RISIS

RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

RISIS

RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES

DOCUMENTATION OF RISIS DATASETS

EUPRO_FP

M. Barber, B. Heller-Schuh, Th. Scherngell, A. Wolfmayr, G. Zahradnik (AIT)

June 2024

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

Outline

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

1 Basic Characteristics

Name and short description of the infrastructure

The development of the EUPRO database has been started by AIT back in 2005 comprising information on R&D projects and all participating organizations funded by the European Framework Programmes (FP). Within RISIS, EUPRO has been extended significantly by including data of other European, and recently also national funding initiatives (https://zenodo.org/record/3934327#.Y3zSz32ZMZU).

Currently EUPRO consists of the following four modules:

- data on FP projects (starting between1984-2025), including Horizon Europe as the main update to the prae-RISIS version of EUPRO,
- data on R&D projects funded by EUREKA (1985-2019), an intergovernmental network supporting market-oriented R&D and innovation projects across all technological sectors,
- data on R&D projects funded through the Joint Technology Initiatives ARTEMIS, ENIAC and ECSEL (JTI) (2008-2014), long-term public-private partnerships, which support large-scale multinational research activities in areas of major interest to European industrial competitiveness and issues of high societal relevance,
- data on R&D actions funded by the European Cooperation in Science and Technology (COST) (1971-2014), a pan-European networking initiative supporting transnational cooperation among researchers, engineers and scholars across Europe, and
- data on nationally funded projects in the so-called NATPRO module of EUPRO (https://zenodo.org/record/4610742#.Y3zTC32ZMZU)

The recently published paper by Scherngell et al (2024) gives an overview on the full EUPRO database infrastructure:

> Scherngell, T., Barber, M., Zahradnik, G., Wolfmayr, A. and Bilalli Shkodra, Xh. (2024): EUPRO - A reference database on project-based R&D collaboration networks. Sci Data 11, 29. https://doi.org/10.1038/s41597-024-03129-y

Aim of the database

The EUPRO database is a significant asset of the Center for Innovation Systems & Policy of AIT used for basic oriented research projects and contract research for national and international customers, such as the European Commission. It

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

facilitates the analysis of participation patterns of organisations in and across different European funding initiatives and the investigation of collaborative network structures, including their evolution over time and the development of the European Research Area (ERA).

Legal name of operating organization

The database is operated and maintained by AIT Austrian Institute of Technology GmbH, located at Giefinggasse 4, 1210 Vienna, Austria.

2 Database content

The documentation for the EUPRO FP database (comprising all four European programme modules), last updated in 2021, can be found here:

> Heller-Schuh, Barbara, Barber, Michael, Bilalli Shkodra, Xheneta, Scherngell, Thomas, Zahradnik, Georg. (2021). Documentation of RISIS datasets: EUPRO (1.0). Zenodo. https://zenodo.org/record/5828177#.Y3zSeX2ZMZU

This documentation focuses on the current EUPRO_FP module (EUPRO_FP_2024) comprising all project data for FP7, H2020 and Horizon Europe, publicly available in May 2024. Moreover, note that NATPRO as new module of EUPRO will have its own documentation (currently under development: https://zenodo.org/record/4610742#.Y3zTC32ZMZU).

2.1 Definition and description of observations

Units and definition of observations

Basically, EUPRO_FP covers information on **projects** (such as project objectives and achievements, project costs, total funding, start and end date, contract type, information on the programme call), **participants** (name of the participating organisation, organisation type, geographical location, and funding) and **project output** (such as title, author/applicant name and DOI or access URLs of project reports, deliverables, publications and patents).

Number of observations

EUPRO_FP_2023 comprises information on ***72,449 projects*** and **392,819 *participations*** *for FP7,H2020 and Horizon Europe*. Note that information on former programmes (FP1-FP6) is still available upon request (https://zenodo.org/record/5828177#.Y3zSeX2ZMZU).Table 1 disaggregates the units of observation by different FPs and other European funding initiatives.

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

Table 1: EUPRO database - number of projects and participations

Programme	Period	Projects	Participations
FP7	2007 - 2013	25,785	140,064
H2020	2014 - 2020	35,386	178,677
Horizon Europe	2021 - 2027	11,278	74,078
Total	**2007 - 2027**	**72,449**	**392,819**

2.2 Data acquisition and processing (e.g. data cleaning)

Where the data are retrieved from

The data on FP projects for FP7, H2020 and Horizon Europe are publicly available and can be downloaded as excel files from the CORDIS projects database (http://cordis.europa.eu/). Basic raw data on FP7, H2020 and Horizon Europe projects, participants, and project outcomes of the current version (EUPRO_FP_2024) were downloaded in May 2024 in xlsx-format.

