project. Sheffield will also contribute to work on the RCF infrastructure.

4.1.7. Partner 7: Fraunhofer Institute for Systems and Innovation Research (FRAUNHOFER)

Partner Profile
Since its inception in 1972, Fraunhofer ISI has been influential in shaping the German and international innovation landscape. Today, Fraunhofer ISI employs about 240 staff members, among them scientists from the fields of natural sciences, engineering, economics and social sciences who each year work on about 370 research projects. The annual budget, nearly 23 million euro in 2014, is mainly earned from contracts performed for national and international public bodies, for industry and for foundations and scientific organizations. ISI's expertise in the area of innovation research is based on the synergy of the technical, economic and social science knowledge of its staff members. In our work we apply not only a broad spectrum of advanced scientific theories, models, methods and social-science measurement instruments, but continually develop them further, utilizing the empirical findings from the research projects conducted. On behalf of our customers we investigate the scientific, economic, ecological, social, organizational, legal and political framework conditions for generating innovations and their implications. We use scientifically based analysis, evaluation and forecasting methods. Fraunhofer ISI's Competence Center "Policy - Industry - Innovation" explores the mode of operation and changes in research and innovation systems at the supranational, national and regional level. It investigates the various institutions, instruments and strategies in industry, science and government, which produce knowledge and technological innovations. For our clients from politics, science and industry we devise concepts to develop new promotional instruments and funding programs, which provide theoretical and methodological approaches to generating strategic knowledge. In this context, we use the latest indicators and evaluation concepts. The objective is to develop successful methods of active policy-making for our customers. In our research, we draw on a broad range of qualitative and quantitative analytical methods which we are continuously further developing. Among these are for instance surveys, document analyses, discursive practices and peer group analyses, but also patent, publication and social network analyses. Our scientific studies, reports and consulting help our clients to understand current processes better, that societal institutions can bear knowledgeable, foresighted and effective influence on the system.

Staff profile			
Name	Rainer FRIETSCH	**Gender**	M
Short CV / description of work experience relevant to the project	Dr. Rainer Frietsch completed a business training in wholesale and import/export trade. Studied social science at the University of Mannheim. Ph.D. in economics at the Karlsruhe Institute of Technology (KIT). Since October 2000 employed at the Fraunhofer Institute for Systems and Innovation Research ISI in Karlsruhe. Rainer Frietsch is deputy head of the Competence Center Policy and Regions and he is Coordinator of the Business Unit Innovation Indicators. He is visiting professor at the Institute of Policy and Management of the Chinese Academy of Sciences. His work focus is on Patent statistics, bibliometrics, economic and social indicators, qualifications and education in the innovation process, technology foresight, methodological foundations of empirical research		
Name	Peter NEUHÄUSLER	**Gender**	M
Short CV / description of work experience relevant to the project	Dr. Peter Neuhäusler studied social sciences, majoring in micro- and macro-sociology, applied social research and quantitative methodology at the University of Mannheim. Graduated in 2008 with a diploma in social sciences. Since February 2009 researcher and since 2012 project manager at the Competence Center Policy and Regions at the Fraunhofer Institute for Systems and Innovation Research ISI in Karlsruhe. Doctoral thesis on the topic of Intellectual Property Rights (IPR) and economic performance written at the Chair of Innovation Economics of the Berlin University of Technology from October 2009 until September 2012. Since October 2012 also research fellow at the Chair of Innovation Economics, Technical University Berlin. His work focus is on industrial and social indicators,		

	patent statistics, bibliometrics, value and benefits of patents and methodological principles of empirical research.		
Name	Patricia HELMICH	**Gender**	F
Short CV / description of work experience relevant to the project	Patricia Helmich studied Computational Linguistics at the University of Stuttgart (Master of Science) and the Saarland University (Bachelor of Science). Since Januar 2016 she has been working as a researcher in the Competence Center Policy - Industry - Innovation at the Fraunhofer Institute for Systems and Innovation Research ISI in Karlsruhe.		
Name	Oliver ROTHENGATTER	**Gender**	M
Short CV / description of work experience relevant to the project	Oliver Rothengatter studied economics at the Freie Universität (FU) Berlin and the Technical University Karlsruhe. From October 2004 until February 2007 he was a student assistant in the Competence Center Policy and Regions at the Fraunhofer Institute for Systems and Innovation Research ISI in Karlsruhe. He has been employed here since September 2007 as a member of the research team in the Business Unit Innovation Indicators.		
Name	Jakob EDLER	**Gender**	M
Short CV / description of work experience relevant to the project	JaKob Edler is Professor of Innovation Policy and Strategy at the Alliance Manchester Business School, University of Manchester and executive director of the Fraunhofer ISI institute. Jakob's research interests lie in the analysis and conceptualisation of governance and policy in science and innovation. He works on the impact of science and innovation policy and is currently leading an EU funded project to analyse evaluation practice in science and innovation policy, and has been the main lead for the database SIPER. Another major focus in recent years has been on demand side innovation policy and public procurement of innovation. Further, Jakob has worked on the conceptualisation of governance of socio-technical systems and governance of responsible research and innovation as well as on internationalisation of research and innovation activities and related governance and policy issues. Jakob is elected member of the German National Academy for Science and Engineering (Acatech), member of the Austrian Council for Research and Technological Development and member of the Royal Society for the Encouragement of Arts, Manufactures and Commerce. Since June 2016 Jakob is President of the association of 17 institutes in research and innovation policy EU-SPRI FORUM.		

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
Composition of inventor teams and technological progress - The role of collaboration between academia and industry	Research Policy 44 (7), 1360–1375	2015	Dornbusch, F., Neuhäusler, P.
The economic impacts of computer-implemented inventions at the European Patent Office	4IP Council, Fraunhofer ISI	2015	Frietsch, R., Neuhäusler, P., Melullis, K.-J., Rothengatter, O., Conchi, S.
Transnational Patents and International Markets	Scientometrics 82, 185–200	2010	Frietsch, R., Schmoch, U.
Patent Families as Macro Level Patent Value Indicators: Applying Weights to Account for Market Differences	Scientometrics 96, 27–49	2010	Neuhäusler, P., Frietsch, R.

Identifying the Technology Profiles of R&D Performing Firms — A Matching of R&D and Patent Data	International Journal of Innovation and Technology Management 14 (01), 1740003	2017	Neuhäusler, P., Frietsch, R., Mund, C., Eckl, V.

Relevant previous projects or activities			
Project name	Collection and analysis of private R&D investment and patent data in different sectors, thematic areas and societal challenges (EC-JRC)	**Dates (start/end)**	2014-2017
Short description	The tasks in the project "Collection and analysis of private R&D investment and patent data in different sectors, thematic areas and societal challenges" was to collect BERD, patent and other business data and make it ready to be put in the Research and Innovation Observatory (RIO). The analytical parts of the project required analyses on the sectoral level (NACE 2 digit). However, one of the main tasks was also to provide data and indicators based on these data sources also for Societal Grand Challenges (SGCs) and Key Enabling Technologies (KETs). In consequence, we had to develop a method to convert sectoral data into technology-based data. In addition, we were asked to fill the data gaps of BERD by sectors, so we had to identify additional (reliable) data sources and to develop a methodology to impute the still missing data.		
Project name	Joint Project: Setting up of a bibliometric competence center for the German Science System; Conducting of sub-tasks: Expected citations and classifications as well as complete coverage of patent applications from universities (including a PhD project) (National contract for Deutsches Zentrum für Luft- und Raumfahrt e.V.)	**Dates (start/end)**	2009-2014
Short description	The Competence Centre for Bibliometrics for German Science funded since late 2008 by the Federal Ministry for Education and Research (BMBF) has constructed a quality controlled in-house data infrastructure for these purposes and uses it for the development and advancement of analytical methods and indicators. The Competence Centre is an association of institutions working together to contribute to the further progress of bibliometrics and its application. The objectives of the Competence Centre are creation, development and curation of a quality assured data infrastructure for bibliometric applications research into appropriate bibliometric methods and indicators investigations on effects and meaningfulness of bibliometric results to support of young bibliometric researchers to offer services, mainly bibliometric analyses, and counselling for the application and interpretations of bibliometric information. The activities of the Competence Centre for Bibliometrics are coordinated and supported by its administrative office, which is run by DZHW.		
Project name	Innovationsindikator (Innovation Indicator 2015-2017) (Association of German Industry)	**Dates (start/end)**	2015-2017
Short description	In order to assess the technological competitiveness of Germany in the global innovation competition, a differentiated analysis and an international comparison of innovation capabilities is necessary. The "Innovation Indicator" exactly aims at providing this comparison. On behalf of acatech – Deutsche Akademie der Technikwissenschaften and the Bundesverband der Deutschen Industrie (BDI), 35 economies are analyzed alongside their innovation orientation and –capabilities. The Innnovation Indicator is provided by the Fraunhofer Institute for Systems and Innovation Research ISI in cooperation with the Zentrum für Europäische Wirtschaftforchung (ZEW). It compares the innovative performance of 35 countries alongside 38 indicators.		
Project name	Results of Public and Private Research: Patents	**Dates**	Since 2008

	(Expertenkommission Forschung und Innovation)	**(start/end)**	
Short description	In this study, commissioned by the Expert Commission Research and Innovation (EFI), which is integrated in the entire reporting system of the Expert Commission, Fraunhofer ISI contributes to the evaluation of the performance of the German science and research system by analyzing patent applications. Patents mirror the output of application-oriented research and development. The dynamics and structures of patenting activity and technological specialization are investigated in an international comparison. The investigation covers the period from 1994 and predominantly deals with international comparisons for a wide array of countries. Transnational patents, separated by 35 high-technology fields including aggregate categories (high-tech, leading-edge technology, other technology) are calculated and analyzed. In addition to country comparisons trends in international co-patents, trademark registrations and patents of German universities and research institutions will be analyzed. In addition, patent statistics at the level of German federal states will be provided. Within three additional modules, global innovations, effects of the German "Excellence Initiative" and secular trends in terms of innovation and technological change are considered.		
Project name	Regional Innovation Monitor Plus (Technopolis Belgium – European Commission)	**Dates (start/end)**	2013-2016
Short description	The overall objective of the RIM-Plus project is to contribute to the competitiveness of European regions by increasing the effectiveness of regional innovation policies and strategies, by continuing the efforts begun in the framework of the Regional Innovation Monitor from 2009-2012. To this end, the inventory of innovation strategies at regional level in Europe that was established in the course of the first Regional Innovation Monitor project, will be maintained, regularly updated in a more proactive manner than before made available to actors involved in developing regional innovation policy measures. Moreover, the so far established channels of communication (workshops, topical publications and newsfeeds) will be extended and further developed. The organisation of workshops has been outsourced to ERRIN, the European Regions Research and Innovation Network. Nonetheless, all project partners attend all workshops, invite relevant speakers and contribute content.		

Relevant infrastructure	
Name	**Description**
SIPER	SIPER is the repository of evaluation reports and policy documents in science, technology and Innovation Policy. It is a unique global dataset that is being used by policy makers, academics and other stakeholders. It has a search interface for combined keyword searches. In RISIS 2, SIPER will be further develop to also include the actual results of the evaluation reports in order to further broaden the scope from evaluation learning in STI policy to policy learning. http://si-per.eu/

Contribution to the project
We run several large-scale databases (e.g. PATSTAT, WoS, Scopus) on our servers. These are implemented as Oracle-SQL databases, but we also use PosgreSQL in parallel. We have a long-lasting experience in handling these kinds of databases so implementing the trademark database will not be a technical challenge. The databases are updated and maintained by two persons dedicated only for data base maintenance and treatment. The trademark data is available for bulk download and comes in a csv or xml format. This can easily be handled. The connection to the RISIS database will be ensured through adequate data transfer formats. We are able to be flexible in this point and adapt to the needs of the data users. Analytical capacities for the set-up and preparation of the trademark database exist as well. We have been working on the conceptual use of trademark databases since more than one decade. The analytical limits

existing so far will be reduced and new potentials will be discovered throughout this project, mainly by matching with other data sources. We also have a long-lasting experience of matching different data sources using similarity measures (e.g. Levenshtein, n-gram). For this purpose we have computer scientists as well as policy and innovation systems analysts in our team so that we can ensure to fulfil the tasks that will occur.

We will maintain and further develop the database SIPER as described above and interact with users as described in the work programme.

4.1.8. Partner 8: Universita della Svizzera Italiana (USI)

Partner Profile
The Laboratory of Science Communication is a research center in the Faculty of Economics of the Università della Svizzera italiana, focused on communication practices in science and on the sociological and economic development of research activities. It is co-directed by prof. Michael Gibbert and prof. Benedetto Lepori. The unit performs both scholarly research and contract research in the field of research policy and evaluation. Main recent projects include the two European Science Foundation projects in Transforming Universities in Europe (TRUE) and academic careers (EUROAC), as well as participation to the DG research contracts on joint and open programs in the European Research Area (JOREP). In the specific field of higher education indicators, the unit has participated to a large number of studies and contracts, including the AQUAMETH-PSR experimental project on data collection on Universities, the U-Map project on characterizing higher education institutions in Europe and, finally, U-MULTIRANK. The unit is currently involved with a central role in two EU-FP7 projects, namely the infrastructure action on Research Infrastructure for Science and Innovation Studies (RISIS) and the IMPACT-EV project on the policy impact of European project in the field of social sciences and humanities. The unit is also coordinating the European Tertiary Education Register contract on behalf of the European Commission.

Staff profile			
Name	Benedetto LEPORI	**Gender**	M
Short CV / description of work experience relevant to the project	Benedetto Lepori is co-director of the laboratory and titular professor in the Faculty of Communication sciences of USI. He is a recognized scholar in the field of research and higher education policy and of S&T indicators, with a specialization on general methodological issues (Lepori, Barré & Filliatreau 2008), on funding indicators (Lepori et al 2008; Lepori 2011) and on higher education indicators (Bonaccorsi et al. 2007). He extensively published on the major journals in the field of S&T indicators (Journal of Informetrics, Research Evaluation), research policy (Research Policy, Science and Public Policy), Evaluation (Evaluation, Research Evaluation), as well as higher education studies (Higher Education, Studies in Higher Education). He is proficient in Italian, English, German and French.		
Name	Barbara ANTONIOLI MANTEGAZZINI	**Gender**	F
Short CV / description of work experience relevant to the project	Barbara Antonioli Mantegazzini holds a Ph.D. in Economics and Law from the University of Bologna. After a post-doctoral fellowship at the LATTs (Laboratoire Techniques, Territoires et Sociétés - École des Ponts ParisTech, Université Paris-Est Marne-la-Vallée, CNRS) she joined the Università della Svizzera italiana, where she's currently Maitre d'Enseignement et de Recherche in Economics. Her research interests include Public management and strategy of public services (energy, water, waste, health), higher education, Small Medium Enterprises, Public private Partnership, Non-profit Sector. Current projects in those areas are the European comparative analysis of funding strategies in higher education institutions (USI) and the NEMo project "Novel E-MObility Grid Model" (SUPSI) within ERA-NET Horizon 2020. She published various articles in national and international journals and reviews in the field of public economics		

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
Conceptualizing and measuring performance	Research Evaluation,	2018	Lepori, B., Reale E.,

orientation of R&D funding	forthcoming		Spinello A. O.
Towards an Infrastructure for Understanding and Interlinking Knowledge Co-Creation in European Research	Lecture Notes in Computer Science (Proceedings of ESWC 2017 Workshop on Scientometrics). Springer Verlag	2017	Maynard, D., Funk, A., & Lepori, B.
Competition for talent. Country and organizational-level effects in the internationalization of European Higher Education Institutions	Research Policy, 44(3), 789-802	2015	Lepori B., Seeber M., Bonaccorsi A.
European Tertiary Education register database	https://www.eter-project.com/		
RISIS register of public research and higher education institutions	http://orgreg.joanneum.at/		

Relevant previous projects or activities			
Project name	RISIS - Research infrastructure for research and innovation policy studies (EC-GA N°313082) http://risis.eu/	**Dates (start/end)**	2014-2018
Project name	PREF - Analysis of national public research funding, by theme and by type of allocation (JRC Contract N° 154321)	**Dates (start/end)**	2015-2016
Project name	IMPACT-EV - Evaluating the Impact and Outcomes of EU SSH Research (EC Contract No. 613202)	**Dates (start/end)**	2014-2017
Project name	ETER - European Tertiary Education Register (EC-ERASMUS) Coordinated by USI https://www.eter-project.com	**Dates (start/end)**	2013-2019
Short description	The European Tertiary Education Register is a database collecting information on Higher Education Institutions (HEIs) in Europe, concerning their basic characteristics and geographical position, educational activities, staff, finances and research activities. ETER is a European Commission initiative, which aims at providing data on Higher Education Institutions in Europe. It is an Erasmus+ project fully financed by the European Commission. It is managed by the Directorate General for Education and Culture of the European Commission, in cooperation with the Directorate General for Research and Innovation, EUROSTAT and the National Statistical Autorities of the participating countries.		
Project name	TRUE - Transforming Universities in Europe (European Science Foundation)	**Dates (start/end)**	2009-2013
Short description	This joint European project aims to clarify how steering and governance of higher education affect essential organisational characteristics of HEIs and in turn how this affects the differentiation of the European HE landscape. This furthermore, will help understand how consciously designed European and national reforms affect individual institutions, systems and presumably their basic functions: education, research and innovation.The whole project is organised in 6 Individual Projects looking to specific topics inside organisational change in Europe; the Swiss project undertaken jointly by the University of Lugano and the University of Lausanne will specifically focus on the role of the internal allocation of resources inside higher education institutions for their strategy, steering and long-term development.		

Relevant infrastructure	
Name	**Description**
RISIS-ETER	RISIS-ETER represents an extension of the Tertiary Education Register (ETER; https://www.eter-project.com/) database that is supported by the European Commission through a service contract. RISIS-ETER provides an environment for enriching ETER with additional data in three respects: integrating additional variables from other RISIS datasets, particularly concerning research output (EU-FP participations, scientific publications, patents); extending the time coverage of ETER; extending the geographical scope of ETER beyond Europe. RISIS-ETER is closely integrated and hosted by the same technical infrastructure as the register of public-sector research and higher education institutions OrgReg, but is a distinct database providing statistical data.

Contribution to the project
USI will contribute in RISIS2 in the three following tasks: • Continuing the development and consolidation of register of public research and higher education institutions (OrgReg) that represents a central facility in the harmonization of organizations within the project. • Maintaining and expanding RISIS-ETER, i.e. the only integrated database of European Higher Education Institutions that combines input data and data on educational activities (for the ETER project) and research and technological output data from RISIS databases. • Developing advanced usage communities and creating a statistical support unit for advanced statistical methods that allow exploiting RISIS data for relevant research policy and scholarly questions.

4.1.9. Partner 9: JOANNEUM RESEARCH Forschungsgesellschaft m.b.H. (JOANNEUM)

Partner Profile
JOANNEUM RESEARCH is a professional innovation and technology provider with a track record of more than 30 years in cutting-edge research at international level. It focuses on applied research, technology development and consulting. *The POLICIES - Institute for Economic and Innovation Research* offers a wide range of thematic and methodological expertise to provide advice and support in the planning of technology and innovation strategies, in regional location decisions and analysis and in assessing the risk and effects of political and business decisions. We leverage our longstanding experience in empirical and model-based research to develop new market positions and expertise in order to successfully open up new markets, especially in cooperation with business partners. POLICIES provides public institutions, industrial partners and international policy makers with background analyses, scientific studies and knowledge intensive consulting services in the areas of science, research, technology and innovation. Our work forms the basis for evidence-based policy. We develop empirical fundamentals and new theoretical and conceptual approaches with a strong focus on research and innovation in the business sector. These concepts are implemented in policy-relevant analyses and recommendations also in corporate consulting services. POLICIES has long lasting experiences in analysing National innovation and Higher Education Systems and in evaluating national and international HEI and STI-Policy measures. This includes in particular proven experience and technical knowledge in research/innovation program or research/innovation policy evaluation (ex-ante impact assessment and ex ante, interim and ex post evaluation), and the use of the following tools and approaches: quantitative research methods including collection, cleaning and archiving of data; data mining techniques; development and implementation of surveys; statistical analysis of data including significance testing and multivariate analysis.

Staff profile			
Name	Michael PLODER	**Gender**	M
Short CV / description of work experience relevant to the project	Michael Ploder senior researcher and head of the research group Technology, Innovation and Policy Consulting at POLICIES – Institute for Economic and Innovation Research of JOANNEUM RESEARCH. He graduated in economics from the Karl-Franzens University in January of 2000. He took up his function as research fellow at the Graz Office of the Institute of Technology and Regional Policy, JOANNEUM RESEARCH in August 2001. His major interests are in economic analysis of technology and innovation, the analysis of sectoral systems, technological and structural change, technology assessment as well as policy evaluation and impact analyses. Michael Ploder participated in a broad range of studies on the regional, national and international level. His recent projects include: the evaluation of the "Overhead-Funding of the Federal Ministry for Education and Research, Germany", Aquameth Austria, EUMIDA and the current two ETER contracts, the evaluation of the "Spitzencluster Wettbewerb" in Germany; the external evaluation of the "Strategic Centers for Science, Technology and Innovation" in Finland. Examples for projects commissioned by the EC are Sector Innovation Watch, RTD Sectors, ERDF Ex post Evaluation 2000-2006. Michael Ploder is member of the European Network of Innovation Indicators Designers (www.enideurope.org) and the Austrian Platform of Research and Technology Evaluation (www.fteval.at). Michael Ploder was the co-coordinator of the two previous ETER contracts.		
Name	Daniel WAGNER-SCHUSTER	**Gender**	M
Short CV / description of work experience relevant to	Daniel Wagner-Schuster graduated in economics from the Karl-Franzens University. He took up his function as research fellow at POLICIES – Institute for Centre for Economic and Innovation Research of JOANNEUM RESEARCH in 2011. His main areas of work are Technology, Higher Education and Science Policy with a particular focus of quantitative methods and data base management. Recent projects in this context are the evaluation of the		

the project	"Overhead-Funding of the Federal Ministry for Education and Research, Germany", the evaluation of the "Spitzencluster Wettbewerb", the "multi-KETs pilot lines project", the "Austrian Research and Technology Report" and was coordination the data collection and infrastructure part of the previous two ETER projects as well as the current RISIS activities at JOANNEUM RESEARCH.

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
European Tertiary Education register database	www.eter-project.com/		
OrgReg - RISIS register of public research and higher education institutions	orgreg.joanneum.at/		
FirmReg – RISIS register of firms	firmreg.joanneum.at		

Relevant previous projects or activities			
Project name	RISIS - Research infrastructure for research and innovation policy studies (EC-GA N°313082) http://risis.eu/	**Dates (start/end)**	2014-2018
Project name	EUMIDA – Feasibility Study for Creating a European University Data Collection (Contract No. RTD/C/C4/2009/0233402, 2009-2010)	**Dates (start/end)**	2009-2010
Short description	EUMIDA was a feasibility study to create a European University Data Collection. The project aimed to build a complete census of European universities and included a pilot data collection with particular emphasis on those universities that are research-active.		
Project name	Establishing a European Tertiary Education Register (ETER) (Contract EAC 2013-0308, 2013-2015) Implementing and Disseminating the European Tertiary Education Register (Contract EAC 2015-0280, 2015-2017) Implementation and further Development of a European Tertiary Education Register (Contract 934533-2017 A08-CH, 2017-2019)	**Dates (start/end)**	2013-2019
Short description	The European Tertiary Education Register is a database collecting information on Higher Education Institutions (HEIs) in Europe, concerning their basic characteristics and geographical position, educational activities, staff, finances and research activities. ETER is a European Commission initiative, which aims at providing data on Higher Education Institutions in Europe. It is an Erasmus+ project fully financed by the European Commission. It is managed by the Directorate General for Education and Culture of the European Commission, in cooperation with the Directorate General for Research and Innovation, EUROSTAT and the National Statistical Authorities of the participating countries.		
Project name	Study on Knowledge Transfer and Open Innovation (No 2012/S 248-408643)	**Dates (start/end)**	2013-2018
Short description	The Study on Knowledge Transfer and Open Innovation aims to identify, analyse and synthesise various ways of knowledge transfer from the perspective of an open innovation		

	paradigm all over Europe at different points in time in a comprehensive, comparable and clear manner. This includes an analysis of state-of-the art practices, emerging trends, the barriers, challenges and success factors for open innovation and knowledge transfer. The output will be a solid EU-wide information base on open innovation and knowledge transfer.

Relevant infrastructure	
Name	**Description**
ETER	The ETER infrastructure is a homogenous system, which is dedicated to the workflow of a yearly data collection as well as to usability. It allows the project team to manage the data structure, data validation and quality as well as user management directly within the application. The user needs are addressed by a large set of possibilities regarding searching, filtering and exporting data. As an additional service, the ETER web application allows also tabulation and visualization in the web application.

Contribution to the project
The contribution of JOANNEUM RESEARCH will focus on the following work packages: • Developing services for the RISIS core facility. • Further developing the existing RISIS infrastructure services OrgReg and FirmReg. • Performing a prototype for building a career dataset. The specific task for JOANNEUM will be to implement a prototype for Austria.

4.1.10. Partner 10: University of Sussex (UOS)

Partner Profile
With over 50 years of experience, SPRU (Science Policy Research Unit) is internationally recognised as a leading centre of research on science, technology and innovation policy. Founded in 1966 by Christopher Freeman, a pioneer of what is now known as innovation studies, SPRU was one of the first interdisciplinary research centres in the field of science and technology policy and management. Today, with over 60 research staff, more than 70 doctoral students, over £6m of on-going Research Council funded projects and over £2m of ongoing European Commission funded projects, as well as the leading journal in its field, Research Policy, SPRU remains at the forefront of new ideas, problem-orientated research, inspiring teaching, and creative, high impact engagement with decision makers across government, business and civil society. Drawing on insights from across the social and natural sciences, engineering and humanities, SPRU's pioneering research is known and respected worldwide and addresses pressing global policy agendas, including the future of industrial policy, inclusive economic growth, the politics of scientific expertise, energy policy, entrepreneurship, and pathways to a more sustainable future. Over time, SPRU has built substantial capacities over many fields of research, including the study and measurement of science, technology and innovation, research and innovation infrastructure, firm innovation and growth. The research team will therefore benefit from exposure to a vibrant research environment and will draw on the knowledge and capacities of colleagues not directly involved in this project, as well as through SPRU's large network of international contacts and frequent eminent visitors.

