RISIS

RESEARCH INFRASTRUCTURE FOR SCIENCE AND INNOVATION POLICY STUDIES

EUPRO
Basic Tutorial

T. Scherngell (AIT), M. Barber (AIT), G. Zahradnik (AIT), B. Heller-Schuh (AIT)

December 2020

Outline

1 Introduction

The EUPRO database comprises information on **R&D projects and all participating organizations** funded by different public R&D funding programmes, mainly the EU FP, but also including COST, EUREKA and JTIs. It has been a significant asset of the AIT Center for Innovation Systems & Policy Innovation, used for basic oriented research projects and contract research for national and international customers, such as the European Commission. EUPRO is maintained and regularly updated (annual additions of new information), and also constantly advanced by additional modules, most importantly the addition of national programmes next to European and transnational ones.

For the funding programmes part of EUPRO, it basically covers information on:

- **projects** (such as project objectives and achievements, project costs, total funding, start and end date, contract type, information on the call), and

- **participations** (standardized name of the participating organization, contact person with contact details, organisation type, and geographical location)

This basic tutorial will provide condensed user-oriented information on how to use EUPRO for different kinds of research questions. It highlights main fields of application, provides an overview on the main tables, and how to use them. It will conclude with some examples of basic data extractions and analyses. For all technical details of the database refer to the EUPRO technical documentation (https://rcf.risis2.eu/dataset/4/metadata).

2 Fields of application and coverage

The investigation of the structure and dynamics of R&D networks has attracted burst of attention in the recent past, not only in the scientific domain, but also in a policy context (see Scherngell 2013 for an overview). From a European policy perspective, research in this direction is great importance for providing a richer and evidence-based picture on – broadly speaking – the evolution of the European Research Area (ERA). Based on its political conception, ERA should in its essence be characterized by an efficient and smooth, European wide diffusion of knowledge, requiring the establishment of sustainable and geographically dispersed networking linkages (see, e.g., DeLanghe et al. 2009). In this context, the characterisation and modelling of cohesion vs. fragmentation processes in European R&D networks along different dimensions, e.g. geographical integration, technological integration or social and institutional integration, is often considered as particularly interesting.

One essential cornerstone of the research stream investigating dynamics of R&D networks is the focus on project-based, publicly funded networks (next to other types, as of for instance, co-publishing or co-patenting networks). In this context, EUPRO has become a backbone of empirical research in this research stream. It has been recently used intensively as a core facility in research activities that investigate the **structure, dynamics and impacts of project-based R&D collaboration**, in particular to grasp and understand the development of the European Research Area (ERA) (see, e.g., Barber and Scherngell 2013, Hoekman et al. 2013, Scherngell and Lata 2013, Wanzenböck et al. 2014,

Wanzenböck et al. 2015, Lata et al. 2015, Lepori et al. 2015, Wanzenböck and Piribauer 2016, Uhlbach et al. 2017, Villard et al. 2017). In a nutshell, these studies focus on the observation and modelling of integration processes in European R&D from different anchor points (e.g. geographical, technological, institutional, etc.), and the complex relationships of publicly funded R&D networks with (regional) knowledge creation and diffusion, as well as regional technological diversification tendencies.

Summarizing the fields of application for which EUPRO has been used (also substantial by users in RISIS), we can distinguish the following main directions:

- 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)

- Observing **FP participation patterns and networking of firms** in specific industries (e.g. pharmaceutical and chemical industries)

- Using EUPRO to quantify and model **impacts of publicly funded R&D networks** on knowledge creation and technological diversification

- Analysis of **country-specific participation patterns in the FP,** with a special focus on topical orientation and main partners

- Investigating **R&D processes at the organisational level,** e.g. concentration processes of organisations (observed by demographic changes), topical/spatial R&D hot spots, etc. (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

- 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.

Such questions and fields of applications have been addressed by mobilizing different parts of EUPRO database. Its perimeter in terms of the different funding programmes covered is given in Table 1; the different modules are illustrated by Figure 1.