How the data are processed in terms of data cleaning

The quality of the raw data extracted from the different programmes websites is not generally sufficient for policy-relevant analyses. AIT has undertaken substantial efforts to improve quality of the data and to retrieve and add missing data. Data quality of data was improved by extending, cleaning and harmonizing the type of participating organizations. Furthermore, organisation names have been linked to organisations listed in OrgReg, the Register of European Public Research and Higher Education Actors, and FirmReg, the list of businesses included in the firms' datasets which are part of RISIS[1].

Data cleaning and harmonisation includes three major steps:

- harmonization of funding scheme
- linking of organisation names to OrgReg and FirmReg entities, and
- regionalisation (i.e., assignment to European NUTS3 regions).

The funding scheme names were harmonised across the different Framework Programmes by aligning different spellings and excluding the Framework Programmes name from the funding scheme name (e.g. HORIZON-IA was changed to IA).

[1] for more information on OrgReg and FormReg, see Lepori, B. (2022). OrgReg Methodological manual. Zenodo. https://doi.org/10.5281/zenodo.6396703 and Camerani, R. and Guerini, M. (2022), FirmReg Handbook. Zenodo, https://doi.org/10.5281/zenodo.6424802.

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

The linking between EUPRO organisation names and the entities in OrgReg and FirmReg was conducted by applying specific matching algorithms developed by AIT. These algorithms use information on the location of the organisations and the organisation name. They are based on statistical properties such as the frequency of adjacent characters in the organisation names and are used to identify similar organisation names that can be attributed to the same organisation. All algorithmically identified name matches are manually checked for accuracy.

For the data regionalisation, we build on the NUTS3 regions (2021 revision) assigned in the CORDIS data, filling in missing values using a multi-stage process. First, we found cases where the address shown for a participation without a region assigned matched the address for a participation with a region assigned, copying the region when found. Second, we used a mapping from postal codes to NUTS3 regions using Eurostats mapping. Where that was unsuccessful, we identified geographical locations from address information, converting latitude and longitude to NUTS3 regions. The regionalisation facilitates all kind of spatial analyses of project-based R&D networks, e.g., the investigation of the network at the level of functional urban areas.

2.3 Information on all variables/indicators

Programmes table

Table 2: Description of variables providing information about programmes

Variable	Description
prgId	internal unique identifier for FP subprogrammes[2]
prgType	code for the names of the specific framework programme types (FP7, H2020 and HEU) in which the subprogrammes were funded[2]
prgName	full name of subprogramme areas in each of the framework programmes (e.g., Specific Programme "Cooperation": Health)[3]
prgAcr	subprogramme acronym (e.g., FP7-HEALTH)[3]
prgShortName	short version of prgName[3]
objective	Conceptual orientation of the programme (only available for H2020 programmes)[3]

[2] introduced and/or processed by AIT

[3] provided by source data set, unchanged

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

Table 3: Description of variables providing information about the hierarchy of programmes (prgCodeClosure)[7]

Variable	Description
ancestorPrgID	Programme identifier of the superordinate programme, link to programmes table[2]
prgID	Programme identifier, link to programmes table[2]
depth	number of hierarchy levels between ancestorprgID and prgID[2]

Projects tables

Table 4: Description of variables providing information about projects

Variable	Description
projectId	unique identifier for each project in the database, identical with unique identifier of all projects in the CORDIS projects database[3]
title	full title of the project[3]
projAcr	(non-unique) project acronym or abbreviation of the project title[3]
startDate, endDate	day, month and year of project start and end[3]
totalProjectCosts	official project costs as indicated in the project proposal[3]
projectECFunding	factual financing contribution of the European Union; since not all projects are financed completely, figures in "projectECFunding" are equal to or smaller than figures in "totalProjectCosts".[3]
fundingSchemeId	funding scheme identifier (corresponding Funding Schemes see Table 6)[2]
masterCall	call identifier for H2020[4,3]
subCall	subcall identifier for FP7,H2020 and HEU[3]
objective	conceptual orientation of the project[3]
DOI	Digital object identifier for H2020 and HEU projects[3]
rcn	unique identifier (record control number) for each project in the database[3]
nature	thematic classification with respect to crises, i.e. crisis preparedness, crises recovery, crisis response[3]

Table 5: Description of variables providing information on webLinks of projects

Variable	Description
projectId	unique identifier for each project in the database, identical with unique identifier of all projects in the CORDIS projects database[3]
physUrl	Website related to project[3]
Status	Status of physUrl indicating physUrl as either invalid, webArchived or legacy[3]