Staff profile			
Name	Roberto Camerani	**Gender**	M
Short CV / description of work experience relevant to the project	Roberto Camerani is Research Fellow at SPRU (Science Policy Research Unit), University of Sussex. He obtained his PhD in Science and Technology Policy from SPRU with a thesis about the diffusion of innovations. He also has a degree in Economics at Bocconi University, and an MSc in Innovation and Industry Analysis. At SPRU he convenes the Innovation in the Creative Economy course, and he has previously convened the courses Introduction to Statistics, and Innovation in Infrastructure. He is currently participating in a large EU funded project about science and technology indicators and research infrastructure for innovation policy studies. He recently participated in a number of research projects as Principal Investigator, Co-Investigator or Research Fellow, funded by the UK AHRC, NESTA, and the European Commission. In the past, he worked at CRIOS (Center for Research on Innovation, Organisation and Strategy), Bocconi University, Milan, and INGENIO (Institute of Innovation and Knowledge Management), Polytechnic University, Valencia. His main research interests include the adoption and diffusion of innovations, science and technology indicators, patent analysis, firm growth and innovation in the creative and digital sectors. His research skills are mostly quantitative. He is familiar with several statistical and econometric techniques and is proficient in the use of a variety of statistical software. Roberto is also experienced in the development and administration of surveys to firms and final consumers.		

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
Do Firm Publish? A Cross-Sectoral Analysis of Corporate Science	To be submitted to Research Policy	2018	Camerani, R., Rotolo, D., Grassano, N.
Scientific Publication Activity of Scoreboard	European Commission,	2017	Rotolo, D., Camerani,

Companies	available at this link		R.
Creative Diversity and Firm Performance	Submitted to Small Business Economics	2018	Camerani, R., Siepel, J., Masucci, M.
The Fusion Effect: The Economic Returns to Combining Arts and Science Skills	NESTA, available at this link	2016	Siepel, J., Pellegrino, G., Camerani, R., Masucci, M
Patterns of Technological Entry in Different Fields: An Analysis of Patent Data	In Malerba, F. (Ed.) Knowledge Intensive Entrepreneurship and Innovation Systems: Evidence from Europe, Routledge, 103-135.	2010	Camerani, R., Malerba, F.

Relevant previous projects or activities			
Project name	RISIS - Research infrastructure for research and innovation policy studies (EC-GA N°313082) http://risis.eu/	**Dates (start/end)**	2014-2018
Project name	PASCO (Publication Activity of Scoreboard Companies) (Funded by: Directorate B - JRC, European Commission)	**Dates (start/end)**	2016-2017
Short description	The PASCO project is an explorative project that aims at increasing understanding of the extent to which large corporates' R&D activity can be traced by examining their scientific publications. To do so, the project complements the conventional set of financial indicators used to assess the R&D investments of large corporates with indicators about the publication activity of these. In particular, the project examines the extent to which firms included in the 2014 EU Industrial R&D Investment Scoreboard are involved in publication activity. The Scoreboard includes 2,500 firms most active in terms of R&D expenditure. These firms account for about 90% of the global private R&D expenditure. On the basis of a novel methodological approach to collect publication data from the Web of Science (WoS) for all Scoreboard firms (2,500 companies) and their subsidiaries (about 570,000 subsidiaries), the project retrieved full bibliographic information about more than 314,000 publications authored by the firms in the sample. By using a number of publication-based indicators, the project provided perspectives on firms' R&D efforts and strategies that are complementary to more traditional financial indicators.		
Project name	A study of innovation and growth in the UK creative industries (commissioned by Nesta, ref PR300)	**Dates (start/end)**	2014-2016
Short description	This research explores the combination of skills in UK companies and its impact on performance. Using official UK data, the study develops indicators on firm innovation and capabilities based on firms' access to a range of internal of external sets of skills. The project focuses on the particular role of arts and science skills, and analyses weather the combination of these two types of skills has an impact on firm performance. The research finds compelling evidence suggesting that firms combining these skills are more likely to grow in the future, are more productive, and are more likely to produce innovations.		

Relevant infrastructure	
Name	**Description**
FirmReg	The Firm Register (FirmReg) is a central facility within RISIS which aims at uniquely identifying and tracking over time all the business included in the firms' datasets which are part of RISIS, allowing the addition of more datasets over time. Currently FirmReg includes more than 60,000 firms, with harmonized data about their demographic characteristics, location, and parent-subsidiary relationships (about 6,000 linkages)
Original databases	Longitudinal dataset including more than 314,000 publications authored by the firms in the 2014 EU Scoreboard (2,500 firms, and about 570,000 subsidiaries) from 2011-2015.
Acquired databases	Web of Science, PATSTAT, Orbis, FAME

Contribution to the project

In RISIS 2, SPRU will bring its consolidated experience in working on firm's R&D, innovation and growth strategies, and on the development of innovation indicators based on financial data, patents, scientific publications, and firm capabilities/skills.
SPRU will also bring the specific competences and knowledge acquired during the RISIS project to work on firm-level data. In particular, SPRU will capitalize on the work undertaken in the design, development and implementation of the first prototype of FirmReg.
SPRU will mainly provide inputs to two work packages:

- In WP 5: SPRU will contribute to consolidate the current prototype of FirmReg developed during the first RISIS project, by extending its perimeter and inclusion criteria, by revising the current structure and variables, by improving its linkages with other datasets, and by developing procedures to allow periodic updates and further extensions in the future.
- In WP 8: SPRU will work with the other project partners on the update of the dataset on fast growth mid-sized firms (Cheetah), and on the analysis of the innovation and growth dynamics of these firms.

Finally, SPRU will also build on its long-standing experience in the field of science and technology policy, and the economics of innovation, and on its attitude to address policy-relevant questions to provide original theoretical and methodological approaches to the overall RISIS2 project.

4.1.11. Partner 11: German Centre for Higher Education Research and Science Studies (DZHW)

Partner Profile
The German Centre for Higher Education Research and Science Studies (DZHW) conducts empirically driven research with practical application in the fields of higher education and science studies. The Institute's research results, data, and analyses support higher education institutions and educational administration in shaping the profile of higher education in Germany and Europe. The research area Research System and Science Dynamics concentrates on the analysis of the research system in a national and international context with a special regard to the interdependency of the different governance, financing, and funding policies. The DZHW's Berlin branch furthermore focusses on the development and practical application of indicator-based assessments for the measurement of research performance and uses bibliometric and scientometric approaches.

Staff profile			
Name	Stefan HORNBOSTEL	**Gender**	M
Short CV / description of work experience relevant to the project	Prof. Dr. Stefan Hornbostel, Director of research area Research System and Science Dynamics, studied Social Sciences at the University of Göttingen. In 1995, he received his PhD from Freie Universität Berlin. He worked at the Universities of Kassel, Cologne, Jena and Dortmund, as well as at the Center of Higher Education Development (CHE – Centrum für Hochschulentwicklung). Stefan Hornbostel served as Director of the Institute for Research Information and Quality Assurance (IFQ) from 2005 to 2015. He was appointed Professor at Humboldt-Universität zu Berlin, Department of Social Sciences (Science Studies) in 2005. Since 2016, he is head of the research area "Research System and Science Dynamics" at German Centre for Higher Education Research and Science Studies (DZHW). He is a member of the advisory board of the Saxon State and University Library Dresden (SLUB), and member of the advisory board for the Centre for Research Quality and Policy Impact Studies (R-Quest), Oslo. His research interests lie in in the field of science studies, bibliometrics, and elite sociology.		
Name	Jakob TESCH	**Gender**	M
Short CV / description of work experience relevant to the project	Jakob Tesch, researcher at DZHW, studied Social Sciences, Media Science and Sociology of Law (B.A.) and Social Sciences at Humboldt-Universität zu Berlin (M.A.). He was a data analyst freelancer for quantitative methods and worked for the Institut für Regionalentwicklung und Strukturplanung, the European School of Management and Technology as well as for Humboldt-Universität zu Berlin. He has started to work at DZHW (formerly iFQ) in December 2011 and has successfully completed a wide range of national and international research and tender projects on indicators for human resources in Science and research careers.		

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
National Academics Panel Study and its predecessor study PROFILE - The German Doctoral Candidates and Doctorate Holders Study	http://datasets.risis.eu/	2017	
Semantic Wiki for the analysis of research careers, containing interlinked information on theories, measurement concepts and data	www.researchcareers.info	2017	

Developing research career indicators using open data: the RISIS infrastructure	Peripheries, frontiers and beyond. Proceedings of the 21st International Conference on Science and Technology Indicators (S. 771-779) Editorial Universitat Politècnica de València	2016	Cañibano, C., Woolley, R., Iversen, E., Hinze, S., Hornbostel, S., & Tesch, J.
A Functional Review of Literature on Research Careers	Ingenio Working Paper Series. 2016-05	2016	Woolley, R., Cañibano, C., & Tesch, J.
Structured Diversity – How structured programs change doctoral education in Germany	Research Evaluation, 4, 292-301	2017	Ambrasat, J., & Tesch, J.

Relevant previous projects or activities			
Project name	RISIS - Research infrastructure for research and innovation policy studies (EC-GA N°313082) http://risis.eu/	**Dates (start/end)**	2014-2018
Project name	ProFile - Promovierendenpanel and National Academics Panel Study (DFG) http://www.promovierendenpanel.de	**Dates (start/end)**	2006-2019
Short description	In general, little is known about doctoral education in Germany. Even the total number of doctoral candidates at German universities can only be estimated; official data do not exist. Therefore, reliable information on the conditions, processes and success of doctoral education is essential. To improve the data available, the former iFQ has set up ProFile, a longitudinal study focusing on the situation of doctoral candidates and their postdoctoral professional careers. Since April 2009, doctoral candidates at different universities and funding institutions have been surveyed at regular intervals. ProFile aims at identifying determinants of postdoctoral career development and providing information on conditions of doctoral education in a comparative perspective. Special attention is paid to the effects of structured doctoral programs, which have increasingly emerged during recent years. Since 2016 DZHW is continuing this project.		
Project name	MERCI – Monitoring European Research Council‘s Implementation of Excellence (EC-FP7)	**Dates (start/end)**	2009-2014
Short description	This project aims to develop and implement instruments for a continuous assessment of the ERC Starting Grants impact on funded researchers. The main item will be a panel study (standardized online interviews) which includes two waves and which is complemented by qualitative interviews and bibliometric studies (triangulation of methods) applied to second stage applicants. The first wave will take place at runtime of the respective projects when grantees will have made sufficient experiences under funding. The second wave will proceed at an adequate time after the end of funding (rejection of the project respective) to gather possible middle-term (long-term) effects. In focus are direct impacts of the Starting Grants on the grantees scientific careers, scientific merits (bibliometric analyses), motives concerning mobility, selection of and relation to the host institution, job satisfaction, soft skills and their prestige gained from being funded by the ERC.		
Project name	SciMo - Determinants and career effects of scientists' international mobility (DFG)	**Dates (start/end)**	2016-2019
Short descrip-	The SciMo project aims to identify factors that influence researchers' international mobili-		

tion	ty, and the positive or negative effects international mobility has on their careers. It also focuses on the issue of how far possible career effects of different types of international mobility depend on physically crossing borders or whether they can be substituted by forms of virtual mobility.

Relevant infrastructure	
Name	**Description**
Competence Centre for Bibliometrics	Affiliated with DZHW's research area 2 is the Competence Centre for Bibliometrics, a compound of seven partners that has been funded by the Federal Ministry of Education and Research since 2008 and provides a qualitative and sustainable dataset for bibliometric analysis. One of the main objectives of the Competence Centre is to make the database a permanent research infrastructure which is open to new services and applications.
Name	**Description**
DZHW Research Data Centre	The German Council of Science and Humanities (Wissenschaftsrat) has often emphasised the need for a Research Data Centre (FDZ) for higher education research and science studies, both for the DZHW and for higher education research in general, to fill the gaps in the research data infrastructure. The DZHW's FDZ is thus available to the whole scientific community as a permanent, reliable research infrastructure feature and will guarantee transparent access to anonymised microdata using a range of data access paths. The scientific use files comprise an important factor in the transfer of research findings into the scientific community; guest researcher places and remote data use are also available. A website with a meta data portal is provided under https://metadata.fdz.dzhw.eu in addition to the user service to ensure that the provision is transparent.

Contribution to the project
DZHW will contribute in RISIS2 to the three following tasks: • Coordinating the development of a prototype database for the study of (non-)academic careers of PhDs (including the work for the German prototype) • Maintaining and continuing to provide access to the ProFile facility • Maintaining and expanding the semantic wiki researchcareers.info

4.1.12. Partner 12: University of Strathclyde (STRATHCLYDE)

Partner Profile
Established more than 200 years ago 'for the good of mankind', the University of Strathclyde is renowned for excellent teaching, research and strong links with industry, government and business. Its roots can be traced back to 1796, when Professor John Anderson's legacy established a 'place of useful learning' — the only higher education institution to be created in Scotland during the Enlightenment. This tradition remains at the heart of its ethos today, as a leading international technological university. University of Strathclyde is situated in the heart of Glasgow – one of the UK's largest cities – and it has a vibrant, international community including more than 20,000 students and 3,500 staff from over 100 nations. Innovation and entrepreneurship is one of the seven research themes of the university, with the Hunter Centre for Entrepreneurship is leading this line of research. Located within the Strathclyde Business School, one of the elite UK business schools, the Hunter Centre is now one of the largest university-based centres of entrepreneurship in Europe and is home to a team of around 20 recognised experts in a range of specialised fields. Endowed by an alumni-entrepreneur and growing rapidly, Hunter Centre for Entrepreneurship leads the entrepreneurship and innovation research in the UK and beyond.

Staff profile			
Name	Abdullah Gök	**Gender**	M
Short CV / description of work experience relevant to the project	Dr Abdullah Gök is joining to the University of Strathclyde as a Senior Lecturer (Associate Professor) and Chancellor's Fellow in June 2018, as part of the University of Strathclyde's prestigious Global Talent Programme. Previously he worked as a Lecturer and Research Fellow at the Alliance Manchester Business School (AMBS) and Manchester Institute of Innovation Research (MIoIR), University of Manchester, where is currently an Honorary Research Fellow. Prior to moving to the UK in 2006, Abdullah worked at The Scientific and Technological Research Council of Turkey (TUBITAK) as a policy researcher. Abdullah's research spans three inter-related areas of innovation management and policy namely: i) emerging technologies and their governance, ii) science, technology and innovation policy and evaluation and iii) social innovation. Besides these three areas, Abdullah utilises novel and big data sources by developing novel methodologies through data science. Abdullah has widely published research articles in these areas and he has recently been building the European Social Innovation Database (ESID 1.0) by using advanced text-mining and machine learning approaches, as part of the KNOWMAK project.		
Name	Yashar Moshfeghi	**Gender**	M
Short CV / description of work experience relevant to the project	Dr Yashar Moshfeghi is a Chancellor's Fellow and Lecturer at the University of Strathclyde since July 2017, as part of the University of Strathclyde's prestigious Global Talent Programme. Dr Moshfeghi leads the Data Science and Analytics (DSA) research group within the Department of Computer and Information Sciences at the University of Strathclyde where homes to internationally recognised researchers in Data Science and particularly Information Retrieval and Machine Learning. Dr Moshfeghi is also an Honorary Lecturer at the University of Glasgow. Previously he worked as a postdoctoral research associate in Big Data within the Urban Big Data Centre at the University of Glasgow, after obtaining his PhD in 2012 from the School of Computing Science at the University of Glasgow and completing a two-year research assistant position in Information Retrieval there from 2012 until 2014. Much of Dr Moshfeghi's research focuses on processing large volumes of heterogeneous data, i.e. Big Data analysis using statistical inference models. In particular, his research interests and expertise are in line with information retrieval and big data processing and analysing data using machine learning and data mining techniques. He		

	is also interested in natural language processing in particular extracting emotion from textual documents, and utilising them to improve information retrieval applications. His interests, however, are extended beyond information retrieval and pertain to rigorously understanding the world of Data Science.

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
Use of Web Mining in Studying Innovation	Scientometrics	2015	Abdullah Gok, Alec Waterworth, Philip Shapira
Graphene Enterprise: Mapping Innovation and Business Development in a Strategic Emerging Technology	Journal of Nanoparticle Research	2016	Philip Shapira, Abdullah Gok, Fatemeh Salehi
Probing "green" industry enterprises in the UK: A new identification approach	Technological Forecasting and Social Change	2014	P. Shapira, A. Gok, E. Klochikhin, M. Sensier
Topic Detection and Tracking on Heterogeneous Information	Journal of Intelligent Information Systems	2017	Long Chen, Huaizhi Zhang, Joemon M Jose, Haitao Yu, Yashar Moshfeghi, Peter Triantafillou
Exploiting Crowdsourced Geographic Information and GIS for Assessment of Air Pollution Exposure During Active Travel	Journal of Transport & Health	2017	Yeran Sun, Yashar Moshfeghi and Zhang Liu
Search Process as Transitions Between Neural States	Proceedings of the 27th International Annual WWW Conference (WWW'18)	2018	Yashar Moshfeghi, and Frank E. Pollick
A Study of Snippet Length and Informativeness: Behaviour, Performance and User Experience	Proceedings of the 40th Annual ACM SIGIR Conference (SIGIR'17)	2017	David Maxwell, Leif Azzopardi and Yashar Moshfeghi

Relevant previous projects or activities			
Project name	KNOWMAK - Knowledge in the making in the European society (EC-GA N° 726992) http://www.knowmak.eu/	**Dates (start/end)**	2017-2019
Project name	UK ESRC Enterprise Research Centre https://www.enterpriseresearch.ac.uk/	**Dates (start/end)**	2013-2018
Short description	The Enterprise Research Centre (ERC) is an independent research centre which aims to provide trusted, robust and independent commentary on SME research and policy. ERC has six core research programmes focussing on entrepreneurial intentions, SME leadership,		

	diversity, finance, innovation and business demographics. Each research programme aims to contribute to our understanding of SME growth and performance. Research methodologies vary between the different research programmes but draw on approaches from economics and econometrics, management science, sociology, geography and psychology. The work of the ERC has a strong comparative and international element. Funding Programme/Body: UK ESRC Funded Value: £4,547,611		
Project name	DiasporaLink (EC-GA N° 645471) http://diasporalink.com/	**Dates (start/end)**	2015-2019
Short description	DiasporaLink is a 4-year exchange program between 24 universities and research institutes representing EU, the Americas, Africa and Australia and will investigate, evaluate and facilitate transnational diaspora entrepreneurship, TDE as driver of development and wealth creation in countries of origin and residence. Funding Programme/Body: EU H2020 EU contribution: EUR 571 500		

Relevant infrastructure	
Name	**Description**
ESID 1.0	European Social Innovation Database, being developed as part of the KNOWMAK projects is currently the most comprehensive and authorative source of social innovative activity in Europe and beyond. Its distinctive methodology benefits from collecting and analyzing online data on social innovation through semi-automated text-mining and machine learning approaches.

Contribution to the project
STRATHCLYDE team has a wealth of experience on delivering research within international research projects. They will contribute to the project for the following tasks: • Developing ESID 2.0 dataset and integrating it into other RISIS datasets • Developing conceptual and theoretical approaches in understanding the social innovative activity in Europe and beyond • Advising other partners in collecting and analyzing data through text mining and machine learning • Support training of new generation of scholar in STI studies; • Dissemination of results achieved and supporting actions toward integration of the STI research community; Development of participative events devoted to produce social and political impact through stakeholders' interactions.

4.1.13. Partner 13: Universita degli Studi di Roma – La Sapienza (SAPIENZA)

Partner Profile
The Department of Computer, Control and Management Engineering Antonio Ruberti (Dipartimento di Ingegneria Informatica, Automatica e Gestionale Antonio Ruberti, in short DIAG) of the University of Roma La Sapienza was established in 1983 as an evolution of the Istituto di Automatica; in 2001 it was named after Antonio Ruberti, the eminent scholar who founded it. DIAG is a center for research and education at the undergraduate and graduate levels in computer, system, and management sciences. Basic research is the main goal of DIAG, with a strong emphasis on interdisciplinary research, on applications that stimulate basic research, and with a specific attention to technology transfer and dissemination of results. The Department undertakes research and analysis for a number of national and international agencies and organizations. See the detailed activities and projects available in its research reports at: http://www.dis.uniroma1.it/research_report. This provides the Department with ample expertise and capacity to take on this project. According to the Law no. 232/2016, the DIAG department has been evaluated as a university department of excellence (2018-2022).

Staff profile			
Name	Cinzia DARAIO	**Gender**	F
Short CV / description of work experience relevant to the project	Cinzia Daraio is an Associate Professor of Management Engineering at the University of Rome "La Sapienza" where she teaches Productivity and Efficiency Analysis, Quantitative Models for Economic Analysis and Management and Economics and Business Organization. She has been invited keynote speaker and/or visiting researcher in several international research centers in Belgium, China, Denmark, France, Germany, Italy, Romania, Spain, Sweden, Switzerland and US. She is specialized in science and technology indicators, higher education microdata and methodological and empirical studies in productivity and efficiency analysis. Over the years she participated in several international and national research projects and expert groups at the European Commission and OECD on these subjects. She is member of the Editorial Board of Scientometrics, International Transactions in Operational Research, Journal of Productivity Analysis, Journal of Data and Information Science, Scholarly Metrics and Analytics and others. She is referee for more than 35 journals. She was project evaluator for the Rustaveli National Science Foundation, external reviewer for the Portuguese Foundation for Science and Technology and member of the Commission of Experts for the Evaluation of the Third Mission Activities of the Italian National Agency for the Evaluation of Universities and Research. She authored more than 150 publications including a monograph an edited book, many articles in international peer-reviewed journals, several chapters and papers presented at conferences and in proceedings. Currently, Professor Cinzia Daraio is the coordinator of the research unit of Sapienza of the RISIS project (FP7). In the current ETER 3 contract service she is the coordinator of the Sapienza team as she was in the two previous ETER 1 and 2 contracts. She is member of the Core Team of the European Tertiary Education Register (ETER 3) Contract, Quality manager and national expert for Italy, Ireland, UK and Malta, as she was for the previous ETER contracts, where she has been the coordinator of the Data validation and data Quality WP. Prof. Cinzia Daraio will coordinate the activity of the Sapienza Research unit in RISIS II.		
Name	Giuseppe CATALANO	**Gender**	M
Short CV / description of work experience relevant to	Giuseppe Catalano is full professor of Management Engineering at the Sapienza University, where he teaches Public Sector Services Management and Economics and Business Administration. His expertise on Public Services Management will be help-		

the project	ful for the organization of the Training on Advanced Benchmarking Models and Techniques. He is working within the ETER III project.		
Name	Renato BRUNI	**Gender**	M
Short CV / description of work experience relevant to the project	Renato Bruni is assistant professor of Operational Research at the Sapienza University, where he teaches Operational Research Methods. He will contribute to the project with his expertise in data imputation and techniques during the Training on Data Quality and to the data quality package. He is working within the ETER III project.		
Name	Giorgio MATTEUCCI	**Gender**	M
Short CV / description of work experience relevant to the project	Giorgio Matteucci assistant professor of management engineering. He will contribute to the project with his expertise during the Training on Data Quality and to the data quality package.		
Name	Domenico LAISE	**Gender**	M
Short CV / description of work experience relevant to the project	Domenico Laise is Associate Professor at the Department of Computer, Control and Management Engineering at the University of Rome "La Sapienza", Italy. His main interests are in Bounded Decision Models, Multicriteria Decision Making, Management Control Models, Organizational Design. He will contribute to the organization of the training on Advanced Benchmarking Models and Techniques.		
Name	Leopold SIMAR	**Gender**	M
Short CV / description of work experience relevant to the project	Leopold Simar is emeritus professor at the Université catholique de Louvain, associated at Sapienza: Quality manager/statistician. Professor Simar is expert in frontier efficiency analysis, in statistics and data analysis with a longstanding experience on European projects on these matters. He participated in EUMIDA as Quality manager and in the current ETER service as Quality Manager in the Sapienza team (as in the previous two ETER contracts). He will contribute to the organization of the training on Advanced Benchmarking Models and Techniques.		
Name	Henk F. MOED	**Gender**	M
Short CV / description of work experience relevant to the project	After being senior scientific advisor at Elsevier, Prof. Henk Moed is now visiting professor at Sapienza. In 2009 he was appointed as professor of research assessment methodology at the University of Leiden, Netherlands, and from 1986 to 2010 he has been senior staff member at the Centre for Science and Technology Studies (CWTS), in the Department (Faculty) of Social Sciences at Leiden University. He will provide advisory activity to the Sapienza team.		
Name	Gianpaolo IAZZOLINO	**Gender**	M
Short CV / description of work experience relevant to the project	Gianpaolo Iazzolino's research interests are mainly in Intellectual Capital and Performance Evaluation. In particular he works on methodologies for analyzing value creation in knowledge-based organizations and the impact of intellectual capital on firm performances. He will contribute to the organization of the training on Advanced Benchmarking Models and Techniques.		

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
Directional Distances and their Robust versions. Computational and Testing Issues	European Journal of Operational Research, 237, 358-369	2014	Daraio C. Simar L.
Explaining Inefficiency in Nonparametric Production Models: the State of the Art	Annals of Operations Research, 214,	2014	Badin L., Daraio C., Simar L.

	5–30		
Rankings and University Performance: a Conditional Multidimensional Approach	European Journal of Operational Research, 244, 918-930	2015	Daraio C., Bonaccorsi A., Simar L.
Grand Challenges in Data Integration. State of the Art and Future Perspectives: An Introduction	Scientometrics, 108 (1), 391-400.	2016	Daraio C., Glänzel W.
Beyond university rankings? Generating new indicators on universities by linking data in open platforms	Journal of the Association for Information Science and Technology.	2017	Daraio C., Bonaccorsi A.
Central Limit Theorems for Conditional Efficiency Measures and Tests of the "Separability" Condition in Nonparametric, Two-Stage Models of Production	The Econometrics Journal.	2017	Daraio C., Simar L., Wilson P.W.
A framework for the assessment of Research and its Impacts	Journal of Data and Information Science, Vol. 2 No. 4, 7–42.	2017	Daraio C.