Table 1: EUPRO perimeter (2020)

Programme	Period	Projects	Participations
FP1	1984 - 1987	3,348	7,972
FP2	1987 - 1991	3,987	19,184
FP3	1990 - 1994	5,474	30,768
FP4	1994 - 1998	14,524	67,831
FP5	1998 - 2002	16,026	78,562
FP6	2002 - 2006	10,100	75,356
FP7	2007 - 2013	25,778	139,682
H2020*	2014 - 2018	20,080	92,484
EUREKA	1985 - 2016	4,853	20,778
JTIs**	2008 - 2014	133	2,612
COST	1971 - 2014	1,132	35,543
Total	**1971 - 2018**	**105,435**	**570,772**

Note: *until December 2018, **Including ARTEMIS (calls 2009-2013), ENIAC (calls 2008-2013), ECSEL (2014)

Figure 2: EUPRO coverage und functionalities by spring 2020
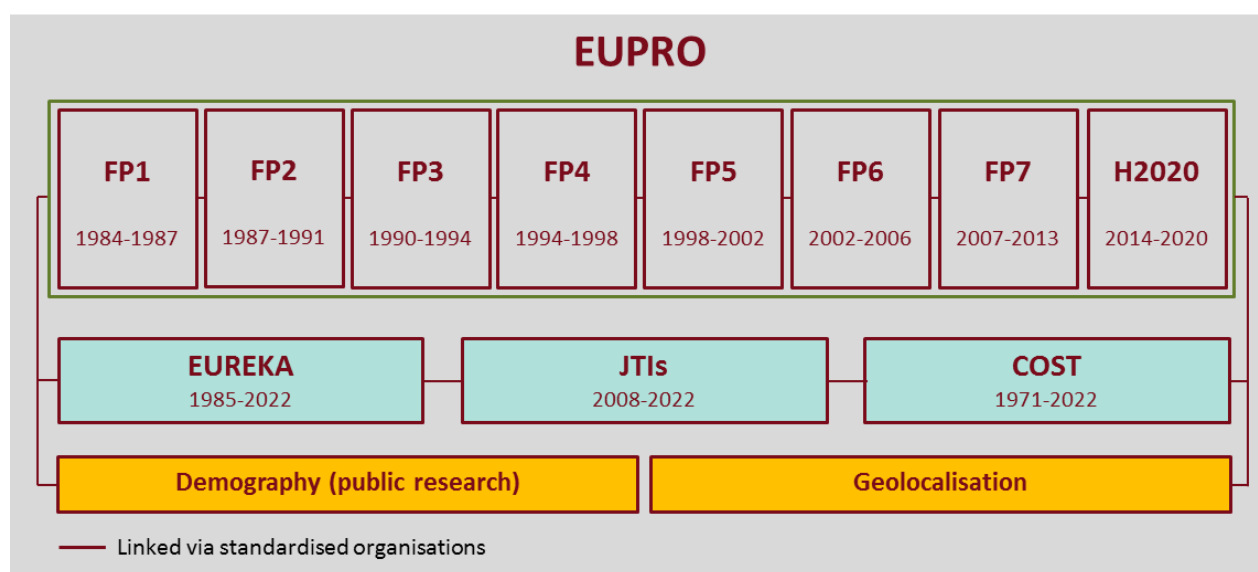
3 Basic User Guide

EUPRO is realized as Microsoft Access and MySQL database, with the Microsoft Access solution being the one offered to users for putting queries in the database, extract data and process them in any other data analytic software tools.

3.1 EUPRO tables

According to the conceptual set-up of EUPRO, we can distinguish three main tables subject to analysis of different kinds

- The **projects** table contains all funded R&D projects under the funding schemes as given in Figure 1. It contains a unique project id (Record Control Number, RCN) basic and relevant information about these funded projects, including the projects title, and the projects objectives (usually an abstract of more than 200 words). Further information usually comprises start- and end-date, duration, projects costs and funding, as well as information on the specific call and/or subprogrammes.

- The **participations** table is linked via the RCN field to the projects table. It lists all individual, i.e. organisational participations to a project. This table functions mainly as junction to the organisations table via an organisations id. Furthermore, it consists of participation specific information, such as the address of the participating organisation. This is central as one organisation can have multiple locations, so for e.g. tracing the geography of research networks we need to treat this at the level of participations rather than organisations.

- The **organisations** table is linked to the participations table via the organisation id and consists all relevant information about them, most importantly the standardised organisation name (after semi-automated cleaning activities), the category of the organisation (most importantly firm, university or research organisation), and its mean seat. Further it consists of an OrgReg id to directly link EUPRO to the RISIS organisation register, and by this, to other RISIS datasets.

