Report on the content and technical structure of the

JOREP
Infrastructure

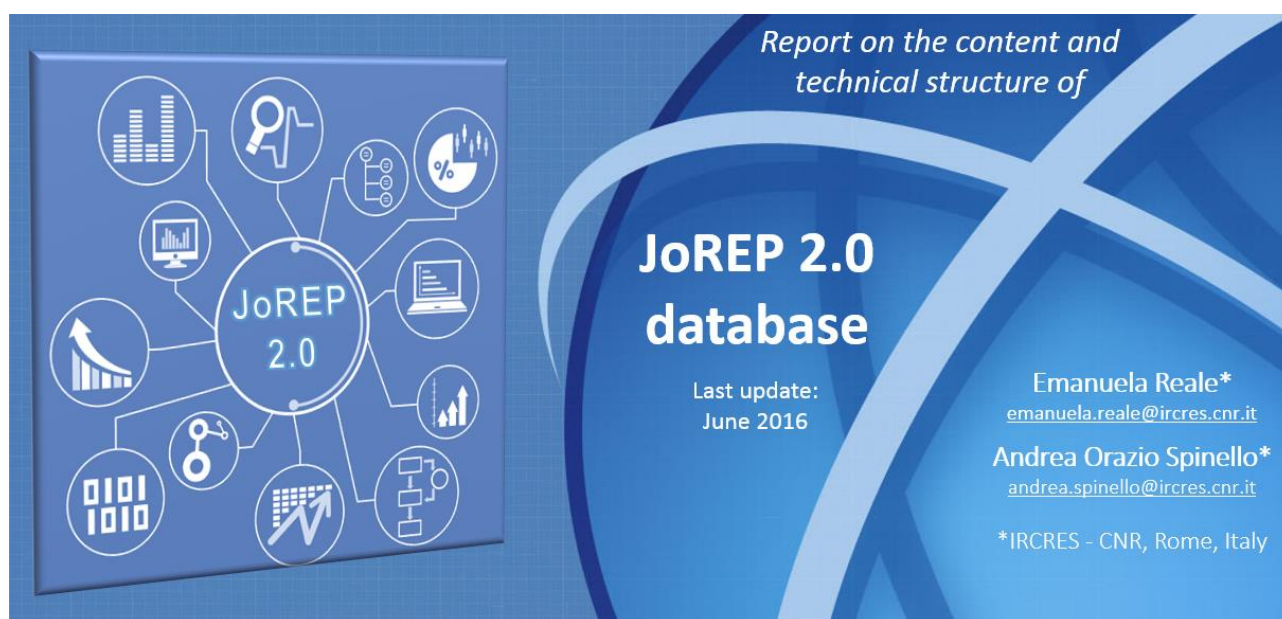RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

Report on the content and technical structure of

JoREP 2.0 database

Last update: June 2016

Emanuela Reale*
emanuela.reale@ircres.cnr.it

Andrea Orazio Spinello*
andrea.spinello@ircres.cnr.it

*IRCRES - CNR, Rome, Italy

Table of contents

1 Basic characteristics

1.1 Name and short description of the infrastructure

Name: JoREP 2.0 - Joint Research Programmes database
JoREP 2.0 is a database on joint R&D programmes, which are (i) European-level publicly funded research initiatives, in principle open to all countries belonging to European Research Area (ERA) either because they are established by the European Union or because they are based on international treaties, and (ii) publicly funded research programmes established by a group of countries through a bilateral/multilateral agreement. It provides a quantitative basis for the monitoring of investments in joint R&D programmes in the countries belonging to the ERA, pointing out the policy rationales behind them and their impact. The main focus is on national funding dedicated to the programmes, according to a clear typology of joint R&D programmes, so that cross-country comparisons would be possible. The set of data aims also at describing when, how and serving what purposes European-level initiatives and bilateral/multilateral joint R&D programmes are combined.

The JoREP 2.0 is a relational database implemented in MS Access, its key characteristics are:
- a standard set of descriptors covering the main alternatives concerning organizational features of joint R&D programmes;
- a group of 152 programmes in the sample; about 65% are European-level initiatives, while the others include bilateral/multilateral programmes;
- several data on the volume of funding channelled through these programmes:
  - European-level research programmes funding for the period 2000-2014;
  - bilateral and multilateral research programmes funding for the period 2000-2009;
  - flows to research performers from both types of programmes for the period 2000-2009.
- a large geographical coverage:
  - for the period 2010-2014, 32 countries covered (EU28 countries plus Israel, Norway, Switzerland, Turkey);
  - for the period 2000-2009, data are available for 11 countries selected in order to describe representative situations in the ERA landscape, which include medium-size countries with a well-developed science basis, large countries, Mediterranean countries and Central and Eastern European Member States (Czech Republic, Denmark, France, Germany, Italy, Netherlands, Norway, Poland, Spain, Switzerland and the United Kingdom).
- a reasonably good data coverage, only few descriptors being problematic because of limited availability.

1.2 Aim of the database

JoREP 2.0 allows understanding and interpreting the strategies adopted by the different actors involved in joint R&D programmes in terms of the underlying logics which frame their behaviour and decisions concerning if and how to establish joint R&D programmes, how to manage and implement it, linking the findings with broader conceptions of how European integration in the field of S&T policies should take place. JoREP 2.0 also supports, using network or spatial analyses methods, the analysis of important ERA dynamics and Europeanization processes through the study of the behaviour of main national actors (i.e. funding agencies).
The current JoREP 2.0 database structure has been designed in order to store raw panel data on joint R&D programmes and a basic set of descriptors of agencies participating to the programmes

and to better manage the volume of data to be collected in future updates as well as to adapt to potential end-user demands for temporal analysis.

1.3   Legal name of operating organization

The operating organization is the IRCRES CNR – Research Institute on Sustainable Economic Growth of the National Research Council of Italy (http://www.ircres.cnr.it)

1.4   Database location and type of access

The database is located at IRCRES CNR in Rome (Italy). The type of access to JoREP 2.0 is 'on site' (the user/researcher comes on site and works on a project using the data).

2   Information on substantive content of *JoREP 2.0*

2.1   Definition and description of observations

The conceptual scheme of JoREP 2.0 considers **joint R&D programmes** as the main unit of analysis, characterized by both internal and contextual features.
Joint R&D programmes are publicly funded research programmes for which at least one of them functions is shared between more than one country (or by regions belonging to more than one country). Observations focus primarily on programme attributes and examine in depth the dynamics of participation and fund allocation, describing the funding models adopted and the flows from funding agencies to performers.

JoREP 2.0 stores data on:

- **99 European-level joint R&D programmes[1]**
  - which launched a call for proposal in 2013 or 2014 and in which at least one of the EU28 countries or Israel, Norway, Switzerland, Turkey participates (data on funding are provided for the period 2010-2014);
  - which launched a call for proposal in 2008 and 2009 and at least one of the following countries – the Czech Republic, Denmark, France, Germany, Italy, the Netherlands, Norway, Poland, Switzerland, Spain, and the United Kingdom – participates (data on funding are provided for the period 2000-2009 and following years until the end of the programme; data on flows to performers are provided for the period 2000-2009).

- **53 Bilateral/multilateral joint R&D programmes** which launched a call for proposal in 2008 and 2009 in which at least one of the following countries – the Czech Republic, Denmark, France, Germany, Italy, the Netherlands, Norway, Poland, Switzerland, Spain, and the United Kingdom – participates (data on funding and on flows to performers are provided for the period 2000-2009).

---

[1] The counting refers to the unique ID codes associated to the programmes. The flexible characteristics of the programmes have imposed the creation of a set of rules in order to manage their demography, as outlined in *par. 4.2.1*. For instance, CIRCLE and CIRCLE 2 have a single ID code in the database, being one the continuation of the other. This applies to 24 programmes in the database, which are not counted twice or more times. Differently, the single actions deriving from disaggregation of large programmes have a proper ID code and are counted separately.

Programmes managed by the European Union (Framework Programmes) or by the EU together with a single member state as well as all programmes supporting research infrastructures and careers are excluded.

Secondary units of analysis in JoREP 2.0 are **funding agencies**, defined as formal organization managing at least one of the programme functions listed below:
- definition of the programme goals and mission;
- preparation and diffusion of the call;
- management of the submission process;
- evaluation and selection process;
- decision on which projects to fund;
- management of contract and payments.

JoREP 2.0 describes a basic set of characteristics of **348 national funding agencies** participating at least at one joint R&D programme included in the sample[2].

2.2 Data acquisition and processing

The first version of JoREP database (v.1.0) – containing data until 31.12.2009 – was based on a data collection performed within the JOREP EC contract[3], and developed by a team composed of national experts that also provided evidences for the data validation at their statistical offices.

In the context of EUFP7 RISIS project[4] WP22 on 'Enlargement of Databases dealing with ERA dynamics', an update of the data until 31.12.2014 – in parallel with an enlargement of the geographical coverage – was developed in order to include the most recent information on European-level joint R&D programmes, creating the current version of JoREP (v.2.0).

2.2.1 Data sources
As displayed in *tab. 1*, **combinations of sources** have been used for the descriptors of the same units of analysis.

Main sources for the data collection have been: ERA-LEARN 2020 (formerly NETWATCH) website; calls for proposal publicly available; joint R&D programme websites; joint R&D programme activity reports and evaluation reports; joint R&D programme leaflets; funding agency websites. Nevertheless, some information has been collected directly from ministries or funding agencies, e.g. through interviews or personnel contact.

