RISISWEEK AGENDA
January 27th-30th, 2020 ISI-Fraunhofer - Karlsruhe

RISIS
RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES

	Monday 27	Tuesday 28	Wednesday 29	Thursday 30
9		**Plenary Session on Datasets** Core and New Datasets Developments (WP5 / WP7 / WP8 / WP9 / WP10) *(T. Scherngell, B. Lepori, M. Guerini, E. Noyons)*	**Plenary Session on RCF and services** (WP4 / WP6) *(M. Barbier, P. Breucker, D. Maynard)*	**Plenary Session** Interaction with the Research community, Communication and Dissemination (WP2) *(E. Reale)*
10				
		Break		
11		**Reflexive Session 1** Identification of 'use cases' (work in 5 groups and collective reporting)	**Reflexive Session 2** Developing integrated solutions to address selected use cases (work in 5 groups and collective reporting)	**Take-away messages from the parallel sessions** Conclusion/Closure *(P. Laredo)*
12	**Arrival** Light lunch			
13	**RISIS overall introduction (WP1)** *(P. Laredo)*	Lunch		
14		**Parallel Sessions** S1: Datastore Design & Specification (P. Breucker) S2: Spatial models (M. Neulandtner, T. Scherngell) S3: Careers *(J. Tesch)*	**Parallel Sessions** S7: Ontology development for missions *(D. Maynard, B. Lepori, P. Laredo)* S8: Open Science *(E. Noyons)* S9: Data quality *(C. Daraio)*	**Governing Board** Feedback from the Advisory Board Year 2 planning Administrative & financial matters Amendment request Discussion on indicators Memorandum of Agreement ESFRI roadmap
	Break			
15	**Presentation & discussion of the long term future of RISIS (WP3)** *(SWG)*			
		Break		
16		**Parallel Sessions** S4: Services integration in the RCF *(D. Maynard, P. Breucker)* S5: Integration of national Programs & Projects *(E. Reale, G. Zahradnik, T. Scherngell, A.Zinilli)* S6: RISIS registers of actors (M. Guerini, B. Lepori)	**Parallel Sessions** S10: RCF Specification and tasks *(P. Breucker)* S11: Panel models *(M. Guerini)* S12: Integration of policy evaluations within RISIS *(E. Wallwaey, S. Buhrer-Topcu)*	
	Break			
17	**New RISIS achievements** ESID/Trademarks Gate/Geography			
20	**Welcome Cocktail**		**Social dinner**	

Monday 27/01

Monday afternoon is dedicated to the state of affairs of RISIS2 (plenary session 1) and the expectations that come from the fact we are supported as an advanced research community research infrastructure. One central aspect of such communities is that they are supposed to last beyond the end of the contract, so after 2022. This is why we created at the kick-off meeting a sustainable working group. Their members will introduce the discussion by identifying the issues at stake and the options they are considering. We expect a large discussion as it deals with the long term engagement of many of us and of our groups or centres (plenary session 2)

At 5pm, we shall have short presentations about recent achievements on datasets and/or services that have just opened or are soon to open

Four are planned for the time being:

▸ The 2 new datasets to open early 2020 ESID on social innovation projects and Trademarks (final name still to define).

▸ The new service on geographical annotation that is now fully open under CORTEXT

▸ The possibilities GATE offers and how it can be accessed and used before its integration in the RCF and with CORTEXT.

VENUE

ISI-Fraunhofer - Karlsruhe
Breslauer Str. 48, 76139 Karlsruhe
(line 4 / tram stop: Glogauer Str.)

Tuesday 28/01

9h00-10h30 Plenary Session on Datasets

It covers the Work Packages 5, 7, 8, 9 & 10
This session will present major developments and advancements of RISIS core datasets, the access to these datasets via the RISIS datasets portal as well as advanced usage cases. First, the status of access (WP8) will be discussed (number and characteristics of access requests in 2019), together with new developments in the portal (e.g. multiple selection in one request). Then, main activities for the core datasets maintenance (WP5) and deepening (WP9) will be discussed, with a focus on the work plans for 2020 and transversal issues. Moreover, the status in development of new datasets (WP10) will be summarized, also introducing a timeline of opening of these new datasets, as well as highlighting the main questions the new datasets are intended to address. Finally, the session will provide an overview on advanced usage cases (WP7) mobilizing the core datasets, and how these usage cases are positioned in the research and policy community.

