RISIS
Research infrastructure for research and innovation policy studies

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

The EUPRO database to measure EU Framework Program (FP) networks: An introduction

Thomas Scherngell
AIT Innovation Systems Department

Course
SOCIAL NETWORK ANALYSIS. INTRODUCTION TO METHODS AND APPLICATIONS TO THE EUPRO DATABASE

LUGANO, FEBRUARY 16-18 2016

Outline

- The focus on (funded) R&D networks
- What is EUPRO?
- Coverage of the dataset
- Data acquisition and processing
- EUPRO database structure and variables
- Quality and accuracy of the data
- Future developments
- EUPRO’s contribution to studying Europeanization
- Appendix: Main publications and references

The importance of R&D networks for innovation

- Wide agreement on the increasing importance of R&D networks
  – defined as a set of organisations inter-linked via joint Research and Development (R&D) activities –
  for successful innovation (see, e.g., Castells 2006), due to
  - *increasing complexity in innovation processes* (Pavitt 2005),
  - *converging technologies* (Granstand 1998),
  - *increasing market pressures*, and *rapidly changing patterns in global demand* (Fischer 2003)
- The integration of external knowledge – transmitted via networks – is an essential ingredient in the innovation processes of firms (Cowan 2004)
- Participation in R&D networks reduces the degree of uncertainty, and provides faster access to different kinds of knowledge (Kogut 1988)

A policy perspective

Having in mind that

i. successful innovation is essential for sustainable economic competitiveness (see, Romer 1990), and
ii. R&D networks are essential for successfully generating innovation

modern STI policies emphasize supporting linkages between innovating actors (for a discussion of major international examples, see Caloghirou et al. 2002)

- The cornerstone of EU policy instruments are the European Framework Programmes (FPs)

- supporting pre-competitive R&D projects, creating a pan-European network of actors performing joint R&D

What is EUPRO?

- EUPRO is a database comprising systematic and comprehensive 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,

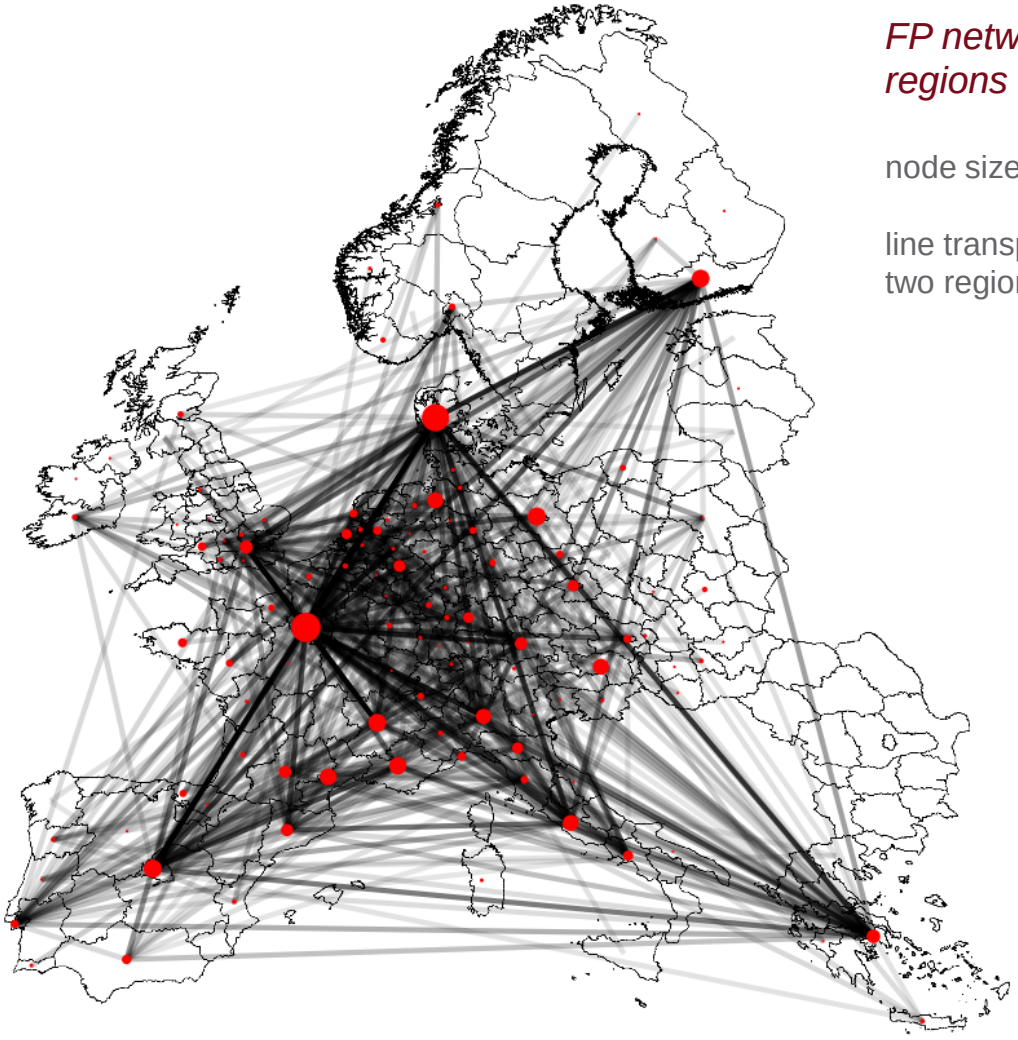
- including systematic information on
  - more than 60,000 collaborative research projects of (FP1-FP7), and
  - more than 60,000 participating organisations,