Relevant previous projects or activities			
Project name	RISIS - Research infrastructure for research and innovation policy studies (EC-GA N°313082) http://risis.eu/	**Dates (start/end)**	2014-2018
Project name	Implementation and Further Development of the European Tertiary Education Register ETER III, Contract No. 934533-2017 A08-CH"	**Dates (start/end)**	2017-2019
Project name	Implement and Disseminate the European Tertiary Education Register (ETER 2015-2017), Contract No. EAC-2015-0280	**Dates (start/end)**	2015-2017
Project name	European Tertiary Education Register (ETER), Contract No. EAC-2013-038	**Dates (start/end)**	2013-2015
Project name	Modelling and Mapping the Dynamics of Knowledge. A Co-Creation Indicators Factory Design, Project H2020 Sapienza 2017 Awards, N. PH11715C8239C105	**Dates (start/end)**	2017-2020
Project name	Italian PRIN 2015 Project Prot. 2015RJARX7 Comparing Governance Regime Changes in Higher Education: systemic performances, national policy dynamics, and institutional responses. A multidisciplinary and mixed methods analysis.	**Dates (start/end)**	2017 - 2020
Project name	Sapienza Big Projects Awards 2015 (C26H15XNFS): Establishing a Knowledge Infrastructure for the Development of Methodologies for the Assessment of Research and its Impacts (KIMAR)	**Dates (start/end)**	2015-2018

Relevant infrastructure	
Name	**Description**

Contribution to the project
The Sapienza unit coordinated by Prof. Cinzia Daraio will contribute its expertise and competences to the project organizing and being responsible for two training activities detailed below. The first Training will be on "Data quality" (M18). The reimbursement of the expenses related to the participants will not be covered by Sapienza but by the training WP. The second Training will be on "Advanced Benchmarking Models and Techniques" (M24). The reimbursement of the expenses related to the participants will not be covered by Sapienza but by the training WP. Moreover, Sapienza will be responsible for and will realize a Data quality package that will be made available on line (M36). Finally, Sapienza will produce at M48 an Analytical Report with full implementation details about a successful application on RISIS data of Advanced Performance Techniques.

4.1.14. Partner 14: Nordic Institute for Studies in Innovation, Research and Education (NIFU)

Partner Profile
NIFU is an independent social science research institute that is organized as a non-profit foundation. First established in 1969, it merged with the STEP-Group (Science Technology, Innovation and Economy) in 2005. Today, NIFU conducts social science research in the areas of innovation, research, and education in Norway for a range of mainly public-sector organizations, including The Research Council of Norway, Innovation Norway, Norwegian ministries including Trade and Industry, Education and Research, Local Government and Regional Development, The European Union, The Nordic Council of Ministers, Nordic Energy Research. Its mission is to provide theoretical and practical insight into the dynamics of these activities, while contributing to relevant policy development. It is a regular contributor to Nordic (Nordic Council of Ministers, Nordic Energy Research, The Nordic Academy for Advanced Study) as well as to international bodies particularly the OECD. It is organized into four main areas : 1) Statistics and Indicators, 2) Studies in Primary and Secondary Education, 3) Studies in Higher Education, 4) Studies in Research and Innovation The four areas fit well with the profile of the RISIS-II infrastructure. In terms of the empirical component, NIFU's longstanding work in the Statistics and Indicators section includes a central role as national producer of public sector R&D statistics (for universities and public research organizations); in addition, the institute compiles IPR statistics (patents, trademarks, and design) for national statistical reports. Combine with its role as a producer of national statistics in focal areas of RISIS-II, NIFU has a strong research profile within the fields of innovation policy and science, where it has carried out numerous projects for national and international governmental and other public-sector bodies. It has participated in the Framework Programmes of the EU Commission since the mid-1990s. It also contributed to the RISIS-I infrastructure.

Staff profile			
Name	Susanne SUNDNES	**Gender**	F
Short CV / description of work experience relevant to the project	Susanne Sundnes is Head of Research, Statistics and Innovation. She leads NIFU's work with R&D statistics in the higher education and the research institute sectors. Internationally she contributes the OECD and Eurostat on R&D statistics. She contributed to the RISIS-I project on the topic of research institutes.		
Name	Eric J. IVERSEN	**Gender**	M
Short CV / description of work experience relevant to the project	Eric J. Iversen is associated with the Statistics and Indicators section, where he has been responsible for development and maintenance of innovation data and indicators since 2011. In the field of innovation studies, his dissertation focused on IPR, and he has worked empirically with patents and trademarks for over 20 years. He has held leading roles in over a dozen projects for Norwegian government agencies, Nordic agencies, and other international organizations including the World Intellectual Property Organization and the EU Commission. He work with Statistics Norway and the Norwegian Patent Office (NIPO) on combined datasets for use by the research community (FlipdOPEN). He was Framework Consortium Leader, FWC DG RTD, Lot 2, Data Collection and performance indicators to monitor European research policy (2012-2016). His work extends to science studies, where he has led or otherwise contributed to a number of national and international (Nordic and EU) empirically oriented projects on researcher mobility and the changing research career. He coordinated and contributed to NIFU's work in the RISIS-I project.		
Name	Asgeir SKÅLHOLT	**Gender**	M
Short CV / description of work experience relevant to	Asgeir Skålholt is a researcher at Nordic studies in Innovation, Research and Education (NIFU). His research centers on the role of skills in the labour market. Within the		

the project	tradition of innovation studies, he has focused on how different types of skills contribute in innovation processes. Skålholt has long experience with the use and analysis of administrative data. He is working on finishing his PhD-thesis on the role of skills in innovation.		
Name	Pål BØRING	**Gender**	M
Short CV / description of work experience relevant to the project	Pål Børing is a senior researcher in the field of economics. He has worked at NIFU since 20013. His work has focused on empirical methods particularly related to the relationship between training and innovation. He participated in the MORE-1 as well as three national and Nordic studies in the area of research mobility.		

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
A conceptual framework for studying science research careers	Journal of Technology Transfer (under acceptance)	2018	C. Cañibano, R.Woolley, E. Iversen, S. Hinze, S. Hornbostel and J. Tesch
Patent, trademarks and the emergence of 'New' Firms: a labor-tracking approach to analyze IPR use among spinoffs and other new firms in Norway	Proceedings 14th Conference on Comparative Analysis of Enterprise Data, Seoul, South Korea: 22-24 September 2017	2017	Iversen E, T Sandven
International mobility: Findings from a survey of researchers in the EU	Science and Public Policy	2015	Børing, P.; Flanagan, K.; Gagliardi, D.; Kaloudis, A.; Karakasidou, A.

Relevant previous projects or activities			
Project name	RISIS - Research infrastructure for research and innovation policy studies (EC-GA N°313082) http://risis.eu/	**Dates (start/end)**	2014-2018
Project name	Internationalization of business investments in R&D and analysis of their economic impact (BERD Flows): Final Analysis Report, (30-CE-0677869/00-21/A4/2014)	**Dates (start/end)**	2015-2017
Project name	Researcher Mobility and the Research Barometer, Norwegian Ministry of Education & Research	**Dates (start/end)**	2013-2014
Project name	Nordic Crossing: Mobility of Researchers and Knowledge Transfer in the Nordic Region: Patterns, Framework Conditions, Incentives, and Trends. NordForsk	**Dates (start/end)**	2012-2013
Project name	Study on international mobility and researchers' career development (SIM ReC), (ERAWATCH NETWORK ASBL) Framework Service Contract Nr-151364-X13	**Dates (start/end)**	2011-2012

Relevant infrastructure
As a producer of national R&D statistics and an institute that has worked with research mobility since the 1990s, NIFU is a primary actor in the production and presentation of national statistics

that are relevant to science and innovation policy. NIFU is authorized as a national statistical contact point, with established and recognized best-practice for data-safety routines.	
Name	**Description**
Flipd OPEN	Open firm-linked patent-data for indicator development: Norwegian patent applicants 1995-2014
Name	**Description**
DIAN	Data on Innovation Analysis at NIFU), includes: • Firm-linked innovation data (CIS/R&D) and IPR data for Norway • Employee-Employer data for Norway

Contribution to the project
NIFU confirms that it is prepared and qualified to contribute principally to the following empirical areas: • IPRs with a focus on Trademarks • Careers/HR

4.1.15. Partner 15: Consejo Superior de Investigaciones Científicas (CSIC)

Partner Profile
INGENIO (CSIC-UPV) is a joint research centre of the Spanish Council for Scientific Research (CSIC) and the Universitat Politècnica de València (UPV), with more than 30 researchers from different professional fields, academic backgrounds and nationalities. INGENIO is internationally known for its research in the economics of innovation, research evaluation and science policy. It is building an increasingly strong reputation in the areas of science and research policy, social innovation and responsible research and innovation (RRI). INGENIO researchers publish in the leading journals, and present their work at the top international conferences, in their respective fields. As a joint CSIC-UPV centre, INGENIO provides a broad range of possibilities to develop research and teaching activities, train postgraduate students, and regularly hosts international researchers and PhD students. The Spanish National Research Council (CSIC) is the largest public research institution in Spain and a top ten institution in terms of the amount of funding received from H2020. CSIC has 10,940 employees, with 3,764 researcher distributed across 70 CSIC institutes and 53 joint institutes. These institutes collaborate frequently internationally, including having participated in 723 FP7 (97 as coordinator) and 110 H2020 actions. Founded in 1971, the Universitat Politècnica de València (UPV) is a public higher education institution that offers modern, flexible degrees designed to meet the demands of contemporary society. UPV has 30,000 students, over 2,500 academics, 17 university research Centres of Excellence. UPV includes the Polytechnic City of Innovation, a multidisciplinary science park where INGENIO is located. UPV has participated in over 100 FP7 projects and 40 H2020 projects, and participates in other major European partnering initiatives (JTIs, PPPs, KICs).

Staff profile			
Name	Jordi MOLAS-GALLART	**Gender**	M
Short CV / description of work experience relevant to the project	Prof. Jordi Molas-Gallart is Research Professor at the Spanish Council for Scientific Research (CSIC) and Director of INGENIO (CSIC-UPV) He was trained as an economist at Universitat Autónoma de Barelona, and obtained an MA in International Relations (The Johns Hopkins University) and a DPhil from the Science Policy Research Unit (University of Sussex). He has led and managed research assignments and policy evaluations for a variety of clients, including the European Commission, the European Parliament, and the UK Economic and Social Research Council. He is President of the European Network of Indicator Designers (ENID), and editor of "Research Evaluation" (journal published by Oxford University Press). Between 2014 and 2016 he served as Chair of the Science Europe Working Group on Research Policy and Programme Evaluation.		
Name	Carolina CAÑIBANO	**Gender**	F
Short CV / description of work experience relevant to the project	Dr. Carolina Cañibano is Research fellow at INGENIO. She was trained as an economist at the Universidad Autónoma of Madrid. She has been a visiting researcher at the OECD (1999), at the University of Sussex (2003), the Georgia Institute of Technology (2006) and the University of Western Sydney (2009). She has participated in various national and international research projects. She has recently lead the Spanish funded project (Prest-Ence Spain) exploring the interconnections between the construction of academic excellence and prestige and contributed to the European project RISIS (Research Infrastructure for Science and Innovation Studies). Carolina's research focuses on the theoretical and empirical investigation of research careers, human capital and economic change from an evolutionary perspective. She has made theoretical, methodological and empirical contributions to the socio-economics of science by addressing research questions regarding the dynamics of geographical mobility and career development of researchers and the knowledge diffusion processes that derive from mobility. She has also contributed to addressing the role of human inten-		

	tionality in the shaping of economic evolution. At present, her theoretical efforts are targeted at providing the micro-foundations for an evolutionary theory of human capital.		
Name	Richard WOOLLEY	**Gender**	M
Short CV / description of work experience relevant to the project	Dr. Richard Woolley has a PhD in sociology and has been research fellow at INGENIO since 2011. Previously he had ten years' experience working on science and innovation policy and projects in Australia and SE Asia. In recent years he has worked primarily on a European projects including: focused on the organization of diagnosis, care and treatment for Rare Disease patients in Europe, conducting research in 11 Member States (EUCERD Joint Action); the development of a conceptual framework and data infrastructure for understanding and monitoring research careers (RISIS project); and on the development of indicators of societal benefit associated with responsible research and innovation (RRI) (MoRRI project). Richard is currently working on new approaches to understanding the societal impact or scientific research (OSIRIS project).		
Name	François PERRUCHAS	**Gender**	M
Short CV / description of work experience relevant to the project	François Perruchas is an engineer in industrial systems management from the University of Angers (France) and holds a Master degree in innovation and technology management from the UTC (Université de Technologie de Compiègne, France). From 2004 to 2006 he worked for the Scientific and Technology Service of the French Embassy in Spain, in charge of the technological monitoring in information technologies in Spain. Since 2006 he has been working in INGENIO (CSIC-UPV) in charge of the institute's scientific databases for research and as a lead software engineer for the information management system. He is also an assistant researcher and he is collaborating in scientific research involving bibliometric studies about patents and scientific publications. His role is to create and maintain bibliographic databases containing the material for the studies, and to develop data processing programs to achieve semi-automatic treatments of the information. He participates in several national and European research projects. Since 2016 he is also a PhD student at the Universitat Politècnica de València. The aim of his thesis is to conduct an empirical analysis of green innovation related to technology, skills and policy.		

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
Research evaluation and the assessment of public value	Arts & Humanities in Higher Education 14(1): 111-126	2015	Molas-Gallart, J.
Tracing 'productive interactions' to identify social impacts: An example from the social sciences	Research Evaluation 20(3): 219-226	2010	Molas-Gallart, J., Tang, P.
A conceptual framework for studying science research careers	Journal of Technology Transfer (under acceptance)	2018	C. Cañibano, R.Woolley, E. Iversen, S. Hinze, S. Hornbostel and J. Tesch
A functional review of literature on research careers	INGENIO Working Paper series, WP 2016-05	2016	R. Woolley, C. Cañibano and J. Tesch
Evolving cohesion metrics of a research network on rare diseases: a longitudinal study over 14 years	Scientometrics, 108(1), 41-56	2016	Amat, C. B., Perruchas, F.

Relevant previous projects or activities			
Project name	RISIS - Research infrastructure for research and innovation policy studies (EC-GA N°313082) http://risis.eu/	**Dates (start/end)**	2014-2018
Project name	MoRRI – Monitoring the Evolution and Benefits of Responsible Research and Innovation (EC-DG RTD)	**Dates (start/end)**	2014-2017
Short description	This project aims at providing scientific evidence, data, analysis and policy intelligence to support DG Research & Innovation's research funding activities and policy-making activities in relation to RRI. INGENIO is a task leader, specifically responsible for the development of indicators of RRI benefits.		
Project name	SIAMPI – Social Impact Assessment Methods for research and funding instruments through the study of Productive Interactions between science and society (EC-FP7)	**Dates (start/end)**	2009-2011
Short description	Our project aims at developing a method for the assessment of the social impact of scientific research. We review current literature including experimental studies and conduct case studies in four different fields (nanotechnology, ICT, health, social sciences) with various grades of social impact, in both national and supranational settings. Our goal is both to enhance insight in social impact assessment and to develop assessment methods.		
Project name	MapRePort – Mapping and assessment of research portfolios (EC-FP7-MSCA)	**Dates (start/end)**	2013-2017
Short description	This project aims to develop an analytical framework for the mapping and assessment of research portfolios. The notion of "research portfolio" is increasingly used by funding agencies. However, there is no clear understanding and very little research on how to define, map and assess a research portfolio. Here we aim to develop a multi-method approach to investigate a research portfolio of a given disease.		
Project name	OSIRIS – Oslo Institute for Research on the Impact of Science (The Research Council of Norway)	**Dates (start/end)**	2016-2023
Short description	The overall objective of the OSIRIS Centre is to study how and under what circumstances research produces effects in society at large – in a way that generates new insights and helps policymakers, users and research organisations to better contribute to generating impact.		

Relevant infrastructure	
Name	**Description**
INGENIO	INGENIO (CSIC- UPV) is a joint research institute with its own budget, human resources and research premises, including dedicated IT team and research management expertise. INGENIO's premises and resources are wholly independent from the partners in the consortium. Researchers have their own workspace, state of the art IT facilities and access required to conduct their work. INGENIO has also its own library devoted to innovation studies, its own databases and software licenses. It is located in the main UPV campus, so researchers also have access to all facilities provided by the University: libraries, databases, e-reviews, etc. Researchers also have access to a full range of services including on campus medical service, technology transfer office and sport services of the University.

Contribution to the project
INGENIO [CSIC-UPV] team has a long track record in the collection, management (including development of software tools, geocoding, etc…), and analysis of various types of data used in the study and evaluation of science and innovation policies. Of relevance to this project INGENIO has for instance developed approaches to the analysis of CV data, and designed tools for the collection of information on academic careers. INGENIO will contribute to the consortium, providing technical development and solutions, and designing data collection and analysis strategies. Specifically, INGENIO will: • Contribute to the design of the RISIS Core Facility and integration of datasets and services • Support the development and integration of tools for advanced geographical analysis • Contribute to the design and development of the PhD holders career database • Contribute to the augmentation of the non-academic career trajectories through a range of methods including CV analysis and surveys.

4.1.16. Partner 16: The Samuel Neaman Institute (SNI)

Partner Profile
The Samuel Neaman Institute was established in 1978 in the Technion at Mr. Samuel Neaman's initiative. It is an independent multi-disciplinary national policy research institute. The activity of the institute is focused on issues in science and technology, education, economy and industry, physical infrastructure and social development which determine Israel's national resilience. National policy research and surveys are executed at the Samuel Neaman Institute and their conclusions and recommendations serve the decision makers at various levels. The policy research is conducted by the faculty and staff of the Technion and scientists from other institutions in Israel and abroad and specialist from the industry. The research team is chosen according to their professional qualifications and life achievements. In many cases the research is conducted by cooperation with governmental offices and in some cases at the initiative of the Samuel Neaman institute and without direct participation of governmental offices. The Samuel Neaman Institute has performed hundreds of exploratory national policy research projects and surveys that serve decision makers and professionals in economy and government. In particular the institute plays an important leading role in outlining Israel's national policies in science, technology and higher education. The Institute supports national projects, such as the Ministry of Industry, Trade & Labor clusters - the MAGNET program in nano-technologies, media, optics and communication, chemistry, energy, environmental and social projects of national importance. The institute organizes also comprehensive seminars in its leading fields of research. SNI's vision is to promote informed national decisions in Israel through research and analysis of well-established information. SNI's mission is be a leading research institute that identifies, formulates and analyzes matters of policy of national importance in the areas of scientific–technological development, economics and social issues in Israel. The Institute's activity is expected to encourage educated public debate, promote and assist the decision-making process of the State of Israel and bring to the final adoption of its recommendations. The Institute focuses primarily on formulating national policies in the fields of science and technology, industry, schooling and higher education, social integration, infrastructure, environment and energy, and other issues of national importance, where the Institute can provide valuable and unique contribution. Most relevant to this project, SNI has been involved in various projects concerning career evaluation in Israel. These include: a research within an EU NoE – PRIME on: Euro-CV, using researchers' CV to study mobility, career patterns and productivity; and tracking the academic and professional careers of Technion Graduates and former Technion Excellence Programs students.

Staff profile			
Name	Daphne GETZ	**Gender**	F
Short CV / description of work experience relevant to the project	Daphne Getz is the head of the CESTIP (Center of Excellence in Science, Technology and Innovation Policies), and has been a senior research fellow at the Samuel Neaman Institute (SNI) since 1996. Dr. Getz is also the director of SNI's information center. Dr. Getz is a specialist in R&D policy, technology and innovation, human resources in science& technology, policies on new and emerging technologies, and relationships between academia, industry and government, among others. She has represented the academia and the Technion (Israel Institute of Technology) in the MAGNET R&D Consortia in many fields (for example: ICT, semiconductors) and represents Israeli academia in several EU and UN projects. She has a D.Sc. from the Technion in Physical Chemistry and has served in several positions related to R&D management in the industry. Over the years, Dr. Getz has initiated numerous projects, including Israeli indicators for Science, Technology and Innovation, evaluation of R&D programs, the evaluation of Israeli R&D outputs using Bibliometric data (publications & patents) and several studies on STI human resources.		
Name	Eran LECK	**Gender**	M

Short CV / description of work experience relevant to the project	Eran Leck is a researcher at the Samuel Neaman Institute for National Policy Research (SNI). Eran holds a Ph.D. and MSc. degrees from the Technion in Urban and Regional Planning and a B.A. degree in Statistics and Geography from the University of Haifa. As a PI, He has led a large group of researchers from SNI in a study focusing on Israel's Science, Engineering, Technology and Innovation (SETI) policy. The study was invited by the Israel Academy of Sciences and Humanities and was published in UNESCO's Country Profiles on Science and Technology. Dr. Leck has been a member in several large scale European projects (FP-7) focusing on innovation policy and innovation statistics (RISIS and PICK-ME). He has considerable experience in innovation statistics, patent statistics and innovation policy evaluation		
Name	Tsipy BUCHNIK	**Gender**	M
Short CV / description of work experience relevant to the project	Tsipy Buchnik, is a researcher at the Samuel Neaman Institute for National Policy Research (SNI) and a PhD student in the Urban and Regional Planning program at the Faculty of Architecture and Town planning, Technion, Israel. Tsipy Buchnik, bachelor in economics and Msc in QA&R. also knowledgeable in computer programming. Tsipy's fields of specialization in SNI are: science, technology and innovation - indicators (ISTI), and Human resources at Science & Technology. Tsipy's knowledge in and experience with statistics, economy, survey, databases and more, are used in her line of work.		
Name	Rinat KLEIN	**Gender**	M
Short CV / description of work experience relevant to the project	Rinat Klein is a researcher at the Samuel Neaman Institute for National Policy Research (SNI). Rinat holds a M.Sc. and B.Sc. degrees from the Technion in Chemical Engineering. As a researcher, Rinat takes part in the researches focusing on Israel's Science, Engineering, Technology and Innovation policy such as RISIS, Globalisation of higher education, Artificial Intelligence and Smart robotics etc. Rinat is a certificated information specialist and takes part in an Israeli MAGNET R&D Consortia data centre.		

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
Predictions for Human Resources in Science and Technology: Models and Indicators	SNI publications	2013	D. Getz, T. Buchnik, B. Zalmanovich, N. Zemer-Batsir
An Innovative Approach for Measuring the Digital Divide in Israel: Digital Trace Data as Means for Formulating Policy Guidelines	SNI publications	2018	Rafaeli, S. Leck, E. and Albo Y.
Evaluation of the Technion Excellence Programs	SNI publications	2017	Leck E, Gilad V, Nathan O., Zalmanovich B, Getz D.
Technion's Contribution to the Israeli Economy through its Graduates	SNI publications	2012	Frenkel, A. and Maital, Shlomo
Researcher's CV Collection in Israel". In EURO-CV: Building New Indicators for Researcher's Careers and Mobility Based on Electronic Curriculum Vitae	PRIME Report	2008	Getz, D. Zalmanovich, B. and Buchnik, T.

Relevant previous projects or activities			
Project name	RISIS - Research infrastructure for research and innovation policy studies (EC-GA N°313082) http://risis.eu/	**Dates (start/end)**	2014-2018
Project name	PICK-ME - Policy Incentives for the Creation of Knowledge:	**Dates**	2011-2014

	Methods and Evidence (Grant number 266959) https://cordis.europa.eu/project/rcn/97563_en.html http://pick-me.carloalberto.org/	**(start/end)**	
Short description	The project Policy Incentives for the Creation of Knowledge: Methods and Evidence (PICK-ME) was aimed at assessing the impact of market demand and public procurement on the patterns of knowledge creation across different regions and sectors. The PICK-ME project relied on different methods of analysis so as to ascertain the impact of demand on: i) the rate and direction of technological change; ii) the evolutionary dynamics of knowledge intensive sectors; iii) mobility + and skill matching in local contexts; iv) sectoral patterns of innovation. The empirical investigation has focused on different units of analysis, ranging from micro-level studies focusing on firms to regional and country level analyses.		
Project name	PREF - Analysis of national public research funding, by theme and by type of allocation (JRC Contract N° 154321)	**Dates (start/end)**	2015-2016
Short description	The aim of the study on national public research funding (PREF) is to collect information and provide an analysis of national public research funding, by theme and by allocation mode (project-based funding versus institutional funding) including an overview of the evolution and current state of public research funding in European and selected non-European Countries. SNI's role was to provide data on Israel and Turkey.		
Project name	The study was commissioned by the Israel Academy of Sciences and was published in UNESCO's series of studies on Science, Engineering, Technology, and Innovation (SETI) policies during 2016.	**Dates (start/end)**	2015-2016
Short description	The aim of the project is to map the SETI system in Israel, to describe it using relevant metrics, and to examine the strengths and weaknesses of Israeli science and technology and innovation policy, while placing special emphasis on policy tools and best practices.		
Project name	UNESCO Science Report 2015, towards 2030 - Chapter on Science and Innovation in Israel.	**Dates (start/end)**	2015-2016
Short description	At UNESCO's request SNI has written a chapter about science, technology, and innovation (STI) in Israel. Experts from about 60 countries described in the report the trends and developments in science, technology, and innovation policy in their country during 2009-2015. The report provides vital information about the countries' priorities and the subjects on which they wish to focus and to continue to develop in the coming years.		
Project name	"Best Practices and Lessons Learned in ICT Sector Innovation: A Case Study of Israel"	**Dates (start/end)**	2015-2016
Short description	SNI was commissioned by the World Bank to prepare a chapter for their World Development Report 2016. The study analyzes and describes how government policies (national and local) have contributed to the development of a vibrant ecosystem that has spurred a high rate of both technological innovation and entrepreneurship, and describes the environment it has created.		

Relevant infrastructure	
Name	**Description**
International and datasets on research outputs and national	Subscription to PATSTAT, Subscription to SciVal (Elsevier), Foreign R&D Centers in Israel Dataset, Israeli Start-ups and VCs dataset, Subscription to IVC (Israeli High-Tech Capital Raising, Israeli High-Tech Exits, Most Active Venture Capital Funds in

datasets on the innovation of firms	Israel). Datasets in Use: WoS (Web of Science); Scopus; START-UP NATION CENTRAL (Start-Up Nation Finder™ is an online, free platform that collects & organizes comprehensive information on the Israeli innovation ecosystem. Software in Use: SPSS, OPINIO – a web-based survey tool which provides a framework for authoring and distributing surveys as well as a range of reporting facilities, Microsoft – SharePoint Platform.