This conception is illustrated by Figure 3 showing the basic scheme of EUPRO. It shows the projects, participations and organisations table as core, a programmes table and the classifications table related to projects, as well as the geography table related to participations:

- The **programmes** table comprises important information about the specific programme a project is funded by. It usually lists call information at different levels (e.g. specific calls per se, or bundles of calls such as sub-programme areas in the FP) that are often of interest for specific type of research questions.

- The **outputs** tables comprise information on project outputs usually in terms of published research in form of reports, deliverables but also publications. The table includes the document type, the title and a link to the documents.

- The **classification** table comprises the assignment of projects to specific transversal topical fields, such as Key Enabling Technologies and or Societal Grand Challeges (SGCs). This is for instance applied in the RISIS-KNOWMAK tool (knowmak.eu) providing indicators on projects in such topics at a regional level in Europe.
- The **geography** table covers the city level coordinates as well as the nomenclature of territorial units for statistics (NUTS) region of each participation.

Figure 3: Basic scheme of EUPRO (full ERM see full documentation)
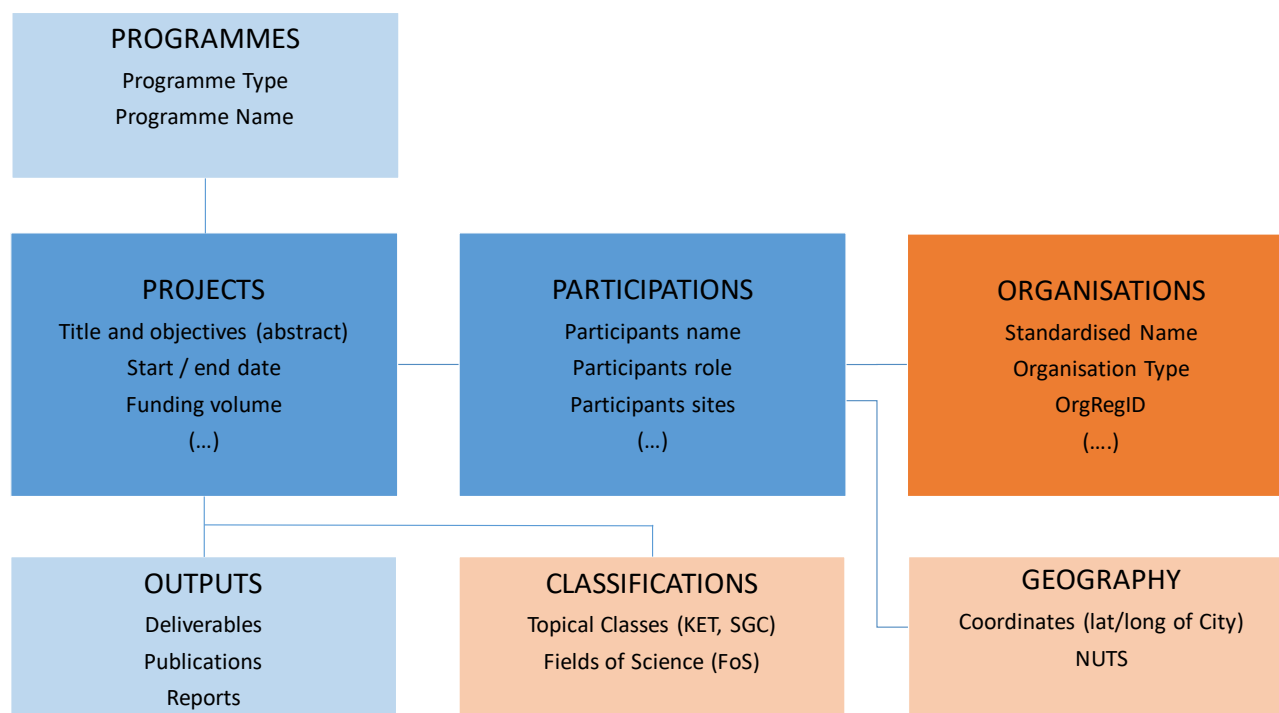
Next to these main tables, EUPRO comprises a number of auxiliary tables, mainly to keep the data more consistent in terms of quality checks and robustness, and to speed up complex queries.