Other approved sources have been: national statistical offices and GBAORD. Estimations for not-available data, if applicable, and assessment data have been produced by national experts.

---

[2] The counting does not consider special 'agency ID codes' used to avoid 'null' values in the *FundingAgenciesList* when the information on national funding agency is not available (see *par. 4.3.4*). The list of funding agencies will be harmonised according to the work developed for the register of public sector organizations (EUFP7 RISIS WP8, *see EUFP7 RISIS DOW*).

[3] JOREP (Investments in joint and open R&D programmes and analysis of their economic impact) was a service contract commissioned by the European Commission [Contract. No. RTD/DirC/C3/2010/SI2.561034] under the Seventh Framework Programme of the European Union for research, technological development and demonstration activities (2007 to 2013).

[4] EUFP7 RISIS project (Research infrastructure for research and innovation policy studies), http://risis.eu

UNIT OF ANALYSIS	SOURCES
Joint R&D Programmes	ERA-LEARN 2020 (formerly NETWATCH) website
	Calls for proposal publicly available
	Joint R&D programmes' websites
	Joint R&D activity reports
	Joint R&D evaluation reports
	Funding agencies websites
	GBOARD
	Direct contacts with agency personnel
	Estimations
Funding agencies	Funding agency websites
	Direct contacts with agency personnel

Table 1: sources for descriptors of the units of analysis of JoREP 2.0

2.2.2 Data cleaning

Data cleaning foresaw exploratory controls focused on the detection of *non-sampling errors*, whom correction required the recognition of systematic errors and random errors:

- a harmonization of the codes of the units of analysis has been accomplished;
- consistency checks between different descriptors have been undertaken in order to guarantee the coherency of data;
- errors committed in the transcriptions of data have been corrected through format checks;
- a check of referential integrity has been implemented as ultimate database safety check for inconsistent data and mechanism for the synchronization of the archives.

Treatment of empty fields - where values should occur – has been implemented so that in case of a complete non-availability of data a 'not available' code has been entered.

When there was a non-applicability of the field, a 'not applicable' code has been entered.

The presence of empty fields on 'amount' fields (data-type: numeric) indicates the non-availability or non-applicability of data (as remarked on 'data status' fields).

The presence of empty fields on 'remark' fields indicates the absence of remarks.

2.3   Information on all variables

JoREP 2.0 includes variables dealing with the two main units of analysis that characterize the joint R&D programmes dynamic: **programmes and funding agencies**.
Each joint R&D programme descriptor refers to a specific reference year in the programme life, whereas funding agencies descriptors are not changing.
Some of the descriptors are matched with 'remarks fields' included in the database as separated fields.

2.3.1 Joint R&D programme descriptors

The following list provides the main characteristics of the descriptors of joint R&D programmes at the *programme level*, at the *participation level* and at the *beneficiaries' level*.

- **Programme identifier** *(prog_ID)*. The code identifying the programme.

- **Programme start year** *(prog_start_year).* The year when the specific programme has been officially created, by signing a specific agreement. This might be earlier than the official launch of the funding scheme, as well as the start of funding to performers.

- **Programme end year** *(prog_end_year).* The year when the specific programme has been officially closed. Note: if the programme has only changed name, reconfirming most part of its features, it is considered still active. The end year refers only to the closing of a programme and its successors.

- **Original programme** *(original_prog).* If the programme has been generated by a merging, by a split or a spin-out, the indication of the original programme from which derives.

- **Successor programme** *(successor_prog).* If the programme has been merged, split or taken over, the indication of the successor programme.

- **Demographic transformation** *(demo_transformation).* Description of relevant demographic processes along the programme life (e.g. change of name, taking-over or spin-out).

- **Establishing authority** *(establishing_authority)* The body – EU, national states, regions, funding agencies – which officially established the programme, either by its own decision or by signing some kind of agreement. This includes political authorities, but also funding agencies when the decision is taken at this level. Example. The D-CH-A Lead agency agreement has been established through a direct agreement between the participating research councils and thus these will be considered as the establishing authorities (rather than the national States involved). A 'remark' field is provided for this descriptor.

- **Name of the programme in English** *(name_prog_eng).* If the main programme language is not English, the official translation if it exists. If there is no official name in English, the name in the national language is used.

- **Type of instrument** *(type_instrument).* Indication of the type of instrument used by the programme (e.g. ERA-NET, ERA-NET plus, JPI, etc.)

- **Programme duration** *(prog_duration).* This variable distinguishes between:
  - Programmes *limited* in time and with one or few calls.
  - *Periodic* programmes without a time limitation, but with irregular calls.
  - *Regular* programmes without a time limitation and regular calls (e.g. yearly or each two years).

- **Project duration** *(proj_duration).* This variable identifies the typical duration of projects funded by the programme, by using the following scale: less than 2 years, 2-4 years, more than 4 years. By typical, we mean that most of the projects are in this duration range. This information is derived from programme descriptions and calls. Exceptions and specific cases will be noted in the associated 'remarks' field.

- **Research topic** *(research_topic).* For classification of programme topics, the Nomenclature for the Analysis and Comparison of Scientific Programmes and Budgets from the Frascati Manual (2007 version) is adopted (*see 2.4.1*). This classification refers to the socio-economic objective of the programme, not to the actual research content. Please notice that category 12 is not applicable for programmes, while investigator-driven programmes should be classified under

category 13. More specific indications on subtopics can be inserted in the associated 'remarks' field.

- **Submission procedure** *(submission_procedure).* Following categories are used:
  - *Single entry point* when proposal are submitted to a single agency.
  - *Parallel submission* when proposal have to be submitted at the same time to two or more agencies (as in many bilateral programmes).

  By submission is meant delivering the whole proposal for the purposes of evaluation and selection. Sending copies for purposes of information is not considered as parallel submission. A 'remarks' field is provided for this descriptor.

- **Selection criteria.** This descriptor provides synthetic information on the importance of the two following criteria in the selection of projects:
  - Scientific quality *(selcri_scientific).*
  - Relevance to strategic or economic priorities *(selcri_relevance).*

  These criteria derives from a national experts assessment based on the information from programme descriptions and calls on the following point scale:

  > 4: is the most important criterion for project selection.
  > 3: it is an important criterion.
  > 2: it is an additional criterion.
  > 1: it is not a relevant criterion.

  Total number of points for the two criteria has to be 5.
  A 'remarks' field is provided for this descriptor.

- **Funding model** *(funding_model).* As joint R&D programmes do not necessarily involve cross border flows of funding and the joint call function can be separated from the funding function, different possible options concerning the management of financial flows in joint R&D programmes exist (ERAC, 2010; ERA LEARN 2020 tool box, 2015). This descriptor specifies how national funding for joint programmes is managed. We distinguish between the following models *Common pot* when all financial resources from participating countries are put in a single pot and used for financing the selected projects, independently of the country where research is performed.
  - *Common pot with return rules*, when some relationships are formally requested between national contributions and funding to national performers. The rule should be stated in some official documents (including statutes, policy briefs, and minutes). This model is applied for example by the European Space Agency.
  - *National pot* (virtual common pot) when financial resources for participating countries are managed separately and devoted to national performers. This model is applied, as an example, by ERA-NET programmes.
  - *Mixed-mode*, where the principle of national return is maintained and most of the resources are managed at national level, but there is a compensation mechanism to fund the best ranked proposals anyway, through top-up funding from national contributions.

  The selection of the funding model will be based on the most important model adopted, e.g. if 90% of national funding is used in national pot model this category should be used. A 'remarks' field is provided for this descriptor.

- **EU contribution** *(EU_contribution).* This descriptor identifies whether the programme is co-funded by the European Union (yes/no).

- **Coordinator country** *(coordinator_country).* ISO code of the country coordinating the programme.

- **Partner countries** *(partner_countries)* This field lists the ISO codes of the partner countries of the programme. This is a multivalue field, so that multiple values stored in this field can be treated independently in potential query tables. A 'remarks' field is provided for this descriptor.

- **Observer countries** *(observer_countries).* This field lists the ISO codes of the countries participating to the programme as observers. The field is filled only if applicable.

- **Mode of integration** *(integration_mode).* This descriptor identifies how the common programmes activities are institutionalized. We distinguish between three categories:
    - Creation of a specific agency, where joint activities are managed by a supranational agency with an enduring and long-term status (*agency*).
    - Management of joint activities by the way of no permanent structures as joint committees, whose existence is specifically related to the programme itself (*coordination*).
    - Management of joint activities through the delegation to a national agency in one of the participating countries (*delegation*).
    - Independent evaluation and selection, where the project is approved only if both parties decide independently to finance it (*independent selection*).
    A 'remarks' field is provided for this descriptor.

- **ERA category** *(ERA_category).* This descriptor provides a general categorization of joint programmes in terms of their relationships with the European research area. It is thus inserted in the database after data collection based on other descriptors. Following three categories are distinguished:
    - *European-level initiatives* are those joint programmes that are in principle open to all ERA countries either because they are established by the European Union or based on international treaties.
    - *Bilateral programmes within the ERA* are joint programmes established by a closed group of countries (not necessarily two) and which include only ERA countries.
    - *Bilateral programmes outside the ERA* are joint programmes established by a closed group of countries (not necessarily two) and which include also countries not belonging to the European Research Area.