11h00-12h30 Reflexive Session 1

The object of the 2 reflexive sessions is to identify use-cases and see how we need to be organised so that users can handle them from A to Z.
In this first session, we shall split in 5 groups
Each group will do an open screening about what are from their viewpoints the 3 use cases that demonstrate best the added value of RISIS and the integration of its datasets and services (1H)
The 5 groups will present their results on one poster with their arguments and we shall select those for further elaboration the following day.

13h30-15h30 Session 1: Datastore design and specifications

This session will present results of a preliminary study conducted on the structure of the RCF datastore. It will be based mainly on the work done by Cristian Martinez on the different technologies and framework regarding the state of the art on data storage, and a draft of the future datastore specifications. A discussion will be held on technical aspects of the RCF datastore. The outcome of this discussion will be a decision on

which technology to put in place with the perspective of a working proof of concept for the beta version of RCF in June.

A second discussion will focus on a more long-term data sustainability approach, in order to prepare the work for the following years, especially on the skills needed.

13h30-15h30 Session 2: Spatial models

This session will focus on one of the advanced methods used to address advanced user communities (WP7). Although spatial data and geographical analyses become increasingly important in STI studies, spatial econometric models that are standard in e.g. economic geography are hardly used up to now in this field. Such models are promising instruments for estimating relationships in spatial data, e.g. the influence of some independent factors on a dependent variable observed for concrete spatial units, most often regions or urban areas. For instance, one may want to estimate the effects of different determinants, such as R&D funding, on different knowledge production activities (e.g. publications or patents) in regions. Spatial regression are models are not only able to get rid of the statistical problems related to such a questions (biased estimates due to spatial autocorrelation), but will also provide an estimate on spatial spill over effects. In the case of spatial interaction (or network) data, spatial interaction models come into usage. The session will introduce the basics of spatial models, and present two use cases, one employing a spatial regression approach, and another a spatial interaction model.

13h30-15h30 Session 3: Careers

This session will present progress of the careers database. We will give a short overview of the data collected and integrated regarding the two cohorts (2010, 2014) of PhD-holders in Norway, Germany, The Netherlands, Spain and Israel that comprise the database. This includes situating the data in the overall population frame. We will then give a presentation on the method used to match PhD-Holders to the Leiden publication database, which is currently being developed using the Netherlands as a test case, which will be expanded to the other countries in the first half of 2020. The remainder of the session will be used to discuss the ways in which the database can be expanded, both in size (cohorts from other countries) and content (adding more data, e.g. from dissertation acknowledgements and EURAXXES) on the current cohorts.

16h00-17h45 Session 4: Services integration in the RCF

This session will be a presentation and discussion around integrated and tested services in the RCF, related to WP6 work. We will discuss technical issues encountered during the first phase of the project, and further integration to come.

We will also discuss the services usability and pertinence regarding existing problems to solve, and differences between similar services (GATE / CORTEXT / CWTS). The aspect of interesting linking between services should also be discussed, to identify potential opportunity of scenarios of dataset processing towards the identified use cases.

The outcome should be a clear roadmap of integration, and the list of services to integrate, with the technical problems to be overcome in the next few months.

16h00-17h45 Session 5: Integration of national programmes and projects

This session will discuss progress, planned activities and problems related to the deepening of the EUPRO database (WP9) in terms of information on national projects (NATPRO), and the development of the new dataset (WP10) on national R&D policy programmes at the level of policy instruments (EFIL). Since the two activities are inter-related, both conceptually and from a data collection perspective, the progress (e.g. status of pilot countries data collection, status of new member states) will be put on the table, and further advanced in the session. In particular, decisions on the further activities in 2020 with respect to the first pilots of NATPRO and EFIL will be discussed in some detail.

