



**Joint and Open REsearch Programmes:
features and potentialities of the JOREP database**

IRCrES-CNR, Rome – November 27-28, 2014

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

Emilia Primeri – Andrea Orazio Spinello

RISIS
Research Infrastructure for research and innovation policy studies

IRCrES
Consiglio Nazionale delle Ricerche
INSTITUTE OF RESEARCH IN ECONOMIC SUSTAINABLE GROWTH

RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

JOREP data collection-main steps

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

RISIS
Research Infrastructure for research and innovation policy studies

IRCrES
Consiglio Nazionale delle Ricerche
INSTITUTE OF RESEARCH IN ECONOMIC SUSTAINABLE GROWTH

RISIS RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

Main sources for data collection

- Web pages of funding agencies
- Programs web site and call texts
- Other descriptive documents (i.e. progress reports, national and regional project reports, flyers, ppt presentations to conferences and meetings)
- Contacts with national agencies experts
- Use of GBAORD data generally limited





RISIS RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

Quality issues

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





RISIS RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

Most recurrent problems in the data collection

- Coverage of ERA NETs and other European level initiatives
- Reference year 2009 (financial data/program start/agreement signature)
- Confidentially problems (limited impact in the final data set)
- Problems of data availability (inclusion or exclusion of programs 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, validation?)




RISIS RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

Main inconsistencies in the data collection (1)

- 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 and reporting of bilateral programmes (cross checking for missing programmes, programs labelled as bilateral instead of joint)




RISIS RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

Main inconsistencies in the data collection (2)

- Incoherencies (e.g. bilateral programmes included in one country and excluded in the other)
- Different classification of open programs (UK-DK)
- Description of funding agencies (agencies participating with funding or with coordination/management tasks only)
- Programme topic/theme classifications (in some cases need to not match this with the thematic organisation of research funding –e.g. UK)




RISIS RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

Cases of exclusion: 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.
PR029	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.
PR030	Executive Bilateral Programme for scientific and technological cooperation Italy and Slovakia	No funding in 2009 for research projects of major relevance.




RISIS RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

Data cleaning

- The original relational JOREP dataset has been subject to a re-engineering process
- 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.




RISIS RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

Quality of data

- No missing data for descriptors of joint R&D programme at programme level and descriptors of funding agencies
- Good coverage of the basic characteristics of the units of analysis
- Only few descriptors that are problematic in terms of availability, such as programme budgets and flows to beneficiaries.
- The share of non-available data is higher for what concerns programme budgets (16% of missing data) and funding to beneficiaries (17,5% of missing data).




RISIS RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

Data comparability

- Homogeneity in the use of definitions by national experts.
- 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.





RISIS RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

Summing up-Main limitations

- Funding data
 - Data about funding to beneficiaries (few data, not for all countries)
 - Programme budget overall data (missing data)
- Data interpretation and application sometimes different
 - Type of programme
 - Funding agency function and involvement in the programme





IRIS - RESEARCH INFRASTRUCTURE FOR RESEARCH AND INNOVATION POLICY STUDIES

Summing up-Main potentialities

- Good coverage
 - Major programmes are covered (along their “lives”)
 - Missing programmes are very few (limited as for the budget or relevance in terms of participation)
- Data comparability
 - Homogeneity of data ensure comparability
 - Standard classifications used (e.g. topics)
- Data reliability
 - Validation process
 - Several checks accomplished