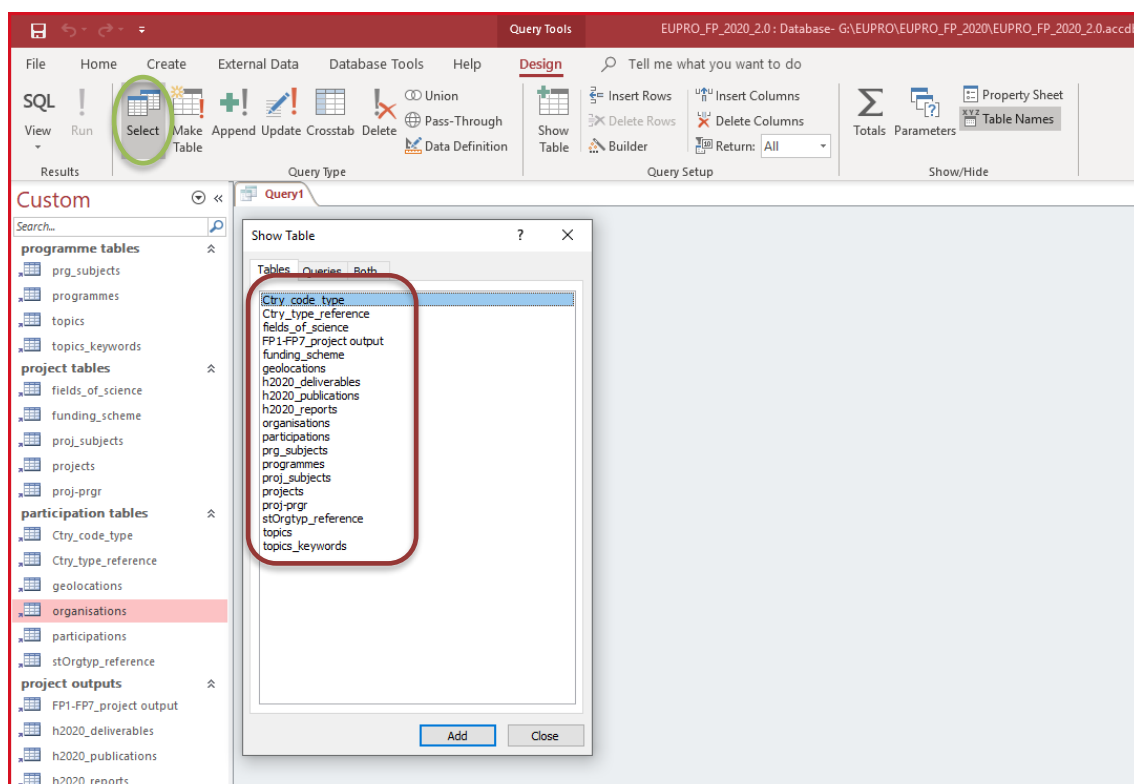### 3.2 Working with queries and selecting variables

