RISIS – EUPRO Dataset Demo

Research infrastructures for the assessment of science, technology and innovation policy

Barbara Heller-Schuh, Michael Barber, Thomas Scherngell

Content

- What is EUPRO?
- EUPRO's contribution to study Europeanization
- Characteristics of the dataset
- Variables
- Preparing opening
- Access and training
- Interoperability
- Timeline 2014

What is EUPRO

- EUPRO is a database comprising information on R&D collaborations within the European Framework Programmes (FPs),

- developed and maintained by AIT by standardising information collected from the EU CORDIS database,

- including systematic information on more than 60,000 collaborative research projects of (FP1-FP7) and more than 70,000 participating organisations,

- such as the geographical location and the organisation type etc.

EUPRO's contribution to study Europeanization I

- Initiatives to foster collaborative R&D in precompetitive research have become a key instrument of STI-policy at the regional, national and supranational levels

- Main examples in Europe: European Framework Programmes (FPs) on Research and Technological Development

- (Co-)funding of thousands of R&D projects supporting transnational collaboration and coordination in research and transnational mobility for training purposes; seven FPs since 1984

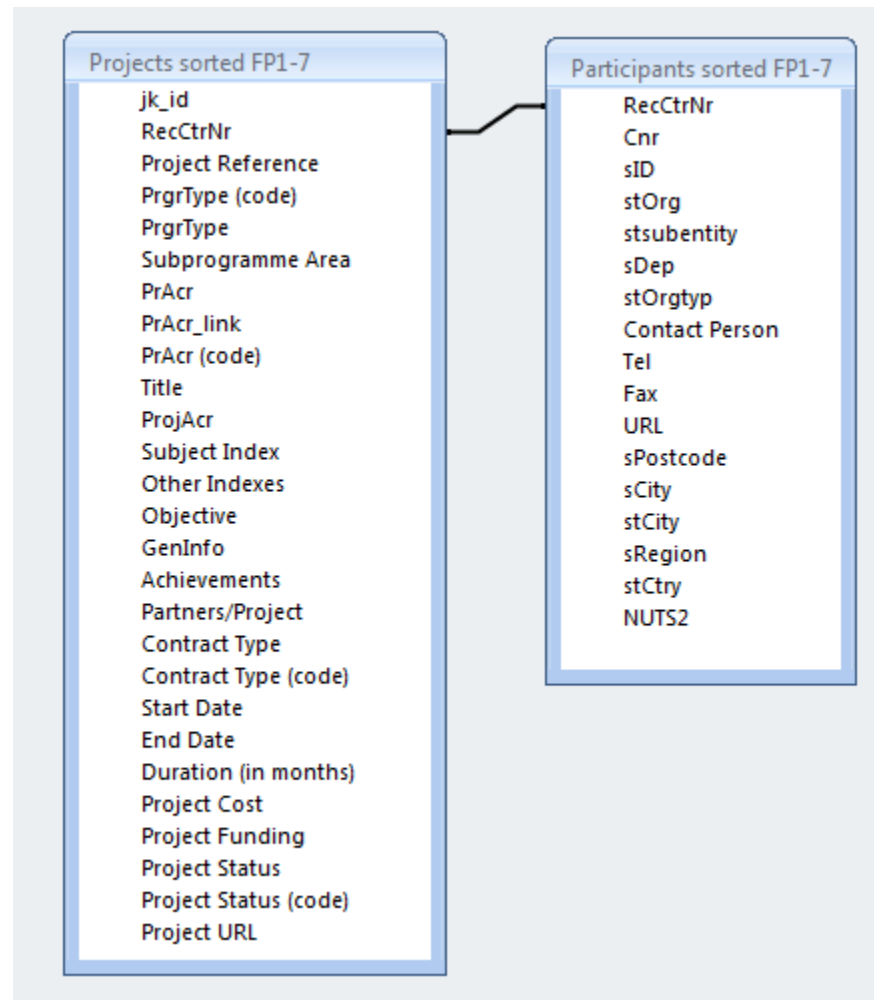# EUPRO's contribution to study Europeanization I

- Key information source for the empirical analysis of network structures and network processes within the European Research Area (ERA)

    - Coverage of a diverse set of countries and regions enables analysis of progress towards ERA from a detailed geographical perspective

    - Investigation of structural properties of collaboration networks provides information for appraising the effectiveness of EU policy in this area

    - Observing the involvement of individual organisations (number of participations, profile of partners, positioning in the collaboration network) supports the identification of key actors in the FPs

