



**RISIS**  
Research infrastructure for research  
and innovation policy studies



## A new data infrastructure for science and innovation studies – RISIS

<http://risis.eu>

Philippe Larédo

ENID conference, Lugano, 2-4 September 2015

UNIVERSITÉ  
— PARIS-EST



### RISIS in brief



- EC infrastructure programme project
- 4 years (2014-2017), 13 partners (from universities & PRO)
- Ambition: build a data infrastructure complementary to present statistical approaches, supporting the development of “positioning indicators” (Lepori et al., 2008)
- 4-year goals:
  - (i) integrate existing fragmented datasets and open on a free-of-cost basis to European researchers (100 projects anticipated)
  - (ii) develop new software platforms to support dataset building and treatments (from heterogeneous & unformatted textual corpuses)

## Contents



### Part 1- Three points to characterise RISIS

- 3 major features – linked to 30 years of research and innovation studies
- Anticipated features of the 'central' facility
- A first series of problem based datasets

### Part 2- Where do we stand and challenges ahead

## 3 major features (1)



- Cope with strong asymmetry in knowledge & innovation dynamics
- e.g. 200 groups represent half of world industrial R&D (IPTS scoreboard)
- 4 implications:
  - 1- central use of public data (whether public or private, free of charge or paying)
  - 2- choice of the organisation (and not the individual) as the central reference unit
  - 3- extensive use of textual data
  - 4- need for a number of 'reference databases' or 'registers' on actors/organisations

### 3 major features (2)



- Take into account the other face of globalisation, the importance of 'place' e.g. 'glocalisation'
- See work on agglomeration dynamics – cf. example of world production of nanoscience concentrated in 200 urban areas
- 2 implications:
  - 1- a central issue for the 'infrastructure': develop an approach of 'relevant spatial units' – very similar to the OECD work on 'functional urban areas'
  - 2- extensive need and tools for geolocalisation at the lowest level of aggregation (e.g. addresses of authors in papers or inventors in patents, etc.)

### 3 major features (3)



- Be responsive to evolving theoretical questions and policy needs
- Our answer: balance developments of targeted datasets and of a 'central facility' enabling problem-based integration of different datasets
- 3 implications:
  - 1- develop common resources available to all dataset developers while insuring technical, legal & cognitive harmonisation
  - 2- progressive process: first 'integrating' a number of existing datasets, second developing new datasets on issues considered critical
  - 3- organise processes supporting opening to new datasets and articulation with other types of datasets

## About the 'central' facility



- A mostly virtual facility based upon 4 interconnected tools
- A central access and 'working space' for researchers
- Support to researchers through extensive training opportunities (more than 15% of total EC budget for this)
- Development of 2 software platforms supporting on-line database construction (SMS , opening planned in 2016) and dataset treatment (Cortext, already in beta version, more than 1000 users, [www.cortext.fr](http://www.cortext.fr))
- The development of shared data resources:
  - On actors: ETER register for universities & PROs, CIB for large firms
  - on places: geolocalisation (as mentioned)
  - on critical data: enrichment for public research use only (publications & patents in a first stage) datasets)

## A first series of problem-based datasets



- 6 themes focused during this creation process, and for each of them, selecting/building focused datasets.
- Innovation dynamics of firms: 3 datasets
  - globalisation of R&D activities of large firms (CIB, see Laurens at al., 2014 for first unexpected results),
  - long-term dynamics of small high tech firms (VICO dataset)
  - innovation characteristics of fast growing mid-size firms (whatever industry, under construction, first results expected in 2016)
- European integration with enlarged EUPRO (a longitudinal dataset of all European projects) and with JOREP (focused on trans-border funding programmes in Europe)

## A first series of problem-based datasets RISIS

- Public research dynamics with the ETER register, and enlarged Leiden ranking
- New 'dominant sciences': A demonstration dataset to develop tools for characterising emerging technologies, based on nanosciences and technologies (IFRIS Nano)
- PhD careers datasets combining a panel-based longitudinal approach (PROFILE) and transversal approaches on mobility (MORE), plus the development of a novel approach enabling integration of a number of national datasets
- R&I policy support tools: SIPER repository of "science and innovation policy evaluations", articulated with the OECD-World Bank Innovation Policy Platform

## Part 2: where do we stand? RISIS

- A coordinated effort for opening existing datasets (objective: transform them from experimental datasets to robust, reliable, documented and coherent datasets)
- Results:
  - 9 datasets to be opened during the second semester of 2015
  - the decision to maintain the transversal team built to follow the use of opened datasets and organise the progressive integration of new/external datasets
  - an important training effort engaged
  - the definition of a computer architecture insuring researchers a generalised distant access and a dedicated working space (opening planned in 2017)

## Part 2: where do we stand? (2)



- The implementation of common tools
  - opening of the RISIS version of CORTEXT (for non specialist users)
  - deepening of reference datasets on European public research
  - development of an 'hybrid' approach of key private databases used (Patstat, WoS)
  - on going comparison of geolocalisation / clusterisation tools
  - on going negotiation with OECD for an agreement on SIPER (interconnection with IPP, shared definitions, joint workshops on policy instruments & mixes), and on geography

## Part 2: where do we stand? (3)



- Interacting with the researcher community
  - the web access for learning about datasets, and proposing projects
  - the periodic training courses: different types for learning about types of data (e.g. patents), datasets & methods
  - considering ENID and its annual conference as a major place for exchange & channel for presenting results derived from the use of the infrastructure
  - an on-going reflection on new modalities (e.g. barcamps)

## Part 2: where do we stand? (4)



- Demonstrators to show the relevance and interest of integration based on problems or questions
- A first choice made to address 4 spatial issues
  - characteristics & dynamics of R&I of European metropolitan areas (in association with OECD, <http://measuringurban.oecd.org>)
  - polarisation processes associated to European and transborder funding
  - functional spatial distribution & hierarchisation of Higher education
  - the role of metropolitan areas in the growth of new technology-based firms

## Organisational challenges



- A starting point: we consider RISIS as a prototype answering the data needs of a specialised social science field (meaning most fields will face similar issues)
- Organisational issues: what lasting forms should we take?
  - Should we be 'stabilised' by being inscribed on the ESFRI roadmap?
  - Is it enough to have an inter-institutional convention (with one managing the centrally shared aspects)?
  - or do we need a legal form? And which one? Can it be shared with others (like 'Fondation de France' that de facto entails hundreds of individual foundations)?