[4] no MasterCall data for FP7 available

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

archivedDate	Date when physUrl was last archived[3]
Type	Type of physUrl, showing whether the URL is a related website, podcast, related news, related print, related story, related video or related social media[3]

Table 6: Description of variables providing information about funding_schemes

Variable	Description
fundingSchemeId	internal unique funding scheme identifier[2]
fundingSchemeCode	abbreviation of funding schemes[5]
fundingSchemeName	different types of contracts which regulate size, financing and funding of the research projects (e.g., RIA - research and innovation action, CP – collaborative project)[3]

Participations tables

Table 7: Description of variables providing information about participations

Variable	Description
participId	internal unique identifier of each participation[2]
projectId	link to projects table (Table 4)[3]
role	role of participant in the project; differentiates between "coordinator", "participant", "partner" (in MSC-Actions), "thirdParty" (commissioned by participant), "internationalPartner (outside Europe) and associatedPartner (participant without right to declare cost or apply for funding) [3]
orgId	link to organisations table (Table 8)[2]
acronym	Abbreviation of the organisation name[3]
endOfParticipation	indicates that organisations left the projects, but also changes of organisations' names or addresses during project run-time[3]
geoId	link to geography table (Table 10)[2]
address, postcode, city	street level address information[3]
ecContribution	amount of overall EC funding on participant level[3]
netEcContribution	amount of net EC funding for each participant and third party in H2020 and HEU; for FP7 programmes netEcContributions are the same as ECcontributions as netEcContributions are not available for FP7 projects.[3]
orgUrl	link to the organisation's website[3]
cordisOrganisationId	9-digit Participant Identification Code used for all FP participants in EU programmes[3]
vatNumber	Value Added Tax Registration Number; unique number that identifies a taxable person (business) or non-taxable legal entity that is registered for VAT[3]

[5] harmonized by AIT

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

nutsAssignmentId	Link to table defining strategy used to identify NUTS3 region for the participation (Table 11)[2]

Table 8: Description of variables providing information about organisations

Variable	Description
orgId	internal unique identifier for each organisation; corresponds with the entries in the field orgId in the participations table (Table 7)[2]
orgName	CORDIS organisation name[3]
orgTypeId	link to organisation type (Table 9)[2]
orgRegId	link to RISIS-OrgReg, the Register of European Public Research and Higher Education Actors[2]
firmRegId	link to RISIS-FirmReg, the Register of businesses included in RISIS firms' datasets[2]
ctryId	link to country[2]

Table 9: Description of variables providing information about the organisation type

Variable	Description
orgTypeId	internal unique identifier for organisation types; corresponds with the entries in the field orgTypeId in the organisations table[2]
orgType	CORDIS activity type[3]
orgTypeDescription	explanation of organisation types[3]

Geography tables

Table 10: Description of variables providing information about the geography of participations

Variable	Description
geoId	internal unique identifier for each geographical location of the participants[2]
ctryId	link to ctry table (Table 12)[2]
NUTS3	Regional EUROSTAT classification using the 2021 NUTS definitions[5]

Table 11: Description of variables providing information about the method used to assign NUTS3 regions to participations

Variable	Description
nutsAssignmentId	internal unique identifier for strategy used to assign a NUTS3 region to a participation[2]
description	Explanation of strategy used to assign a NUTS3 region[2]

Table 12: Description of variables providing information about the country of origin

Variable	Description
ctryId	internal unique identifier for each participating country[2]

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

ctryCode_eu	standardised country codes of the participating organisational units; used by the European Commission[6]; in the case of multinational organisations the participating national branches are listed[2]
ctryCode_iso	standardised country codes of the participating organisational units; abbreviations are given ISO 3166-1 Alpha-2 country codes[7]; in the case of multinational organisations the participating national branches are listed[2]
ctryName	official name of countries[2]
ctryTypeId	link to ctryTypes table[2]

Table 13: Description of variables providing information about the country type

Variable	Description
ctryTypeId	internal unique identifier for each participating country[2]
ctryType	differentiation between EU member states, associated countries and third countries (including subcategories) – abbreviations[2]
ctryTypeDesc	differentiation between EU member states, associated countries and third countries (including subcategories) – full text[2]

Project outputs tables

Table 14: Description of variables providing information about deliverables[8]

Variable	Description
cordisDelivId	Identifier for each project deliverable[3]
projectId	link to projects table (Table 4)[3]
title	deliverable title[3]
description	short description of the content[3]
deliverableType	Documents, reports; open data; Websites, patent fillings, videos etc.[3]
url	direct link to download the document (only for FP7)[3]
lastUpdateDate	date of last update[3]