Characteristics of the dataset

Programmes	Period	Number of projects	Number of participations	Number of organisations
FP1	1984 - 1987	3.283	7.929	1.982
FP2	1987 - 1991	3.878	19.068	4.406
FP3	1990 - 1994	5.528	31.378	7.051
FP4	1994 - 1998	14.562	67.858	18.635
FP5	1998 - 2002	16.034	78.728	22.105
FP6	2002 - 2006	9.709	72.012	19.479
FP7 (July 2012)	2007 - 2013	10.866	60.135	15.355
Total (July 2012)		**63.860**	**337.108**	**78.336**

- Standardisation procedure
  - Harmonisation of organisation names (heterogenous spellings, different languages)
  - Harmonisation and assignment of organisation types
  - Regionalisation of geographical data to NUTS3 level

EUPRO database structure

Projects sorted FP1-7
- jk_id
- RecCtrNr
- Project Reference
- PrgrType (code)
- PrgrType
- Subprogramme Area
- PrAcr
- PrAcr_link
- PrAcr (code)
- Title
- ProjAcr
- Subject Index
- Other Indexes
- Objective
- GenInfo
- Achievements
- Partners/Project
- Contract Type
- Contract Type (code)
- Start Date
- End Date
- Duration (in months)
- Project Cost
- Project Funding
- Project Status
- Project Status (code)
- Project URL

Participants sorted FP1-7
- RecCtrNr
- Cnr
- sID
- stOrg
- stsubentity
- sDep
- stOrgtyp
- Contact Person
- Tel
- Fax
- URL
- sPostcode
- sCity
- stCity
- sRegion
- stCtry
- NUTS2

EUPRO project variables

RecCtrNr (data type: number)
- unique identifier (record control number) for each project in the database, identical with unique identifier of all projects in the CORDIS projects database

Project Reference (data type: number)
- project index, probably for internal use in the European Commission, structure and usage not clear

Prgr Type (code) (data type: number)
- code (1-7) for the names of the specific framework programme types FP1 to FP7

Subprogramme Area (data type: memo)
- full name of the subprogramme areas in each of the framework programmes

PrAcr (data type: text)
- subprogramme acronyms of the programmes (e.g., FP6-FOOD or FP6-INCO)

Title (data type: text)
- title of the project (full length)

ProjAcr (data type: text)
- (not-unique) project acronym or abbreviation of the project title

EUPRO project variables

Subject Index (data type: text)
- one or more of 45 standardized keywords;

Other Indexes (data type: memo)
- additional keywords to characterize project contents, freely chosen by the project team

Objective, GenInfo (general information), Achievements (data type: memo)
- distinction of these fields not clearly defined, in practice all of these fields (may) contain objectives and/or an abstract of the project.

Partners/Project (data type: number)
- number of participants in each project; additionally generated information, calculated with the help of table "participants sorted"

Contract type (data type: text)
- different types of contracts which regulate size, financing and funding of the research projects

Start Date, End Date (data type: date)
- day, month and year of project start and ending

Duration (in months) (data type: number)
- duration of the project in months

EUPRO project variables

Project Cost (data type: number)

- official project costs

Project Funding (data type: funding)

- financing contribution of the European Union; since not all projects are financed completely in the Additional Costs Model, figures in "Project Funding" are equal or smaller than figures in "Cost".

Project Status (date type: text)

- status of the projects (accepted, completed, execution) found in the CORDIS projects database in January, 2012

Project URL (data type: text)

- official website of the project

EUPRO participants variables

RecCtrNr (data type: number)

- unique identifier (record control number) for each project in the database, identical with unique identifier of all projects in the CORDIS projects database; corresponds with the entries in the field RecCtrNr in the table "projects sorted"

Cnr (data type: number)

- unique identifier (control number) assigned internally by systems research, as project participants are not uniquely indexed in the CORDIS projects database; all project-relevant information is indicated with "1", prime contractor with "2", and remaining participants with "3", "4", etc.

sID (data type: number)

- internal number of standardised project participants (ambiguous number, assigned only for standardisation procedures!)

stOrg (data type: text)

- standardised organisation name

Contact Person (data type: text)

- name of contact person, without name affixes like titles, etc.

sEmail (data type: text)

- electronic mail address for the contact person

sAddress (data type: text)

EUPRO participants variables

<u>sPostcode</u> (data type: text)

<u>sCity</u> (data type: text)

<u>sReg</u> (data type: text)

- region according to the CORDIS projects database

<u>stCtry</u> (data type: text)

- standardised country; country abbreviations are given as ISO Alpha-3 country codes

<u>stOrgtype</u> (data type: text)

- standardised organisation type; available types are: IND (industry), EDU (universities and other educational institutions), ROR (publicly funded research organisations), GOV (governmental institutions), CON (consultants), NCL (non-commercial and non-profit organisations), and OTH (rest group)

<u>NUTS1-NUTS3</u> (data type: text)

- region code according to the NUTS classification scheme

Preparing for opening

- Ongoing raw data retrieval from CORDIS project website

- Need for further cleaning and validation of data
  - Especially needed for organisation names, etc.