    - **Programme type** *(prog_type).* This categorization has been introduced to distinguish the main organizational settings of joint programmes and is based on a set of other descriptions. Following categories are distinguished
        - *Integrated programmes* are those characterized by the existence of a supranational agency (integration mode: agency); they are further divided into integrated programmes with integration of funding (funding model: common pot) and without integration of funding (funding model: national pot).
        - *Coordinated programmes* are those characterized by lighter coordination modes (integration mode: coordination or delegation) and by single-entry point submission. They are further divided into coordinated programmes with delegation (coordination mode: delegation), coordinated programmes with integration of funding (coordination mode: coordination; funding model: real pot) and coordinated programmes without integration of funding (coordination mode: coordination; funding model: national pot).

- *Collaborative programmes* are those characterized by independent selection (collaborative programmes, independent programmes) or those characterized by coordination and parallel submission (collaborative programmes, parallel programmes).

Participation level

- **Reference Year** *(reference_year).* The calendar year to which the amount refers.

- **Participating country** *(participating_country)* Indication of one country participating to a programme included in the JoREP 2.0 perimeter[5]. It corresponds with the national state (or EU) from where funding originates (in case of regional budgets the relevant country).

- **National role** *(national_role).* This descriptor identifies the situation of national participants in the programme, as well as the availability of funds:
  - *Full participation*, if research groups from the considered country can participate to all programme activities without restrictions; in case of programmes with national pot, this means also that full funding is available (e.g. for research purposes).
  - *Full participation with restricted funding*, if research groups from the considered country can participate to all programme activities without restrictions, but availability of funding is restricted to coordination and networking activities.
  - *Limited participation* if research group from the country can participate with limitations, e.g. as external partners or not taking a coordination role.
  - A 'remarks' field is provided for this descriptor.

- **Funding agency** *(funding_agency_ID).* The funding agency receiving the funding amount specifically for this programme. If for a programme there is more than one agency receiving a share of the budget, the amount for each agency should be entered separately. A 'remarks' filed is provided for this descriptor.

- **Agency functions.** For each agency, indication is requested if it manages at least one of following functions:
  - Definition of the programme goals and mission *(agencyfunction_mission);*
  - Preparation and diffusion of the call *(agencyfunction_call);*
  - Management of the submission process *(agencyfunction_submission);*
  - Evaluation and selection process (*agencyfunction_evaluation*);
  - Decision on which projects to fund *(agencyfunction_fundingdecision);*
  - Management of contract and payments *(agencyfunction_payments).*

  Three options are provided for each field: 'yes', 'no' and 'participation to collective decision' in case a joint committee of agencies representatives takes decisions.

  This list include only the agencies taking the lead for each of these functions; other bodies cooperating in these functions (e.g. providing advice or helping in peer-review processes) will be included in the remark section. In cases where at national or European level different functions are managed by different agencies there will be more than one participating agency per country. If there is a two stages submission procedure, answers refer to the final selection stage and provide information on other stages in the remarks section.

- **Origin of funding** *(origin_of_funding).* The exact origin of the budget, including the European Union budget, the national budgets and regional budgets. This variable provides the specific name of the funding source (e.g. French research ministry).

---

[5] 'Perimeter' means the set of programmes that comply with the definition chosen for joint R&D programmes.

- **Budget source category** (*budget_source_category*). The classification of the budget source by distinguishing between national budget and regional budget.

- **Budgeting** *(budgeting).* The type of budgeting conforming to the following categories:
  - *Specific budget line*, when funding to joint programmes is explicitly part of a specific budgetary line which can be identified in GBAORD data. In these cases, the level of funds transferred to the funding agency should be derived directly from GBAORD data and comply with the EUROSTAT pilot data on transnational funded programmes.
  - *Earmarked budget*, when there is no specific budgeting line, but it is specified (for example in political decision or strategic documents) that part of the general transfer to a funding agency should be used for the participation to specific programmes. In these cases, data will have to be estimated from these policy documents (possibly by averaging over different years).
  - *Delegated budget*, when there might be a general political decision to participate, but the decision on the level of funding is completely delegated to the agency. In this case, the volume of funds will have to be derived from data of the funding agency and should match the one in the funding to beneficiaries table.

- **Currency** *(currency)* The currency in which the data are expressed. Standard ISO codes is used.

- **Amount** *(amount)*. The funding for the whole year expressed in currency units.

- **Budget data source category** *(data_source_category_participation)*. This variable provides a categorization of data sources as follows:
  - GBAORD or other budgetary data.
  - Information from funding agency.
  - Other source (specified in the following variable).
  - National expert estimate (will be clarified under remarks).

- **Budget data source** *(data_source_participation)*. The exact source of the budget data provided.

- **Potential beneficiary sectors** *(beneficiary_sector_GOV / beneficiary_sector_HE / beneficiary_sector_PNP / beneficiary_sector_PRIVATE / beneficiary_sector_ABROAD)*. This descriptor identifies the performing sectors that are legally entitled to get funding from the programme. We use the sectorial classification of the Frascati Manual (*see 2.4.1*), presented in the form of 5 different fields with binary 'data type'. A 'remarks' field is provided for this descriptor.

- **Beneficiary sector data source** *(beneficiary_sector_data_source)*. The exact source of the data provided on beneficiary sectors.

Beneficiaries' level

- **Reference Year** *(reference_year)*. The calendar year to which the amount refers.

- **Beneficiary country** *(beneficiary_country).* Country the funding amount refers to.

- **Funding agency** *(funding_agency_ID).* The code of the funding agency the funding originates from. If for a programme there is more than one funding agency, the corresponding amounts is entered separately.

- **Currency** *(currency).* The national currency used. Standard ISO currency codes is used.

- **Amount (Public/Private/Total)** *(amount_public/amount_private/amount_total).* For each year, the volume of funding transferred to public and private beneficiaries and the total. The amount for the whole year is expressed in currency units.

- **Status of data on the beneficiaries** *(data_status_beneficiary).* The field indicates the availability/non-availability of data on the beneficiary.

- **Data source category** *(data_source_category_beneficiary).* This variable provides a categorization of data sources as follows:
  - GBAORD or other budgetary data.
  - Information from funding agency.
  - Other source (specified in the following variable).
  - National expert estimate (clarified under remarks).

- **Data source** *(data_source_beneficiary).* Please indicate the exact source from were data has been retrieved.

2.3.2 Funding agency descriptors
The following list provides the main characteristics of the descriptors of funding agencies.

- **National agency identifier** *(funding_agency_ID)* The code identifying the national agency in the format of XX-NN-Acronym of the agency, where XX is the ISO code for the country and NN is the number of funding agencies for the same country running from 01 as long as necessary.

- **Country** *(agency_country).* For national and regional agencies only, the country where the agency is established.

- **Acronym** *(agency_acronym).* The official acronym of the funding agency, if available.

- **Name of the agency in official language** *(agency_name_nat_lang).* The full name in the language of establishment. For international agencies, the official English name should be used.

- **Name of the agency in English language** *(agency_name_eng_lang).* Full name of the agency in English, e.g. the one adopted in policy documents or on the agency website (if available).

- **Status of the agency** *(agency_status)* We distinguish between following categories.
  - *National agency* established by a single country.
  - *European agency*, established through European law (e.g. European Interest Groups).
  - *Intergovernmental agency* established by an international treaty between national states (possibly including also the European Union). These can be at the international, European or (country) regional level (e.g. Nordic Council).
  - *International non-governmental association* e.g. established through an agreement between national or regional funding agencies.
  - *Regional agencies* established by a regional authority.

  We notice that this categorization refers to the authority of establishment and not to the geographical space where these agencies fund research.

- **Agency website** *(agency_website)* The official website of the agency, (if available the link to the English section). This information is inserted to quickly retrieve additional information for the purposes of analysis.

- **Total budget** *(agency_tot_budget_2009). (note: provisionally provided only for funding agencies stored in JoREP 1.0)*. The total budget of the agency for research project funding is provided for the year 2009 as a rough measure of the size of the programmes it manages. Data is provided in national currency at current prices. For agencies performing other functions than research funding (e.g. ministries) or funding their own research centres, this value should refer to project funding research only.

- **Geographical level** *(agency_geo_level)*. Agencies are distinguished between supranational, national and regional. This distinction refers to the institutional embedding, not to the funding activities; e.g. a regional agency, funded under regional law, might support also research outside the region.

- **Agency classification/agency domain** *(agency_classification/agency_domain)*. The classification of funding agencies is two-level, the first one refers to the position with respect to the State, while the second one specifies more precisely the domain of activity.
  At the first level distinguish between following categories:
  - ➢ *Governmental agencies* are agencies that are functionally part of the public administration, meaning for example division of ministries, ministerial committees, etc. Typical examples at the European level are DG research (managing the European FP), at national level research ministries. These are divided between:
    - National research/science ministry
    - National sectorial ministry (e.g. energy)
    - Regional government (undivided in subcategories)
  - ➢ *Independent agencies* are agencies that have functionally a large degree of independence from the State in managing their activities and selecting the projects to be funded; in some cases this might be realized by a specific legal status g ranting autonomy. A key criterion to distinguish the two types of agencies is if the State (e.g. ministry) retains the right to take the final decision on granting money to specific projects. These are divided between:
    - Innovation agency, whose mission and funding are oriented towards innovation and creation of economic value;
    - Research councils, whose funding is mainly oriented towards basic research and which have strong connection to the academic community (for example in the composition of decision-making committee);
    - Sectorial agency – related to specific topic (energy, environment, etc.), e.g. sectorial regulatory agencies or sectorial funding agencies;
    - Intergovernmental agency created by international treaty (ESA);
    - EU-implementation agency based on EU law (e.g. the agency managing AAL);
    - International non-governmental association (European Science Foundation).
  - ➢ *Performers* are organizations whose main mission is to perform R&D activities, even if might host some funding agencies activities. These are divided between:
    - Public research organizations (PRO) assuming also a function in funding;
    - Private research organizations

2.3.3 Countries' descriptors
The following list provides the main characteristics of the descriptors of countries.

- **Country name** (*country_name*). The list of world countries.