Table 15: Description of variables providing information about publications[9]

Variable	Description
cordisPublId	Identifier for each project publication[3]
projectId	link to projects table (Table 4)[3]
title	publication title[3]

[6] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Country_codes

[7] https://www.iso.org/obp/ui/#search/code/

[8] data only for H2020 and HEU available

[9] data only for FP7 and H2020

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

authors, journalTitle, journalNumber, publishedYear, publishedPages, doi	bibliographic information in the case of open access publications[3]
url	direct link to download the document[3]
isPublishedAs	type of publication (Peer reviewed articles, Conference proceedings, Thesis dissertations, etc.)[3]
lastUpdateDate	date of last update[3]

Table 16: Description of variables providing information about patents[8]

Variable	Description
applicationId	Unique identifier for each project patent application[3]
projectId	link to projects table (Table 4)[3]
title	Patent title[3]

Table 17: Description of variables providing information about reports[8]

Variable	Description
cordisReportId	Unique identifier for each project report[3]
projectId	link to projects table (Table 4)[3]
title	report title[3]
relatedFile	link to related illustrations, images, announcements, etc.[3]
lastUpdateDate	date of last update[3]

Thematic classifications tables

Table 18: Description of variables describing the thematic orientation (FoS) of projects (fosTerms)

Variable	Description
fosCode	unique identifier of each Fields of Science (FoS) Term[3]
fosTerm	verbal representation of fosCode[3]
hierarchyCode	placement of fosCode in the FoS taxonomy[10,3]

Table 19: Description of variables providing information about the hierarchy of FoS (fosCodeClosure)

Variable	Description
ancestorFosCode	fosCode of the superordinate Fos term[3]
fosCode	link to fosTerms table
depth	number of hierarchy levels between ancestorFosCode and fosCode[3]

[10] for a representation of the EuroSciVoc Taxonomy see: https://op.europa.eu/en/web/eu-vocabularies/concept-scheme/-/resource?uri=http://data.europa.eu/8mn/euroscivoc/40c0f173-baa3-48a3-9fe6-d6e8fb366a00

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

Table 20: Description of variables providing information about project topics (topics)

Variable	Description
topicId	internal unique identifier for each topic[2]
topicCode	topic identifier within subprogrammes[3]
topicName	name of topic within subprogrammes[3]
objective	conceptual orientation of the topic (only for H2020 and HEU)[3]
keywords	Keywords related to topic(only for H2020 and HEU)[3n]
call	Call topic relates to(only for H2020 and HEU)[3n]

Table 21: Description of variables providing information about programme Gate-Knowmak classification (kmClasses)

Variable	Description
classId	Internal unique identifier for each Class[2]
persistentId	Gate-Knowmak unique identifier[3]
label	Class name[3]
description	Detailed description of the class[3]

Table 22: Description of variables providing information about the hierarchy of Gate-Knowmak classification

Variable	Description
ancestorClassId	ClassId of the superordinate class[2]
classId	link to kmClasses table[2]
depth	number of hierarchy levels between ancestorClassId and classId[2]

Table 23: Link table of Gate-Knowmak classification and projects

Variable	Description
projectId	ClassId of the superordinate class[2]
classId	link to kmClasses table[2]
Score	Evaluation of the quality of fit of the class to the project, based on the project title and objective[2]
knowmak	Score sufficiently high, relative to the median value of scores for that class, to be included in the Knowmak web tool (https://knowmak.eu)[2]

2.4 Sectorial, temporal and geographical coverage

Information on the sectorial classifications used

Table 24: Organisation type

OrgTypes	Description
HES	Higher or Secondary Education Establishments
REC	Research Organisations

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

PRC	Private for-profit entities (excluding Higher or Secondary Education Establishments)
PUB	Public bodies (excluding Research Organisations and Secondary or Higher Education Establishments)
OTH	Other

Information on the temporal coverage used

EUPRO module	First call year	Last call year
FP 2024	1984	2024

Information on the geographical coverage and classifications used

Since we have information on the geographical location of the project participants in the EUPRO database, we can analyse their geographical distribution across Europe at the country-level as well as on the regional level by assigning organisation to European NUTS regions[11] using NUTS classification revision 2021[12].