Contribution to the project
Most relevant to this project, SNI has been involved in various projects concerning career evaluation in Israel. These include: a research within an EU NoE – PRIME on: Euro-CV, using researchers' CV to study mobility, career patterns and productivity; and tracking the academic and professional careers of Technion Graduates and former Technion Excellence Programs students.

4.1.17. Partner 17: Technology Centre of the Czech Academy of Sciences (TC CAS)

Partner Profile
Technology Centre of the Czech Academy of Sciences (TC CAS) was established in 1994. It is a private non-profit institution with around 70 employees. TC CAS carries out analytical and strategic studies for Czech state administration, particularly in the area of science, technology, and innovation. It is experienced in the use of various methodologies for development of studies and analyses for research, development and innovation support, such as microdata and macrodata analyses, field research and surveys, expert interviews, bibliometry analyses or technology assessment and foresight. TC CAS coordinates a national project "Science and Technology for Society" focused on improving the quality of strategic decision-making in research by systematic monitoring and assessment of technological and socio-economic trends, assessing potential impacts of new technologies on society, and evaluating the results and impacts of research policies. In this respect, it cooperates with governing bodies and funding agencies responsible for research and innovation policy on the national and regional levels. In addition, TC CAS acts as the National Information Centre for European Research and it provides technology transfer services being the leader of the Czech part of the Enterprise Europe Network. TC CAS is a contact entity for Czech industry in several renowned international organisations of which the Czech Republic is a member, e.g. Industrial Liaison Officer for the European Southern Observatory (ESO), European Spallation Source (ESS), European Organization for Nuclear Research (CERN), Joint Institute for Nuclear Research in Dubna (JINR). TC CAS also runs the Czech Liaison Office for Research, Development, and Innovation in Brussels.

Staff profile			
Name	Michal PAZOUR	Gender	M
Short CV / description of work experience relevant to the project	Ph.D. from the University of Economics in Prague, head of the Strategic Studies Department at the Technology Centre of the Czech Academy of Sciences. He specializes in research and innovation policy evaluation, analysis and design. He coordinates national projects and studies evaluating research and innovation policies and exploring the opportunities and measures to strengthen national innovation systems. He coordinated elaboration of core background documents for governmental reform of the Czech RDI system and for the Czech RDI Policy 2009 – 2015. He was also involved in innovation policy studies for the European Commission, the European Parliament and OECD. He was member of several expert groups, e.g. expert group of the European Commission on Strategic Foresight for Research an Innovation Policy in Horizon 2020, UNECE expert group in the project "Financing for Innovative Development" or expert group for methodology assessment of R&D and innovation statistics provided by the Czech Statistical Office.		
Name	Ondrej PECHA	Gender	M
Short CV / description of work experience relevant to the project	Ph.D. in kinanthropology from the Charles University in Prague. He has former experiences in biostatistics as a co-author of several medical articles. Ondrej joined Technology Centre ASCR in 2012. He works as a front-end developer of visualizations and bibliometric data analyst. More information about Ondrej is also available on his personal web page (http://www.onpecha.cz/).		
Name	Kristyna MEISLOVA	**Gender**	F
Short CV / description of work experience relevant to the pro-	MA in Social Geography and Regional Development at the Faculty of Science of the Charles University. Kristyna joined Technology Centre CAS in 2012. She works as an analyst for research and technology development in the Strategic Studies Department. She is interested in technology trends research, foresight and firm innovativeness. She		

ject	is also keen on developing innovative analytical methods, especially big data analysis and machine learning.

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
Visualization of ERC grants and proposals 2007-2016. Open-access on-line data visualization tool.	http://svizualizace.tc.cas.cz	2018	Pecha, O., Frank, D., Fedorová, P., Pazour, M.
Visualization of cooperation among Czech research organizations 2010-2014	http://svizualizace.tc.cas.cz	2017	Pecha, O., Vondrák, T., Kučera, Z., Pazour, M.
Bibliometric Analysis of FP7 Publications Co-authored by Researchers from seven EU Countries	Echo, 2017(3), 15-30	2017	Vaněček, J., Pecha, O., Albrecht, V.
Patent activities of the Czech enterprises – an international comparison	Ergo, 12(1), 3-17	2017	Kučera, Z., Vondrák, T., Pecha, O.
Analysis of the Czech and Hungarian physiology publications 1994-2011	Scientometrics, 105(2), 991-1003	2015	Pecha, O., Vaněček, J.

Relevant previous projects or activities			
Project name	VATES - Science and Technology for Society (Czech Ministry of Education, Youth and Sports) https://www.strast.cz/en/projects/projects-list/vates	**Dates (start/end)**	2016-2020
Short description	This project aimed at systematically developing quantitative and qualitative methods for analysing structured and unstructured data and establishing a system for continuous monitoring and evaluation of technology and socio-economic trends using quantitative and qualitative methods of social and technology foresight, in order to improve the quality of strategic decision-making in research and innovation on European, national, regional, and institutional levels.		
Project name	CIMULACT - Citizen and Multi-Actor Consultation on Horizon 2020 (EC-GA-H2020) http://www.cimulact.eu/	**Dates (start/end)**	2015-2018
Short description	This project aimed at active including and engaging citizens and scientific stakeholders in the outlook and debate on relevant STI issues, increasing scientific literacy and shared understanding between scientific stakeholders, policy-makers and citizens, and producing policy options for Horizon2020 based on societal needs and demands.		
Project name	PACITA - Parliaments and Civil Society in Technology Assessment (EC-FP7) http://www.pacitaproject.eu/	**Dates (start/end)**	2011-2015
Short description	This project aimed at increasing European wide capacity and enhancing the institutional foundation for knowledge based policy-making on issues involving science, technology and innovation using a range of methods such as cross-disciplinary expert studies, stakeholder involvement, citizen consultation and parliamentary discourse.		
Project name	Observatory NANO - European Observatory on Nanotech-	**Dates**	2008-2012

	nologies (EC-FP7)	**(start/end)**	
Short description	This project aimed at creating European observatories for nanotechnologies that would provide comprehensive scientific and economic analyses across various technology sectors.		
Project name	iKnow - Interconnecting knowledge for the early identification of issues, events and developments ((EC-FP7) http://community.iknowfutures.eu/	**Dates (start/end)**	2008-2011
Short description	This project aimed at creating a conceptual and methodical framework to identify and analyse the so-called wildcards and weak signals, and to apply this framework while evaluating the impact of identified wildcards and weak signals on the future of the ERA.		

Relevant infrastructure
Extended database of the national RDI Information System (running on PostgreSQL) Acquired databases: Web of Science, InCites, PATSTAT, MagnusWeb (financial data of all Czech companies), Database of European technology centres active in Key Enabling Technologies Used software: SPSS, Statistica, Stata, GIS, R, VOSviewer, Pajek, Gephi, NETDRAW, Inkscape Used JavaScript libraries: D3.js, jQuery, Angular, Crossfilter.js, Queue

Contribution to the project
TC has experiences with advanced visualization tools for S&T policy that will be used in the project. As an example, additional enhancements of the existing visualization tools used in Cortext by using D3.js will be introduced. TC will contribute to the development of a novel conceptual framework for the identification of projects responding to social challenges and/or having the potential to introduce social innovation and to design a content-based filtering system for datasets on social innovation related projects. The developed framework and methodology will enable to carry out analyses of these projects and contribute to better planning of new instruments focused on social challenges. This approach will be verified on selected social challenges. For these purposes datasets developed in the RISIS I and RISIS II projects will be complemented by new data sources. TC will also contribute to the collection of novel data on funding (PREF+) and with proposing data collection approaches for R&I projects at national level (NATPRO), in particular as related to the Central and Eastern European Countries. TC will also be involved in drafting policy briefs using the EUPRO and NATPRO datasets.

4.1.18. Partner 18: Athena Research and Innovation Center (ATHENA RC)

Partner Profile
Athena RC conducts research in Information, Communication, and Computational Science and Technology, including all areas of data science (such as open data, big data, data analytics, data management, large-scale information systems, and cloud technologies), automation, robotics, signal processing, artificial intelligence, networking and digital communication, modeling, and simulation. Research and development is carried out at the level of both information technology itself and specific applications. Computational sciences form a strong component of the Athena RC activities, including but not limited to computational linguistics, archaeology, engineering, medicine, biology, and mechanics. Athena RC operates in three cities (Athens, Patras and Xanthi) and implements its research and innovation strategy through its research institutes and units: • Institute for Language and Speech Processing (ILSP) – http://www.ilsp.gr/ • Industrial Systems Institute (ISI) – http://www.isi.gr/ • Institute for the Management of Information Systems (IMIS) – http://www.imis.athena-innovation.gr/ • Technology Clusters Initiative Unit (Corallia) – http://www.corallia.org/ • Space Programmes Unit (SPU) – http://www.athena-spu.gr/ • Robot Perception and Interaction Unit (RPI) – http://www.athenarc.gr/en/institutesunits/87.html • Environmental and Networking Technologies and Applications Unit (ENTA) Athena plays a major role in the Greek & European data infrastructure area with extensive expertise regarding the design, development and operation of data infrastructures in a broad range of subject areas (HELIX, the Greek Research Infrastructure, CLARIN-EL, DYAS-DARIAH, META-SHARE, ELIXIR-EL, INSTRUCT-EL). On the national level it operates GeoData, a Greek Open Data Government portal, and is an Excellence Center for Open Source Geospatial Information Software (GIS). Athena's team is in the leadership of OpenAIRE (www.openaire.eu), and coordinates/participates in related H2020 projects: OpenMinTeD (open e-Infrastructure for text and data mining), FutureTDM (policy and legal frameworks for EU text and data mining), OpenUP (review-disseminate-assess processes for Open Science), Data4Impact (research assessment indicators for health area), and an active participant in EC tenders for evaluation of FP7 and H2020.

Staff profile			
Name	Yannis IOANNIDIS	**Gender**	M
Short CV / description of work experience relevant to the project	Yannis Ioannidis (male) is the President and General Director of the "Athena" Research and Innovation Center as well as a Professor at the Department of Informatics and Telecommunications of the University of Athens. He holds a Diploma in Electrical Engineering from the National Technical University of Athens, an MSc in Applied Mathematics from Harvard University, and a Ph.D. degree in Computer Science from the University of California-Berkeley. His research interests include database and information systems, data infrastructures & digital repositories, scalable data processing, data and text analytics, and personalization and social networks, topics on which he has published over 150 articles in leading journals and conferences. His work is often inspired by and applied to data management problems arising in other fields, such as life sciences, physical sciences, biodiversity, and cultural heritage. He has been a (co-)coordinator of all the OpenAIRE projects and a partner in many other research projects, including the currently active OpenMinTeD, OpenUP, BlueBRIDGE, and the Human Brain Project. Yannis is an ACM and IEEE Fellow, a member of Academia Europaea, and a recipient of several research and teaching awards. He is the Greek delegate to the European Strategy Forum on Research Infrastructures (ESFRI), a member of the ESFRI Executive Board, and the ESFRI representative to the e-Infrastructures Reflection Group (e-IRG).		
Name	Natalia MANOLA	**Gender**	F

Short CV / description of work experience relevant to the project	Natalia Manola (female) is a research associate in Athena RC and the University of Athens, Department of Informatics & Telecommunications. She holds a Physics degree from the University of Athens, and an MS in Electrical and Computing Engineering from the University of Wisconsin at Madison, USA. She has several years of employment as a Software Engineer and Architect employed in the Bioinformatics sector. From 2009 she has served as the managing director of OpenAIRE, a key EC e-Infrastructure supporting open access and open science in scholarly communication. She has technically coordinated ESPAS, a data infrastructure on near earth science, and is currently coordinating OpenMinTeD a H2020 e-Infrastructure project on Text and Data Mining for scientific literature, and she is participating in OpenUP, a H2020 project that investigates the review-dissemination-impact aspects of research and Data4Impact a H2020 project for assessing new research indicators for health area. She has served on the EC's Future Emerging Technologies (FET) advisory group (2013-2017) and serves on EC's Open Science Policy Platform. Her research interests include the topics of e-Infrastructures management, scientific data management, data curation and validation, complex data visualization, open science policies.		
Name	Prodromos TSIAVOS	**Gender**	M
Short CV / description of work experience relevant to the project	Prodromos Tsiavos is a research affiliate in Athena RC, the Head of Digital Development at the Onassis Cultural Centre and a Senior Research Fellow at The Media Institute at University College London (UCL)/ BBC. He read law and Information Systems in Athens and London and holds a PhD in Law and Information Systems from the LSE. He has over 100 publications and talks on legal and business aspects of open data, Free/ Open Source Software, open hardware and open innovation/ fabrication. Prodromos has extensive experience (since 2002) in the area of IPR and LRs, personal data protection and big data regulation. More specifically, he has participated in the EnCoRe Project with LSE and Oxford University on issues of personal and sensitive data protection and the management and revocation of consent. With Oslo University on issues of Big Data regulation theory. With UCL on issues of IPR clearance on big data, with emphasis on audiovisual material meta-data. With Athena on issues of Language Resource and IPR (copyright and sui-generis right) management (METASHARE, CLARIN (EU and EL), CEF.AT projects). Prodromos has worked as an adviser in a number of public sector bodies as well as companies in the cultural and creative industry.		

Relevant publications, and/or products, services			
Title of publication	**Journal**	**Year**	**Author**
Data curation in the openaire scholarly communication infrastructure	Information Standards Quarterly, 25(3):13{19, Fall 2013	2013	J. Schirrwagen, P. Manghi, N. Manola, L. Bolikowski, N. Rettberg, B. Schmidt
OpenAIREplus: the European Scholarly Communication Data Infrastructure	D-Lib Magazine 18(9/10)	2012	P. Manghi, L. Bolikowski, N. Manola, J. Schirrwagen, T. Smith
Content Visualization of Scientific Corpora Using an Extensible Relational Database Implementation	TPDL Workshop 2013: 101-112	2013	T. Giannakopoulos, El. Stamatogiannakis, I. Foufoulas, H. Dimitropoulos, N. Manola, Y. E. Ioannidis
Supervised Content Visualization of Scientific Publications: A Case Study on the ArXiv Dataset	IIS 2013: 206-211	2013	T. Giannakopoulos, H. Dimitropoulos, Om. Metaxas, N. Manola, Y. E. Ioannidis

Relevant previous projects or activities			
Project name	OpenAIRE2020 - Open Access/Open Knowledge e-Infrastructure (EC-GA N° 643410) ATHENA RC Coordinator https://www.openaire.eu	**Dates (start/end)**	2015-2019
Short description	50 partners, from all EU countries, and beyond, collaborate to work on this large-scale initiative that aims to promote open scholarship and substantially improve the discoverability and reusability of research publications and data. The initiative brings together professionals from research libraries, open scholarship organisations, national e-Infrastructure and data experts, IT and legal researchers, showcasing the truly collaborative nature of this pan-European endeavor. A network of people, represented by the National Open Access Desks (NOADs), will organise activities to collect H2020 project outputs, and support research data management. Backing this vast outreach, is the OpenAIRE platform, the technical infrastructure that is vital for pulling together and interconnecting the large-scale collections of research outputs across Europe. The project will create workflows and services on top of this valuable repository content, which will enable an interoperable network of repositories (via the adoption of common guidelines), and easy upload into an all-purpose repository (via Zenodo).		
Project name	OpenMinTeD – Text and data mining e-Infrastructure for scientific literature and related data (EC-GA N° 654021) ATHENA RC Coordinator openminted.eu/	**Dates (start/end)**	2015-2018
Short description	OpenMinTeD aspires to enable the creation of an infrastructure that fosters and facilitates the use of text mining technologies in the scientific publications world, builds on existing text mining tools and platforms, and renders them discoverable and interoperablethrough appropriate registriesand a standards-based interoperability layer, respectively. It supports training of text mining users and developers alike and demonstrates the merits of the approach through several use cases identified by scholars and experts from different scientific areas, ranging from generic scholarly communication to literaturerelated tolife sciences, food and agriculture, and social sciences and humanities.		
Project name	Data4Impact – Research assessment indicators for health related science (EC-GA N° 770531) http://www.data4impact.eu/	**Dates (start/end)**	2017-2019
Short description	Recent technological developments in data mining, data treatment and data analysis have been rapid and far reaching, offering new dimensions and opportunities for performance analytics in various domains. The introduction of new technologies and initiatives, including open access mechanisms and social media/online media, has been generating increasing volumes of new data on the research domain. Data4Impact aims to capitalise on these developments and utilise big data approaches to improve the monitoring of research and innovation performance and assessment of the societal impact in the Health, Demographic Change and Wellbeing Societal Challenge.		
Project name	OpenUP – OPENing UP new methods, indicators and tools for peer review, dissemination of research results, and impact measurement (EC-GA N°710722) openup-h2020.eu/	**Dates (start/end)**	2016-2018
Short description	Through analysis, consultation, hands-on engagement with researchers, publishers, institutions and funders, industry and citizens, OpenUP will a) define a framework that defines roles and processes, benefits and opportunities, b) validate the proposed mechanisms		

	through a series of pilots involving researchers from four scientific communities (Life Sciences, Social Sciences, Arts & Humanities, Energy), and c) come up with practical policy recommendations and guidelines to be used by EU, national and institutional policymakers at different settings. OpenUP will engage with all stakeholders via a series of outreach and training events, and the creation of an Open Information Hub, a collaborative web based Knowledge Base that will host a catalogue of open tools/services, methodologies, best practices from various disciplines or settings, success stories, reports. This increased level of engagement and knowledge will feed into the development of research and innovation policies that aim to support and complement Open Science.

Relevant infrastructure	
Name	**Description**
HELIX	Greek National Digital Infrastructures for Research: Athena, together with GRNET, will develop and operate the Greek National Research Infrastructure.
META-SHARE	Open network of repositories for sharing and exchanging language data, tools and related web services, coordinated by ILSP Europe-wide.
CLARIN-EL	Greek node of the CLARIN infrastructure
DYAS-DARIAH	Greek research infrastructure network for humanities (Greek DARIAH node) ELIXIR: technology partner in ELIXIR–EXCELERATE for the Greek Elixir node
Excellence Center for Open Source Geospatial Information Software	Implement the "Services to develop and promoted Open Source Software", through training and dissemination actions, across ten thematic domains.

Contribution to the project
Athena RC is a co-lead of OpenAIRE (www.openaire.eu) and the role in RISIS-2 is to bring in and provide OpenAIRE data as an additional dataset and also to build a Research Initiative Dashboard to act as the basis for measuring the impact of RISIS-2 infrastructure (**WP6**)

4.2. Third parties involved in the project (including use of third party resources)

Does the participant plan to subcontract certain tasks (please note that core tasks of the action should not be sub-contracted)	Y
• **UPEM may hire subcontractors for the following expenses : Internal networking (RISIS Weeks, WP1 : 40.000€), Collective legal expertise (WP3: 34.312,50€), Access to private Databases linked to the hybrid model (WP1, 30.000€)** • **CNR may hire subcontractors for the following expenses : Communication and interactions activities (WP2) : Website (20.000€), Printed materials (15.000€), support to ENID conferences (30.000€) and other core conferences (35.000€), trainings (20.000€) and webinars/video-based activities (40.000€)** • **USI may hire subcontractors for support to external experts on specific methods (WP6, 18.000€)** **Each subcontract shall be selected either on the best value for money (considering the quality of the service proposed, i.e. the best price-quality ratio) or on the lowest price**	
Does the participant envisage that part of its work is performed by linked third parties	Y
LISIS Laboratory is a Joint Research Unit (JRU) managed by UPEM – Université Paris-Est Marne-la-Vallée (PIC 999896662) but gathering human and financial resources also from INRA (PIC 999993274), CNRS (PIC 999997930) and ESIEE Engineering School, owned by the Paris Chamber of Commerce (CCIP-IdF, PIC 950323357) and CNRS. The LISIS JRU is an administrative entity created by the signature of a contract between these higher education and research establishment. Thus, UPEM is the beneficiary and the partner of the project but may charge costs incurred by the third parties to INRA, CNRS and ESIEE in carrying out the project under the article 14 of the Grant Agreement.	
Does the participant envisage the use of contributions in kind provided by third parties (Articles 11 and 12 of the General Model Grant Agreement)	**N**

The beneficiaries shall base their subcontracts either on the best value for money (considering the quality of the service proposed, i.e. the best price-quality ratio) or on the lowest price

5. Ethics and Security

5.1. Ethics

In what follows we present the overall situation, address potential critical situations, define RISIS access procedures and the way they cover ethical aspects. This provides an initial analysis of the situation. As mentioned by the ethical summary report we reinforce the role of the ethics committee by requiring that an annual report is made, identifying potential issues and proposing ways to address them. This report will be presented to the Governing Board for decisions to be taken. The REC will monitor their implementation. The annual report will be annexed to project overall periodic reports and the periodic report will present the solutions developed to address issues raised.

5.1.1. Overall situation

Our data management concept acknowledges the EU General Data Protection Regulation (GDPR). In general, we refer to Article 5 (1, b) of the GDPR and claim scientific research purposes for the data treatment in this project.

In RISIS2, we will not process individual data in general. We do not conduct surveys and most of the processed data is at the level of organisations/companies. However, there will be some data sources, which contain individual data. These data are publicly available and therefore free of access to anyone (next to subscriptions or payment of fees). Among these data sources are patents with information on individual inventors and representatives (names, addresses), publications with names and institutional addresses of authors as well as trademark data with information on representatives (names, addresses). In addition, we also compile a data set of career path information of individual researchers. This data will be extracted from bibliometric data and web sources.

As already mentioned, most of the data treated and produced in the context of this project will not be on the individual level, but on the level of organisations. The existing and continued modules of the RISIS project are at the level of companies or public research organisations. The integration and matching of different data sources takes place on the organisational and/or company level. It is also the main level of data collection, of provision and of analysis. For some of the new data sets it will be necessary to process data on the level of individuals (researchers/inventors), but we will not publish individual data or data aggregated at the level of individuals at any time. In the cases where the data processing at the individual level is necessary, we will use clear names only for matching of data sources. A very limited number of people will have access to this data for the purpose of processing. Immediately after the matching of the data it will be anonymised (by using system-free IDs) and the clear names together with the IDs (key table / concordance) will be stored separately in a secret data file with additional access protection.

5.1.2. Sensitive cases in the project

We have a number of cases that might become sensitive in the project.

Publications, patents and trademarks

Patents, trademarks and scientific publication datasets contain information on individuals – authors, inventors and representatives. The data sources we use are publicly available and the authors/inventors/representatives have agreed to the publication of this information. However, the databases as such and in particular the possibility of developing individual performance measures make them a sensitive

subject to the EU General Data Protection Regulation (GDPR). As we have said, nearly all our treatments do not consider individuals as such, but their institutional affiliation and address to build aggregated information. We do not consider that the circulation of this information enters in the GDPR. It is only when we aggregate information from different sources that such issues may arise. In these cases, we address the rules and regulations of the GDPR first and foremost by anonymisation of the data that we store in RISIS2 databases accessible to external researchers. We will use system-free IDs for the data we exchange among partners and with our customers/users. However, at any time we will be able to provide information on the data of individuals upon request ("right to access") and we will also be able to delete individuals from the data set, if the particular person wishes to ("right to be forgotten", Article 17).

Academic patents

RISIS2 will develop a project for enriching university patenting data with patents taken by other organisations while including academic researchers. We employ a method to identify academic patents based on a matching of names of scientific authors (research-active university staff) from Scopus with inventor names from the PATSTAT database. This allows us to draw a more complete picture of the patent output of universities. The matching of patent and publication data will occur on the level of individual inventors/authors so that their affiliation – this information is missing in the patents, but available in publications – can be assigned to the patent. The information on the individual is just an intermediate step of the assignment of the affiliation and is of no direct interest to the project or its analytical aims. The individual level is just the way to link the data. After the matching the data will be aggregated at the level of organizations/universities and the information on the individual is of no further interest. However, it could be possible to calculate individual performance measures or even identify the names of the persons who invented patents that were not directly filed by the particular university. For this reason we will separate the clear names from the original data and store in an additionally protected (encrypted) file. We thereby guarantee that the data is anonymised, but that upon request of individuals we can ensure the 'right to access' as well as the 'right to be forgotten'.

Careers of PhD researchers

RISIS2 will develop an important project of understanding the role of the PhD in society, focusing on non-academic careers (estimated at 70% of total PhD) for which we know little about where they go and through which path, and even less on the role of the PhD in such trajectories. For this we combine different public sources (no surveys) on doctoral dissertations, on publications, and on publicly available CVs. That combination requires that data is fully anonymised after linkages have been made. The two operators of the project, DZHW and NIFU are already operating such datasets at a large scale and are specialists in insuring robust anonymisation. To follow GDPR rules, we will keep the clear names from the original data and store in an additionally protected (encrypted) file. We thereby guarantee that the data is anonymised, but that upon request of individuals we can ensure the 'right to access' as well as the 'right to be forgotten'.

5.1.3. Access procedures

The research infrastructure is dedicated to the access by researchers to the datasets developed by RISIS. This is done through a peer-reviewed process that enables to judge about the project, its objectives and the conditions of use of the data. A process of 'authorisation' is being developed to insure that researchers only access the elements of datasets they require for their project, limiting possibilities of inadequate use. We also require that researchers get clearance from the ethical committee of their institution. In case this committee does not exist, the RISIS committee (see below) will have to consider projects directly. Further-

more, researchers to access the data must sign the RISIS charter that incorporates ethical rules of use. Finally, in cases judged potentially problematic, access will be only physical (and not virtual) to enable the operator of the dataset to control the effective use made. This has already smoothly operated in RISIS1.

5.1.4. Overseeing of ethical procedures: the RISIS ethical Committee

RISIS will create a three persons ethical Committee (REC). It will be chaired by an 'independent individual with expertise in the ethical requirements of H2020'. This committee will have 3 functions: review the ethical procedures put in place by the different datasets, and in particular 'sensitive case' datasets; review access procedures to be sure that required anonymisation or access at relevant aggregate levels are insured; and consider individual projects before selection (when there is no ethical committee at the researcher's institution), during and after operation (when an operator considers that there may be ethical issues pending). The committee can be seized bottom-up by researchers from member organisations and by researchers having asked for access.

The Committee can be asked by the CMC to look for given situations, and it can ask the CMC to act and address given situations. It establishes an annual report that it presents to the Governing Board and that is annexed to the overall periodic reports of the project. The reports include the ways the problems have been addressed and solved.