- **Country ISO code** (*country*). Country codes in ISO 3166-1 alpha 2 (*see 2.4.1*).

- **EC-ATC classification scheme** *(EC_ATC_classification)* The descriptor distinguishes the inclusion of the country between three categories: EU28 countries, countries associated to EU28 and third countries.

- **JoREP sample country** *(JoREP_country).* The descriptor is binary and points out the inclusion/non-inclusion of the country in the JoREP perimeter.

- **Latitude** *(latitude)* The latitude of the capital city of the country.

- **Longitude** *(longitude).* The latitude of the capital city of the country.

- **GERD (Gross domestic expenditure on R&D) as a percentage of GDP** *(gerdbygdp)* (period 2005-2014, source: EUROSTAT).

- **Graduates on population** *(gradbypop).* The ratio between total number of graduates and total population for each country considered (period 2005-2014, source: EUROSTAT).

- **Researchers on population** *(reserbypop).* The ratio between total number of researchers and total population for each country considered (period 2005-2014, source: EUROSTAT).

- **H-Index** *(h_index).* The h-index is an author-level metric that attempts to measure both the productivity and citation impact of the publications of a scholar. It accounts the number of articles produced in a country (h) that have received at least h citations (period 2010-2014, source SCOPUS).

- **Population** *(population).* The total population of the country in the considered year.

2.4 Sectorial, temporal and geographical coverage

2.4.1 Classifications used
JoREP 2.0 adopted the Nomenclature for the Analysis and Comparison of Scientific Programmes and Budgets from the Frascati Manual (OECD, 2002) in order to classify research topics, as showed in tab. 2.

01	Exploration and exploitation of the earth
02	Environment
03	Exploration and exploitation of space
04	Transport, telecommunication and other infrastructures
05	Energy
06	Industrial production and technology
07	Health
08	Agriculture
09	Education
10	Culture, recreation, religion and mass media
11	Political and social systems, structures and processes
12	General advancement of knowledge: R&D financed from General University Funds (GUF)

13	General advancement of knowledge: R&D financed from other sources than GUF
14	Defence

Table 2: classification of research topics used in JoREP 2.0

The performing sectors that are legally entitled to get funding from the programme have been classified using sectorial classification of the Frascati Manual (OECD, 2002), contained in *tab. 3*.

GOV	Government sector: Research institutes/governmental institutions with R&D which are mainly financed and controlled by the government.
HEI	Higher education institutions
PNP	Nonmarket, private non-profit institutions serving households/ the general public.
Private	Business enterprise sector: firms/organisations/institutions whose primary activity is the market production of goods or services, including the private non-profit institutions mainly serving the business enterprise sector.
Abroad	

Table 3: performing sector classification used in JoREP 2.0

Countries are classified according to the International Standard for country codes (ISO 3166-1 alpha 2)[6]. The EC-ATC classification scheme, which distinguishes the inclusion of the country between EU28, countries associated to EU28 and third countries is also included as a country descriptor.

2.4.2 Temporal coverage
Annual data refer to the 2000-2014 period:

- Data on funding and on internal features of European-level joint R&D programmes which launched a call for proposal in 2013 or 2014 and where at least one of the EU28 countries or Israel, Norway, Switzerland, Turkey participates are provided for the period 2010-2014;

- Data on funding and on flows to performers of European-level joint R&D programmes and bilateral/multilateral programmes which launched a call for proposal in 2008 and 2009 and at where least one of the following countries – the Czech Republic, Denmark, France, Germany, Italy, the Netherlands, Norway, Poland, Switzerland, Spain, and the United Kingdom – participates are provided for the period 2000-2009. Data on internal features refers to the year 2009.

Descriptors refer to the last day of reference year. Financial data refer to the calendar year of each year.

2.4.3 Geographical coverage
The geographical coverage of JoREP 2.0 includes 32 countries (EU28 countries plus Israel, Norway, Switzerland, Turkey) for the period 2010-2014; 11 countries (Czech Republic,

---

[6] http://www.iso.org/iso/country_codes.htm

Denmark, France, Germany, Italy, Netherlands, Norway, Poland, Spain, Switzerland and the United Kingdom) are also covered for the period 2000-2009.

2.5 Quality and accuracy of data

In the current version of the database, descriptors of joint R&D programmes at programme level and descriptors of funding agencies (except for funding agencies total budget), revealed a very good coverage of the basic characteristics of the units of analysis.
Nonetheless, these are the features that characterize the quality of participation and budget data of the two data collections (see 2.2):

- Data on European-level initiatives and bilateral/multilateral programmes for the period 2000-2009 (11 countries covered), despite the completeness of the data is rather good in general, have a few descriptors that revealed problematic in terms of availability during the data collection, such as programme budgets and flows to beneficiaries. The share of non-available data is higher for what concerns programme budgets - 18% of missing data - and funding to beneficiaries - 22% of missing data.

- Data on European-level initiatives for the period 2010-2014 (32 countries covered), at participation level, for beneficiary sectors are provided for the programme in general and not punctually for each country. Funding agencies functions are provisionally not available for art.185/JTI/COST/EUREKA/ESA programmes. Data on origin of funding will be added in the next releases of the database.

In terms of data comparability, in some cases, there might have been slight differences in their application across countries, which do not affect comparability. Some quality issues concern the thematic classification – as it is not always easy to fit the different programmes into the classification categories – and the functions of funding agencies.

3   Legal issues encountered and access conditions

3.1 Legal issues concerning access of the database

IRCRES CNR – Research Institute on Sustainable Economic Growth of the National Research Council of Italy is the owner of the JoREP 2.0 database.

3.2 Owner of raw data

According to the EUFP7 RISIS legal specificities, the databases can be fully exploited under the EUEUFP7 RISIS activities for the aims of the project (opening of the infrastructure to external users for research purposes, integration of the infrastructure with other similar).

3.3 Current practice for opening up of the database to external users

Opening to external uses is be based on 'site visits' at IRCRES CNR.

4   Technical structure of *JoREP 2.0*

4.1   Information on the data base system

Being the structure of joint R&D programmes complex, an appropriate system for data management is required which integrate and allow several level of analysis.
Data are implemented in **MS Access 2013**, considering the potential of the software in handling several sets of information from different archives logically related to each other and in creating custom views of data. Currently, no future technical changes about the database system are planned. Nevertheless, the possibility to export new releases of the database to another software for a better fruition is not excluded.

4.2   Technical variable definition

Table 4 summarizes **data-type** of the sets of variables in JoREP 2.0:

DIMENSION	VARIABLES	LEVEL	DATA TYPE
Joint R&D Programme	*prog_ID*	Programme-level	Short text (Code)
	prog_start_year	Programme-level	Short text
	prog_end_year	Programme-level	Short text
	original_prog	Programme-level	Short text
	successor_prog	Programme-level	Short text
	demo_transformation	Programme-level	Short text
	establishing_authority	Programme-level	Short text
	reference_year	Programme-level	Numeric/Multivalue
	name_prog_eng	Programme-level	Short text
	type_instrument	Programme-level	Short text
	prog_duration	Programme-level	Short text
	proj_duration	Programme-level	Short text
	research_topic	Programme-level	Short text
	submission_procedure	Programme-level	Short text
	selcri_scientific/selcri_relevance	Programme-level	Scale (Numeric)
	funding_model	Programme-level	Short text
	EU_contribution	Programme-level	Binary
	coordinator_country	Programme-level	Short text
	partner_countries	Programme-level	Short text/Multivalue
	observer_countries	Programme-level	Short text
	integration_mode	Programme-level	Short text
	ERA_category	Programme-level	Short text
	prog_type	Programme-level	Short text
	reference_year	Participation-level	Date
	participating_country	Participation-level	Short text (Code)
	national_role	Participation-level	Short text
	funding_agency_ID	Participation-level	Short text (Code)
	agency_function (disaggregated)	Participation-level	Short text
	origin_of_funding	Participation-level	Short text
	budget_source_category	Participation-level	Short text
	budgeting	Participation-level	Short text
	currency	Participation-level	Short text
	amount	Participation-level	Numeric
	budget_data_status	Participation-level	Short text
	budget_data_source_category	Participation-level	Short text
	budget_data_source	Participation-level	Short text

	beneficiary_sector (disaggregated)	Participation-level	Binary
	beneficiary_sector_data_source	Participation-level	Short text
	reference_year	Beneficiary-level	Date
	beneficiary_country	Beneficiary-level	Short text (Code)
	funding_agency_ID	Beneficiary-level	Short text
	currency	Beneficiary-level	Short text
	amount_public/amount_private/amount total	Beneficiary-level	Numeric
	data_status_beneficiary	Beneficiary-level	Short text
	data_source_category_beneficiary	Beneficiary-level	Short text
	data_source_beneficiary	Beneficiary-level	Short text
Funding agency	*funding_agency_ID*	Agency-level	Code
	agency_country	Agency-level	Short text (Code)
	agency_acronym	Agency-level	Short text
	agency_name_nat_lang	Agency-level	Short text
	agency_name_eng_lang	Agency-level	Short text
	agency_status	Agency-level	Short text
	agency_website	Agency-level	Short text
	agency_tot_budget_2009	Agency-level	Numeric
	agency_geo_level	Agency-level	Short text
	agency_classification	Agency-level	Short text
	agency_domain	Agency-level	Short text
Country	*country_name*	Country-level	Short text
	country (ISO code)	Country-level	Short text (Code)
	EC_ATC_classification	Country-level	Short text
	JoREP country	Country-level	Binary
	latitude	Country-level	Numeric
	longitude	Country-level	Numeric
	gerdbygdp	Country-level	Numeric
	gradbypop	Country-level	Numeric
	reserbypop	Country-level	Numeric
	h_index	Country-level	Numeric
	population	Country-level	Numeric

Table 4: data-type of the variables stored in JoREP 2.0

4.2.1 Definition of identifiers and demography rules

A set of specific records – containing some unique pieces of information identified by a code – has been designed for the joint R&D programmes and the funding agencies which represent the units of analysis in JoREP 2.0.