16h00-17h45 Session 6: RISIS registers of actors

This session will discuss progress, planned activities and problems related to the two registers (OrgReg and FirmReg). For OrgReg, the focus will be in particular on the activities related to the extension of the coverage (to other countries as well as other organizations), the status of the deployment of the new version, and the integration of organizational indicators. For FirmReg, the priority is to consolidate its structure, in terms of main tables, unit of analysis (group vs. subsidiaries) and key attributes to be included, and to identify a common procedure/approach to identify firms in datasets other than CIB, VICO and Cheetah.

Wednesday 29/01

9h00-10h30 Plenary Session on the RCF and services (WP4 and WP6)

This session will present major developments and advancements of RISIS services.

There will be first a presentation of the user orientation of the RCF, knowing that 3 sessions are dedicated to its technical development and 2 reflexive sessions on how to identify interesting use cases that will serve to architecture user access and working.

A second part will present the services, their advancement and the developments anticipated with their workplans. They cover our positioning triplet: actors (later treated in a parallel session on RISIS registers), geography (see demonstration Monday evening) and themes of interest, with GATE ontology building (see session on missions). They also cover the further development of CORTEXT for semantic analysis, of visualisation tools (with VosViewer as a first step) and with access to open access datasets (via Open Aire). The discussion should focus on the developments anticipated and their sequencing.

11h00-12h30 Reflexive Session 2

This second session will see groups work on different use cases. The objective is to identify options for addressing the use case, to select one and discuss its implications in term of activities to integrate and of dashboard for the RCF.

13h30-15h30 Session 7: Ontology development for missions

RISIS now integrates the Knowmak interactive tool that enables policy analysts and policymakers to visualise the situation of different emerging technologies and different societal grand challenges. It also enables researchers that are looking only for aggregated panel data to download them. It has raised a lot of interest and RISIS is pushed to develop it widely, which requires that we think how we organise (and even reorganise) for doing it.

One central issue is that we need to evaluate the ways we have built ontologies using the GATE Platform, and the development work needed to make them more robust. This is all the more important that we have planned a number of experiments in RISIS2 (WP6) and that the

Commission is interested in testing our approach on the new missions that will arise from the Mission Boards that have been created and should be known before Summer 2020.

13h30-15h30 Session 8: Open Science

Even though Open Acess tends to be the central focus of most discussions on Open Science, this umbrella term is much more complex than it seems. As defined by the European Commission, OS represents a new approach to the scientific process; one that is based on cooperative work and new ways of diffusing knowledge. Of course, different communities might interpret these concepts in distinct ways, so Fecher and Friesike (2014) have accommodated established approaches in five Open Science schools of thought: Public, Measurement, Democratic, Pragmatic and Infrastructure.

As it can be expected, RISIS has a clear relation to the Infrastructure School of OS, as this intends to achieve more efficient research processes by creating openly available platforms, tools and services for scientists. However, infrastructures should serve a purpose and the development of new digital technologies, and collaborative tools should aim towards the primary assumption of the Pragmatic School of OS: scientists working together will make knowledge-creation more efficient.

In this session, we are going to debate RISIS role as a breeding ground for such Open Science initiatives and their relation to open research data, intellectual property rights, scientific production repositories, and to other infrastructure-based efforts. As a consequence, we hope to better inform our role as actors of the systemic change in the way science and research is done.

With that as a starting point, our belief is that the discussion around the Infrastructure School can develop into other schools beyond the Pragmatic one. For example, the Democratic School can relate to how information systems should be open to the public and how data should be freely accessible (like the ConECTI example discussed last week). On the other hand, the Measurement School can also relate very well, as infrastructure provides new tools to evaluate science (the esPPIral system I briefly mentioned last week is an example of that: a repository designed to make scientific production access easier on evaluation committees, while also providing openness to the public as well).

13h30-15h30 Session 9: Dataset quality

The quality of data and of related information is crucial to add value and improve the awareness and better exploitation of the available data, enhancing data quality-aware empirical investigations for studies in Research, Education and Innovation.