- Valuable information available, such as
  - the geographical location
  - the organisation type
  - the instrument/call
  - etc.

Coverage of FPs in the dataset (until July 2012)

Programmes	Period	Number of projects	Number of participations	Number of organisations
FP1	1984 - 1987	3,283	7,929	2,012
FP2	1987 - 1991	3,878	19,068	4,512
FP3	1990 - 1994	5,528	31,378	7,183
FP4	1994 - 1998	14,562	67,858	18,992
FP5	1998 - 2002	16,034	78,728	22,505
FP6	2002 - 2006	9,709	72,012	19,823
FP7 (July 2012)	2007 - 2013	10,866	60,135	15,551
Total (July 2012)		**63,860**	**337,108**	**62,506**

The European network of R&D cooperation 2013



FP networks 2013 across European NUTS-2 regions

node size corresponds to regional degree centrality

line transparency with the number of joint projects between two regions

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 II

- 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

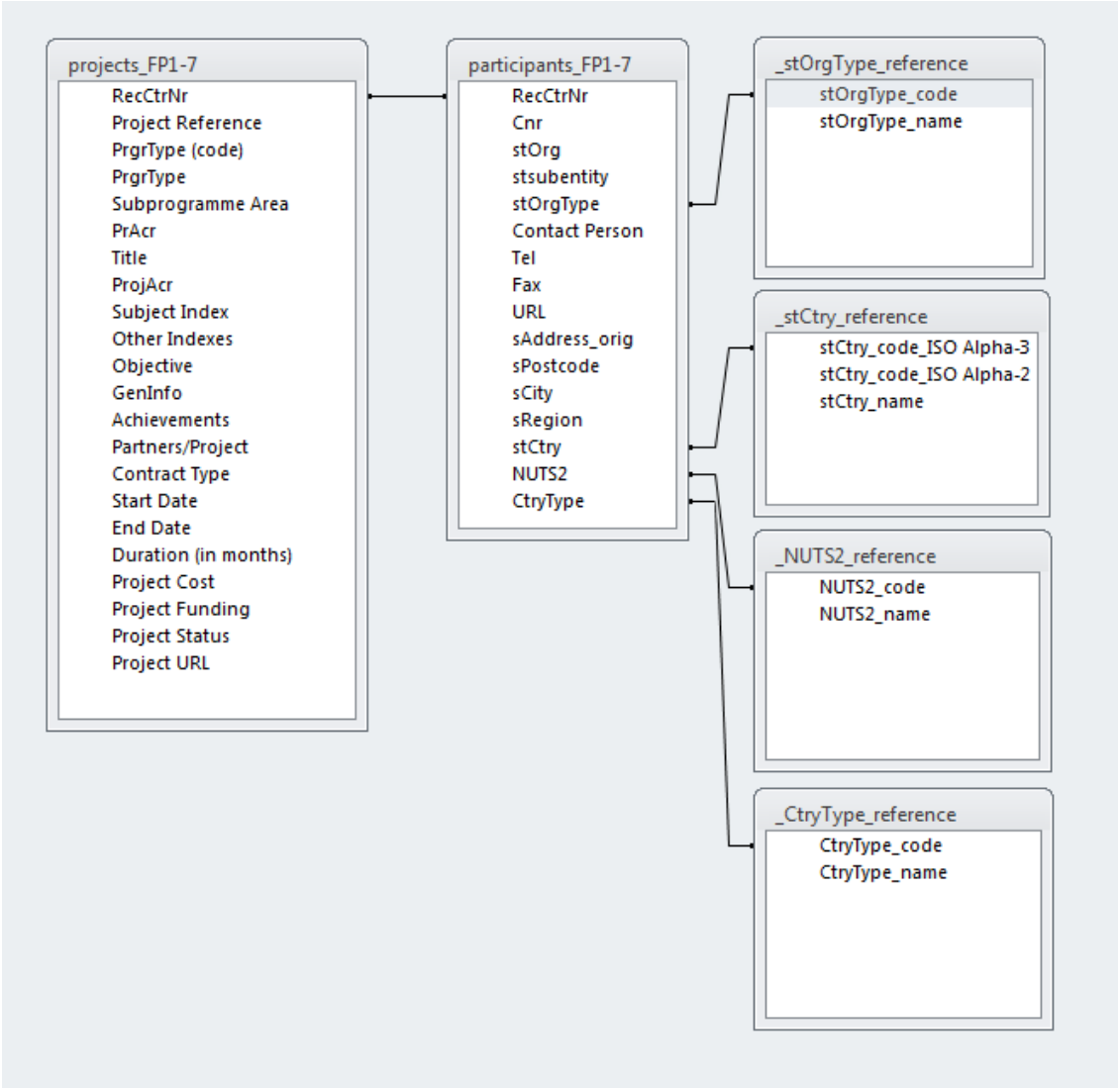
Data acquisition and processing

- The core data source for the construction of the EUPRO database is the CORDIS projects database. It contains detailed information on funded projects and project participants of FPs

- The quality of the raw data extracted from the CORDIS projects database is not generally sufficient for policy-relevant analyses

- Data cleaning and standardisation includes three major steps:
  - Harmonisation of organisation names (heterogeneous spellings, different languages): usage of automated matching procedures, followed by manual cleaning of recommended matches
  - Harmonisation and assignment of organisation types: not available in CORDIS anymore; organisation types extended, cleaned and harmonised
  - Regionalisation of geographical data to NUTS2 level: development and usage of city-2-NUTS table

EUPRO database structure (v7.4.11)



EUPRO project variables

Variable	Description