The flexible characteristics of programmes and funding agencies have imposed the creation of a **set of rules** in order to manage their demography, as outlined in the table 5:

Demographic Event		Decision about identifier (ID)
Creation of a new joint R&D programme.	**Closure** of an existing joint R&D Programme.	New independent identifier code (ID) assigned in the case of entry.
Entry of a new funding agency.	**Modification of status** of a funding agency	Identifiers of disappearing units are reserved and not reused.

Merger between two joint R&D programmes / funding agencies in a new joint R&D programme/funding agency (with new name).	**Split** of an existing joint R&D programme / funding agency into two or more independent joint R&D programmes /funding agencies	New independent identifier (ID) assigned analogously to entry. Identifiers of antecedent programmes/agencies are reserved and not reused.
Take-over of one programme by another one (death of the programme which was taken over, the taking over programme continues to exist). The new programme retains the identity of the programme taking over.	**Spin-out (Spin-off).** Split-off of a part of a programme to become a separate programme. Old programme exists unchanged, new programme has new name	In the case of take-overs, the ID of the dominant programme taking over another will be held. The identity of the programme taken over is reserved and not reused. In the case of spin-outs, the ID of the original programme will continue. Analogously to a birth, a new ID is created for the new spinning out.

Table 5: demographic events and decisions about identifiers.

While the **demography of the funding agencies** will be available according to the work developed in EUFP7 RISIS WP8, information on the **demography of the programmes** is traced through some fields inserted in the *ProgrammeCatalogue* table as shown in table 6:

EVENT	FIELD	INFORMATION CONTENT
Creation	*Start year*	Year of creation of the programme
Closure	*End year*	Year of closure of the programme
Merger	*Original programme*	- Indication of the single programmes merged appears in the *Original programme* field related to the programme born from the merging;
	Successor programme	- Indication of the programme born from the merging appears in the *successor programme* fields related with the merged programmes.
Split	*Successor programme*	- Indication of the new programmes born after the split appears in the *successor programme* field related to programme subject to split;
	Original programme;	- Indication of the programme subject to split appear in the *original programme* fields related with the new programmes
Take-over	*Successor programme;*	- Indication of the programme taking over in the *successor programme* field of the programme taken-over.
	Demographic transformation	- Indication of the process of take-over (in *demographic transformations* field) associated to the programme takin-over (in *ProgrammeHistory* table)
Spin-out	*Original programme;*	- Indication of the original programme in the original programme field of the spun-out programme.
	Demographic transformation	- Indication of the process of spin-out (in demographic transformations field) associated to the programme subject to the spin-out process (in the *ProgrammeHistory* table)

Table 6: traceability of the demographic events regarding joint R&D programmes

4.3   Description of the Relational Model of *JoREP 2.0*

The design of JoREP 2.0 database was created in order to better manage the volume of data to be collected in future updates as well as to adapt to potential end-user demands for temporal analysis. It is based on the nature of the main unit of analysis it concerns: joint R&D programmes. Specifically, two features drove the creation of the model:

- A joint R&D programme is a **multilevel entity**, characterized by several or highly diversified attributes. Programme descriptors refer to three different levels: one is the level of the programme – including e.g. research topics; the other is the level of participation – regarding e.g. financial participation of countries and the third one is the level of beneficiaries – concerning e.g. budget allocated to performers.

- Programmes can **change over time** (*see 4.2.1*): they can undergo specific transformations, potentially each year. Data reveal variability both *within* and *between* programmes, considering internal changes within the programme along the years or participation and allocation dynamics that might change in the same time.

Thus, due to the **multilevel structure** of the joint R&D programmes and considering the need to handle **panel descriptors**, data have been stored into distinct manageable tables within a **relational structure:**
- Allowing to easily browse specific information referring to all levels;
- Accounting for the storage and visualization of longitudinal data.
- Assuring maximum possibility of querying.

4.3.1 Structure of the database
JoREP 2.0 rests on a solid bridge-structure *(fig.1)* including:

- ✓ **Five entity sets**, autonomous archives designed as **parent tables**, divided in:
    - two tables containing the **units of analysis** including
        - i.   basic and stationary data about joint R&D Programmes;
        - ii. minimal descriptors of the funding agencies;
    - three **auxiliary sets** containing
        - iii. a list of countries *source of funding for/beneficiary of* the programmes;
        - iv. a list of reference years, as support for panel data collection;
        - v.   an auxiliary list of currencies

- ✓ **Five longitudinal data tables**, designed as **child tables,** configuring as
    - **three junction-archives** among the entity sets, containing in-depth data on the historical transformation of the programmes and focusing on the dynamics of the three levels of the joint R&D programmes:
        - a. the internal characteristics of programmes;
        - b. the National participations in term of role and allocated budget;
        - c. the funding received by the beneficiaries of the programme;
    - an **auxiliary set** of panel countries' descriptors;
    - a **support panel archive** containing exchange rates.

- ✓ A web of **one-to-many relationships (1:M)**, in order to create links between parent and child tables – through unique identifiers – synching up data.
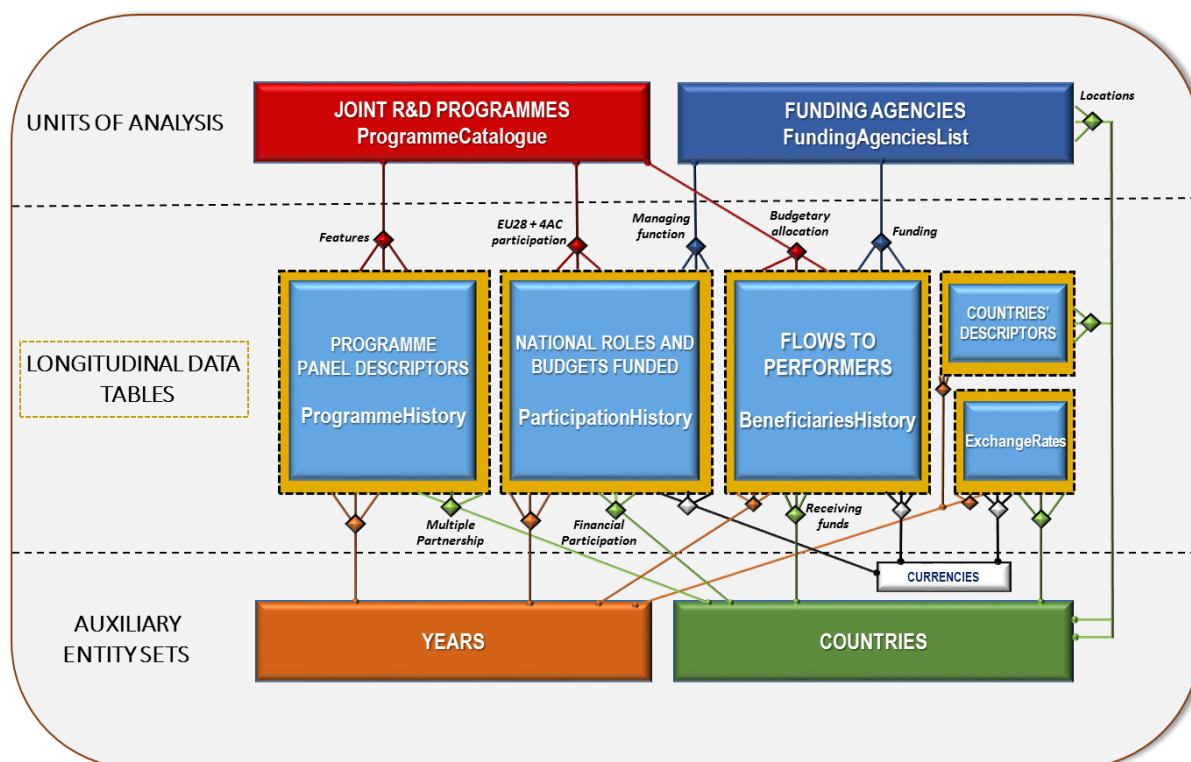
Figure 1: the bridge-structure of JoREP 2.0.

The relational model has been studied both in order to register the **demography** of the programmes and in order to follow the **history** of the programmes by year, pointing out the significant internal changes (at the programme-level) and the peculiar features which distinguish the participation and allocation dynamics (at the participation-level and beneficiaries-level).