After the Special Session on: Data Quality issues in Research, Education and Innovation studies: state of the art, practices in use and needs held during the EUSPRI 2019 Conference, the 5-7 June 2019 Rome (Italy), the Sapienza team has analysed the RISIS database technical documents. This activity is now used for the development of a Data Quality Questionnaire that will be sent out 10 January 2020 to each database manager, to be compiled on line by 20 January 2020.

During this session, we will discuss the outcome of the Data Quality Questionnaire, analyse existing Data Quality practices and emerging needs of the RISIS datasets.

This session will be helpful for the development of the activities that will be included in the Data Quality for Research and Higher Education Studies Methodological Course, organized by Sapienza University of Rome, Rome, in September 2020 (16-18 September 2020).

16h-17h45 Session 10: RCF specification and tasks

This session will present the current Roadmap for the RCF, its goals and milestones, and the planning of human resources needed and how they are integrated in the team.

The specification mission done by Cristian Martinez on API specification will be presented, to lead the discussion on the RCF internal structure.

We will discuss specifically the details of the future user workspace, services and datasets that are planned to be online for the beta version in June 2020.

The outcome of this session should be a specific task groups and assignments for the next year, and what management tools are necessary for the team integration.

16h-17h45 Session 11: Panel models

This session will focus on one of the advanced methods used to address advanced user communities (WP7). Panel data (a.k.a. longitudinal or cross-sectional time-series data) are repeated measures on the same entities over time. The entities could be, e.g., countries, companies, universities or individuals. The use of panel data allows for richer models and estimation methods as well as the identification of causal effects under weaker assumptions compared to cross-sectional data. In RISIS, a number of dataset are characterized by a panel data structure (e.g., the firm level datasets, CWTS on publications, ETER on universities), therefore a better understanding on the main issues related to the treatment and analysis of this kind of data could be of interest of various internal and external RISIS dataset users. The session will introduce the basics of panel models (e.g. FE vs. RE models), and present use cases and examples based on VICO.

16h-17h45 Session 12: Integration of policy evaluation within RISIS

Policy evaluations are one important dimension in the RISIS core datasets in form of the SIPER database. SIPER is going through significant advancements and developments at the moment, both technically and content-wise. In particular, a better integration of SIPER in the overall RISIS data nexus is high on the agenda. Potential entry points would be the organization registers but also the topical annotation services developed within RISIS. This session will discuss these issues in a broader context of RISIS, feeding into the workplans for SIPER maintenance and deepening for the coming years.

Thursday 30/01

13.00-17h00
Governing Board meeting

9h00-10h30 - Plenary Session on Interaction with the Research community, Communication and Dissemination

The session shall start with a short presentation of activities carried out and foreseen under WP2 (training and communication), with a special focus on the website development in order to collect suggestions for improving the most important mean for RISIS general presentation (30 minutes from 9.00 to 9.30).

Interactions with the research community, neighbouring communities, stakeholders and policy makers is the second topic to be addressed through:

one presentation based on WP7 dissemination activities to the research community,

one presentation based on WP2 dissemination activities to neighbouring fields, stakeholders and policy makers,

plus, discussion on present and future strategies (45 minutes from 9,30 to 10,15).

The last part of the session is devoted to a short information about the RISIS KPI indicators, including developments for the ESFRI proposal (Networking, Excellence, Access, Engagement 15 minutes).

The agenda is extremely heavy and requires the presence of members until the end of the session.

- We have to get the feedback from those members of the advisory board present.

- We have to consider Year 2 planning and administrative and financial issues (amendment request in particular).

- We have to discuss the development of the project and the indicators/KPI that we shall put in place.

- We have to discuss the developments proposed by the SWG on indirect contracting and about the Memorandum of Agreement for preparing the long-term sustainability of RISIS, and we need to have a discussion on the situation of RISIS towards the ESFRI roadmap.

11h00-12.00 Closing session

The closing session will propose a return from the 2 reflexive sessions about our vision of what type of services RISIS wants to promote. And I wish we discuss the take away messages that will come from each session and will be presented as a set of posters for all of us to look at before the session.