For RISIS users, EUPRO is set up as mirror of the original MySQL version in MS Access. Accordingly, using EUPRO requires basic knowledge in MS Access and how to set up SQL queries there. Basis for queries are the EUPRO tables as described above that are accordingly appearing in the MS Access tables view under the **Home** tab:
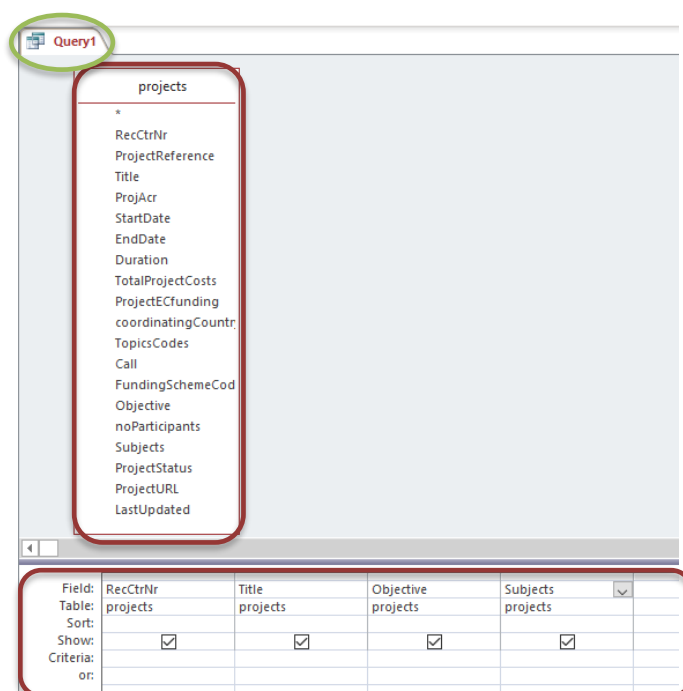
Figure 4: Screenshot basic table view
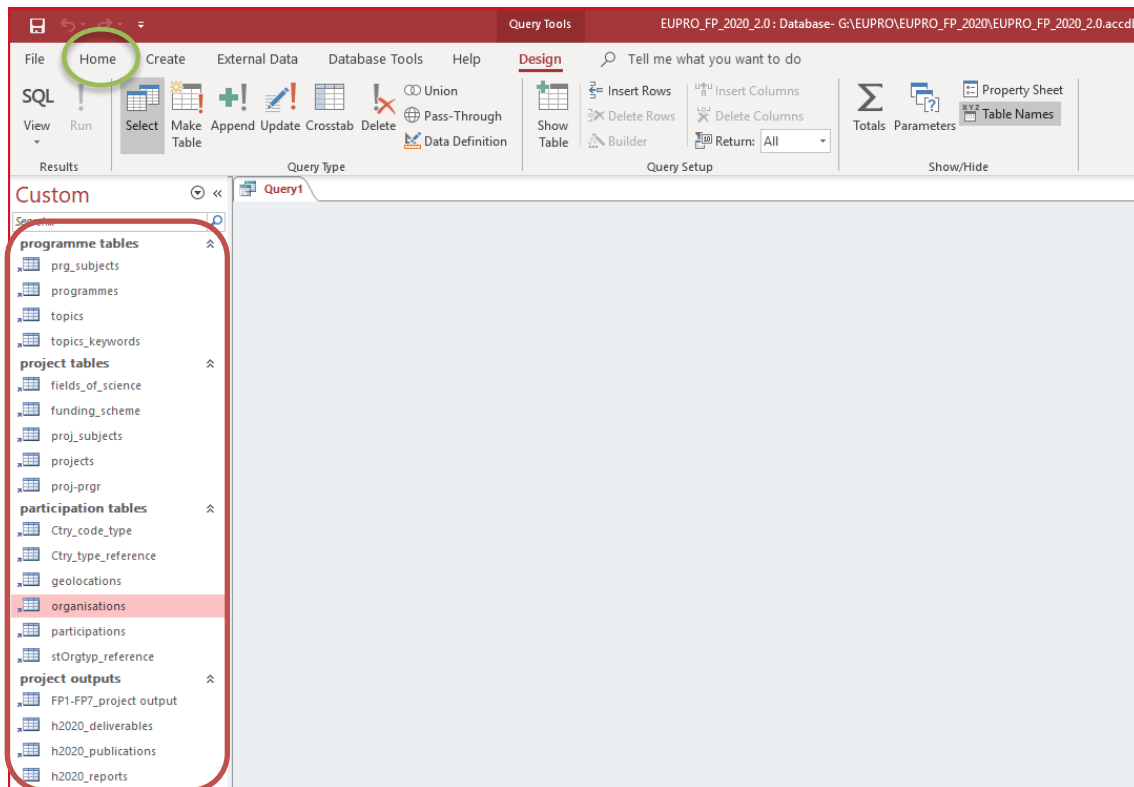
In order to place the query of interest, one needs to switch to the **Query view** where you can set up a query under **Create / Query Design**, and then **Select** the tables that are needed for your purpose and the question you address.

Figure 5: Screenshot Query Design View

![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

Figure 6: Screenshot for selecting variables in a query

After selecting your tables, they will appear with all variables (i.e. columns) the table consists in the **Query View** (default labelled as **Query 1**). In the example in Figure 6, the projects table has been selected, showing up all the variables, while 4 variables have been selected by clicking on them in the table. These variables are accordingly shown in the variable selection frame at the bottom of the query view.