As showed above (*fig. 1*), all the theoretical *many-to-many relationships* which link the main entities involved in the programme life – *joint R&D programmes, funding agencies, countries participating/beneficiating, years*– have been treated by creating **junction tables in one-to-many relationship (1:M) with single entity sets**.

So-built relationships envisage that each record in the parent tables (showed at the top and at the bottom of the figure), is related to one or more records contained in the child tables (at the centre of the figure).

All the 1:M relationships have been implemented in the relational model by inserting common fields marked as **primary keys** on the *1 side* and as **foreign keys**[7] in the *M side,* as showed in MS Access relationship window (*fig.2*).
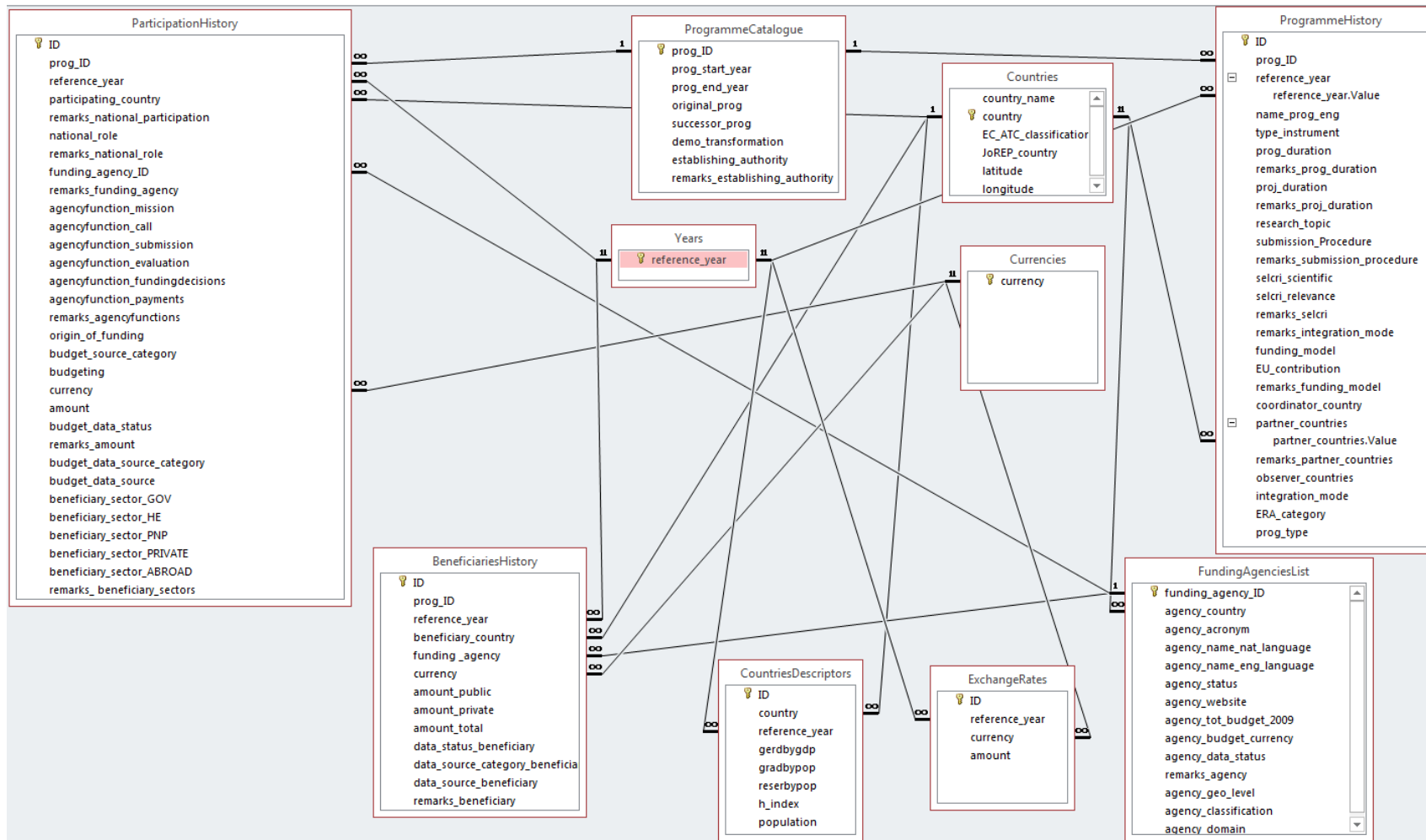

Figure 2: the relational model in the MS Access layout.

---

[7] The foreign key is a field whose values match the primary key values in the related table. Foreign key values refers to an existing valid record in the entity set (Alexander, Kusleika, 2013).

22

Primary keys in the parent tables and foreign keys in the child tables automatically verify against each other. All their links represents the connections establishing data integrity between tables (Alexander, Kusleika, 2013). **Referential integrity** has been enforced in order to allow manageable future updates of the database and to prevent potential errors so that:

- When adding a new record to a child table, if a foreign key value is entered, it must exist in the related primary key field of the parent table;

- When changing a record in a parent table if the primary key is changed, the change must be cascaded to all foreign key valued records in any related child tables. Otherwise, the change to the parent table must be prohibited.

- When deleting a parent table record then related foreign key records in child tables must either be cascade deleted or deleted from child tables first.

4.3.2 The units of analysis tables

The individual joint R&D programmes and the single funding agencies – representing the **units of analysis** in JoREP 2.0 – are handled in two autonomous **parent tables**, as unique and distinct entities (forming single records), depending on primary keys.

- *Programme catalogue*

The table contains a primary key (a code) that uniquely identify each joint R&D programme as a single object. Further, it provides a minimal set of attributes in order to identify the main unit of analysis. The few descriptors focuses on the *basic* and *constant* information concerning the joint R&D programmes, including the start and end years and the establishing authorities. This set of information is considered as stationary (*see tab.7*).
Each record in the *ProgrammeCatalogue* has been linked with functionally dependent rows of longitudinal data tables, named *ProgrammeHistory, ParticipationHistory, BeneficiariesHistory*.

- *Funding agencies list*

It contains a list of formal organizations that manage functions for joint R&D programmes. Each agency is identifiable by a code contained in the table.
JoREP 2.0 collects a set of descriptors of the agencies as background information for the analysis.
Each record in the *FundingAgenciesList* has been linked with a functionally dependent row of longitudinal data tables, named *ParticipationHistory* and *BeneficiariesHistory*.

4.3.3 Auxiliary entity sets in the relational scheme

Two relevant aspects in the programme life – countries participating/beneficiating and single years in which programme exist – form single records in two main **auxiliary entity sets**, depending on primary keys.

- *Countries*

It contains the complete list of National states (associated to their ISO country code) with the indication of the inclusion/non-inclusion in the JoREP geographical coverage. In case of inclusion, the country is matched with the indication of latitude and longitude of its capital city.

Countries are intended both as **source of funding** and **location of performers**. The former is the case of the 1:M relationship with the longitudinal table about participation, in which the *1 side* will only link JoREP 2.0 countries; the latter is the case of the 1:M relationship with the longitudinal table about beneficiaries. Countries are also the references for the funding agencies contained in JoREP 2.0.

Each record in *Countries* has been linked with a functionally dependent row of the longitudinal table of *Countries Descriptors* and with all the single values contained in the multivalue field named *Partner Countries* in the table *Programme History*.

- *Years*

Years a simple list of the reference years for each programme.

Each year is essential to review the evolution of the programmes under all levels. Each record in *Years* has been linked with a functionally dependent row of all longitudinal data table in the database.

A further auxiliary parent table, named *Currencies*, has been envisaged in order to favour monetary conversions.

- *Currencies*

Currencies entity set has been created in order to contain a list of currencies. Each row is linked with the 'currency' fields of the two longitudinal tables regarding participation and beneficiaries.

4.3.4 Child tables and their relationships

Junction tables, configuring as **child tables**, has the task of linking together two or more entity sets – via foreign key –, permitting to browse, select, join, divide information in the database. These tables collect **longitudinal panel data**, being the constant link with the temporal dimension (the 'years' entity set).

All the other relevant **disaggregation** are integrated in the sets of panel data, giving the possibility to follow the dynamic component of the programme **over all levels**.

Thus, a single record in the longitudinal data tables links together one programme with a single or a multiple disaggregation.

- *Programme history.*

Potentially, a joint R&D programme might vary in some **internal characteristics** along the programme life (including e.g. the name of the programme). The table about programme history hosts potentially variable descriptors of joint R&D programmes in order to keep track of the characteristics of the programmes – situated at the 'programme level' – in their evolution. The interest focuses on the descriptors that identify the organizational features of the programmes, including e.g. the research topics, partner countries and the integration mode (*see tab. 7*).

Descriptor	Level	Type	
Programme start year	Programme-level	Stationary	
Programme end year	Programme-level	Stationary	
Original programme	Programme-level	Stationary	**PARENT TABLE**
Successor programme	Programme-level	Stationary	
Establishing authority	Programme-level	Stationary	
Reference year	-	-	
Name of the programme	Programme-level	Variable	
Type of instrument	Programme-level	Variable	
Programme duration	Programme-level	Variable	
Project duration	Programme-level	Variable	
Research topic	Programme-level	Variable	
Submission procedure	Programme-level	Variable	
Selection criteria	Programme-level	Variable	**LONGITUDINAL**
Funding model	Programme-level	Variable	**TABLE**
EU contribution	Programme-level	Variable	
Coordinator country	Programme-level	Variable	
Partner countries	Programme-level	Variable	
Observer countries	Programme-level	Variable	
Mode of integration	Programme-level	Variable	
ERA category	Programme-level	Variable	
Programme type	Programme-level	Variable	

Table 7: the split among stationary and variable descriptors at the programme level.

