Data without Boundaries
A short Overview

Coordination : Roxane Silberman
CNRS/Réseau Quetelet

Presented by
Stefan Bender / Roxane Silberman

Berlin, German Data Forum, April 18/19

SEVENTH FRAMEWORK PROGRAMME

DwB

Outline:

1) Introduction

2) Context

3) New Conditions to Build

4) Project Architecture

5) Conclusions

6) Future Events 2013 - 2014

Toward a European Research Infrastructure

- A four-year EU-funded FP7-13 project (2011-2015)
- Aims:
  - ➢ Linking the capacity of the research community with the important resources of the official micro data in Europe
  - ➢ Enhancing researchers access to official micro data in Europe
    - ▪ Surveys and administrative datasets, combined files
    - ▪ Focus on confidential (highly detailed) data
    - ▪ Focus on crossing national boundaries

- Mechanism = Coordination of existing infrastructures
  - ➢ CESSDA Data Archives, and the ESS (NSIs coordinated by Eurostat, ECB)
  - ➢ Based on volunteers

Partners

- **Coordination**: Roxane SILBERMAN (CNRS/Réseau Quetelet)

- 29 partners
  - ➢ 1/3 CESSDA Archives: CNRS/RQ, GESIS, NSD, SND, FSD, DANS, UKDA, FORS, EKKE, CIS, RODA
  - ➢ 1/3 NSIs and Statistical departments: ONS, CBS, INSEE/GENES, SORS, IAB, SCB, DESTATIS, CSIC, CNPS-INS
  - ➢ 1/3 Universities: URV, UL, UPC, ULL, SOTON, UoMan, CED (IPUMS)
  - ➢ + MT (SME)

From Current Situation ...

- **Access to official statistics** both anonymized and highly detailed is still uneven in Europe, both at national and at European levels

- **Access to Eurostat** highly anonymized datasets is still burdensome

- **Increasing level of anonymization** does not meet the researchers needs

- Though crucial for comparative Research, **crossing borders** is even worse:
  - ➢ different legal frameworks, institutional arrangements and criteria for accreditation,
  - ➢ different providers (NSIs, Archives),
  - ➢ different modes of access (no access, safe centres, remote execution, remote access),
  - ➢ different languages,
  - ➢ different views about security, anonymization, output checking...

SEVENTH FRAMEWORK PROGRAMME

DwB

... To DwB Project Main Issues

- **Building a central point of access**: what are the available data? How can they be accessed?

- **Metadata standards and interoperability**: NSIs tend to use SDMX as a standard for metadata exchange, CESSDA Archives use DDI as a standard for documentation

- **Legal issues and accreditation**: towards a European accreditation

- **Servicing the use of OS data**: provide tools (format, routines for harmonization), train the researchers for using European micro data

- **Technical, standardization and methodological issues** in developing a European distributed remote access both for national and for European micro data, flexible to national institutional arrangements (NSI or data archives as provider): propose and implement a test case

SEVENTH FRAMEWORK PROGRAMME

DwB

New Conditions to Build

- ## At national level

  - ➢ Strong cooperation between Data Archives and NSIs in some countries
  - ➢ Changes in the legal framework in several countries
  - ➢ Increasing numbers of RDCs providing in site access, remote execution or remote access
  - ➢ Some RDCs providing access to foreign researchers

- ## At European level

  - ➢ The ESFRI roadmap and the CESSDA ERIC process
  - ➢ Projects and initiatives within the framework of the European Statistical System (ESSnet, WGSC …)
  - ➢ Discussions about a new EC Regulation on European micro data access for researchers expected about 2012/2013

Three Blocks, Twelve Work Packages

- **Block 1:** Access Facilities (WP3, WP4, WP9, WP10 and WP11)

- **Block 2:** Front Office (WP5, WP7, WP8 and WP12)

- **Block 3:** Enlarging Cooperation (WP6)

  + **WP1** (Project Management)
  + **WP2** (Internal & External Communication)

Block 1: Access

- **Legal issues and accreditation**: examine current national situation, agree on best practices & common standards, test an accreditation pilot and suggest changes in the legal framework (WP3)

- **Technical issues in developing RA and SDC procedures**: discuss and agree on security standards, architecture for a distributed remote access (WP3 and WP4)

- **Methodological issues**: anonymization and output checking (WP11)

- **Implement a case study** for a distributed European remote access (WP4) building on current possibilities for national micro-data conditional to a change in 2012 on new possibilities for Eurostat micro-data

- Immediately offer **transnational access** through open calls for researchers to access data either on site or remotely (WP9 and WP10)

Block 2: Front Office

- Provide a single point of access within the context of the CESSDA portal and the current CESSDA ERIC process (WP12)

- Discuss standards (SDMX and DDI) and develop tools to harmonize metadata (WP7)

- Devise techniques to harvest NSI metadata by CESSDA (WP8)

- Service the use of the OS micro data : Improve metadata (including translations issues), formats and provide routines for OS and Eurostat micro data (WP5)

Block 3: Enlarging Cooperation

- All WPs work in cooperation and aim at identifying best practices, agreeing on standards and building on volunteers while bridging the different communities (NSIs, Archives, Researchers)

- Yet the long term success requires involving the whole ESS, the whole CESSDA , and the researchers who are the final users , also making bridges with non European partners (WP6)
  - ➤ European Data Access Forum and regional workshops on data access
  - ➤ Users conferences
  - ➤ Training activities
  - ➤ Staff visits in RDCs where remote access solutions exist

To Summarize …

- A challenging project:
  - Need to build trust and common understanding between NSIs, Archives and Research Communities
  - Need to agree on standards, provide a model and implement a pilot
  - Need to enlarge cooperation and strong coordination with other initiatives & ongoing discussions

- A crucial step toward a European research infrastructure within the context of the CESSDA ERIC:
  - Building a single point of entry,
  - Paving the way for a European accreditation,
  - Enhancing access to anonymized official data,
  - Providing a flexible infrastructure for accessing confidential data

Some Major Steps in 2013-2014

- **Continuous calls** for transnational access to highly detailed microdata:
  - ➤ France, Germany, Netherlands, UKDA RDCs
  - ➤ Give support to comparative research projects that required transnational access
  - ➤ A business case for a European accreditation and a distributed remote access

- **Staff Visits in RDCs** providing remote access to confidential OS microdata
  - ➤ Directed to NSIs and data archives that have not implemented any RDC yet and RDCs interested in shifting to or implementing a RA solution.
  - ➤ More info: http://dwbproject.org/events/visits.html

- **24 & 25 April 2013** Regional Workshop in Eastern Europe (Slovenia):
  - ➤ To foster and/or trigger cooperation between NSIs, data archives, researchers and other stakeholders in 15 countries of this region
  - ➤ More info: http://dwbproject.org/events/regional_workshop1.html

Thanks for Listening

Contact: email.adress@dwb.org

Website: http://www.dwbproject.org/