5.2.Security

The activities developed in RISIS2 do NOT raise security issues and are NOT concerned by 'EU-classified information' as background or result

ESTIMATED BUDGET FOR THE ACTION

Associated with document Ref. Ares(2018)6225458 - 04/12/2018

	Estimated eligible[1] costs (per budget category)										EU contribution			Additional information		
	A. Direct personnel costs				B. Direct costs of subcontracting	*[C. Direct costs of fin. support]*	D. Other direct costs		E. Indirect costs[2]	F. Special unit costs	Total costs	Reimbursement rate %	Maximum EU contribution[3]	Maximum grant amount[4]	Information for indirect costs	Information for auditors	Other information:
	A.1 Employees (or equivalent) A.2 Natural persons under direct contract A.3 Seconded persons *[A.6 Personnel for providing access to research infrastructure]*		A.4 SME owners without salary A.5 Beneficiaries that are natural persons without salary				D.1 Travel D.2 Equipment D.3 Other goods and services *[D.4 Costs of large research infrastructure]*	D.5 Costs of internally invoiced goods and services		F.1 "Costs for providing trans-national access to research infrastructure" **					Estimated costs of in-kind contributions not used on premises	Declaration of costs under Point D.4	Estimated costs of beneficiaries/ linked third parties not receiving funding/ international partners
Form of costs[6]	Actual	Unit[7]	Unit[8]		Actual	Actual	Actual	Unit[9]	Flat-rate[10] 25%	Unit[12]							
	a	Total b	No hours	Total c	d	*[e]*	f	Total g	h = 0,25 x (a +b+c+f+g +*[i1]*[13]+*[i2]*[13]-n)	Total i1	j = a+b+c+d +*[e]*+f+g+h +*[i1]*+*[i2]*	k	l	m	n	Yes/No	
1. UPEM	480 000.00	0.00	0.00	0.00	104 312.50	0.00	136 000.00	0.00	154 000.00	0.00	874 312.50	100.00	874 312.50	874 312.50	0.00	No	n/a
- INRA	336 000.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	84 000.00	0.00	420 000.00	100.00	420 000.00	420 000.00	0.00	No	n/a
- CNRS	95 400.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	23 850.00	0.00	119 250.00	100.00	119 250.00	119 250.00	0.00	No	n/a
- CCIP	144 000.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	36 000.00	0.00	180 000.00	100.00	180 000.00	180 000.00	0.00	No	n/a
Total beneficiary	1 055 400.00	0.00			104 312.50	0.00	136 000.00	0.00	297 850.00	0.00	1 593 562.50		1 593 562.50	1 593 562.50	n/a	n/a	0.00
2. AIT	599 200.00	0.00	0.00	0.00	0.00	0.00	110 400.00	0.00	177 400.00	0.00	887 000.00	100.00	887 000.00	887 000.00	0.00	No	n/a
3. POLIMI	576 600.00	0.00	0.00	0.00	0.00	0.00	80 000.00	0.00	164 150.00	0.00	820 750.00	100.00	820 750.00	820 750.00	0.00	No	n/a
4. CNR	650 650.00	0.00	0.00	0.00	160 000.00	0.00	50 000.00	0.00	175 162.50	0.00	1 035 812.50	100.00	1 035 812.50	1 035 812.50	0.00	No	n/a
5. LEIDEN	489 600.00	0.00	0.00	0.00	0.00	0.00	48 000.00	0.00	134 400.00	0.00	672 000.00	100.00	672 000.00	672 000.00	0.00	No	n/a
6. USFD	396 000.00	0.00	0.00	0.00	0.00	0.00	35 000.00	0.00	107 750.00	0.00	538 750.00	100.00	538 750.00	538 750.00	0.00	No	n/a
7. FRAUNHOFER	0.00	433 800.00	0.00	0.00	0.00	0.00	52 000.00	0.00	121 450.00	0.00	607 250.00	100.00	607 250.00	607 250.00	0.00	No	n/a
8. USI	240 800.00	0.00	0.00	0.00	18 000.00	0.00	73 000.00	0.00	78 450.00	0.00	410 250.00	100.00	410 250.00	410 250.00	0.00	No	n/a
9. JOANNEUM	175 000.00	0.00	0.00	0.00	0.00	0.00	29 000.00	0.00	51 000.00	0.00	255 000.00	100.00	255 000.00	255 000.00	0.00	No	n/a
10. UOS	149 500.00	0.00	0.00	0.00	0.00	0.00	19 000.00	0.00	42 125.00	0.00	210 625.00	100.00	210 625.00	210 625.00	0.00	No	n/a
11. DZHW	122 000.00	0.00	0.00	0.00	0.00	0.00	16 000.00	0.00	34 500.00	0.00	172 500.00	100.00	172 500.00	172 500.00	0.00	No	n/a
12. STRATHCLYDE	178 200.00	0.00	0.00	0.00	0.00	0.00	16 000.00	0.00	48 550.00	0.00	242 750.00	100.00	242 750.00	242 750.00	0.00	No	n/a
13. SAPIENZA	126 500.00	0.00	0.00	0.00	0.00	0.00	25 000.00	0.00	37 875.00	0.00	189 375.00	100.00	189 375.00	189 375.00	0.00	No	n/a
14. NIFU	184 100.00	0.00	0.00	0.00	0.00	0.00	17 000.00	0.00	50 275.00	0.00	251 375.00	100.00	251 375.00	251 375.00	0.00	No	n/a
15. CSIC	159 500.00	0.00	0.00	0.00	0.00	0.00	23 000.00	0.00	45 625.00	0.00	228 125.00	100.00	228 125.00	228 125.00	0.00	No	n/a
16. SNI	84 000.00	0.00	0.00	0.00	0.00	0.00	20 000.00	0.00	26 000.00	0.00	130 000.00	100.00	130 000.00	130 000.00	0.00	No	n/a
17. TC CAS	70 000.00	0.00	0.00	0.00	0.00	0.00	15 000.00	0.00	21 250.00	0.00	106 250.00	100.00	106 250.00	106 250.00	0.00	No	n/a
18. ATHENA RC	99 000.00	0.00	0.00	0.00	0.00	0.00	14 000.00	0.00	28 250.00	0.00	141 250.00	100.00	141 250.00	141 250.00	0.00	No	n/a
Total consortium	5 356 050.00	433 800.00			282 312.50	0.00	778 400.00	0.00	1 642 062.50	0.00	8 492 625.00		8 492 625.00	8 492 625.00			0.00

[1] See Article 6 for the eligibility conditions.
[2] Indirect costs already covered by an operating grant (received under any EU or Euratom funding programme; see Article 6.5.(b)) are ineligible under the GA. Therefore, a beneficiary/linked third party that receives an operating grant during the action's duration cannot declare indirect costs for the year(s)/reporting period(s) covered by the operating grant, unless it can demonstrate that the operating grant does not cover any costs of the action (see Article 6.2.E).
[3] This is the theoretical amount of EU contribution that the system calculates automatically (by multiplying all the budgeted costs by the reimbursement rate). This theoretical amount is capped by the 'maximum grant amount' (that the Commission decided to grant for the action) (see Article 5.1).
[4] The 'maximum grant amount' is the maximum grant amount decided by the Commission. It normally corresponds to the requested grant, but may be lower.
[5] Depending on its type, this specific cost category will or will not cover indirect costs. Specific unit costs that include indirect costs are: costs for energy efficiency measures in buildings, access costs for providing trans-national access to research infrastructure and costs for clinical studies.
[6] See Article 5 for the forms of costs.
[7] Unit : hours worked on the action; costs per unit (hourly rate) : calculated according to the beneficiary's usual accounting practice.
[8] See Annex 2a 'Additional information on the estimated budget' for the details (costs per hour (hourly rate)).
[9] Unit and costs per unit : calculated according to the beneficiary's usual accounting practices.
[10] Flat rate : 25% of eligible direct costs, from which are excluded: direct costs of subcontracting, costs of in-kind contributions not used on premises, direct costs of financial support, and unit costs declared under budget category F if they include indirect costs (see Article 6.2.E).

ESTIMATED BUDGET FOR THE ACTION

[11] See Annex 2a 'Additional information on the estimated budget' for the details (units, costs per unit).
[12] See Annex 2a 'Additional information on the estimated budget' for the details (units, costs per unit, estimated number of units, etc).
[13] Only specific unit costs that do not include indirect costs.
[14] See Article 9 for beneficiaries not receiving funding.
[15] Only for linked third parties that receive funding.

ANNEX 2a

ADDITIONAL INFORMATION ON THE ESTIMATED BUDGET

- **Instructions and footnotes in blue will not appear in the text generated by the IT system (since they are internal instructions only).**
- **For options *[*in square brackets*]*: the applicable option will be chosen by the IT system. Options not chosen will automatically not appear.**
- **For fields in [grey in square brackets] (even if they are part of an option as specified in the previous item): IT system will enter the appropriate data.**

⚠ **Transitory period: Until SyGMa fully supports Annex 2a, you must prepare it manually (using this template by choosing and deleting the options/entering the appropriate data).**
For the 'unit cost tables': either fill them out manually or use currently existing tables from Annex 1 or the proposal.
The document can then be uploaded in SyGMa and attached to the grant agreement.

Unit cost for SME owners/natural beneficiaries without salary

1. Costs for a *[*SME owner*][*beneficiary that is a natural person*]* not receiving a salary

Units: hours worked on the action

Amount per unit ('hourly rate'): calculated according to the following formula:

{the monthly living allowance for researchers in MSCA-IF actions / 143 hours}
multiplied by
{country-specific correction coefficient of the country where the beneficiary is established}

The monthly living allowance and the country-specific correction coefficients are set out in the Work Programme (section 3 MSCA) in force at the time of the call:

- for calls *before* Work Programme 2018-2020:
  - for the monthly living allowance: **EUR 4 650**
  - for the country-specific correction coefficients: see Work Programme 2014-2015 and Work Programme 2016-2017 (available on the Participant Portal Reference Documents page)
- for calls *under* Work Programme 2018-2020:
  - for the monthly living allowance: **EUR 4 880**
  - for the country-specific correction coefficients: see Work Programme 2018-2020 (available on the Participant Portal Reference Documents page)

[additional OPTION for beneficiaries/linked third parties that have opted to use the unit cost (in the proposal/with an amendment): For the following beneficiaries/linked third parties, the amounts per unit (hourly rate) are fixed as follows:

- beneficiary/linked third party [short name]: EUR [insert amount]
- beneficiary/linked third party [short name]: EUR [insert amount]

[same for other beneficiaries/linked third parties, if necessary] *]*

Estimated number of units: see Annex 2

Energy efficiency measures unit cost

2. Costs for energy efficiency measures in buildings

Unit: m^2 of eligible 'conditioned' (i.e. built or refurbished) floor area

Amount per unit*: see (for each beneficiary/linked third party and BEST table) the 'unit cost table' attached

* Amount calculated as follows:
{EUR 0.1 x estimated total kWh saved per m^2 per year x 10}

Estimated number of units: see (for each beneficiary/linked third party and BEST table) the 'unit cost table' attached

Unit cost table (energy efficiency measures unit cost)[1]

Short name beneficiary/linked third party	BEST No	Amount per unit	Estimated No of units	Total unit cost (cost per unit x estimated no of units)

[1] Data from the 'building energy specification table (BEST)' that is part of the proposal and Annex 1.

Research infrastructure unit cost

3. Access costs for providing trans-national access to research infrastructure

Units[2]: see (for each access provider and installation) the 'unit cost table' attached

Amount per unit*: see (for each access provider and installation) the 'unit cost table' attached

* Amount calculated as follows:

$$\frac{\text{average annual total access cost to the installation (over past two years[3])}}{\text{average annual total quantity of access to the installation (over past two years[4])}}$$

Estimated number of units: see (for each access provider and installation) the 'unit cost table' attached

Unit cost table (access to research infrastructure unit cost)[5]

Short name access provider	Short name infrastructure	Installation		Unit of access	Amount per unit	Estimated No of units	Total unit cost (cost per unit x estimated no of units)
		No	Short name				

Clinical studies unit cost

4. Costs for clinical studies

Units: patients/subjects that participate in the clinical study

Amount per unit*: see (for each sequence (if any), clinical study and beneficiary/linked third party) the 'unit cost table' attached

* Amount calculated, for the cost components of each task, as follows:

For **personnel costs**:

For personnel costs of doctors: 'average hourly cost for doctors', i.e.:

{certified or auditable total personnel costs for doctors for year N-1

{1720 * number of full-time-equivalent for doctors for year N-1}

multiplied by

estimated number of hours to be worked by doctors for the task (per participant)**}**

For personnel costs of other medical personnel: 'average hourly cost for other medical personnel', i.e.:

{certified or auditable total personnel costs for other medical personnel for year N-1

{1720 * number of full-time-equivalent for other medical personnel for year N-1}

---

2 Unit of access (e.g. beam hours, weeks of access, sample analysis) fixed by the access provider in proposal.

3 In exceptional and duly justified cases, the Commission/Agency may agree to a different reference period.

4 In exceptional and duly justified cases, the Commission/Agency may agree to a different reference period.

5 Data from the 'table on estimated costs/quantity of access to be provided' that is part of the proposal and Annex 1.

multiplied by
estimated number of hours to be worked by other medical personnel for the task (per participant)**}**

For personnel costs of technical personnel: **'**average hourly cost for technical personnel', i.e.:

{certified or auditable total personnel costs for technical personnel for year N-1

{1720 * number of full-time-equivalent for technical personnel for year N-1}
multiplied by
estimated number of hours to be worked by technical personnel for the task (per participant)**}**

'total personnel costs' means actual salaries + actual social security contributions + actual taxes and other costs included in the remuneration, provided they arise from national law or the employment contract/equivalent appointing act

For **consumables**:

For each cost item: 'average price of the consumable', i.e.:

{{certified or auditable total costs of purchase of the consumable in year N-1

total number of items purchased in year N-1}
multiplied by
estimated number of items to be used for the task (per participant)**}**

'total costs of purchase of the consumable' means total value of the supply contracts (including related duties, taxes and charges such as non-deductible VAT) concluded by the beneficiary for the consumable delivered in year N-1, provided the contracts were awarded according to the principle of best value- for-money and without any conflict of interests

For **medical equipment**:

For each cost item: 'average cost of depreciation and directly related services per unit of use', i.e.:

{{ certified or auditable total depreciation costs in year N-1 + certified or auditable total costs of purchase of services in year N-1 for the category of equipment concerned}

total capacity in year N-1
multiplied by
estimated number of units of use of the equipment for the task (per participant)**}**

'total depreciation costs' means total depreciation allowances as recorded in the beneficiary's accounts of year N-1 for the category of equipment concerned, provided the equipment was purchased according to the principle of best value for money and without any conflict of interests + total costs of renting or leasing contracts (including related duties, taxes and charges such as non-deductible VAT) in year N-1 for the category of equipment concerned, provided they do not exceed the depreciation costs of similar equipment and do not include finance fees

For **services**:

For each cost item: 'average cost of the service per study participant', i.e.:

{certified or auditable total costs of purchase of the service in year N-1

total number of patients or subjects included in the clinical studies for which the service was delivered in year N-1**}**

'total costs of purchase of the service' means total value of the contracts concluded by the beneficiary (including related duties, taxes and charges such as non-deductible VAT) for the specific service delivered in year N-1 for the conduct of clinical studies, provided the contracts were awarded according to the principle of best value for money and without any conflict of interests

For **indirect costs**:

{{{cost component 'personnel costs' + cost component 'consumables' + cost component 'medical equipment'}

minus

{costs of in-kind contributions provided by third parties which are not used on the beneficiary's premises + costs of providing financial support to third parties (if any)}**}**

multiplied by

25%**}**

The estimation of the resources to be used must be done on the basis of the study protocol and must be the same for all beneficiaries/linked third parties/third parties involved.

The year N-1 to be used is the last closed financial year at the time of submission of the grant application.

Estimated number of units: see (for each clinical study and beneficiary/linked third party) the 'unit cost table' attached

Unit cost table: clinical studies unit cost[6]

Task, Direct cost categories	**Resource per patient**	**Costs year N-1 Beneficiary 1 [short name]**	**Costs year N-1 Linked third party 1a [short name]**	**Costs year N-1 Beneficiary 2 [short name]**	**Costs year N-1 Linked third party 2a [short name]**	**Costs year N-1 Third party giving in-kind contributions 1 [short name]**
Sequence No. 1						
Task No. 1 Blood sample						
(a) **Personnel costs:** - Doctors	n/a					
- Other Medical Personnel	Phlebotomy (nurse), 10 minutes	8,33 EUR	11,59 EUR	10,30 EUR	11,00 EUR	9,49 EUR
- Technical Personnel	Sample Processing (lab technician), 15 minutes	9,51 EUR	15,68 EUR	14,60 EUR	15,23 EUR	10,78 EUR
(b) **Costs of consumables**:	Syringe	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
	Cannula	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
	Blood container	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
(c) **Costs of medical equipment**:	Use of -80° deep freezer, 60 days	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
	Use of centrifuge, 15 minutes	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
(d) **Costs of services**	Cleaning of XXX	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
(e) **Indirect costs** (25% flat-rate)		XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
Task No. 2						
…						
Amount per unit (unit cost sequence 1):		**XX EUR**	**XX EUR**	**XX EUR**	**XX EUR**	**XX EUR**
Sequence No. 2						
Task No. 1						

[6] Same table as in proposal and Annex 1.

XXX						
(a) **Personnel costs:** - Doctors	XXX	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
- Other Medical Personnel	XXX	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
- Technical Personnel	XXX	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
(b) **Costs of consumables**:	XXX	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
	XXX	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
	XXX	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
(c) **Costs of medical equipment**:	XXX	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
	XXX	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
(d) **Costs of services**	XXX	XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
(e) **Indirect costs** (25% flat-rate)		XX EUR	XX EUR	XX EUR	XX EUR	XX EUR
Task No. 2						
…						
Amount per unit (unit cost sequence 2):		**XX EUR**	**XX EUR**	**XX EUR**	**XX EUR**	**XX EUR**
…						
Amount per unit (unit cost entire study):		**XX EUR**	**XX EUR**	**XX EUR**	**XX EUR**	**XX EUR**

1

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY GMBH (AIT), established in GIEFINGGASSE 4, WIEN 1210, Austria, VAT number: ATU14703506, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('2')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

POLITECNICO DI MILANO (POLIMI), established in PIAZZA LEONARDO DA VINCI 32, MILANO 20133, Italy, VAT number: IT04376620151, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('3')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

CONSIGLIO NAZIONALE DELLE RICERCHE (CNR), established in PIAZZALE ALDO MORO 7, ROMA 00185, Italy, VAT number: IT02118311006, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('4')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

UNIVERSITEIT LEIDEN (LEIDEN), established in RAPENBURG 70, LEIDEN 2311 EZ, Netherlands, VAT number: NL809778282B01, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('5')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

THE UNIVERSITY OF SHEFFIELD (USFD), established in FIRTH COURT WESTERN BANK, SHEFFIELD S10 2TN, United Kingdom, VAT number: GB648238808, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('6')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V. (FRAUNHOFER), established in HANSASTRASSE 27C, MUNCHEN 80686, Germany, VAT number: DE129515865, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('7')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

UNIVERSITA DELLA SVIZZERA ITALIANA (USI), established in VIA LAMBERTENGHI 10 A, LUGANO 6904, Switzerland, VAT number: CH507591, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('8')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

JOANNEUM RESEARCH FORSCHUNGSGESELLSCHAFT MBH (JOANNEUM), established in LEONHARDSTRASSE 59, GRAZ 8010, Austria, VAT number: ATU28781306, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('9')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

THE UNIVERSITY OF SUSSEX (UOS), established in SUSSEX HOUSE FALMER, BRIGHTON BN1 9RH, United Kingdom, VAT number: GB692712320, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('10')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

DEUTSCHES ZENTRUM FUR HOCHSCHUL- UND WISSENSCHAFTSFORSCHUNG GMBH (DZHW), established in LANGE LAUBE 12, HANNOVER 30159, Germany, VAT number: DE291239300, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('11')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

UNIVERSITY OF STRATHCLYDE (STRATHCLYDE), established in Richmond Street 16, GLASGOW G1 1XQ, United Kingdom, VAT number: GB261339762, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('12')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

UNIVERSITA DEGLI STUDI DI ROMA LA SAPIENZA (SAPIENZA), established in Piazzale Aldo Moro 5, ROMA 00185, Italy, VAT number: IT02133771002, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('13')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

NORDISK INSTITUTT FOR STUDIER AV INNOVASJON, FORSKNING OG UTDANNING (NIFU), established in ØKERNVEIEN 9, OSLO 0608, Norway, VAT number: NO976073169MVA, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('14')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS (CSIC), established in CALLE SERRANO 117, MADRID 28006, Spain, VAT number: ESQ2818002D, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('15')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

THE SAMUEL NEAMAN INSTITUTE FOR ADVANCED STUDIES IN SCIENCE AND TECHNOLOGY OF THE TECHNION LTD (PSC) (SNI), established in TECHNION CITY, HAIFA 32000, Israel, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('16')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

TECHNOLOGICKE CENTRUM AKADEMIE VED CESKE REPUBLIKY (TC CAS), established in Ve Struhach 1076/27, PRAHA 160 00, Czech Republic, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('17')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

ATHINA-EREVNITIKO KENTRO KAINOTOMIAS STIS TECHNOLOGIES TIS PLIROFORIAS, TON EPIKOINONION KAI TIS GNOSIS (ATHENA RC), established in ARTEMIDOS 6 KAI EPIDAVROU, MAROUSSI 151 25, Greece, VAT number: EL999723442, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary No ('18')

in Grant Agreement No 824091 ('the Agreement')

between UNIVERSITE DE MARNE LA VALLEE **and** the European Union ('the EU'), represented by the European Commission ('the Commission'),

for the action entitled 'European Research Infrastructure for Science, technology and Innovation policy Studies 2 (RISIS 2)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement it in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

ⓘ print format A4 landscape

MODEL ANNEX 4 FOR H2020 GENERAL MGA — MULTI

Associated with document Ref. Ares(2018)6225458 - 04/12/2018

FINANCIAL STATEMENT FOR *[BENEFICIARY [name]/ LINKED THIRD PARTY [name]]* FOR REPORTING PERIOD [reporting period]

	Eligible[1] costs (per budget category)													Receipts	EU contribution			Additional information
	A. Direct personnel costs				B. Direct costs of subcontracting	*[C. Direct costs of fin. support]*	D. Other direct costs			E. Indirect costs[2]	*[F. Costs of …]*		Total costs	Receipts	Reimbursement rate %	Maximum EU contribution[3]	Requested EU contribution	Information for indirect costs :
	A.1 Employees (or equivalent) A.2 Natural persons under direct contract A.3 Seconded persons *[A.6 Personnel for providing access to research infrastructure]*		A.4 SME owners without salary A.5 Beneficiaries that are natural persons without salary			*[C.1 Financial support]* *[C.2 Prizes]*	D.1 Travel D.2 Equipment D.3 Other goods and services	*[D.4 Costs of large research infrastructure]*	D.5 Costs of internally invoiced goods and services		*[F.1 Costs of …]*	*[F.2 Costs of …]*		Receipts of the action, to be reported in the last reporting period, according to Article 5.3.3				Costs of in-kind contributions not used on premises
Form of costs[4]	Actual	Unit	Unit		Actual	Actual	Actual	Actual	Unit	Flat-rate[5] 25%	Unit		*[Unit][Lump sum]*					
	a	Total b	No hours	Total c	d	*[e]*	f	*[g]*	Total h	i=0,25 x (a+b+ c+f+*[g]* + h+ *[j 1]*[6] +*[j2]*[6] -p)	No units	Total *[j1]*	Total *[j2]*	k = a+b+c+d+*[e]* +*f* + *[g]* +h+ i + *[j1]* +*[j2]*	l	m	n	o	p
[short name beneficiary/linked third party]																		

The beneficiary/linked third party hereby confirms that:
The information provided is complete, reliable and true.
The costs declared are eligible (see Article 6).
The costs can be substantiated by adequate records and supporting documentation that will be produced upon request or in the context of checks, reviews, audits and investigations (see Articles 17, 18 and 22).
For the last reporting period: that all the receipts have been declared (see Article 5.3.3).

ⓘ Please declare all eligible costs, even if they exceed the amounts indicated in the estimated budget (see Annex 2). Only amounts that were declared in your individual financial statements can be taken into account lateron, in order to replace other costs that are found to be ineligible.

[1] See Article 6 for the eligibility conditions

[2] The indirect costs claimed must be free of any amounts covered by an operating grant (received under any EU or Euratom funding programme; see Article 6.2.E). If you have received an operating grant during this reporting period, you cannot claim indirect costs unless you can demonstrate that the operating grant does not cover any costs of the action.

[3] This is the *theoretical* amount of EU contribution that the system calculates automatically (by multiplying the reimbursement rate by the total costs declared). The amount you request (in the column 'requested EU contribution') may be less,

[4] See Article 5 for the forms of costs

[5] Flat rate : 25% of eligible direct costs, from which are excluded: direct costs of subcontracting, costs of in-kind contributions not used on premises, direct costs of financial support, and unit costs declared under budget category F if they include indirect costs (see Article 6.2.E)

[6] Only specific unit costs that do not include indirect costs

ANNEX 5

MODEL FOR THE CERTIFICATE ON THE FINANCIAL STATEMENTS

- **For options [*in italics in square brackets*]: choose the applicable option. Options not chosen should be deleted.**
- **For fields in [grey in square brackets]: enter the appropriate data**

TABLE OF CONTENTS

Terms of Reference for an Independent Report of Factual Findings on costs declared under a Grant Agreement financed under the Horizon 2020 Research and Innovation Framework Programme

This document sets out the '**Terms of Reference (ToR)**' under which

[OPTION 1: [insert name of the beneficiary] ('the Beneficiary')] [OPTION 2: [insert name of the linked third party] ('the Linked Third Party'), third party linked to the Beneficiary [insert name of the beneficiary] ('the Beneficiary')]

agrees to engage

[**insert legal name of the auditor**] ('the Auditor')

to produce an independent report of factual findings ('the Report') concerning the Financial Statement(s)[1] drawn up by the *[Beneficiary] [Linked Third Party]* for the Horizon 2020 grant agreement [insert number of the grant agreement, title of the action, acronym and duration from/to] ('the Agreement'), and

to issue a Certificate on the Financial Statements' ('CFS') referred to in Article 20.4 of the Agreement based on the compulsory reporting template stipulated by the Commission.