RecCtrNr	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	(not-unique) project index, for internal use in the European Commission (matches with Project Id in CORDA)
PrgrType (code)	code (1-7) for the names of the specific framework programme types FP1 to FP7
PrgrType	contains the corresponding names of the framework programmes (e.g., Seventh Framework Programme)
Subprogramme Area	full name of the subprogramme areas in each of the framework programmes
PrAcr	subprogramme acronym in the framework programmes (e.g., FP7-Health)
Title	Full title of the project
ProjAcr	(non-unique) project acronym or abbreviation of the project title
Subject Index	one or more of 68 standardized keywords allocated by the European Commission; caution: allocation of subject indices seems sometimes arbitrary – check reliability of contents of this variable before usage
Other Indexes	additional keywords to characterize project contents, freely chosen by the project team
Objective, GenInfo Achievements	distinction of these fields not clearly defined, in practice any of these fields may contain objectives or an abstract of the project.
Partners/Project	number of participants in each project; additional generated information, calculated with the help of table “participants_FP1-7”
Contract type	different types of contracts which regulate size, financing and funding of the research projects (e.g., STREP - Specific Targeted Research Project)
Start Date, End Date	day, month and year of project start and end
Duration (in months)	duration of the project in months (data provided by cordis)
Project Cost	official project costs
Project Funding	financing contribution of the European Union; since not all projects are financed completely, figures in “Project Funding” are equal to or smaller than figures in “Project Cost”.
Project Status	current status of the project (accepted, completed, execution) given in the CORDIS projects database in July, 2012
Project URL	official website of the project