As displayed in *fig. 3*, 1:M relationships link the longitudinal table with two parent table:
- each programme is linked with *ProgrammeHistory* junction table through the ProgrammeID (primary key)/ProgrammeID (foreign key) relationship;
- temporal dimension is linked with the panel table through Year (primary key)/Reference Year (foreign key).

Note: the field *Reference Year* in *ProgrammeHistory* is treated as multivalue (*fig. 3*), so that multiple values stored in the field can be treated independently in potential query tables.
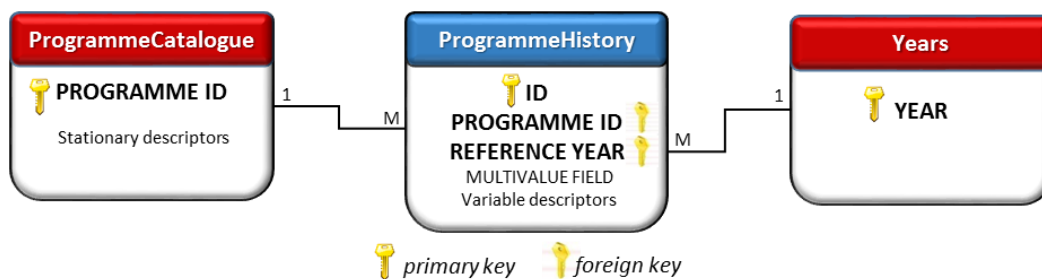
Figure 3: 1:M relationships regarding ProgrammeHistory.

The 1:M relationship between programmes and partner countries (a descriptor of *ProgrammeHistory*) was treated typing the 'partner countries' field as multivalue (*fig. 4*), so that multiple values stored in the field can be treated independently in potential query tables. The set of the possible values of the field is contained in the *Country* table.
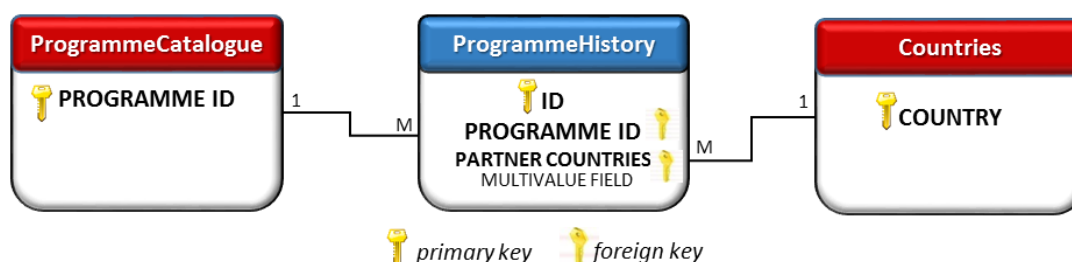
Figure 4: the 1:M relationship between programmes and partner countries.

- *ParticipationHistory*

A key task of JoREP 2.0 is to store complete data on the funding flows related to joint R&D programmes. *ParticipationHistory* collects data on the national participation of JoREP 2.0 countries to the joint R&D programmes from the catalogue for each reference year.

The dynamic of participation points out the funding agency executing at least one of the programme function (if applicable), including also the budget allocated to the programme by the agency or by only the Country (if applicable).

The set of descriptors is displayed in *tab. 8*. The list does not include fields on information about data source and remarks.

Descriptor	Level
Reference year	-
Participating country	Participation-level
National role	Participation-level
Funding agency	Participation-level
Agency functions	Participation-level
Origin of funding	Participation-level
Budget source category	Participation-level
Budgeting	Participation-level
Currency	Participation-level
Amount	Participation-level
Potential beneficiary sectors	Participation-level

Table 8: descriptors of joint R&D programmes at participation level

Panel data are stored – for each year of reference – linking all the programmes to the countries from where participation originates and to funding agencies involved (if applicable).

Because of the presence of a value in the 'country' field and in the 'funding agency' field, two special cases deserve attention:

- *Case 1:* a country can participate with a supranational agency instead of a national agency. The supranational agency will be inserted in the rows of the participation referred to a country.
- *Case 2:* no funding agency might be involved in the national participation or the name of the funding agency is not available. Special code will be used to indicate the absence of some value in order to avoid 'null' value in the *FundingAgenciesList*[8]

---

[8] For instance, referring to the row of a country, a special code in the 'agency' field indicates that the country does not participate to the programme with a funding agency.

Possible combinations between 'country', 'agency' and 'budget' in the participation table are showed in the *tab. 9*:

TYPE OF PARTICIPATION	FIELD CONTENTS		
	Country	Agency	Budget
A country participates *with a National funding agency*, financing / not financing the programme	ISO code of the country	National agency ID	Amount funded
A country participates *with a supranational funding agency*, financing/not financing the programme	ISO code of the country	Supranational agency ID	Amount funded
A country participates *without a National funding agency*, financing / not financing the programme	ISO code of the country	Indication of no agency (special code will be used)	Amount funded

Table 9: descriptors of joint R&D programmes at participation level

Four tables share common attributes with the junction table, enabling the entity sets to be linked together (as showed in *fig. 5*):

- o each programme is linked with the *ParticipationHistory* through the ProgrammeID/ProgrammeID 1:M relationship;
- o temporal dimension is linked with the *ParticipationHistory* through the Year/Reference Year 1:M relationship;
- o countries are linked with the *ParticipationHistory* through Country/Country 1:M relationship;
- o funding Agencies are linked with the *ParticipationHistory* with Funding Agency ID/Funding Agency ID 1:M relationship.
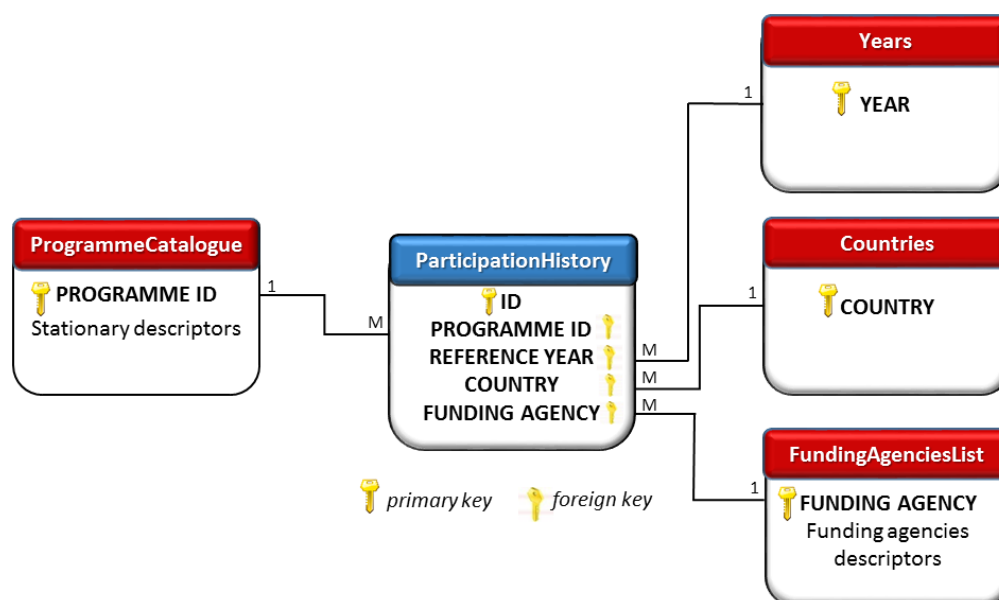
Figure 5: 1:M relationships regarding ParticipationHistory.

- ● *BeneficiariesHistory*

Each country can beneficiate of a budget for many programmes, through funding agencies. Further, the budget received by a national agency may change by year.

Thus, beneficiaries' history stores a data collection of the budget transferred from each funding agencies to the performers. Data are collected by year for the countries of the performing organization and disaggregated by funding agency managing resources for the programme.

Descriptor	Level
Reference Year	-
Beneficiary country	Beneficiary-level
Funding agency	Beneficiary-level
Currency	Beneficiary-level
Amount (public/private/total)	Beneficiary-level

Table 10: descriptors of joint R&D programmes at beneficiary level.
The list does not include fields on information about data sources and remarks.