- Using matching and assignment algorithms
  - For further assigning organisation types and regions
  - For interlinking with other databases (preliminary actions have been undertaken for interlinking with EUMIDA/ETER)

- Documentation
  - Preliminary version available
  - Has to be updated against the background of RISIS requirements

Access and training

- The EUPRO database is located at AIT (Microsoft Access 2010)

- The integration and opening of EUPRO in RISIS shall enable wider access to and use of selected subsets of EUPRO data, and facilitate collaborative research with AIT staff.

- A complete opening of the database, in particular full access to EUPRO raw data, is not foreseen, not the last in order to avoid an uncontrolled proliferation of data sets and thus to ensure quality control.

- Precise rules for access to EUPRO will be defined as an essential precondition for the opening for EUPRO. In general, access will be possible for joint projects and scientific publications.

- In addition, the provision of specific extracts of EUPRO, by way of hands-on access at AIT premises, is planned.

- Due to the size and complexity of EUPRO, a dedicated local training and close collaboration with AIT staff is foreseen, usually requiring local presence for some 2-3 weeks.

Interoperability

- To inter-link different databases, the names of the organisations and entities provide a bridge between the databases. The harmonisation of these entities across different data bases presents significant challenges:
    - organisation name might feature abbreviations, problems with accented characters, variations in punctuation or spacing between words, differences in case, and simple typos like misspellings.

- AIT has devoted significant resources to the development of matching algorithms for improving interoperability within and between databases, e.g., by identifying matches based on statistical properties such as the frequency of adjacent characters in the strings.

- The integration of such tools and methods in RISIS is possible when precise rules for the usage of the algorithms are defined.

Project References

- F&E Transport - Leistungsfähigkeit und Entwicklungspfade von Foschung und Entwicklung im österreichischen Transportsektor im europäischen Kontext (Horizon 2020),
Austrian Federal Ministry for Transport, Innovation and Technology, 01/14-12/14

- Data preparation for the identification of R&D networks related to European water research,
Rathenau Institute, 02/13-03/13

- Data preparation and scientific advisory for the geography of mobile phone R&D networks,
University of Toulouse, 06/11-08/11

- Analyse der F&E-Netzwerke österreichischer Universitäten im 5., 6. und 7. Rahmenprogramm,
Austrian Federal Ministry for Science and Research, 11/10-06/11

- Analyse der F&E-Netzwerke österreichischer Universitätsinstitute im 7. Rahmenprogramm,
Austrian Federal Ministry for Science and Research, 11/10-06/11

- Vergleichende Untersuchung der Publikations- und F&E-Netzwerke österreichischer Universitätsinstitute,
Austrian Federal Ministry for Science and Research, 11/10-06/11

- Beteiligung, Positionierung und Vernetzung österr. Akteure im 7. RP,
Austrian Research Promotion Agency (FFG), 07/10-11/10

- The spatial and temporal evolution of R&D Networks,
Austrian Science Fund (FWF), 08/09-01/12

- Network Analysis Study on participations in Framework Programmes,
European Commission, JRC, 09/08-03/09

Project References

- NEMO - Network Models, Governance and R&D Collaboration Networks,
  European Commission, DG Research, 09/06-11/09

- CEERails - Integration zentral- und osteuropäischer Bahntechnologiestrategien,
  Austrian Federal Ministry for Transport, Innovation and Technology/Austrian Research Promotion Agency (FFG),11/07-01/09

- RailNet - Analyse und Konzeption von kollaborativen Netzwerken in der Bahnforschung,
  Austrian Federal Ministry for Transport, Innovation and Technology/Austrian Research Promotion Agency (FFG), 01/06-07/06

- Analyse der Forschungsfelder in den Rahmenprogrammen der Europäischen Union als Grundlage für die Identifizierung von Emerging Technologies,
  DaimlerChrysler AG, 02/06-06/06

- Systemforschung im Urbanen Raum - Teilprojekt Wissensnetzwerke,
  ARGE Systemforschung im urbanen Raum, 01/05-12/06