The Agreement has been concluded under the Horizon 2020 Research and Innovation Framework Programme (H2020) between the Beneficiary and *[OPTION 1: the European Union, represented by the European Commission ('the Commission')][OPTION 2: the European Atomic Energy Community (Euratom,) represented by the European Commission ('the Commission')][OPTION 3: the [Research Executive Agency (REA)] [European Research Council Executive Agency (ERCEA)] [Innovation and Networks Executive Agency (INEA)] [Executive Agency for Small and Medium-sized Enterprises (EASME)] ('the Agency'), under the powers delegated by the European Commission ('the Commission').]*

The *[Commission] [Agency]* is mentioned as a signatory of the Agreement with the Beneficiary only. The *[European Union][Euratom][Agency]* is not a party to this engagement.

1.1 Subject of the engagement

The coordinator must submit to the *[Commission][Agency]* the final report within 60 days following the end of the last reporting period which should include, amongst other documents, a CFS for each beneficiary and for each linked third party that requests a total contribution of EUR 325 000 or more, as reimbursement of actual costs and unit costs calculated on the basis of its usual cost accounting practices (see Article 20.4 of the Agreement). The CFS must cover all reporting periods of the beneficiary or linked third party indicated above.

The Beneficiary must submit to the coordinator the CFS for itself and for its linked third party(ies), if the CFS must be included in the final report according to Article 20.4 of the Agreement.

The CFS is composed of two separate documents:

- The Terms of Reference ('the ToR') to be signed by the *[Beneficiary] [Linked Third Party]* and the Auditor;

[1] By which costs under the Agreement are declared (see template 'Model Financial Statements' in Annex 4 to the Grant Agreement).

- The Auditor's Independent Report of Factual Findings ('the Report') to be issued on the Auditor's letterhead, dated, stamped and signed by the Auditor (or the competent public officer) which includes the agreed-upon procedures ('the Procedures') to be performed by the Auditor, and the standard factual findings ('the Findings') to be confirmed by the Auditor.

If the CFS must be included in the final report according to Article 20.4 of the Agreement, the request for payment of the balance relating to the Agreement cannot be made without the CFS. However, the payment for reimbursement of costs covered by the CFS does not preclude the Commission *[Agency,]* the European Anti-Fraud Office and the European Court of Auditors from carrying out checks, reviews, audits and investigations in accordance with Article 22 of the Agreement.

1.2 Responsibilities

The *[Beneficiary] [Linked Third Party]:*

- must draw up the Financial Statement(s) for the action financed by the Agreement in compliance with the obligations under the Agreement. The Financial Statement(s) must be drawn up according to the *[Beneficiary's] [Linked Third Party's]* accounting and book-keeping system and the underlying accounts and records;
- must send the Financial Statement(s) to the Auditor;
- is responsible and liable for the accuracy of the Financial Statement(s);
- is responsible for the completeness and accuracy of the information provided to enable the Auditor to carry out the Procedures. It must provide the Auditor with a written representation letter supporting these statements. The written representation letter must state the period covered by the statements and must be dated;
- accepts that the Auditor cannot carry out the Procedures unless it is given full access to the *[Beneficiary's] [Linked Third Party's]* staff and accounting as well as any other relevant records and documentation.

The Auditor:

- *[Option 1 by default:* is qualified to carry out statutory audits of accounting documents in accordance with Directive 2006/43/EC of the European Parliament and of the Council of 17 May 2006 on statutory audits of annual accounts and consolidated accounts, amending Council Directives 78/660/EEC and 83/349/EEC and repealing Council Directive 84/253/EEC or similar national regulations*]*.
- *[Option 2 if the Beneficiary or Linked Third Party has an independent Public Officer:* is a competent and independent Public Officer for which the relevant national authorities have established the legal capacity to audit the Beneficiary*]*.
- *[Option 3 if the Beneficiary or Linked Third Party is an international organisation:* is an [internal] [external] auditor in accordance with the internal financial regulations and procedures of the international organisation*]*.

The Auditor:

- must be independent from the Beneficiary *[and the Linked Third Party]*, in particular, it must not have been involved in preparing the *[Beneficiary's] [Linked Third Party's]* Financial Statement(s);
- must plan work so that the Procedures may be carried out and the Findings may be assessed;
- must adhere to the Procedures laid down and the compulsory report format;
- must carry out the engagement in accordance with this ToR;
- must document matters which are important to support the Report;
- must base its Report on the evidence gathered;
- must submit the Report to the *[Beneficiary] [Linked Third Party]*.

The Commission sets out the Procedures to be carried out by the Auditor. The Auditor is not responsible for their suitability or pertinence. As this engagement is not an assurance engagement, the Auditor does not provide an audit opinion or a statement of assurance.

1.3 Applicable Standards

The Auditor must comply with these Terms of Reference and with[2]:

- the International Standard on Related Services ('ISRS') 4400 *Engagements to perform Agreed-upon Procedures regarding Financial Information* as issued by the International Auditing and Assurance Standards Board (IAASB);
- the *Code of Ethics for Professional Accountants* issued by the International Ethics Standards Board for Accountants (IESBA). Although ISRS 4400 states that independence is not a requirement for engagements to carry out agreed-upon procedures, the *[Commission][Agency]* requires that the Auditor also complies with the Code's independence requirements.

The Auditor's Report must state that there is no conflict of interests in establishing this Report between the Auditor and the Beneficiary *[and the Linked Third Party]*, and must specify - if the service is invoiced - the total fee paid to the Auditor for providing the Report.

1.4 Reporting

The Report must be written in the language of the Agreement (see Article 20.7).

Under Article 22 of the Agreement, the Commission*[, the Agency],* the European Anti-Fraud Office and the Court of Auditors have the right to audit any work that is carried out under the action and for which costs are declared from [*the European Union*] [*Euratom*] budget. This includes work related to this engagement. The Auditor must provide access to all working papers (e.g. recalculation of hourly rates, verification of the time declared for the action) related to this assignment if the Commission *[, the Agency]*, the European Anti-Fraud Office or the European Court of Auditors requests them.

1.5 Timing

The Report must be provided by *[*dd Month yyyy*]*.

1.6 Other terms

[*The [Beneficiary] [Linked Third Party] and the Auditor can use this section to agree other specific terms, such as the Auditor's fees, liability, applicable law, etc. Those specific terms must not contradict the terms specified above.*]

[legal name of the Auditor]	[legal name of the *[Beneficiary][Linked Third Party]*]
[name & function of authorised representative]	[name & function of authorised representative]
[dd Month yyyy]	[dd Month yyyy]
Signature of the Auditor	Signature of the *[Beneficiary][Linked Third Party]*

[2] Supreme Audit Institutions applying INTOSAI-standards may carry out the Procedures according to the corresponding International Standards of Supreme Audit Institutions and code of ethics issued by INTOSAI instead of the International Standard on Related Services ('ISRS') 4400 and the Code of Ethics for Professional Accountants issued by the IAASB and the IESBA.

Independent Report of Factual Findings on costs declared under Horizon 2020 Research and Innovation Framework Programme

(To be printed on the Auditor's letterhead)

To
[name of contact person(s)], [Position]
[[*Beneficiary's*] [*Linked Third Party's*] name]
[Address]
[dd Month yyyy]

Dear [Name of contact person(s)],

As agreed under the terms of reference dated [dd Month yyyy]

with *[OPTION 1: [insert name of the beneficiary] ('the Beneficiary')] [OPTION 2: [insert name of the linked third party] ('the Linked Third Party'), third party linked to the Beneficiary [insert name of the beneficiary] ('the Beneficiary')],*

we

[name of the auditor] ('the Auditor'),

established at

[full address/city/state/province/country],

represented by

[name and function of an authorised representative],

have carried out the procedures agreed with you regarding the costs declared in the Financial Statement(s)[3] of the *[Beneficiary] [Linked Third Party]* concerning the grant agreement [insert grant agreement reference: number, title of the action and acronym] ('the Agreement'),

with a total cost declared of
[total amount] EUR,

and a total of actual costs and unit costs calculated in accordance with the [*Beneficiary's*] [*Linked Third Party's*] usual cost accounting practices' declared of

[sum of total actual costs and total direct personnel costs declared as unit costs calculated in accordance with the *[Beneficiary's] [Linked Third Party's]* usual cost accounting practices] EUR

and **hereby provide our Independent Report of Factual Findings ('the Report')** using the compulsory report format agreed with you.

The Report

Our engagement was carried out in accordance with the terms of reference ('the ToR') appended to this Report. The Report includes the agreed-upon procedures ('the Procedures') carried out and the standard factual findings ('the Findings') examined.

[3] By which the Beneficiary declares costs under the Agreement (see template 'Model Financial Statement' in Annex 4 to the Agreement).

The Procedures were carried out solely to assist the *[Commission] [Agency]* in evaluating whether the *[Beneficiary's] [Linked Third Party's]* costs in the accompanying Financial Statement(s) were declared in accordance with the Agreement. The *[Commission] [Agency]* draws its own conclusions from the Report and any additional information it may require.

The scope of the Procedures was defined by the Commission. Therefore, the Auditor is not responsible for their suitability or pertinence. Since the Procedures carried out constitute neither an audit nor a review made in accordance with International Standards on Auditing or International Standards on Review Engagements, the Auditor does not give a statement of assurance on the Financial Statements.

Had the Auditor carried out additional procedures or an audit of the *[Beneficiary's] [Linked Third Party's]* Financial Statements in accordance with International Standards on Auditing or International Standards on Review Engagements, other matters might have come to its attention and would have been included in the Report.

Not applicable Findings

We examined the Financial Statement(s) stated above and considered the following Findings not applicable:

Explanation (to be removed from the Report):

*If a Finding was not applicable, it must be marked as '**N.A.**' ('Not applicable') in the corresponding row on the right-hand column of the table and means that the Finding did not have to be corroborated by the Auditor and the related Procedure(s) did not have to be carried out.*

The reasons of the non-application of a certain Finding must be obvious i.e.

i) if no cost was declared under a certain category then the related Finding(s) and Procedure(s) are not applicable;

ii) if the condition set to apply certain Procedure(s) are not met the related Finding(s) and those Procedure(s) are not applicable. For instance, for 'beneficiaries with accounts established in a currency other than euro' the Procedure and Finding related to 'beneficiaries with accounts established in euro' are not applicable. Similarly, if no additional remuneration is paid, the related Finding(s) and Procedure(s) for additional remuneration are not applicable.

List here all Findings considered not applicable for the present engagement and explain the reasons of the non-applicability.

….

Exceptions

Apart from the exceptions listed below, the *[Beneficiary] [Linked Third Party]* provided the Auditor all the documentation and accounting information needed by the Auditor to carry out the requested Procedures and evaluate the Findings.

Explanation (to be removed from the Report):

- *If the Auditor was not able to successfully complete a procedure requested, it must be marked as '**E**' ('Exception') in the corresponding row on the right-hand column of the table. The reason such as the inability to reconcile key information or the unavailability of data that prevents the Auditor from carrying out the Procedure must be indicated below.*
- *If the Auditor cannot corroborate a standard finding after having carried out the corresponding procedure, it must also be marked as '**E**' ('Exception') and, where possible, the reasons why the Finding was not fulfilled and its possible impact must be explained here below.*

List here any exceptions and add any information on the cause and possible consequences of each exception, if known. If the exception is quantifiable, include the corresponding amount.

….

Example (to be removed from the Report):

1. *The Beneficiary was unable to substantiate the Finding number 1 on … because ….*
2. *Finding number 30 was not fulfilled because the methodology used by the Beneficiary to calculate unit costs was different from the one approved by the Commission. The differences were as follows: …*
3. *After carrying out the agreed procedures to confirm the Finding number 31, the Auditor found a difference of _____ EUR. The difference can be explained by …*

Further Remarks

In addition to reporting on the results of the specific procedures carried out, the Auditor would like to make the following general remarks:

Example (to be removed from the Report):

1. *Regarding Finding number 8 the conditions for additional remuneration were considered as fulfilled because …*
2. *In order to be able to confirm the Finding number 15 we carried out the following additional procedures: ….*

Use of this Report

This Report may be used only for the purpose described in the above objective. It was prepared solely for the confidential use of the *[Beneficiary] [Linked Third Party]* and the [*Commission*] [*Agency*], and only to be submitted to the *[Commission] [Agency]* in connection with the requirements set out in Article 20.4 of the Agreement. The Report may not be used by the *[Beneficiary] [Linked Third Party]* or by the *[Commission] [Agency]* for any other purpose, nor may it be distributed to any other parties. The *[Commission] [Agency]* may only disclose the Report to authorised parties, in particular to the European Anti-Fraud Office (OLAF) and the European Court of Auditors.

This Report relates only to the Financial Statement(s) submitted to the *[Commission] [Agency]* by the *[Beneficiary] [Linked Third Party]* for the Agreement. Therefore, it does not extend to any other of the *[Beneficiary's] [Linked Third Party's]* Financial Statement(s).

There was no conflict of interest[4] between the Auditor and the Beneficiary *[and Linked Third Party]* in establishing this Report. The total fee paid to the Auditor for providing the Report was EUR _____ (including EUR_____ of deductible VAT).

We look forward to discussing our Report with you and would be pleased to provide any further information or assistance.

[legal name of the Auditor]
[name and function of an authorised representative]
[dd Month yyyy]
Signature of the Auditor

[4] A conflict of interest arises when the Auditor's objectivity to establish the certificate is compromised in fact or in appearance when the Auditor for instance:
- was involved in the preparation of the Financial Statements;
- stands to benefit directly should the certificate be accepted;
- has a close relationship with any person representing the beneficiary;
- is a director, trustee or partner of the beneficiary; or
- is in any other situation that compromises his or her independence or ability to establish the certificate impartially.

Agreed-upon procedures to be performed and standard factual findings to be confirmed by the Auditor

The European Commission reserves the right to i) provide the auditor with additional guidance regarding the procedures to be followed or the facts to be ascertained and the way in which to present them (this may include sample coverage and findings) or to ii) change the procedures, by notifying the Beneficiary in writing. The procedures carried out by the auditor to confirm the standard factual finding are listed in the table below.

If this certificate relates to a Linked Third Party, any reference here below to 'the Beneficiary' is to be considered as a reference to 'the Linked Third Party'.

The 'result' column has three different options: 'C', 'E' and 'N.A.':

- 'C' stands for 'confirmed' and means that the auditor can confirm the 'standard factual finding' and, therefore, there is no exception to be reported.
- 'E' stands for 'exception' and means that the Auditor carried out the procedures but cannot confirm the 'standard factual finding', or that the Auditor was not able to carry out a specific procedure (e.g. because it was impossible to reconcile key information or data were unavailable),
- 'N.A.' stands for 'not applicable' and means that the Finding did not have to be examined by the Auditor and the related Procedure(s) did not have to be carried out. The reasons of the non-application of a certain Finding must be obvious i.e. i) if no cost was declared under a certain category then the related Finding(s) and Procedure(s) are not applicable; ii) if the condition set to apply certain Procedure(s) are not met then the related Finding(s) and Procedure(s) are not applicable. For instance, for 'beneficiaries with accounts established in a currency other than the euro' the Procedure related to 'beneficiaries with accounts established in euro' is not applicable. Similarly, if no additional remuneration is paid, the related Finding(s) and Procedure(s) for additional remuneration are not applicable.

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
A	**ACTUAL PERSONNEL COSTS AND UNIT COSTS CALCULATED BY THE BENEFICIARY IN ACCORDANCE WITH ITS USUAL COST ACCOUNTING PRACTICE**		
	The Auditor draws a sample of persons whose costs were declared in the Financial Statement(s) to carry out the procedures indicated in the consecutive points of this section A. (*The sample should be selected randomly so that it is representative. Full coverage is required if there are fewer than 10 people (including employees, natural persons working under a direct contract and personnel seconded by a third party), otherwise the sample should have a minimum of 10 people, or 10% of the total, whichever number is the highest)* The Auditor sampled _____ people out of the total of _____ people.		

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
A.1	**PERSONNEL COSTS** For the persons included in the sample and working under an employment contract or equivalent act (general procedures for individual actual personnel costs and personnel costs declared as unit costs) To confirm standard factual findings 1-5 listed in the next column, the Auditor reviewed following information/documents provided by the Beneficiary: o a list of the persons included in the sample indicating the period(s) during which they worked for the action, their position (classification or category) and type of contract; o the payslips of the employees included in the sample; o reconciliation of the personnel costs declared in the Financial Statement(s) with the accounting system (project accounting and general ledger) and payroll system; o information concerning the employment status and employment conditions of personnel included in the sample, in particular their employment contracts or equivalent; o the Beneficiary's usual policy regarding payroll matters (e.g. salary policy, overtime policy, variable pay); o applicable national law on taxes, labour and social security and o any other document that supports the personnel costs declared. The Auditor also verified the eligibility of all components of the retribution (see Article 6 GA) and recalculated the personnel costs for employees included in the sample.	1) The employees were i) directly hired by the Beneficiary in accordance with its national legislation, ii) under the Beneficiary's sole technical supervision and responsibility and iii) remunerated in accordance with the Beneficiary's usual practices.	
		2) Personnel costs were recorded in the Beneficiary's accounts/payroll system.	
		3) Costs were adequately supported and reconciled with the accounts and payroll records.	
		4) Personnel costs did not contain any ineligible elements.	
		5) There were no discrepancies between the personnel costs charged to the action and the costs recalculated by the Auditor.	
	Further procedures if 'additional remuneration' is paid To confirm standard factual findings 6-9 listed in the next column, the Auditor: o reviewed relevant documents provided by the Beneficiary (legal form, legal/statutory	6) The Beneficiary paying "additional remuneration" was a non-profit legal entity.	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	obligations, the Beneficiary's usual policy on additional remuneration, criteria used for its calculation, the Beneficiary's usual remuneration practice for projects funded under national funding schemes…); o recalculated the amount of additional remuneration eligible for the action based on the supporting documents received (full-time or part-time work, exclusive or non-exclusive dedication to the action, usual remuneration paid for projects funded by national schemes) to arrive at the applicable FTE/year and pro-rata rate (see data collected in the course of carrying out the procedures under A.2 'Productive hours' and A.4 'Time recording system'). *'ADDITIONAL REMUNERATION' MEANS ANY PART OF THE REMUNERATION WHICH EXCEEDS WHAT THE PERSON WOULD BE PAID FOR TIME WORKED IN PROJECTS FUNDED BY NATIONAL SCHEMES.* *IF ANY PART OF THE REMUNERATION PAID TO THE EMPLOYEE QUALIFIES AS "ADDITIONAL REMUNERATION" AND IS ELIGIBLE UNDER THE PROVISIONS OF ARTICLE 6.2.A.1, THIS CAN BE CHARGED AS ELIGIBLE COST TO THE ACTION UP TO THE FOLLOWING AMOUNT:* *(A) IF THE PERSON WORKS FULL TIME AND EXCLUSIVELY ON THE ACTION DURING THE FULL YEAR: UP TO EUR 8 000/YEAR;* *(B) IF THE PERSON WORKS EXCLUSIVELY ON THE ACTION BUT NOT FULL-TIME OR NOT FOR THE FULL YEAR: UP TO THE CORRESPONDING PRO-RATA AMOUNT OF EUR 8 000, OR* *(C) IF THE PERSON DOES NOT WORK EXCLUSIVELY ON THE ACTION: UP TO A PRO-RATA AMOUNT CALCULATED IN ACCORDANCE TO ARTICLE 6.2.A.1.*	7) The amount of additional remuneration paid corresponded to the Beneficiary's usual remuneration practices and was consistently paid whenever the same kind of work or expertise was required.	
		8) The criteria used to calculate the additional remuneration were objective and generally applied by the Beneficiary regardless of the source of funding used.	
		9) The amount of additional remuneration included in the personnel costs charged to the action was capped at EUR 8,000 per FTE/year (up to the equivalent pro-rata amount if the person did not work on the action full-time during the year or did not work exclusively on the action).	
	Additional procedures in case "unit costs calculated by the Beneficiary in accordance with its usual cost accounting practices" is applied: Apart from carrying out the procedures indicated above to confirm standard factual findings 1-5 and, if applicable, also 6-9, the Auditor carried out following procedures to confirm standard	10) The personnel costs included in the Financial Statement were calculated in accordance with the Beneficiary's usual cost accounting practice. This methodology was consistently	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	factual findings 10-13 listed in the next column: o obtained a description of the Beneficiary's usual cost accounting practice to calculate unit costs;. o reviewed whether the Beneficiary's usual cost accounting practice was applied for the Financial Statements subject of the present CFS; o verified the employees included in the sample were charged under the correct category (in accordance with the criteria used by the Beneficiary to establish personnel categories) by reviewing the contract/HR-record or analytical accounting records; o verified that there is no difference between the total amount of personnel costs used in calculating the cost per unit and the total amount of personnel costs recorded in the statutory accounts; o verified whether actual personnel costs were adjusted on the basis of budgeted or estimated elements and, if so, verified whether those elements used are actually relevant for the calculation, objective and supported by documents.	used in all H2020 actions.	
		11) The employees were charged under the correct category.	
		12) Total personnel costs used in calculating the unit costs were consistent with the expenses recorded in the statutory accounts.	
		13) Any estimated or budgeted element used by the Beneficiary in its unit-cost calculation were relevant for calculating personnel costs and corresponded to objective and verifiable information.	
	<u>For natural persons included in the sample and working with the Beneficiary under a direct contract other than an employment contract, such as consultants (no subcontractors).</u> To confirm standard factual findings 14-17 listed in the next column the Auditor reviewed following information/documents provided by the Beneficiary: o the contracts, especially the cost, contract duration, work description, place of work, ownership of the results and reporting obligations to the Beneficiary; o the employment conditions of staff in the same category to compare costs and; o any other document that supports the costs declared and its registration (e.g. invoices, accounting records, etc.).	14) The natural persons worked under conditions similar to those of an employee, in particular regarding the way the work is organised, the tasks that are performed and the premises where they are performed.	
		15) The results of work carried out belong to the Beneficiary, or, if not, the Beneficiary has obtained all necessary rights to fulfil its obligations as if those	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
		results were generated by itself.	
		16) Their costs were not significantly different from those for staff who performed similar tasks under an employment contract with the Beneficiary.	
		17) The costs were supported by audit evidence and registered in the accounts.	
	<u>For personnel seconded by a third party and included in the sample (not subcontractors)</u> To confirm standard factual findings 18-21 listed in the next column, the Auditor reviewed following information/documents provided by the Beneficiary: o their secondment contract(s) notably regarding costs, duration, work description, place of work and ownership of the results; o if there is reimbursement by the Beneficiary to the third party for the resource made available (in-kind contribution against payment): any documentation that supports the costs declared (e.g. contract, invoice, bank payment, and proof of registration in its accounting/payroll, etc.) and reconciliation of the Financial Statement(s) with the accounting system (project accounting and general ledger) as well as any proof that the amount invoiced by the third party did not include any profit; o if there is no reimbursement by the Beneficiary to the third party for the resource made available (in-kind contribution free of charge): a proof of the actual cost borne by the Third Party for the resource made available free of charge to the Beneficiary such as a statement of costs incurred by the Third Party and proof of the registration in the Third Party's accounting/payroll;	18) Seconded personnel reported to the Beneficiary and worked on the Beneficiary's premises (unless otherwise agreed with the Beneficiary).	
		19) The results of work carried out belong to the Beneficiary, or, if not, the Beneficiary has obtained all necessary rights to fulfil its obligations as if those results were generated by itself..	
		If personnel is seconded against payment: 20) The costs declared were supported with documentation and recorded in the	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	○ any other document that supports the costs declared (e.g. invoices, etc.).	Beneficiary's accounts. The third party did not include any profit.	
		If personnel is seconded free of charge: 21) The costs declared did not exceed the third party's cost as recorded in the accounts of the third party and were supported with documentation.	
A.2	**PRODUCTIVE HOURS** To confirm standard factual findings 22-27 listed in the next column, the Auditor reviewed relevant documents, especially national legislation, labour agreements and contracts and time records of the persons included in the sample, to verify that: ○ the annual productive hours applied were calculated in accordance with one of the methods described below, ○ the full-time equivalent (FTEs) ratios for employees not working full-time were correctly calculated. If the Beneficiary applied method B, the auditor verified that the correctness in which the total number of hours worked was calculated and that the contracts specified the annual workable hours. If the Beneficiary applied method C, the auditor verified that the 'annual productive hours' applied when calculating the hourly rate were equivalent to at least 90 % of the 'standard annual workable hours'. The Auditor can only do this if the calculation of the standard annual workable	22) The Beneficiary applied method [*choose one option and delete the others*] [**A**: 1720 hours] [**B**: the 'total number of hours worked'] [**C**: 'standard annual productive hours' used correspond to usual accounting practices]	
		23) Productive hours were calculated annually.	
		24) For employees not working full-time the full-time equivalent (FTE) ratio was correctly applied.	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	hours can be supported by records, such as national legislation, labour agreements, and contracts. *BENEFICIARY'S PRODUCTIVE HOURS' FOR PERSONS WORKING FULL TIME SHALL BE ONE OF THE FOLLOWING METHODS:* ***A**. 1720 ANNUAL PRODUCTIVE HOURS (PRO-RATA FOR PERSONS NOT WORKING FULL-TIME)* ***B**. THE TOTAL NUMBER OF HOURS WORKED BY THE PERSON FOR THE BENEFICIARY IN THE YEAR (THIS METHOD IS ALSO REFERRED TO AS 'TOTAL NUMBER OF HOURS WORKED' IN THE NEXT COLUMN). THE CALCULATION OF THE TOTAL NUMBER OF HOURS WORKED WAS DONE AS FOLLOWS: ANNUAL WORKABLE HOURS OF THE PERSON ACCORDING TO THE EMPLOYMENT CONTRACT, APPLICABLE LABOUR AGREEMENT OR NATIONAL LAW PLUS OVERTIME WORKED MINUS ABSENCES (SUCH AS SICK LEAVE OR SPECIAL LEAVE).* ***C**. THE STANDARD NUMBER OF ANNUAL HOURS GENERALLY APPLIED BY THE BENEFICIARY FOR ITS PERSONNEL IN ACCORDANCE WITH ITS USUAL COST ACCOUNTING PRACTICES (THIS METHOD IS ALSO REFERRED TO AS 'STANDARD ANNUAL PRODUCTIVE HOURS' IN THE NEXT COLUMN). THIS NUMBER MUST BE AT LEAST 90% OF THE STANDARD ANNUAL WORKABLE HOURS.* *'ANNUAL WORKABLE HOURS' MEANS THE PERIOD DURING WHICH THE PERSONNEL MUST BE WORKING, AT THE EMPLOYER'S DISPOSAL AND CARRYING OUT HIS/HER ACTIVITY OR DUTIES UNDER THE EMPLOYMENT CONTRACT, APPLICABLE COLLECTIVE LABOUR AGREEMENT OR NATIONAL WORKING TIME LEGISLATION.*	*If the Beneficiary applied method B.* 25) The calculation of the number of 'annual workable hours', overtime and absences was verifiable based on the documents provided by the Beneficiary. 25.1) The Beneficiary calculates the hourly rates per full financial year following procedure A.3 (method B is not allowed for beneficiaries calculating hourly rates per month).	
		If the Beneficiary applied method C. 26) The calculation of the number of 'standard annual workable hours' was verifiable based on the documents provided by the Beneficiary.	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
		27) The 'annual productive hours' used for calculating the hourly rate were consistent with the usual cost accounting practices of the Beneficiary and were equivalent to at least 90 % of the 'annual workable hours'.	
A.3	**HOURLY PERSONNEL RATES** I) For unit costs calculated in accordance to the Beneficiary's usual cost accounting practice (unit costs): If the Beneficiary has a "Certificate on Methodology to calculate unit costs " (CoMUC) approved by the Commission, the Beneficiary provides the Auditor with a description of the approved methodology and the Commission's letter of acceptance. The Auditor verified that the Beneficiary has indeed used the methodology approved. If so, no further verification is necessary. If the Beneficiary does not have a "Certificate on Methodology" (CoMUC) approved by the Commission, or if the methodology approved was not applied, then the Auditor: o reviewed the documentation provided by the Beneficiary, including manuals and internal guidelines that explain how to calculate hourly rates; o recalculated the unit costs (hourly rates) of staff included in the sample following the results of the procedures carried out in A.1 and A.2. II) For individual hourly rates: The Auditor: o reviewed the documentation provided by the Beneficiary, including manuals and internal guidelines that explain how to calculate hourly rates;	28) The Beneficiary applied [*choose one option and delete the other*]: [Option I: "Unit costs (hourly rates) were calculated in accordance with the Beneficiary's usual cost accounting practices"] [Option II: Individual hourly rates were applied] *For option I concerning unit costs and if the Beneficiary applies the methodology approved by the Commission (CoMUC):* 29) The Beneficiary used the Commission-approved methodology to calculate hourly rates. It corresponded to the organisation's usual cost accounting practices and was applied consistently for all	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	○ recalculated the hourly rates of staff included in the sample (recalculation of all hourly rates if the Beneficiary uses annual rates, recalculation of three months selected randomly for every year and person if the Beneficiary uses monthly rates) following the results of the procedures carried out in A.1 and A.2; ○ (only in case of monthly rates) confirmed that the time spent on parental leave is not deducted, and that, if parts of the basic remuneration are generated over a period longer than a month, the Beneficiary has included only the share which is generated in the month. *"UNIT COSTS CALCULATED BY THE BENEFICIARY IN ACCORDANCE WITH ITS USUAL COST ACCOUNTING PRACTICES":* *IT IS CALCULATED BY DIVIDING THE TOTAL AMOUNT OF PERSONNEL COSTS OF THE CATEGORY TO WHICH THE EMPLOYEE BELONGS VERIFIED IN LINE WITH PROCEDURE A.1 BY THE NUMBER OF FTE AND THE ANNUAL TOTAL PRODUCTIVE HOURS OF THE SAME CATEGORY CALCULATED BY THE BENEFICIARY IN ACCORDANCE WITH PROCEDURE A.2.* *HOURLY RATE FOR INDIVIDUAL ACTUAL PERSONAL COSTS:* *IT IS CALCULATED FOLLOWING ONE OF THE TWO OPTIONS BELOW:* *A) [OPTION BY DEFAULT] BY DIVIDING THE ACTUAL ANNUAL AMOUNT OF PERSONNEL COSTS OF AN EMPLOYEE VERIFIED IN LINE WITH PROCEDURE A.1 BY THE NUMBER OF ANNUAL PRODUCTIVE HOURS VERIFIED IN LINE WITH PROCEDURE A.2 (FULL FINANCIAL YEAR HOURLY RATE);* *B) BY DIVIDING THE ACTUAL MONTHLY AMOUNT OF PERSONNEL COSTS OF AN EMPLOYEE VERIFIED IN LINE WITH PROCEDURE A.1 BY 1/12 OF THE NUMBER OF ANNUAL PRODUCTIVE HOURS VERIFIED IN LINE WITH PROCEDURE A.2.(MONTHLY HOURLY RATE).*	activities irrespective of the source of funding.	
		For option I concerning unit costs and if the Beneficiary applies a methodology not approved by the Commission: 30) The unit costs re-calculated by the Auditor were the same as the rates applied by the Beneficiary.	
		For option II concerning individual hourly rates: 31) The individual rates re-calculated by the Auditor were the same as the rates applied by the Beneficiary. 31.1) The Beneficiary used only one option (per full financial year or per month) throughout each financial year examined. 31.2) The hourly rates do not include additional remuneration.	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
A.4	**TIME RECORDING SYSTEM** To verify that the time recording system ensures the fulfilment of all minimum requirements and that the hours declared for the action were correct, accurate and properly authorised and supported by documentation, the Auditor made the following checks for the persons included in the sample that declare time as worked for the action on the basis of time records: o description of the time recording system provided by the Beneficiary (registration, authorisation, processing in the HR-system); o its actual implementation; o time records were signed at least monthly by the employees (on paper or electronically) and authorised by the project manager or another manager; o the hours declared were worked within the project period; o there were no hours declared as worked for the action if HR-records showed absence due to holidays or sickness (further cross-checks with travels are carried out in B.1 below) ; o the hours charged to the action matched those in the time recording system. *ONLY THE HOURS WORKED ON THE ACTION CAN BE CHARGED. ALL WORKING TIME TO BE CHARGED SHOULD BE RECORDED THROUGHOUT THE DURATION OF THE PROJECT, ADEQUATELY SUPPORTED BY EVIDENCE OF THEIR REALITY AND RELIABILITY (SEE SPECIFIC PROVISIONS BELOW FOR PERSONS WORKING EXCLUSIVELY FOR THE ACTION WITHOUT TIME RECORDS).*	32) All persons recorded their time dedicated to the action on a **daily/ weekly/ monthly** basis using a **paper/computer-based** system. (*delete the answers that are not applicable)*	
		33) Their time-records were authorised at least monthly by the project manager or other superior.	
		34) Hours declared were worked within the project period and were consistent with the presences/absences recorded in HR-records.	
		35) There were no discrepancies between the number of hours charged to the action and the number of hours recorded.	
	<u>If the persons are working exclusively for the action and without time records</u> For the persons selected that worked exclusively for the action without time records, the Auditor verified evidence available demonstrating that they were in reality exclusively dedicated to the action and that the Beneficiary signed a declaration confirming that they have worked exclusively for the action.	36) The exclusive dedication is supported by a declaration signed by the Beneficiary and by any other evidence gathered.	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
B	**COSTS OF SUBCONTRACTING**		
B.1	**The Auditor obtained the detail/breakdown of subcontracting costs and sampled _____ cost items selected randomly** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest*). To confirm standard factual findings 37-41 listed in the next column, the Auditor reviewed the following for the items included in the sample: o the use of subcontractors was foreseen in Annex 1; o subcontracting costs were declared in the subcontracting category of the Financial Statement; o supporting documents on the selection and award procedure were followed; o the Beneficiary ensured best value for money (key elements to appreciate the respect of this principle are the award of the subcontract to the bid offering best price-quality ratio, under conditions of transparency and equal treatment. In case an existing framework contract was used the Beneficiary ensured it was established on the basis of the principle of best value for money under conditions of transparency and equal treatment). In particular, i. if the Beneficiary acted as a contracting authority within the meaning of Directive 2004/18/EC (or 2014/24/EU) or of Directive 2004/17/EC (or 2014/25/EU), the Auditor verified that the applicable national law on public procurement was followed and that the subcontracting complied with the Terms and Conditions of the Agreement. ii. if the Beneficiary did not fall under the above-mentioned category the Auditor verified that the Beneficiary followed their usual procurement rules and respected the Terms and Conditions of the Agreement..	37) The use of claimed subcontracting costs was foreseen in Annex 1 and costs were declared in the Financial Statements under the subcontracting category.	
		38) There were documents of requests to different providers, different offers and assessment of the offers before selection of the provider in line with internal procedures and procurement rules. Subcontracts were awarded in accordance with the principle of best value for money. *(When different offers were not collected the Auditor explains the reasons provided by the Beneficiary under the caption "Exceptions" of the Report. The Commission will analyse this information to evaluate whether these costs might be accepted as eligible)*	
		39) The subcontracts were not awarded to other Beneficiaries	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	For the items included in the sample the Auditor also verified that: o the subcontracts were not awarded to other Beneficiaries in the consortium; o there were signed agreements between the Beneficiary and the subcontractor; o there was evidence that the services were provided by subcontractor;	of the consortium.	
		40) All subcontracts were supported by signed agreements between the Beneficiary and the subcontractor.	
		41) There was evidence that the services were provided by the subcontractors.	
C	**COSTS OF PROVIDING FINANCIAL SUPPORT TO THIRD PARTIES**		
C.1	**The Auditor obtained the detail/breakdown of the costs of providing financial support to third parties and sampled _____ cost items selected randomly** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest*). The Auditor verified that the following minimum conditions were met: a) the maximum amount of financial support for each third party did not exceed EUR 60 000, unless explicitly mentioned in Annex 1; b) the financial support to third parties was agreed in Annex 1 of the Agreement and the other provisions on financial support to third parties included in Annex 1 were respected.	42) All minimum conditions were met	