The four tables share common attributes with the junction table enabling the entity sets to be linked together (as showed in *fig. 6*):
- o each programme is linked with the history through Programme ID/Programme ID 1:M relationship;
- o each beneficiary country is linked with the *BeneficiariesHistory* table through the Country/Country 1:M relationship;
- o temporal dimension is linked with the *BeneficiariesHistory* through the Year/Reference Year 1:M relationship;
- o Funding Agencies are linked with the *BeneficiariesHistory* with Funding Agency ID/Funding Agency ID 1:M relationship.
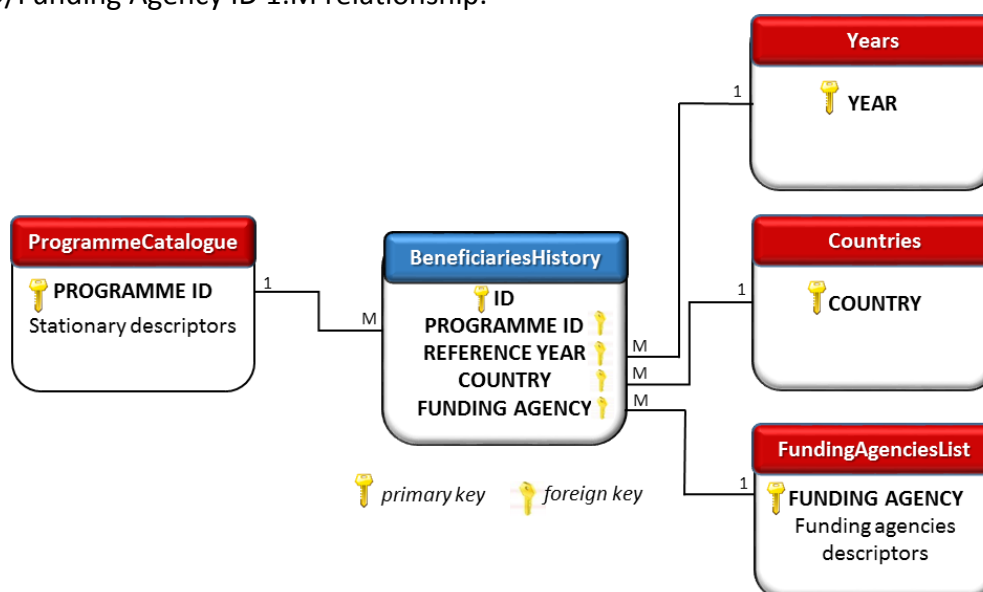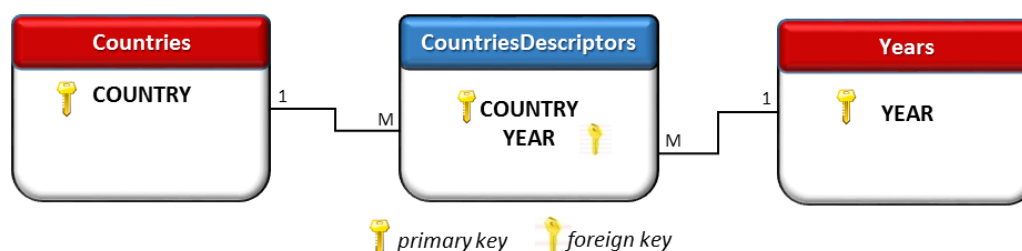

Figure 6: 1:M relationships regarding BeneficiariesHistory

- • *Countries' descriptors*
  A longitudinal table of countries' descriptors has been introduced (*see tab.11*) in order to favour the match of data of country participations to joint R&D programmes with internal and variable-by-year features of the 32 countries which represent the geographical coverage of JoREP 2.0. Linkages between countries' descriptors table and 'countries' and 'years' tables are showed in *fig. 7*.

28

| Descriptor	Level
Reference Year	-
GERD as a percentage of GDP	Country-level
Graduates on population	Country-level
Researcher on population	Country-level
H-Index	Country-level
Population	Country-level

Table 11: descriptors of JoREP countries

Figure 7: 1:M relationships regarding Countries'descriptors

- *Exchange rates*

A longitudinal table of exchange rates has been introduced in order to favour the conversion of monetary value in Euros. The table arose from the link between the 'currencies' table and 'years' entity set (*fig.8*).

Note: budget data for the period 2010-2014 are all provided in Euros.
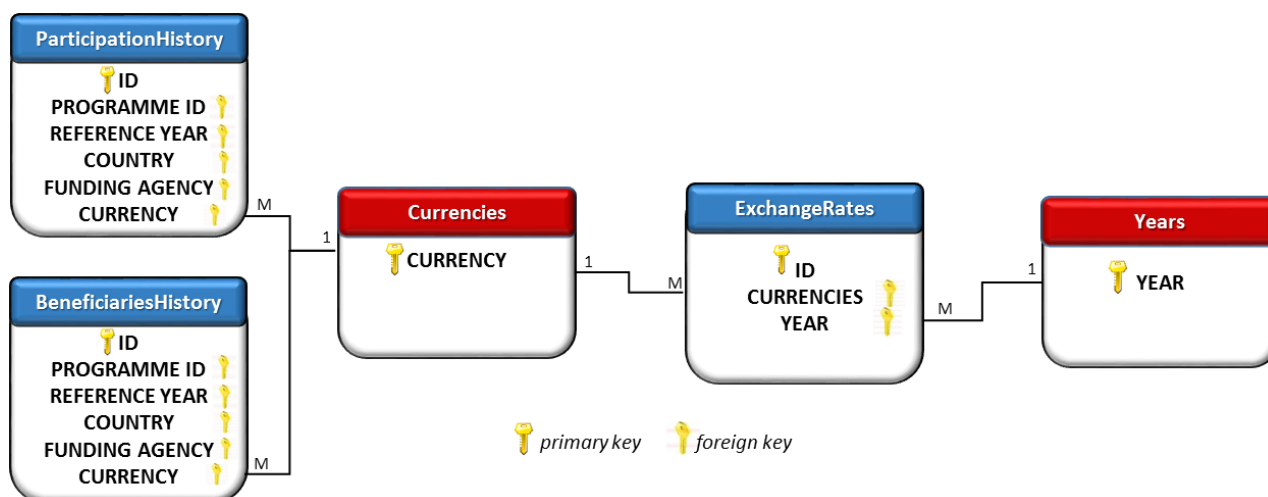
Figure 8: 1:M relationships regarding ExchangeRates

4.4 Interfaces for access and to other infrastructures

Currently no interfaces with other infrastructures are foreseen.

5 Opening of the database and further planning

- *First opening of the database (December 2015)*

First opening of the database occurred in December 2015 and concerned version 1.0 of JoREP, with baseline 2009. In preparation for the opening, JoREP has been subject to a check of all the available information. First concrete steps towards opening were presented during the 'RISIS WEEK 2015' conference (Rome, January 26-29, 2015).

- *Opening of the 2.0 database version (June 2016)*

The following list presents the main actions which were implemented in conclusion of RISIS WP22 updating activities, in preparation for the opening of the 2.0 version of the JoREP database (June 2016):

- Updating of European-level programmes data until 2014, according to the activities to be developed in EUFP7 RISIS WP22. Newest data on large programmes (e.g. EUREKA) were collected following internal disaggregation of the programmes in order to collect more specific information;

- Enlargement of geographical coverage. Due to the activities of WP22, it was foreseen that the coverage of the database would be enlarged in order to expand the geographical coverage from 11 countries to EU28 countries plus 4 associated countries;

- Utility and feasibility checks of the list of variables. Data redundancy was limited to the necessary for the integration of tables within the relational database (e.g. repetition of information in multiple fields was treated clustering information in unique fields).

- *Further planning and technical changes*

The design of JoREP database has been planned to maintain the focus on the amount spent for joint R&D programmes and their organizational characteristics, but it was also studied with an outlook to future enlargement processes, according to the analysis of ERA dynamics.

The relational model is ready to be expanded in order to potentially host data on institutional funding and project funding[9] at country level. The mentioned data can be retrieved from the PREF database[10]. Furthermore an integration of data which could be derived from other EUFP7 RISIS infrastructures can be implemented.

Further technical changes could be applied in order to enhance the new structure of the database:

- All the 'source fields' 'remark fields' contained in the tables of the database will be removed and replaced with metadata or flags;

- Funding agencies codes will be re-designed in a new format according to the work developed for the register of public sector organizations (EUFP7 RISIS WP8).

- *Changing of legal conditions*

No changing legal conditions for accessing the database or parts of the database are foreseen.

---

[9] The distinction between 'institutional funding' (institutional block funding) versus 'project funding' is presented and discussed in OECD (2012).

[10] The aim of the PREF-Public Research Funding study is to collect information on and provide an analysis of national public research funding by theme and by type of allocation.

References

- Alexander M., Kusleika D. (2013), Access 2013 Bible, Wiley, Indianapolis IN

- ERAC – European Research Area Committee (2010), Joint Programming in research 2008-2010 and beyond, Report of the High Level Group on Joint Programming to the Council, November 2010

- EUFP7 RISIS DOW, available at http://risis.eu/wp-content/uploads/2014/08/DOW-RISIS-313082-2013-11-20-1.pdf

- ISO 3166 country codes, available at http://www.iso.org/iso/country_codes.htm

- ERA LEARN tool box (2015), available at https://www.era-learn.eu/manuals-tools/call-implementation/call-planning/call-process-and-administration/funding-modes

- Lepori B., Reale E. (2013), Handbook on data collection on Joint and Open Research Programmes. Annex to the Final Report (JOREP), available at http://ec.europa.eu/research/innovation-union/index_en.cfm?pg=other-studies

- OECD (2002), Frascati Manual. Proposed Standard Practice for Surveys on Research and Experimental Development, Paris, OECD

- OECD (2012), Issue brief: public sector research funding, available at http://www.oecd.org/innovation/policyplatform/48136600.pdf

- Reale E., Lepori B., Nedeva M., Thomas D., Primeri E., Chassagneux E., Laredo P. (2013), Investments in Joint and Open Research Programmes and analysis of their economic impact (JOREP). Final Report, available at http://ec.europa.eu/research/innovation-union/index_en.cfm?pg=other-studies