EUPRO participants variables

Variable	Description
RecCtrNr	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 projects table
Cnr	unique identifier (control number) assigned internally by AIT, 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.
stOrg	standardised organisation name; the EUPRO database currently covers a period of more than 30 years during which organisations have changed to mergers, acquisitions and divestitures. At the moment organisations are labeled by their most recent valid name.
stsubentity	standardised identifier for organisational subunits; incomplete raw data on the participating subunits for FP1 to FP6; with the introduction of the Participant Identification Code (PIC) in FP7 such raw data is no longer available
stOrgType	standardised organisation type
Contact Person	name of contact person, without name affixes like titles, etc.
Tel	phone number of contact person
Fax	fax number of contact person
URL	official website of the organisation
sAddress_orig	complete address information
sPostcode	extracted postal code
sCity	extracted city
sRegion	extracted European Regions
stCtry	standardised country codes; country abbreviations are given as ISO 3166-1 Alpha-3 country codes
NUTS2	region code according to the NUTS -2 classification scheme (revision 2010)
CtryType	type of FP association of the country in terms of eligibility for funding (FP7)

Access

- The EUPRO database is located at AIT (current version still in Microsoft Access 2007)

- EUPRO is open within the RISIS project which shall enable wider access to and use of EUPRO data, and facilitate collaborative research with AIT staff

- EUPRO can be accessed via RISIS (datasets.risis.eu), based on RISIS rules and local access conditions
  - Possibility funded for on-site visits in RISIS with support from local staff in data preparation
  - Possibility for joint publication activities

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.

Future developments

- Organisation table (supports linking with other databases)
- Project table (concordance tables classifying programmes, contract types, project size and project funding)
- Demography of HEIs
- Update Coverage until December 2013

- Which kind of information is available for FP7/H2020 at CORDIS website?
  - Access to project data remains unchanged; additional information on project results available (report summaries, documents, publications)
  - Participants data: organisation name, EU contribution (also supplemented for FP7 participants), sparse information on the geographical location of organisations (only country names), organisation type (only available through Open Data portal)

➢ Potentially new release Spring 2016

Complementary dataset for the course: ETER

- ETER (European Tertiary Education Register) is a project promoted by the EC aiming to build a complete register of higher education institutions in Europe,
- providing data on the number of students, graduates, international doctorates, staff, fields of education, income and expenditure as well as descriptive information on their characteristics, and
- to develop a more complete set of indicators characterizing HEIs according to their main activities; to extend the coverage of the EUMIDA dataset to cover all European HEIs
- ETER currently includes 2673 HEIs across 36 countries; for 2254 HEIs data are available. The project is run in close cooperation with National Statistical Authorities, which provided most data included in ETER
- Data available for years 2011 and 2012 (2013 and 2014 to be published. 2008 data from EUMIDA available for researchers.

ETER: Substantive Content

- The dataset includes following main groups of variable
  - Institutional descriptors: the name of the institution, the foundation year, type of institution
  - Geographical descriptors: NUTS region, the city of the main seat and its postcode, geocoding , information on other campuses.
  - numbers of students and graduates divided by ISCED-2011 level, by gender, fields of education, nationality and mobility.
  - HEI expenditures (personnel, non personnel, capital) and revenues (core, third-party and fees).
  - the number of staff, divided between academic and non-academic, as well as on the number of professors
  - research activities: PhD students and graduates, R&D expenditures.
- Availability of data is excellent for descriptors and geographical information, very good for students and graduates (with a few breakdowns missing), reasonable for staff data and limited for financial data.
- Additional data for research purposes: EU-FP participations, publications (Leiden Ranking).

ETER: data access

- Most ETER data can be downloaded from the ETER website and reused without limitations:
  - http://eter.joanneum.at/imdas-eter/
  - A new website which allows tailored extractions to be released in June 2016.