D	OTHER ACTUAL DIRECT COSTS		
D.1	**COSTS OF TRAVEL AND RELATED SUBSISTENCE ALLOWANCES** **The Auditor sampled _____ cost items selected randomly** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is the highest*). The Auditor inspected the sample and verified that: o travel and subsistence costs were consistent with the Beneficiary's usual policy for travel. In this context, the Beneficiary provided evidence of its normal policy for travel costs (e.g. use of first class tickets, reimbursement by the Beneficiary on the basis of actual costs, a lump sum or per diem) to enable the Auditor to compare the travel costs charged with this policy; o travel costs are correctly identified and allocated to the action (e.g. trips are directly linked to the action) by reviewing relevant supporting documents such as minutes of meetings, workshops or conferences, their registration in the correct project account, their consistency with time records or with the dates/duration of the workshop/conference; o no ineligible costs or excessive or reckless expenditure was declared (see Article 6.5 MGA).	43) Costs were incurred, approved and reimbursed in line with the Beneficiary's usual policy for travels.	
		44) There was a link between the trip and the action.	
		45) The supporting documents were consistent with each other regarding subject of the trip, dates, duration and reconciled with time records and accounting.	
		46) No ineligible costs or excessive or reckless expenditure was declared.	
D.2	**DEPRECIATION COSTS FOR EQUIPMENT, INFRASTRUCTURE OR OTHER ASSETS** **The Auditor sampled _____ cost items selected randomly** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is the highest*). For "equipment, infrastructure or other assets" [from now on called "asset(s)"] selected in the sample the Auditor verified that: o the assets were acquired in conformity with the Beneficiary's internal guidelines and procedures;	47) Procurement rules, principles and guides were followed.	
		48) There was a link between the grant agreement and the asset charged to the action.	
		49) The asset charged to the action was traceable to the accounting records and the underlying documents.	

	○ they were correctly allocated to the action (with supporting documents such as delivery note invoice or any other proof demonstrating the link to the action) ○ they were entered in the accounting system; ○ the extent to which the assets were used for the action (as a percentage) was supported by reliable documentation (e.g. usage overview table); The Auditor recalculated the depreciation costs and verified that they were in line with the applicable rules in the Beneficiary's country and with the Beneficiary's usual accounting policy (e.g. depreciation calculated on the acquisition value). The Auditor verified that no ineligible costs such as deductible VAT, exchange rate losses, excessive or reckless expenditure were declared (see Article 6.5 GA).	50) The depreciation method used to charge the asset to the action was in line with the applicable rules of the Beneficiary's country and the Beneficiary's usual accounting policy. 51) The amount charged corresponded to the actual usage for the action. 52) No ineligible costs or excessive or reckless expenditure were declared.	
D.3	**COSTS OF OTHER GOODS AND SERVICES** **The Auditor sampled _____ cost items selected randomly** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest*). For the purchase of goods, works or services included in the sample the Auditor verified that: ○ the contracts did not cover tasks described in Annex 1; ○ they were correctly identified, allocated to the proper action, entered in the accounting system (traceable to underlying documents such as purchase orders, invoices and accounting); ○ the goods were not placed in the inventory of durable equipment; ○ the costs charged to the action were accounted in line with the Beneficiary's usual accounting practices; ○ no ineligible costs or excessive or reckless expenditure were declared (see Article 6 GA). In addition, the Auditor verified that these goods and services were acquired in conformity with	53) Contracts for works or services did not cover tasks described in Annex 1. 54) Costs were allocated to the correct action and the goods were not placed in the inventory of durable equipment. 55) The costs were charged in line with the Beneficiary's accounting policy and were adequately supported. 56) No ineligible costs or excessive or reckless expenditure were declared. For internal invoices/charges only the cost element was charged, without any mark-ups.	

	the Beneficiary's internal guidelines and procedures, in particular: o if Beneficiary acted as a contracting authority within the meaning of Directive 2004/18/EC (or 2014/24/EU) or of Directive 2004/17/EC (or 2014/25/EU), the Auditor verified that the applicable national law on public procurement was followed and that the procurement contract complied with the Terms and Conditions of the Agreement. o if the Beneficiary did not fall into the category above, the Auditor verified that the Beneficiary followed their usual procurement rules and respected the Terms and Conditions of the Agreement. For the items included in the sample the Auditor also verified that: o the Beneficiary ensured best value for money (key elements to appreciate the respect of this principle are the award of the contract to the bid offering best price-quality ratio, under conditions of transparency and equal treatment. In case an existing framework contract was used the Auditor also verified that the Beneficiary ensured it was established on the basis of the principle of best value for money under conditions of transparency and equal treatment); *SUCH GOODS AND SERVICES INCLUDE, FOR INSTANCE, CONSUMABLES AND SUPPLIES, DISSEMINATION (INCLUDING OPEN ACCESS), PROTECTION OF RESULTS, SPECIFIC EVALUATION OF THE ACTION IF IT IS REQUIRED BY THE AGREEMENT, CERTIFICATES ON THE FINANCIAL STATEMENTS IF THEY ARE REQUIRED BY THE AGREEMENT AND CERTIFICATES ON THE METHODOLOGY, TRANSLATIONS, REPRODUCTION.*	57) Procurement rules, principles and guides were followed. There were documents of requests to different providers, different offers and assessment of the offers before selection of the provider in line with internal procedures and procurement rules. The purchases were made in accordance with the principle of best value for money. *(When different offers were not collected the Auditor explains the reasons provided by the Beneficiary under the caption "Exceptions" of the Report. The Commission will analyse this information to evaluate whether these costs might be accepted as eligible)*	
D.4	**AGGREGATED CAPITALISED AND OPERATING COSTS OF RESEARCH INFRASTRUCTURE** The Auditor ensured the existence of a positive ex-ante assessment (issued by the EC Services) of the cost accounting methodology of the Beneficiary allowing it to apply the guidelines on direct costing for large research infrastructures in Horizon 2020.	58) The costs declared as direct costs for Large Research Infrastructures (in the appropriate line of the Financial Statement) comply with the methodology described in the positive ex-ante assessment report.	

	In the cases that a positive ex-ante assessment has been issued *(see the standard factual findings 58-59 on the next column)***,** The Auditor ensured that the beneficiary has applied consistently the methodology that is explained and approved in the positive ex ante assessment; ***In the cases that a positive ex-ante assessment has NOT been issued*** *(see the standard factual findings 60 on the next column),* The Auditor verified that no costs of Large Research Infrastructure have been charged as direct costs in any costs category; ***In the cases that a draft ex-ante assessment report has been issued with recommendation for further changes*** *(see the standard factual findings 60 on the next column),* • The Auditor followed the same procedure as above (when a positive ex-ante assessment has NOT yet been issued) and paid particular attention (testing reinforced) to the cost items for which the draft ex-ante assessment either rejected the inclusion as direct costs for Large Research Infrastructures or issued recommendations.	59) Any difference between the methodology applied and the one positively assessed was extensively described and adjusted accordingly. 60) The direct costs declared were free from any indirect costs items related to the Large Research Infrastructure.	
D.5	**Costs of internally invoiced goods and services** **The Auditor sampled cost items selected randomly** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest*). To confirm standard factual findings 61-65 listed in the next column, the Auditor: ○ obtained a description of the Beneficiary's usual cost accounting practice to calculate costs of internally invoiced goods and services (unit costs); ○ reviewed whether the Beneficiary's usual cost accounting practice was applied for the Financial Statements subject of the present CFS; ○ ensured that the methodology to calculate unit costs is being used in a consistent manner, based on objective criteria, regardless of the source of funding; ○ verified that any ineligible items or any costs claimed under other budget categories, in particular indirect costs, have not been taken into account when calculating the costs of	61) The costs of internally invoiced goods and services included in the Financial Statement were calculated in accordance with the Beneficiary's usual cost accounting practice. 62) The cost accounting practices used to calculate the costs of internally invoiced goods and services were applied by the Beneficiary in a consistent manner based on objective criteria regardless of the source of funding. 63) The unit cost is calculated using the actual costs for the good or service recorded in the Beneficiary's accounts, excluding any ineligible cost or costs included in other	

	internally invoiced goods and services (see Article 6 GA); ○ verified whether actual costs of internally invoiced goods and services were adjusted on the basis of budgeted or estimated elements and, if so, verified whether those elements used are actually relevant for the calculation, and correspond to objective and verifiable information. ○ verified that any costs of items which are not directly linked to the production of the invoiced goods or service (e.g. supporting services like cleaning, general accountancy, administrative support, etc. not directly used for production of the good or service) have not been taken into account when calculating the costs of internally invoiced goods and services. ○ verified that any costs of items used for calculating the costs internally invoiced goods and services are supported by audit evidence and registered in the accounts.	budget categories. 64) The unit cost excludes any costs of items which are not directly linked to the production of the invoiced goods or service. 65) The costs items used for calculating the actual costs of internally invoiced goods and services were relevant, reasonable and correspond to objective and verifiable information.	
E	**USE OF EXCHANGE RATES**		
E.1	a) For Beneficiaries with accounts established in a currency other than euros **The Auditor sampled _____ cost items selected randomly and verified that the exchange rates used for converting other currencies into euros were in accordance with the following rules established in the Agreement** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest)*: *COSTS RECORDED IN THE ACCOUNTS IN A CURRENCY OTHER THAN EURO SHALL BE CONVERTED INTO EURO AT THE AVERAGE OF THE DAILY EXCHANGE RATES PUBLISHED IN THE C SERIES OF OFFICIAL JOURNAL OF THE EUROPEAN UNION (https://www.ecb.int/stats/exchange/eurofxref/html/index.en.html), DETERMINED OVER THE CORRESPONDING REPORTING PERIOD.* *IF NO DAILY EURO EXCHANGE RATE IS PUBLISHED IN THE OFFICIAL JOURNAL OF THE EUROPEAN UNION FOR THE CURRENCY IN QUESTION, CONVERSION SHALL BE MADE AT THE AVERAGE OF THE MONTHLY ACCOUNTING RATES ESTABLISHED BY THE COMMISSION AND PUBLISHED ON ITS WEBSITE (http://ec.europa.eu/budget/contracts_grants/info_contracts/inforeuro/inforeuro_en.cfm),*	66) The exchange rates used to convert other currencies into Euros were in accordance with the rules established of the Grant Agreement and there was no difference in the final figures.	

	DETERMINED OVER THE CORRESPONDING REPORTING PERIOD.		
	b) For Beneficiaries with accounts established in euros **The Auditor sampled _____ cost items selected randomly and verified that the exchange rates used for converting other currencies into euros were in accordance with the following rules established in the Agreement** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest)*: *COSTS INCURRED IN ANOTHER CURRENCY SHALL BE CONVERTED INTO EURO BY APPLYING THE BENEFICIARY'S USUAL ACCOUNTING PRACTICES.*	67) The Beneficiary applied its usual accounting practices.	

[legal name of the audit firm]
[name and function of an authorised representative]
[dd Month yyyy]
<Signature of the Auditor>

ANNEX 6

MODEL FOR THE CERTIFICATE ON THE METHODOLOGY

- **For options [*in italics in square brackets*]: choose the applicable option. Options not chosen should be deleted.**
- **For fields in [grey in square brackets]: enter the appropriate data.**

TABLE OF CONTENTS

Terms of reference for an audit engagement for a methodology certificate in connection with one or more grant agreements financed under the Horizon 2020 Research and Innovation Framework Programme

This document sets out the '**Terms of Reference (ToR)**' under which

[OPTION 1: [insert name of the beneficiary] ('the Beneficiary')] [OPTION 2: [insert name of the linked third party] ('the Linked Third Party'), third party linked to the Beneficiary [insert name of the beneficiary] ('the Beneficiary')]

agrees to engage

[**insert legal name of the auditor**] ('the Auditor')

to produce an independent report of factual findings ('the Report') concerning the *[Beneficiary's] [Linked Third Party's]* usual accounting practices for calculating and claiming direct personnel costs declared as unit costs ('the Methodology') in connection with grant agreements financed under the Horizon 2020 Research and Innovation Framework Programme.

The procedures to be carried out for the assessment of the methodology will be based on the grant agreement(s) detailed below:

[title and number of the grant agreement(s)] ('the Agreement(s)')

The Agreement(s) has(have) been concluded between the Beneficiary and *[OPTION 1: the European Union, represented by the European Commission ('the Commission')][OPTION 2: the European Atomic Energy Community (Euratom,) represented by the European Commission ('the Commission')][OPTION 3: the [Research Executive Agency (REA)] [European Research Council Executive Agency (ERCEA)] [Innovation and Networks Executive Agency (INEA)] [Executive Agency for Small and Medium-sized Enterprises (EASME)] ('the Agency'), under the powers delegated by the European Commission ('the Commission').]*.

The *[Commission] [Agency]* is mentioned as a signatory of the Agreement with the Beneficiary only. The *[European Union] [Euratom] [Agency]* is not a party to this engagement.

1.1 Subject of the engagement

According to Article 18.1.2 of the Agreement, beneficiaries *[and linked third parties]* that declare direct personnel costs as unit costs calculated in accordance with their usual cost accounting practices may submit to the *[Commission] [Agency]*, for approval, a certificate on the methodology ('CoMUC') stating that there are adequate records and documentation to prove that their cost accounting practices used comply with the conditions set out in Point A of Article 6.2.

The subject of this engagement is the CoMUC which is composed of two separate documents:

- the Terms of Reference ('the ToR') to be signed by the *[Beneficiary] [Linked Third Party]* and the Auditor;

- the Auditor's Independent Report of Factual Findings ('the Report') issued on the Auditor's letterhead, dated, stamped and signed by the Auditor which includes; the standard statements ('the Statements') evaluated and signed by the *[Beneficiary] [Linked Third Party]*, the agreed-upon procedures ('the Procedures') performed by the Auditor and the standard factual findings

('the Findings') assessed by the Auditor. The Statements, Procedures and Findings are summarised in the table that forms part of the Report.

The information provided through the Statements, the Procedures and the Findings will enable the Commission to draw conclusions regarding the existence of the *[Beneficiary's] [Linked Third Party's]* usual cost accounting practice and its suitability to ensure that direct personnel costs claimed on that basis comply with the provisions of the Agreement. The Commission draws its own conclusions from the Report and any additional information it may require.

1.2 Responsibilities

The parties to this agreement are the *[Beneficiary] [Linked Third Party]* and the Auditor.

The *[Beneficiary] [Linked Third Party]:*

- is responsible for preparing financial statements for the Agreement(s) ('the Financial Statements') in compliance with those Agreements;
- is responsible for providing the Financial Statement(s) to the Auditor and enabling the Auditor to reconcile them with the *[Beneficiary's] [Linked Third Party's]* accounting and bookkeeping system and the underlying accounts and records. The Financial Statement(s) will be used as a basis for the procedures which the Auditor will carry out under this ToR;
- is responsible for its Methodology and liable for the accuracy of the Financial Statement(s);
- is responsible for endorsing or refuting the Statements indicated under the heading 'Statements to be made by the Beneficiary/ Linked Third Party' in the first column of the table that forms part of the Report;
- must provide the Auditor with a signed and dated representation letter;
- accepts that the ability of the Auditor to carry out the Procedures effectively depends upon the *[Beneficiary] [Linked Third Party]* providing full and free access to the *[Beneficiary's] [Linked Third Party's]* staff and to its accounting and other relevant records.

The Auditor:

- *[Option 1 by default: is qualified to carry out statutory audits of accounting documents in accordance with Directive 2006/43/EC of the European Parliament and of the Council of 17 May 2006 on statutory audits of annual accounts and consolidated accounts, amending Council Directives 78/660/EEC and 83/349/EEC and repealing Council Directive 84/253/EEC or similar national regulations].*
- *[Option 2 if the Beneficiary or Linked Third Party has an independent Public Officer: is a competent and independent Public Officer for which the relevant national authorities have established the legal capacity to audit the Beneficiary].*
- *[Option 3 if the Beneficiary or Linked Third Party is an international organisation: is an [internal] [external] auditor in accordance with the internal financial regulations and procedures of the international organisation].*

The Auditor:

- must be independent from the Beneficiary *[and the Linked Third Party]*, in particular, it must not have been involved in preparing the Beneficiary's *[and Linked Third Party's]* Financial Statement(s);
- must plan work so that the Procedures may be carried out and the Findings may be assessed;
- must adhere to the Procedures laid down and the compulsory report format;
- must carry out the engagement in accordance with these ToR;
- must document matters which are important to support the Report;
- must base its Report on the evidence gathered;
- must submit the Report to the *[Beneficiary] [Linked Third Party]*.