4 Some basic examples

4.1 Select a sample of interesting projects: The example of Aerospace research

- **Problem / objective**: A user wants to study aerospace research activities in Europe funded by the FP.

- **Preparation**: Define the tables you need for the query, and define your search strategy. In this case, a list of aerospace related keywords is necessary to identify aerospace related projects, e.g. aerospace, aeronautics, aviation, turbine, etc.

- The set-up of a keyword list can be refined by an interactive exploration in the Datasheet view (which can be simply reached by clicking on a **table** left-hand). This is often useful to check/browse how keywords work and which projects are selected by a specific keyword.

Figure 7: Explore / Browse the Data Sheet View

Query: Select the table needed (projects table), and search for aerospace projects by putting the selected keyword with e.g. *Like*"*aero*" in the query in the bottom of the Query View. Note that the "*" stands for a search of "aero" within words, i.e. projects having aerospace or aeronautics will both be selected. Multiple keywords may be used by using the operator OR (e.g. Like "*aero* OR Like "*Aviation*").

- **Extraction**: Store the query and export or copy and paste the data view of your selection to other software tools for further analysis and visualisation (e.g. MS Excel)

4.2 Select a sample of participating organisations: The example of firms in Austria

- **Problem / objective**: Identify H2020 participations of firms located in Austria, and analyse their location
- **Preparation**: Define the tables needed that comprise the necessary information (participations table, organisations table)

Figure 9: Prepare tables for a query on a specific organization type in a specific geographical area

- **Query**: Select the participations related to i) an address in *Austria*, ii) programme *8* (=H2020), and ii) the organisation type *IND*; add the city as relevant addition to your selection

Figure 10: Specify the query limiting the selection to the research focus (firms in Austria, H2020)
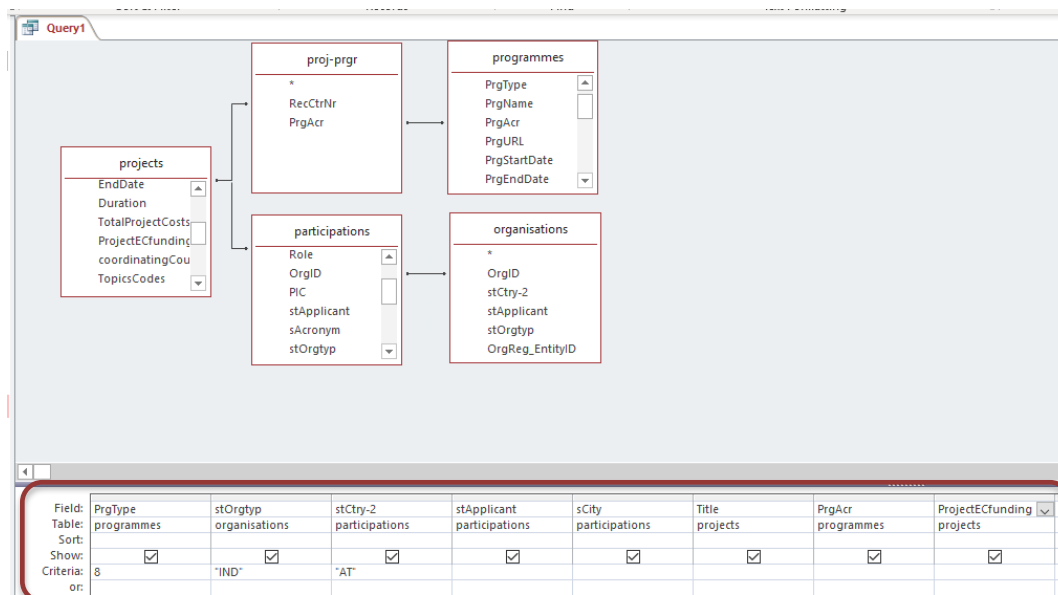
- **Extraction**: Store the query and export or copy and paste the data view of your selection to other software tools for further analysis and visualisation (e.g. MS Excel)

5 Published use cases

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

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., 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

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

Lepori, B., Heller-Schuh, B., Scherngell, T. and Barber, M. (2014): Understanding factors influencing participation in European programs of Higher Education Institutions. STI conference 2014, 3-5, September, Leiden, Netherlands

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

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