EUPRO covers participations from the following countries[13]:

- *EU 27 Member States*
- *European Free Trade Association (EFTA)*
- *United Kingdom*
- *EU candidate countries*
- *Potential candidates (all other Western Balkan countries)*
- *European Neighbourhood Policy (ENP)-East countries*
- *European Neighbourhood Policy (ENP)-South countries*
- *Other European countries*
- *Non-European countries*
- *Overseas Countries and Territories linked to the Member States (EU 27 OCT)*
- *Overseas Countries and Territories linked to UK (UK OCT)*

[11] including the analogous territorial descriptions for Switzerland and Norway

[12] History of NUTS, https://ec.europa.eu/eurostat/web/nuts/history(accessed: 07/08/2024)

[13] The classification is based on the typology of Eurostat: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Country_codes (last accessed: 07.08.2024) and extended by the categories EU 27 OCT and UK OCT. Countries not included in Eurostat's categorisation are added manually to the fitting categories..

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

2.5 Quality and accuracy of data

Information on the number of missing values[14]

Table 25: Number and ratio of missing values of FP project data

Variable	Missing values	
	Count	Ratio
projectId	0	0.0%
rcn	0	0.0%
title	0	0.0%
projAcr	0	0.0%
startDate	129	0.2%
endDate	129	0.2%
totalProjectCosts	5,411[15]	7.5%
projectECFunding	2	0.0%
fundingSchemeId	0	0.0%
masterCall	25,785[16]	35.6%
subCall	0	0.0%
objective	2	0.0%
DOI	25,785	35.6%

Table 26: Number and ratio of missing values of FP participation data

Variable	Missing values	
	Count	Ratio
participId	0	0.0%
projectId	0	0.0%
role	0	0.0%
orgId	0	0.0%
endofParticipation	2,426	0.6%
geoId	6	0.0%
address	1,617	0.4%
Postcode	3,813	1.0%
city	893	0.2%
ecContribution	27,357	7.0%
netEcContribution	8,369	2.1%

[14] This section covers information on the main variables of the two main tables in each of the data sets, projects and participations. Information on variables of the remaining tables can be provided upon request.

[15] Those which had total ProjectCost=0, except for one (projectId= 304096) which had totalProjectCost=0 and projectECfunding=0

[16] No MasterCall data available for FP7, FP7 data only includes subCall data

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

orgUrl	108,736	27.7%
acronym	80,924	20.6%
cordisOrganisationId	472	0.1%
vatNumber	72,963[17]	18.6%

Estimation of data quality issues with respect to data acquisition, reliability of retrieving system

In general, the CORDIS database constitutes a reliable resource on all R&D projects and participations. However, in some few cases information is incomplete or inconsistent, as for instance the address data, project costs, project funding and funding on the participants' level.

3 Technical Specifications

3.1 Information on the data base system

Current data base system used

The EUPRO database is realised as Microsoft Access 2016 database.

As part of the RCF Platform, EUPRO is one of the datasets that has been incorporated. For that purpose, an API – Application Programming Interface has been developed for the transfer of the data to the Platform.

Planned future technical changes concerning data base system

While having a MS Access version of the EUPRO database has its benefits, for server-side web programming, an appropriate relational database management system (RDBMS) and a database server needs to be used. Therefore, a MySQL version for the four modules of the EUPRO database will be provided as well.

[17] No DOI data available for FP7

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

3.2 Technical variable definition

Labelling and data type of all variables[18]

Table 27: Data type of variables providing information about projects

Variable	Data type
projectId	Number
rcn	Number
title	Text
projAcr	Text
startDate	Date
endDate	Date
totalProjectCosts	Number
projectECFunding	Number
fundingSchemeId	Number
MasterCall	Text
subCall	Text
objective	Long Text
DOI	Text

Table 28: Data type of variables providing information about participations

Variable	Data type
participId	Number
projectId	Number
role	Text
orgId	Number
endofParticipation	Text
geoId	Number
address	Text
postcode	Text
city	Text
ecContribution	Number
netEcContribution	Number
orgUrl	Text
acronym	Text
cordisOrganisationId	Text
vatNumber	Text
nutsAssignmentId	Number

[18] This section covers information on the variables of the two main tables in each of the data sets, projects and participations. Information on variables of the remaining tables can be provided upon request.

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

3.3 Description of the Entity Relationship Model

FP 2024 currently consists of 24 tables (Figure 1). The relation between *projects*, *programmes*, *participations*, and project outputs (*reports*, *deliverables*, *publications, patents*) and thematic classifications (*fosTerms*, *topics, kmClasses*) is realised via projectId, which is a unique identifier for each project in the database and identical to the unique identifier for the project in the CORDIS projects database. The thematic classification of projects is provided in the *fosTerms* table, linked by fosCode to the *projects* table. Also, information on FP *topics* are linked by topicId to the *projects* table.