- Few data are restricted only for research purposes
  - Under a non-disclosure agreement

- Additional data concerning research (EU-FP, publications) available from the RISIS project
  - As well as options for site visits to work with datasets (ETER, EUPRO, etc.)
  - http://datasets.risis.eu/

Appendix

Quality and accuracy of data (project table)

Variable	Missing values	
	Count	Ratio
RecCtrNr	none	0%
Project Reference	none	0%
PrgrType (code)	none	0%
PrgrType	none	0%
Subprogramme Area	15,296	24%
PrAcr	none	0%
Title	1	0%
ProjAcr	24,271	38%
Subject Index	3	0%
Other Indexes	54,540	85%
Objective	20,592	32%
GenInfo	43,383	68%
Achievements	55,940	88%
Partners/Project	none	0%
Contract type	13,759	22%
Start Date	2,144	3%
End Date	2,955	5%
Duration (in months)	8,863	14%
Project Cost	27,422	43%
Project Funding	22,919	36%
Project Status	none	0%
Project URL	58,497	92%

Quality and accuracy of data (organisation table)

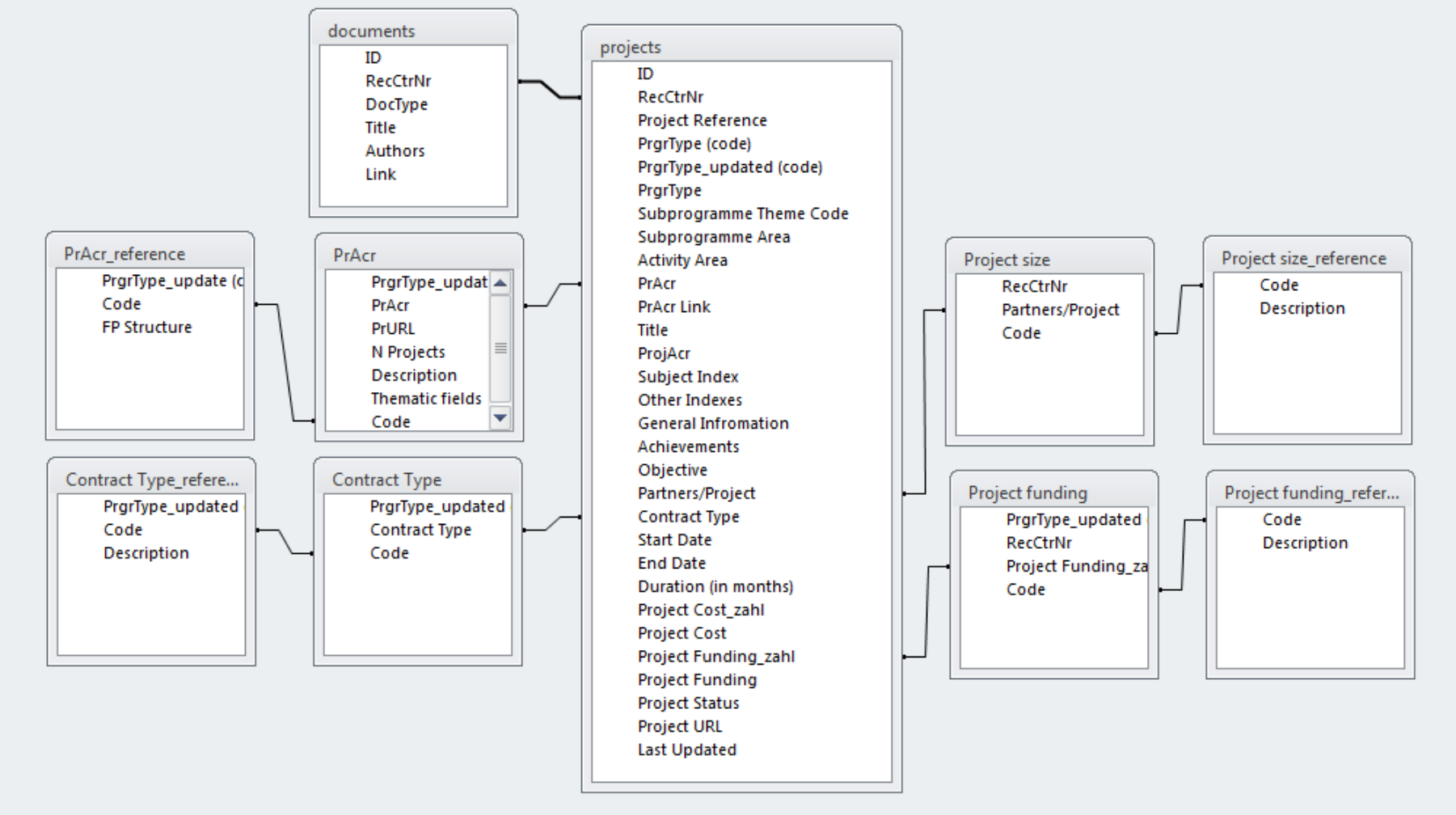
Variable	Missing values	
	Count	Ratio
RecCtrNr	0	0%
Cnr	0	0%
stOrg	3,680	1%
stsubentity	287,245	85%
stOrgType	4,515	1%
Contact Person	68,423	20%
Tel	133,451	40%
Fax	142,477	42%
URL	244,810	73%
sAddress_orig	1,227	0%
sPostcode	129,267	38%
sCity	14,216	4%
sRegion	54,495 (34,655 in EU 28)	16% 11%
stCtry	1,062	0%
NUTS2	13,679	4%
CtryType	0	0%