The Commission sets out the Procedures to be carried out and the Findings to be endorsed by the Auditor. The Auditor is not responsible for their suitability or pertinence. As this engagement is not an assurance engagement the Auditor does not provide an audit opinion or a statement of assurance.

1.3 Applicable Standards

The Auditor must comply with these Terms of Reference and with[1]:

- the International Standard on Related Services ('ISRS') 4400 *Engagements to perform Agreed-upon Procedures regarding Financial Information* as issued by the International Auditing and Assurance Standards Board (IAASB);
- the *Code of Ethics for Professional Accountants* issued by the International Ethics Standards Board for Accountants (IESBA). Although ISRS 4400 states that independence is not a requirement for engagements to carry out agreed-upon procedures, the Commission requires that the Auditor also complies with the Code's independence requirements.

The Auditor's Report must state that there was no conflict of interests in establishing this Report between the Auditor and the Beneficiary *[and the Linked Third Party]* that could have a bearing on the Report, and must specify – if the service is invoiced - the total fee paid to the Auditor for providing the Report.

1.4 Reporting

The Report must be written in the language of the Agreement (see Article 20.7 of the Agreement).

Under Article 22 of the Agreement, the Commission, *[the Agency]*, the European Anti-Fraud Office and the Court of Auditors have the right to audit any work that is carried out under the action and for which costs are declared from *[the European Union] [Euratom]* budget. This includes work related to this engagement. The Auditor must provide access to all working papers related to this assignment if the Commission*[, the Agency]*, the European Anti-Fraud Office or the European Court of Auditors requests them.

1.5 Timing

The Report must be provided by [dd Month yyyy].

1.6 Other Terms

[The [Beneficiary] [Linked Third Party] and the Auditor can use this section to agree other specific terms, such as the Auditor's fees, liability, applicable law, etc. Those specific terms must not contradict the terms specified above.]

[legal name of the Auditor]	[legal name of the [Beneficiary] [Linked Third Party]]
[name & title of authorised representative]	[name & title of authorised representative]
[dd Month yyyy]	[dd Month yyyy]
Signature of the Auditor	Signature of the *[Beneficiary] [Linked Third Party]*

[1] Supreme Audit Institutions applying INTOSAI-standards may carry out the Procedures according to the corresponding International Standards of Supreme Audit Institutions and code of ethics issued by INTOSAI instead of the International Standard on Related Services ('ISRS') 4400 and the Code of Ethics for Professional Accountants issued by the IAASB and the IESBA.

Independent report of factual findings on the methodology concerning grant agreements financed under the Horizon 2020 Research and Innovation Framework Programme

(To be printed on letterhead paper of the auditor)

To
[name of contact person(s)], [Position]
[[Beneficiary's] [Linked Third Party's] name]
[Address]
[dd Month yyyy]

Dear [Name of contact person(s)],

As agreed under the terms of reference dated [dd Month yyyy]

with *[OPTION 1: [insert name of the beneficiary] ('the Beneficiary')] [OPTION 2: [insert name of the linked third party] ('the Linked Third Party'), third party linked to the Beneficiary [insert name of the beneficiary] ('the Beneficiary')]*,

we

[name of the auditor] ('the Auditor'),

established at

[full address/city/state/province/country],

represented by

[name and function of an authorised representative],

have carried out the agreed-upon procedures ('the Procedures') and provide hereby our Independent Report of Factual Findings ('the Report'), concerning the *[Beneficiary's] [Linked Third Party's]* usual accounting practices for calculating and declaring direct personnel costs declared as unit costs ('the Methodology').

You requested certain procedures to be carried out in connection with the grant(s)

[title and number of the grant agreement(s)] ('the Agreement(s)').

The Report

Our engagement was carried out in accordance with the terms of reference ('the ToR') appended to this Report. The Report includes: the standard statements ('the Statements') made by the *[Beneficiary] [Linked Third Party]*, the agreed-upon procedures ('the Procedures') carried out and the standard factual findings ('the Findings') confirmed by us.

The engagement involved carrying out the Procedures and assessing the Findings and the documentation requested appended to this Report, the results of which the Commission uses to draw conclusions regarding the acceptability of the Methodology applied by the *[Beneficiary] [Linked Third Party]*.

The Report covers the methodology used from [dd Month yyyy]. In the event that the *[Beneficiary] [Linked Third Party]* changes this methodology, the Report will not be applicable to any Financial Statement[1] submitted thereafter.

The scope of the Procedures and the definition of the standard statements and findings were determined solely by the Commission. Therefore, the Auditor is not responsible for their suitability or pertinence.

Since the Procedures carried out constitute neither an audit nor a review made in accordance with International Standards on Auditing or International Standards on Review Engagements, we do not give a statement of assurance on the costs declared on the basis of the *[Beneficiary's] [Linked Third Party's]* Methodology. Had we carried out additional procedures or had we performed an audit or review in accordance with these standards, other matters might have come to its attention and would have been included in the Report.

Exceptions

Apart from the exceptions listed below, the *[Beneficiary] [Linked Third Party]* agreed with the standard Statements and provided the Auditor all the documentation and accounting information needed by the Auditor to carry out the requested Procedures and corroborate the standard Findings.

List here any exception and add any information on the cause and possible consequences of each exception, if known. If the exception is quantifiable, also indicate the corresponding amount.

…..

Explanation of possible exceptions in the form of examples (to be removed from the Report):

i. the [Beneficiary] [Linked Third Party] did not agree with the standard Statement number … because…;

ii. the Auditor could not carry out the procedure … established because …. (e.g. due to the inability to reconcile key information or the unavailability or inconsistency of data);

iii. the Auditor could not confirm or corroborate the standard Finding number … because ….

Remarks

We would like to add the following remarks relevant for the proper understanding of the Methodology applied by the [Beneficiary] [Linked Third Party] or the results reported:

Example (to be removed from the Report):

Regarding the methodology applied to calculate hourly rates …

Regarding standard Finding 15 it has to be noted that …

The [Beneficiary] [Linked Third Party] explained the deviation from the benchmark statement XXIV concerning time recording for personnel with no exclusive dedication to the action in the following manner: …

Annexes

Please provide the following documents to the auditor and annex them to the report when submitting this CoMUC to the Commission:

[1] Financial Statement in this context refers solely to Annex 4 of the Agreement by which the Beneficiary declares costs under the Agreement.

1. Brief description of the methodology for calculating personnel costs, productive hours and hourly rates;
2. Brief description of the time recording system in place;
3. An example of the time records used by the [*Beneficiary] [Linked Third Party*];
4. Description of any budgeted or estimated elements applied, together with an explanation as to why they are relevant for calculating the personnel costs and how they are based on objective and verifiable information;
5. A summary sheet with the hourly rate for direct personnel declared by the [*Beneficiary*] [*Linked Third Party*] and recalculated by the Auditor for each staff member included in the sample (the names do not need to be reported);
6. A comparative table summarising for each person selected in the sample a) the time claimed by the [*Beneficiary*] [*Linked Third Party*] in the Financial Statement(s) and b) the time according to the time record verified by the Auditor;
7. A copy of the letter of representation provided to the Auditor.

Use of this Report

This Report has been drawn up solely for the purpose given under Point 1.1 Reasons for the engagement.

The Report:

- is confidential and is intended to be submitted to the Commission by the [*Beneficiary*] [*Linked Third Party*] in connection with Article 18.1.2 of the Agreement;
- may not be used by the [*Beneficiary*] [*Linked Third Party*] or by the Commission for any other purpose, nor distributed to any other parties;
- may be disclosed by the Commission only to authorised parties, in particular the European Anti-Fraud Office (OLAF) and the European Court of Auditors.
- relates only to the usual cost accounting practices specified above and does not constitute a report on the Financial Statements of the [*Beneficiary*] [*Linked Third Party*].

No conflict of interest[2] exists between the Auditor and the Beneficiary *[and the Linked Third Party*] that could have a bearing on the Report. The total fee paid to the Auditor for producing the Report was EUR _____ (including EUR _____ of deductible VAT).

We look forward to discussing our Report with you and would be pleased to provide any further information or assistance which may be required.

Yours sincerely

[legal name of the Auditor]
[name and title of the authorised representative]
[dd Month yyyy]
Signature of the Auditor

[2] A conflict of interest arises when the Auditor's objectivity to establish the certificate is compromised in fact or in appearance when the Auditor for instance:
- was involved in the preparation of the Financial Statements;
- stands to benefit directly should the certificate be accepted;
- has a close relationship with any person representing the beneficiary;
- is a director, trustee or partner of the beneficiary; or
- is in any other situation that compromises his or her independence or ability to establish the certificate impartially.

Statements to be made by the Beneficiary/Linked Third Party ('the Statements') and Procedures to be carried out by the Auditor ('the Procedures') and standard factual findings ('the Findings') to be confirmed by the Auditor

The Commission reserves the right to provide the auditor with guidance regarding the Statements to be made, the Procedures to be carried out or the Findings to be ascertained and the way in which to present them. The Commission reserves the right to vary the Statements, Procedures or Findings by written notification to the Beneficiary/Linked Third Party to adapt the procedures to changes in the grant agreement(s) or to any other circumstances.

If this methodology certificate relates to the Linked Third Party's usual accounting practices for calculating and claiming direct personnel costs declared as unit costs any reference here below to 'the Beneficiary' is to be considered as a reference to 'the Linked Third Party'.

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
A. Use of the Methodology I. The cost accounting practice described below has been in use since *[*dd Month yyyy*]*. II. The next planned alteration to the methodology used by the Beneficiary will be from [dd Month yyyy*]*.	**Procedure:** ✓ The Auditor checked these dates against the documentation the Beneficiary has provided. **Factual finding:** 1. The dates provided by the Beneficiary were consistent with the documentation.
B. Description of the Methodology III. The methodology to calculate unit costs is being used in a consistent manner and is reflected in the relevant procedures. *[Please describe the methodology your entity uses to calculate personnel costs, productive hours and hourly rates, present your description to the Auditor and annex it to this certificate]* *[If the statement of section "B. Description of the methodology" cannot be endorsed by the Beneficiary or there is no written methodology to calculate unit costs it should be listed here below and reported as exception by the Auditor in the main Report of Factual Findings:* - *…]*	**Procedure:** ✓ The Auditor reviewed the description, the relevant manuals and/or internal guidance documents describing the methodology. **Factual finding:** 2. The brief description was consistent with the relevant manuals, internal guidance and/or other documentary evidence the Auditor has reviewed. 3. The methodology was generally applied by the Beneficiary as part of its usual costs accounting practices.

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
C. **Personnel costs** General IV. The unit costs (hourly rates) are limited to salaries including during parental leave, social security contributions, taxes and other costs included in the remuneration required under national law and the employment contract or equivalent appointing act; V. Employees are hired directly by the Beneficiary in accordance with national law, and work under its sole supervision and responsibility; VI. The Beneficiary remunerates its employees in accordance with its usual practices. This means that personnel costs are charged in line with the Beneficiary's usual payroll policy (e.g. salary policy, overtime policy, variable pay) and no special conditions exist for employees assigned to tasks relating to the European Union or Euratom, unless explicitly provided for in the grant agreement(s); VII. The Beneficiary allocates its employees to the relevant group/category/cost centre for the purpose of the unit cost calculation in line with the usual cost accounting practice; VIII. Personnel costs are based on the payroll system and accounting system. IX. Any exceptional adjustments of actual personnel costs resulted from relevant budgeted or estimated elements and were based on objective and verifiable information. *[Please describe the 'budgeted or estimated elements' and their relevance to personnel costs, and explain how they were reasonable and based on objective and verifiable information, present your explanation to the Auditor and annex it to this certificate].* X. Personnel costs claimed do not contain any of the following ineligible costs: costs related to return on capital; debt and debt service charges; provisions for future losses or debts; interest owed; doubtful debts; currency exchange losses; bank costs charged by the Beneficiary's bank for transfers from the Commission/Agency; excessive or reckless expenditure; deductible VAT or costs incurred during suspension of the implementation of the action. XI. Personnel costs were not declared under another EU or Euratom grant	**Procedure:** *The Auditor draws a sample of employees to carry out the procedures indicated in this section C and the following sections D to F.* *[The Auditor has drawn a random sample of 10 employees assigned to Horizon 2020 action(s). If fewer than 10 employees are assigned to the Horizon 2020 action(s), the Auditor has selected all employees assigned to the Horizon 2020 action(s) complemented by other employees irrespective of their assignments until he has reached 10 employees.].* For this sample: ✓ the Auditor reviewed all documents relating to personnel costs such as employment contracts, payslips, payroll policy (e.g. salary policy, overtime policy, variable pay policy), accounting and payroll records, applicable national tax , labour and social security law and any other documents corroborating the personnel costs claimed; ✓ in particular, the Auditor reviewed the employment contracts of the employees in the sample to verify that: i. they were employed directly by the Beneficiary in accordance with applicable national legislation; ii. they were working under the sole technical supervision and responsibility of the latter; iii. they were remunerated in accordance with the Beneficiary's usual practices; iv. they were allocated to the correct group/category/cost centre for the purposes of calculating the unit cost in line with the Beneficiary's usual cost accounting practices; ✓ the Auditor verified that any ineligible items or any costs claimed under other costs categories or costs covered by other types of grant or by other grants financed from the European Union budget have not been taken into account when calculating the personnel costs; ✓ the Auditor numerically reconciled the total amount of personnel costs used to calculate the unit cost with the total amount of personnel costs recorded in the statutory accounts and the payroll system.

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
(including grants awarded by a Member State and financed by the EU budget and grants awarded by bodies other than the Commission/Agency for the purpose of implementing the EU or Euratom budget in the same period, unless the Beneficiary can demonstrate that the operating grant does not cover any costs of the action). <u>If additional remuneration as referred to in the grant agreement(s) is paid</u> XII. The Beneficiary is a non-profit legal entity; XIII. The additional remuneration is part of the beneficiary's usual remuneration practices and paid consistently whenever the relevant work or expertise is required; XIV. The criteria used to calculate the additional remuneration are objective and generally applied regardless of the source of funding; XV. The additional remuneration included in the personnel costs used to calculate the hourly rates for the grant agreement(s) is capped at EUR 8 000 per full-time equivalent (reduced proportionately if the employee is not assigned exclusively to the action). *[If certain statement(s) of section "C. Personnel costs" cannot be endorsed by the Beneficiary they should be listed here below and reported as exception by the Auditor in the main Report of Factual Findings:* *- …]*	✓ to the extent that actual personnel costs were adjusted on the basis of budgeted or estimated elements, the Auditor carefully examined those elements and checked the information source to confirm that they correspond to objective and verifiable information; ✓ if additional remuneration has been claimed, the Auditor verified that the Beneficiary was a non-profit legal entity, that the amount was capped at EUR 8 000 per full-time equivalent and that it was reduced proportionately for employees not assigned exclusively to the action(s). ✓ the Auditor recalculated the personnel costs for the employees in the sample. **Factual finding:** 4. All the components of the remuneration that have been claimed as personnel costs are supported by underlying documentation. 5. The employees in the sample were employed directly by the Beneficiary in accordance with applicable national law and were working under its sole supervision and responsibility. 6. Their employment contracts were in line with the Beneficiary's usual policy; 7. Personnel costs were duly documented and consisted solely of salaries, social security contributions (pension contributions, health insurance, unemployment fund contributions, etc.), taxes and other statutory costs included in the remuneration (holiday pay, thirteenth month's pay, etc.); 8. The totals used to calculate the personnel unit costs are consistent with those registered in the payroll and accounting records; 9. To the extent that actual personnel costs were adjusted on the basis of budgeted or estimated elements, those elements were relevant for calculating the personnel costs and correspond to objective and verifiable information. The budgeted or estimated elements used are: — (indicate the elements and their values). 10. Personnel costs contained no ineligible elements; 11. Specific conditions for eligibility were fulfilled when additional

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
	remuneration was paid: a) the Beneficiary is registered in the grant agreements as a non-profit legal entity; b) it was paid according to objective criteria generally applied regardless of the source of funding used and c) remuneration was capped at EUR 8 000 per full-time equivalent (or up to up to the equivalent pro-rata amount if the person did not work on the action full-time during the year or did not work exclusively on the action).
D. Productive hours XVI. The number of productive hours per full-time employee applied is *[delete as appropriate]*: A. 1720 productive hours per year for a person working full-time (corresponding pro-rata for persons not working full time). B. the total number of hours worked in the year by a person for the Beneficiary C. the standard number of annual hours generally applied by the beneficiary for its personnel in accordance with its usual cost accounting practices. This number must be at least 90% of the standard annual workable hours. <u>If method B is applied</u> XVII. The calculation of the total number of hours worked was done as follows: annual workable hours of the person according to the employment contract, applicable labour agreement or national law plus overtime worked minus absences (such as sick leave and special leave). XVIII. 'Annual workable hours' are hours during which the personnel must be working, at the employer's disposal and carrying out his/her activity or duties under the employment contract, applicable collective labour agreement or national working time legislation. XIX. The contract (applicable collective labour agreement or national working time legislation) do specify the working time enabling to calculate the annual workable hours.	**Procedure (same sample basis as for Section C: Personnel costs):** ✓ The Auditor verified that the number of productive hours applied is in accordance with method A, B or C. ✓ The Auditor checked that the number of productive hours per full-time employee is correct. ✓ If method B is applied the Auditor verified i) the manner in which the total number of hours worked was done and ii) that the contract specified the annual workable hours by inspecting all the relevant documents, national legislation, labour agreements and contracts. ✓ If method C is applied the Auditor reviewed the manner in which the standard number of working hours per year has been calculated by inspecting all the relevant documents, national legislation, labour agreements and contracts and verified that the number of productive hours per year used for these calculations was at least 90 % of the standard number of working hours per year. **Factual finding:** <u>General</u> 12. The Beneficiary applied a number of productive hours consistent with method A, B or C detailed in the left-hand column. 13. The number of productive hours per year per full-time employee was accurate. <u>If method B is applied</u> 14. The number of 'annual workable hours', overtime and absences was

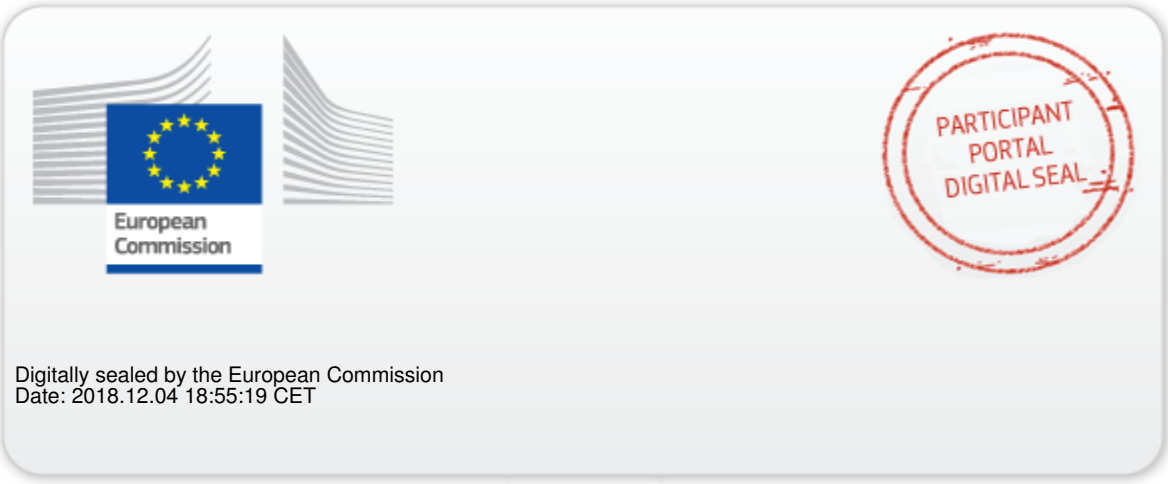
Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
<u>If method C is applied</u> XX. The standard number of productive hours per year is that of a full-time equivalent. XXI. The number of productive hours per year on which the hourly rate is based i) corresponds to the Beneficiary's usual accounting practices; ii) is at least 90% of the standard number of workable (working) hours per year. XXII. Standard workable (working) hours are hours during which personnel are at the Beneficiary's disposal preforming the duties described in the relevant employment contract, collective labour agreement or national labour legislation. The number of standard annual workable (working) hours that the Beneficiary claims is supported by labour contracts, national legislation and other documentary evidence. *[If certain statement(s) of section "D. Productive hours" cannot be endorsed by the Beneficiary they should be listed here below and reported as exception by the Auditor:* - *…]*	verifiable based on the documents provided by the Beneficiary and the calculation of the total number of hours worked was accurate. 15. The contract specified the working time enabling to calculate the annual workable hours. <u>If method C is applied</u> 16. The calculation of the number of productive hours per year corresponded to the usual costs accounting practice of the Beneficiary. 17. The calculation of the standard number of workable (working) hours per year was corroborated by the documents presented by the Beneficiary. 18. The number of productive hours per year used for the calculation of the hourly rate was at least 90% of the number of workable (working) hours per year.
E. Hourly rates The hourly rates are correct because: XXIII. Hourly rates are correctly calculated since they result from dividing annual personnel costs by the productive hours of a given year and group (e.g. staff category or department or cost centre depending on the methodology applied) and they are in line with the statements made in section C. and D. above. *[If the statement of section 'E. Hourly rates' cannot be endorsed by the Beneficiary they should be listed here below and reported as exception by the Auditor:* - *…]*	**Procedure** ✓ The Auditor has obtained a list of all personnel rates calculated by the Beneficiary in accordance with the methodology used. ✓ The Auditor has obtained a list of all the relevant employees, based on which the personnel rate(s) are calculated. For 10 employees selected at random (same sample basis as Section C: Personnel costs): ✓ The Auditor recalculated the hourly rates. ✓ The Auditor verified that the methodology applied corresponds to the usual accounting practices of the organisation and is applied consistently for all activities of the organisation on the basis of objective criteria irrespective of the source of funding. **Factual finding:**

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
	19. No differences arose from the recalculation of the hourly rate for the employees included in the sample.
F. Time recording XXIV. Time recording is in place for all persons with no exclusive dedication to one Horizon 2020 action. At least all hours worked in connection with the grant agreement(s) are registered on a **daily/weekly/monthly** basis *[delete as appropriate]* using a **paper/computer-based system** *[delete as appropriate]*; XXV. For persons exclusively assigned to one Horizon 2020 activity the Beneficiary has either signed a declaration to that effect or has put arrangements in place to record their working time; XXVI. Records of time worked have been signed by the person concerned (on paper or electronically) and approved by the action manager or line manager at least monthly; XXVII. Measures are in place to prevent staff from: i. recording the same hours twice, ii. recording working hours during absence periods (e.g. holidays, sick leave), iii. recording more than the number of productive hours per year used to calculate the hourly rates, and iv. recording hours worked outside the action period. XXVIII. No working time was recorded outside the action period; XXIX. No more hours were claimed than the productive hours used to calculate the hourly personnel rates. *[Please provide a brief description of the time recording system in place together with the measures applied to ensure its reliability to the Auditor and annex it to the*	**Procedure** ✓ The Auditor reviewed the brief description, all relevant manuals and/or internal guidance describing the methodology used to record time. The Auditor reviewed the time records of the random sample of 10 employees referred to under Section C: Personnel costs, and verified in particular: ✓ that time records were available for all persons with not exclusive assignment to the action; ✓ that time records were available for persons working exclusively for a Horizon 2020 action, or, alternatively, that a declaration signed by the Beneficiary was available for them certifying that they were working exclusively for a Horizon 2020 action; ✓ that time records were signed and approved in due time and that all minimum requirements were fulfilled; ✓ that the persons worked for the action in the periods claimed; ✓ that no more hours were claimed than the productive hours used to calculate the hourly personnel rates; ✓ that internal controls were in place to prevent that time is recorded twice, during absences for holidays or sick leave; that more hours are claimed per person per year for Horizon 2020 actions than the number of productive hours per year used to calculate the hourly rates; that working time is recorded outside the action period; ✓ the Auditor cross-checked the information with human-resources records to verify consistency and to ensure that the internal controls have been effective. In addition, the Auditor has verified that no more hours were charged to Horizon 2020 actions per person per year than the number of productive hours per year used to calculate the hourly rates, and verified that

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
present certificate[1]]. *[If certain statement(s) of section "F. Time recording" cannot be endorsed by the Beneficiary they should be listed here below and reported as exception by the Auditor:* - *…]*	no time worked outside the action period was charged to the action. **Factual finding:** 20. The brief description, manuals and/or internal guidance on time recording provided by the Beneficiary were consistent with management reports/records and other documents reviewed and were generally applied by the Beneficiary to produce the financial statements. 21. For the random sample time was recorded or, in the case of employees working exclusively for the action, either a signed declaration or time records were available; 22. For the random sample the time records were signed by the employee and the action manager/line manager, at least monthly. 23. Working time claimed for the action occurred in the periods claimed; 24. No more hours were claimed than the number productive hours used to calculate the hourly personnel rates; 25. There is proof that the Beneficiary has checked that working time has not been claimed twice, that it is consistent with absence records and the number of productive hours per year, and that no working time has been claimed outside the action period. 26. Working time claimed is consistent with that on record at the human-resources department.

[1] The description of the time recording system must state among others information on the content of the time records, its coverage (full or action time-recording, for all personnel or only for personnel involved in H2020 actions), its degree of detail (whether there is a reference to the particular tasks accomplished), its form, periodicity of the time registration and authorisation (paper or a computer-based system; on a daily, weekly or monthly basis; signed and countersigned by whom), controls applied to prevent double-charging of time or ensure consistency with HR-records such as absences and travels as well as it information flow up to its use for the preparation of the Financial Statements.

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
[official name of the *[Beneficiary] [Linked Third Party]**]*** ***[name and title of authorised representative]*** ***[dd Month yyyy]*** ***<Signature of the*** *[Beneficiary] [Linked Third Party]>*	***[official name of the Auditor]*** ***[name and title of authorised representative]*** ***[dd Month yyyy]*** ***<Signature*** *of the Auditor>*



This document is digitally sealed. The digital sealing mechanism uniquely binds the document to the modules of the Participant Portal of the European Commission, to the transaction for which it was generated and ensures its integrity and authenticity.

Any attempt to modify the content will lead to a breach of the electronic seal, which can be verified at any time by clicking on the digital seal validation symbol.