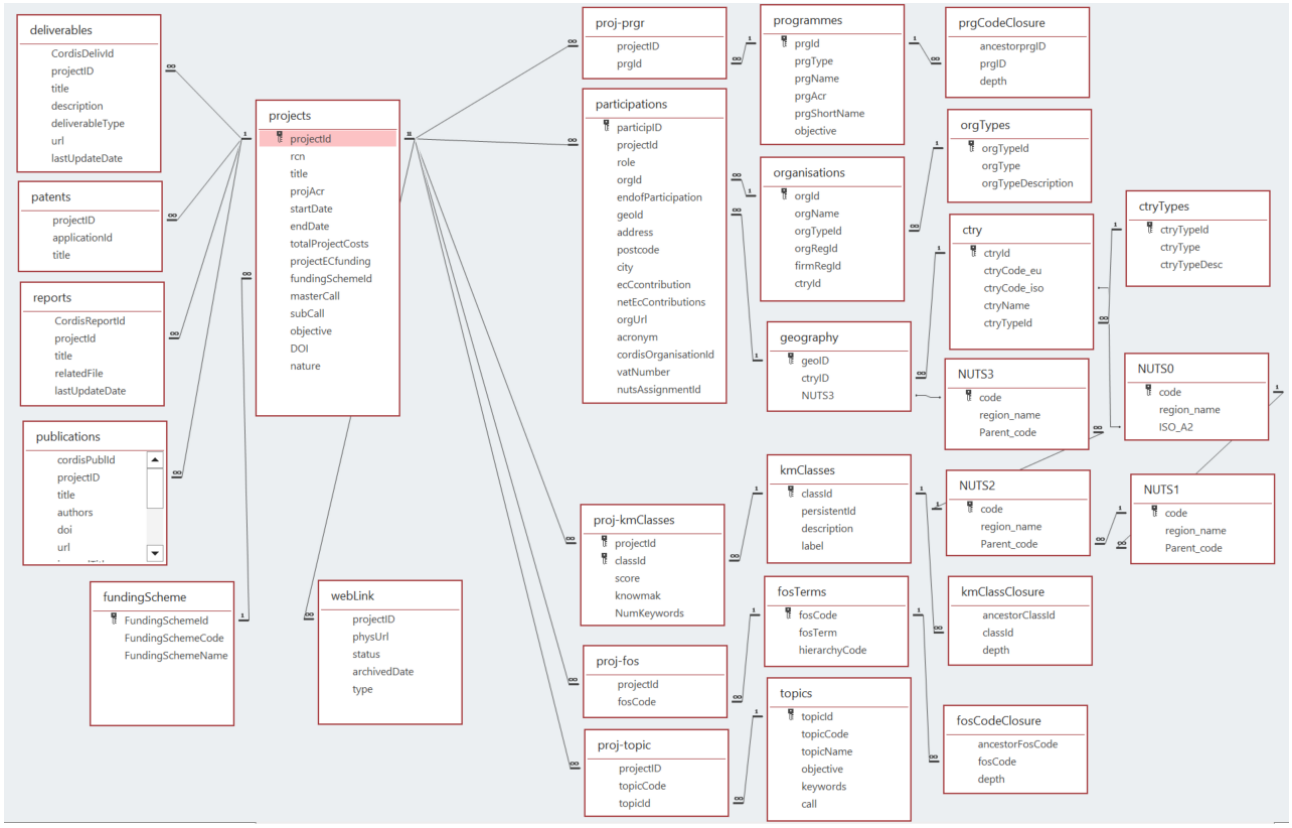
The *participations* table is linked by geoId to the *geography* table, which represents the NUTS3 region (where available and appropriate) and country. By the table *ctryTypes*, countries of origin are assigned to different country categories (EU27, associated countries and third countries).

Every organisation is linked by orgId to the *organisations* table, which provides the link to RISIS-OrgReg, the Register of European Public Research and Higher Education Actors, to RISIS-FirmReg, the Register of businesses included in RISIS firms' datasets, and to the *orgTypes* table, which differentiates the organisations in higher and secondary education establishments, research organisations, public bodies, etc.

Data of the scientific output of each project is comprised in *reports*, *deliverables*, *publications, and patents*.

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

Figure 1: FP Entity Relationship Model



3.4 Interfaces for access and to other infrastructures

Technical information on interfaces with other infrastructures

Integration of EUPRO within RISIS has been core in order to increase the scientific value of EUPRO for cross-dataset empirical analyses, on the one hand, and to be able to gain from RISIS developed facilities, such as geolocalisation tools, for the further advancement of EUPRO, on the other hand. Inter-operability with other datasets is considered as a key element for the further establishment and sustainable attractivity of EUPRO for new research endeavours, in particular those relating to the investigation of impacts of publicly funded R&D networks.

