

Exploring JoREP 2.0: features and potential uses in the studies on Europeanisation of research activities

IRCrES-CNR, Rome – May 25-26, 2017

**JoREP DATA COLLECTION:
SOURCES, QUALITY AND LIMITATIONS**

Andrea Orazio Spinello



RISIS

Research infrastructure for research
and innovation policy studies



Consiglio Nazionale delle Ricerche

IRCrES

ISTITUTO di RICERCA sulla CRESCITA ECONOMICA SOSTENIBILE
RESEARCH INSTITUTE on SUSTAINABLE ECONOMIC GROWTH

JoREP data collection – two waves

- ❑ 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, and developed by a team composed of national experts
- ❑ In the context of EUFP7 RISIS project, an update of the data until 31.12.2014 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)**

Two waves of data collection:

- ✓ Data over 2000-2009, performed in 2011
- ✓ Data over 2010-2014, performed in 2015/2016




New features of JoREP at a glance

- In June 2016 the 2.0 version of JoREP database was opened

JoREP 2.0 - Joint Research Programmes database



JoREP 2.0 is a database on joint R&D programmes. It provides a quantitative basis for the monitoring of investments in joint R&D programmes in the countries belonging to the European Research Area (ERA), pointing out the policy rationales behind them and their impact. The set of data aims at describing when, how and serving what purposes European-level initiatives and bilateral/multilateral joint R&D programmes are combined. JoREP 2.0 also supports the analysis of important ERA dynamics and Europeanization processes through the study of the behaviour of main national actors (i.e. funding agencies).

 Visit Request

- What's new in JoREP 2.0:
 - Inclusion in the database of **new European-level joint R&D programmes**;
 - Collection of European-level joint R&D programmes data at programme level and participation level **from 2010 to 2014**;
 - **Disaggregation of data on large programmes** (e.g. EUREKA) in order to collect more specific information;
 - Enlargement of geographical coverage for data on funding from 11 countries to **EU28 countries plus 4 associated countries** (CH; IL; NO; TR);
 - Creation of new table of indicators at the **country level** for spatial analysis



Key characteristics of JoREP 2.0 / 1

- 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.



Key characteristics of JoREP 2.0 / 2

- A large geographical coverage:
 - for the period 2010-2014, 32 countries covered (EU28 countries plus CH; IL; NO; TR);
 - 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 (CH, CZ, DE, DK, ES, FR, IT, NL, NO, PL, UK).



JoREP data collection main steps

- Definition of the perimeter
- Pilot data collection (first wave) in Switzerland and Norway
- National experts draw a comprehensive list of programmes and agencies described through a limited set of descriptors
- Validation of programmes and perimeter
- Selection of programmes to be included in the restricted perimeter for the full data collection
- Production of a standard template in Excel-transfer of data in Access
- Quality checks, cross checking

**From 2014 onwards enlargement according
to the aims of the RISIS project
CENTRALISATION OF THE COLLECTION**



Main sources for data collection

UNIT OF ANALYSIS	SOURCES
<p>Joint R&D Programmes</p>	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
	GBARD
	Direct contacts with agency personnel
<p>Funding agencies</p>	Funding agency websites
	Direct contacts with agency personnel



Most recurrent problems in the data collection

- Coverage of ERA-Nets and other European level initiatives
- Reference year (financial data/programme start/agreement signature)
- Confidentially problems (limited impact in the final database)
- Problems of data availability (inclusion or exclusion of programmes from the list) especially with respect to funding data back in time (e.g. changes in the funding agency, yearly funding information not always available) and flows to beneficiaries (estimations as solution)
- Data validation (info collected from the web)



Quality issues – FIRST WAVE

- Extensive data checks and cross checks to ensure high data quality
- Standardisation of descriptors and procedures for data collection (e.g. Excel sheet with drop down lists)
- Guidelines to help national expert in understanding descriptors and interpreting programme data the same way
- Review of national reports on data collection – common solution to problems
- Data validation to be ensured by national experts



Within the RISIS project...first step

- The original relational JOREP dataset has been subject to a re-engineering process (implementation of a new relational structure -> creation of JoREP)
- A control of the quality of data was implemented by IRCrES-CNR.
- Data cleaning foresaw exploratory controls focused on the detection of non-sampling errors, whom correction required the recognition of systematic errors and random errors.



Quality issues – SECOND WAVE

- 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
- Check of referential integrity has been implemented as ultimate database safety check for inconsistent data and mechanism for the synchronization of the archives



Main inconsistencies in the data collection

- Inconsistencies between descriptors especially when they are interdependent (a certain value on one variable makes another variable meaningless or inconsistent)
- No reason or explanation for missing data: textual explanation needed
- Flags missing: need to use standardized flags to annotate characteristics of data (e.g. 'Not available', 'Not applicable' etc.)
- Inconsistencies in terms of classification



Cases of exclusion (FIRST WAVE): some examples from Italy

Identification code	Name of programme	Motivation for exclusion from the perimeter
PEU006	ERANET ERA ENWEALTH	Italy did not take part to 2008/2009 joint first experimental call.
PEU013	ERANET ERASYS BIO	The national partner, Province of Trento, did not participate with funding to the first joint call issued on 2009.
PEU015	ERANET EURONANOMED	Italy participated only to the call launched in 2011.
PEU032	ERANET WOODWISDOM	Italy did not participate to 2009 joint call but to calls issued in 2006, 2010 and 2011 only.
PEU023	ERANET MATERA	Participation since 2010
PEU044	Art. 185- Europe-Developing Countries Clinical Trials Partnership (EDCTP)	Almost only small individual grants are funded.
PEU045	EFDA-European Fusion Development Agreement	No joint activities for research funded.
PRO29	Bilateral programme Italy – Sweden	The framework agreement was signed in 2001 and came into force in 2007. Nevertheless the first joint call for proposals was launched in September 2010.
PEU037	COST actions	They do not fund research activities rather coordination of joint actions and this comment has been added in the remark section. Descriptors only have been collected.
PRO30	Executive Bilateral Programme for scientific and technological cooperation Italy and Slovakia	No funding in 2009 for research projects of major relevance.



Cases of exclusion (SECOND WAVE)

- JoREP 2.0 stores data on European-level joint R&D programmes which launched a call for proposal in 2013 or 2014 and in which at least one of the EU28 countries or CH, IL, NO, TR participates. For this reason **JPI MYBL** (first call in 2015) was excluded.
- The JoREP definition of joint R&D programmes covers publicly funded research programmes for which at least one of them functions is shared between more than a single country. European programmes whose R&D activities are supported by only an EU contribution and private partners were excluded. It's the case of the JTIs **CLEAN SKY, FCH** and **IMI**.



Accuracy and completeness of FIRST WAVE

- Data on European-level initiatives and bilateral/multilateral programmes for the period 2000-2009 (11 countries covered)
- Completeness of the data is good in general
- 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.



Accuracy and completeness of SECOND WAVE

- 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



Data comparability

- Homogeneity in the use of definitions by national experts; from 2014 centralisation of the collection
- 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.
- Data collection 2010-2014 more complete than 2000-2009. New data reliable and harmonized