Training

- Due to the size and complexity of EUPRO, local assistance and support in close collaboration with AIT staff is foreseen

- A formal training course is offered via RISIS in Lugano, February 2016

Future database structure: Projects Table



Publications

List of publications using EUPRO (2005-2013)

Wissensbilanz	2005-2013
Books	2
Book sections	13
Publications in refereed journals	12
Contributions in proceedings of international conferences	32
(Final) Reports of contract reserach projects	26
(Invited) Talks	30
Other scientic publications	13
Diploma theses	6
Master theses	2 (+1)*
Doctoral theses	3 (+2)**

Note:
* 1 master thesis ongoing
** 2 doctoral thesis ongoing

Selected publications in journals (SSCI)

Lata, R., Scherngell, T. and Brenner, T. (2015): 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

Wanzenböck, I., Scherngell, T. and Lata, R. (2015): Embeddedness of European regions in EU funded R&D networks: A spatial econometric perspective, Regional Studies 49, 1685-1705

Lepori, B, Veglio, V., Heller-Schuh, B., Scherngell, T. and Barber, M. (2015): Participations to European Framework Programs of Higher Education Institutions and their association with organizational characteristics, Scientometrics [forthcoming]

Wanzenböck, I., Scherngell, T. and Lata, R. (2014): Embeddedness of European regions in EU funded R&D networks: A spatial econometric perspective, *Regional Studies* [forthcoming]

Wanzenböck, I., Scherngell, T. and Brenner, T. (2014): 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* [forthcoming]

Scherngell, T. and Lata, R. (2013): 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

Hoekman, J., Scherngell. T., Frenken, K. and Tijssen, R. (2013): Acquisition of European research funds and its effect on international scientific collaboration, *Journal of Economic Geography* 13, 23-52

Barber, M.J. and Scherngell, T. (2013): Is the European R&D network homogenous? Distinguishing relevant network communities using graph theoretic and spatial interaction modeling approaches, *Regional Studies* 47, 1283-1298

Scherngell, T. and Barber, M. (2011): Distinct spatial characteristics of industrial and public research collaborations: Evidence from the 5th EU Framework Programme, *The Annals of Regional Science* 46, 247-266

Barber, M.J., Fischer, M.M. and Scherngell, T. (2011): The community structure of R&D cooperation in Europe: evidence from a social network perspective, submitted to *Geographical Analysis* [forthcoming]

Paier, M. and Scherngell, T. (2011): Determinants of collaboration in European R&D networks: Empirical evidence from a discrete choice model, *Industry and Innovation* 18(1), 89-104

Scherngell, T. and Barber, M. (2009): Spatial interaction modelling of cross-region R&D collaborations. Empirical evidence from the 5th EU Framework Programme, *Papers in Regional Science* 88, 531-546

Project References I

- An analysis of the role and engagement of universities with regard to participation in the Framework Programmes,
  European Commission, 03/15-10/15
- Urban Research in the European Framework Programmes,
  JPI Urban Europe, 10/14-03/14
- 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

Project References II

- 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
- 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