The following integration activities has been conducted during RISIS II:

- The link to **RISIS-OrgReg** enables the direct inter-linking with other RISIS core datasets featuring public research organisation, in particular LeidenRank and RISIS-ETER. At the moment, organisations included in EUPRO_FP_2024 are linked to RISIS-OrgReg by orgRegId in the *organisations* table. Accordingly, companies included in the RISIS firms'datasets are linked to **RISIS-FirmReg** in the *organisations* table by firmregId.

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

- With the application of the new geolocalisation tools developed in RISIS to EUPRO, we have been able to analyse the spatio-temporal development of FP networks in much more flexible manner and integrate it in geographical terms with other RISIS datasets. One promising example in this context has been a combination of EUPRO with the **Nano S&T dataset**, investing Nanoscience and -technology networks in publications and patents (coming from Nano S&T dataset) with Nanoscience and –technology FP networks (coming from EUPRO) (see Villard et al. 2017).
- EUPRO has been inter-linked directly at the organisational level with **RISIS-ETER** in order to analyse the relationships between several characteristics of European Higher Education Institutions (HEIs) coming from RISIS-ETER, and their participation patterns in the EU FP coming from EUPRO. First research endeavors in this direction have been conducted within RISIS (see Lepori et al. 2015), further research projects in this direction are currently under way.
- **JoREP** has – in contrast to EUPRO – a programme perspective on joint transnational research programmes, i.e. it gathers systematic information on a series of transnational programmes from 2000 to 2014. In EUPRO we have established a direct link (programme identifier) of programmes available in EUPRO and also in JoREP. These are at the current stage EUREKA and JTIs, and enables the relational investigation of networks of performers in these programmes (coming from EUPRO), and programme characteristics (coming from JoREP).
- **RISIS-KNOWMAK** draws upon project data from the EUPRO FP module to compute and visual indicators of knowledge creation in the European Research Area (see Lepori et al. 2017). Projects are annotated[19] with topical classes relevant for Key Enabling Technologies (KET) and Societal Grand Challenges (SGC), include detailed subclasses. The link to RISIS-KNOWMAK has been further extended to include annotation of all projects based on their descriptions, including FP projects outside the scope of RISIS-KNOWMAK as well as EUREKA projects. With the full annotation, alternative indicator definitions may be used, and specific projects may be associated with KETs and SGCs.

Integration with RCF

The EUPRO version that is made available for access to researchers in RISIS is fully incorporated in RCF, under the condition of controlled access and that security of usage is given (i.e. access for selected users with a concrete research project to

[19] https://gate.ac.uk/projects/knowmak

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

the parts of the dataset needed for the research). Note that underlying cleaning and standardisation data (e.g. name variants) will not be made available via RCF. Linking to other datasets in the RCF will be realized via the RISIS registers (providing the respective identifiers to the registers in EUPRO). There is an API (Application Programming Interface) developed for the transfer of the EUPRO data to the RCF Platform.

4 Scientific use cases and main references

Summarizing the research activities of external researchers accessing EUPRO via RISIS, we can distinguish the following main research directions for which EUPRO has been mobilized:

- Observing and characterizing **structure and dynamics of knowledge creation** and networks, disaggregated across different topics and/or geographical spaces (e.g., for climate change, biodiversity, Nanoscience; done by researchers from University of Coimbra, Université Paris-Est Marne-la-Vallée)
- Observing **FP participation patterns and networking of firms** in specific industries (e.g., pharmaceutical and chemical industries, done by researcher from Université Paris-Est Marne-la-Vallée based on the inter-linking of EUPRO with CIB)
- Using EUPRO to quantify and model **impacts of publicly funded R&D networks** on knowledge creation and technological diversification (e.g., done by researchers from Utrecht University, Université Paris-Est Marne-la-Vallée)
- Analysis of **country-specific participation patterns in the FP**, with a special focus on topical orientation and main partners (e.g., done for Israel by researchers from the Samuel Neaman Institute), increasingly as compared to nationally funded projects recorded in the new NATPRO module of EUPRO collecting respective information at country level (see https://zenodo.org/record/6549168#.Y3zQ_H2ZMZU)
- Observing **R&D hot spots** and activities funded by FP projects, at an organizational level in specific topics and geographical locations across Europe (e.g., on marine biotechnology, done by researchers from the EMBRIC project)
- Using real-world network data to test **novel statistical models for dynamic network analysis**, and to test these models in concrete empirical applications (done by researchers from University of Lugano)
- Tracing and investigating **characteristics of universities in terms of FP funding**, e.g., relation of FP funding to other university characteristics, disciplinary background of funding, etc. (done by researchers from Vrije University Brussels)

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

Selected recent references to publications using EUPRO

Barber, M., Guffarth, D. and Scherngell, T. (2015): Determinants of spatial distribution of European aerospace: contrasting R&D and supply. 55th Congress of the European Regional Science Association, 26-29 August, Lisbon, Portugal, 2015

Ciffolilli, A., & Muscio, A. (2018). Industry 4.0: national and regional comparative advantages in key enabling technologies. European Planning Studies, 26(12), 2323-2343.

Freitas, F. and Carrozza, C. (2017): Tracing research and development impacts using geographic data and a FP7 dataset. 2017 Annual Conference of the EU-SPRI Forum, June 7-9, Vienna Austria

Heller-Schuh, B., Reale, E. Scherngell, T., Spinello, A., Varinetti, E., Zahradnik, G. and Zinilli, A. (2022): Insights into project-based R&D funding from RISIS datasets: Some evidence from EFIL and NATPRO, RISIS Policy Brief, Issue 10 / zenodo, DOI: 10.5281/zenodo.6549168

Heller-Schuh, B., Lepori, B. and Neuländtner, M. (2020): Mergers and acquisitions in the public research sector. Toward a comprehensive typology. ***Research Evaluation***, https://doi.org/10.1093/reseval/rvaa015

Heringa, P. W., Hessels, L. K., & van der Zouwen, M. (2016). The influence of proximity dimensions on international research collaboration: an analysis of European water projects. Industry and Innovation, 23(8), 753-772.

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

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*** 3, 2149-2178

Lepori, B. and Guerini, M. (2017): KNOWMAK Manual. KNOWMAK project deliverable D5.2. https://project.knowmak.eu/wp-content/uploads/2018/02/KNOWMAK-D5.2-KNOWMAK-Manual.pdf

Muscio, A., & Ciffolilli, A. (2018). Technological diversity in Europe: Empirical evidence from agri-food research projects. Regional Studies, 52(3), 374-387.

Neuländtner, M. and Scherngell, T. (2022) R&D networks and their effects on knowledge exploration versus knowledge exploitation: Evidence from a spatial econometric perspective. ***Industry and Innovation***, [forthcoming]

Neuländtner M. and Scherngell T. (2020) Geographical or Relational: What drives technology-specific R&D collaboration networks? ***Annals of Regional Science***, https://doi.org/10.1007/s00168-020-01002-5

Robinson, D., Schoen, A., Laurens, P. and Laredo, P. (2017): Developing global and local STI indicators for profiling the territorial embedding of marine biotechnology research centres. STI 2017, Paris [Fr], 6-8 September

Uhlbach, W.-H., Balland, P.-A. and Scherngell, T. (2017): The effects of the European Union Framework Programmes on the technological diversification of European regions. 2017 Annual Conference of the EU-SPRI Forum, June 7-9, Vienna Austria

Uhlbach, W.-H., Balland, P.-A. and Scherngell, T. (2017): R&D Policy and Technological Trajectories of Regions: Evidence from the EU Framework Programmes. STI 2017, Paris [Fr], 6-8 September

Uhlbach, W. H., Balland, P. A., and Scherngell, T. (2022) Public R&D funding and new regional specialisations: The contingent role of technological relatedness. ***Industry and Innovation***, [forthcoming]

This project is funded by the European Union under Horizon2020 Research and Innovation Programme Grant Agreement n°824091

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*** 53, 337-368

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

Wanzenböck, I. and Piribauer, P. (2017): R&D networks and regional knowledge production in Europe. Evidence from a space-time model. ***Papers in Regional Science*** [doi 10.1111/pirs.12236]

Wanzenböck, I., Scherngell, T. and Dünser, M. (2017): Impacts of EU funded R&D networks on the generation of Key Enabling Technologies: Empirical evidence from a regional perspective, EMAEE Conference 2017, Strasbourg [FR], 31 May 2017

Wanzenböck, I., Scherngell, T. and Dünser, M. (2017): RTI policy for co-creation activities and its effects on Key Enabling Technologies: Regional evidence from the 7th EU Framework Programme. 2017 Annual Conference of the EU-SPRI Forum, June 7-9, Vienna Austria

Wanzenböck, I., Neuländtner, M. and Scherngell, T. (2020): Impacts of EU funded R&D networks on the generation of Key Enabling Technologies: Empirical evidence from a regional perspective. ***Papers in Regional Science***, https://doi.org/10.1111/pirs.12473

Villard, L., Perruchas, F., Scherngell, T., Barber M. and Laredo, P. (2017): The role of European Programmes in the European distribution of knowledge, the case of nanoscience and technology. STI 2017, Paris [Fr], 6-8 September