Data without Boundaries

Some 1st outputs

Roxane Silberman
DwP Principal Investigator
CNRS/Réseau Quetelet

Luxembourg, WGSC Eurostat, November 21, 2012

Outline

• Short overview of DwB project objectives
• Front office:
  ➢ Short overview of state of play and 1st outputs
    ➢ Focus on
      ▪ European Service Centre for OS
      ▪ CIMES
• Access: 1st outputs
  ➢ Short overview of state of play and 1st outputs
  ➢ Focus on
    ▪ Researchers needs
    ▪ European Research Accreditation for national microdata
    ▪ European Remote Access Network (ERAN) for national microdata
• Next developments in 2013
1. Short overview of DwB project objectives

- Just to remind
- Objective: enhance researchers’ access to official microdata in Europe
  - Focus on transnational access within the ERA
  - Focus on confidential microdata (same definition as in the draft regulation i.e. without direct identification and SUF)
  - Focus on national microdata
- Contribute to the construction of a European data infrastructure, coordinating 2 existing infrastructures: CESSDA (Data archives European network) and the ESS
  - Building on 28 volunteers partners from the different communities, archives and NSIs/NSAs
  - Involving the whole communities in the discussions
- Project started in May 2011
- DwB and DARA: Many common issues, coordination necessary yet differences
  - DwB focus is on national microdata (many not available in European microdata)
  - Agreements from all member states not necessary, clusters possible

Main issues

- Front office
  - Build a central point of access where the researchers can discover what data are available, get friendly metadata and know where and how and where to get accreditation and access the microdata (including where and how to get access
  - Metadata standards issues (DDI/SDMX)
- Access
  - Facilitate transnational and multiple researchers accreditations
  - Propose a model and a pilot for a European remote access network allowing researchers to conduct comparative research projects based on national microdata
  - Include methodological issues about SDC, output checking …
- Enlarge discussions and agreements to the whole communities (ESS, Archives, researchers, other stakeholders)
  - EDAF (2012, 2014), regional workshops, users conferences, training courses, staff visits in RDCs
2. Front office
Short overview of state of play and 1st outputs

- A proposal (report) providing a framework for a European Service Centre for OS microdata (ESC-OS) linked to the CESSDA portal
- A metadata schema to collect and structure information about available microdata from OS at national level: CIMES
- Compatible with the draft for a metadata model incorporating DDI and SDMX that could be used for harvesting NSIs metadata on CESSDA portal.
- Short focus on ESC-OS and on CIMES

European Service Centre for OS (ESC-OS)

Overall objectives
- To provide a comprehensive overview of available official data in Europe (national and European, include administrative data) building on ESS and CESSDA existing infrastructures i.e. avoiding duplication
- To provide structured metadata and user friendly routines for European and national official microdata
- A specific service yet linked to the CESSDA portal to avoid silos
- Could later be used as a single point of entry for accreditation processes and access building on existing rules and procedures at national and European levels and providing a path for progressive harmonisation
- The European Service Center could also be responsible for coordinating issues of data access and establishing an ongoing exchange between data providers, archives and the research community
- Draft report available on DwB website, enlarged discussion to continue (EDAF, other DwB workshops), feedbacks welcome
ESC-OS

Figure 1: Structure of the ESC-OS webservice

Institutional framework for ESC-OS
CIMES

CIMES: Centralising and Integrating Metadata from European Statistics

• CIMES is a web based tool for inputting structured metadata
• It allows for entering hierarchically structured metadata
• In the future it will allow the import and export of structured XML
• Does not include the variable level for national microdata
  ➢ Provide links to websites (NSIs, archives …)
  ➢ Provide information and links for accreditation process and access
• It will be developed to integrate metadata of European surveys at the level of variable (EU-SILC, EU-LFS, AES, CIS, SES…)
• By the time period of DwB, we hope to cover main microdata on important topics from all members states

![Metadata schema diagram]
3. Access

Short overview of state of play and 1st outputs

- Immediate support for transnational access via regular calls supported by 7 RDCs from 4 countries
  - Provide inputs about researchers needs and coordination issues between RDCs
- Accreditation
  - An almost complete database of accreditation rules and procedures in Europe showing important and encouraging similarities,
  - To be made available for internal use and later for external use (linked to CIMES)
  - To be updated
- A consultation process to start about how to move towards a European accreditation for national microdata
- An ongoing audit of legal frameworks (and interpretation of law) concerning researchers access (also transnational) to confidential microdata
- European remote access network
  - An overview of existing remote access solutions (to be updated)
  - Common principles yet technical solutions differ and other differences (legal/technical)
  - An ongoing survey on researchers experience and needs about remote access
  - Discussions about agreements for a pilot have started
  - First set of discussions about a European remote access network
  - Several papers on methodological issues for managing risk of access to data
Researchers needs

- First lessons
  - From transnational access offered via DwB calls
  - First qualitative interviews
  - To be completed with the survey results
- Multiple accreditations problematic
  - Coordination issues between RDCs in case multiple accesses required
- Importance of transnational access to national administrative microdata for comparative research
- Detailed information about variables not available
- Researchers need to combine microdata from different countries in order to run a single analysis
  - Data have to move out of one of the countries

European accreditation for national microdata (1)

- Overview of current accreditation rules and procedures:
  - Transnational access increasingly open for confidential microdata
  - Information required for applications broadly similar yet differently implemented in the application forms + language issues if not translated
  - Accreditation rules and procedures follow similar principles: safe researcher, safe institution, safe project
  - Main difference: separated institution accreditation needed or not
  - Differences in institutional implementation of procedures (difficult to change in some cases)
European accreditation for national microdata (2)

- How to move towards a European accreditation?
  - An information form providing all information requested by the countries application forms for researcher application
    - The researchers could fill once the form and information could be transferred into the different application forms
    - The information could be saved to be updated by the researcher for a next application
    - A pragmatic step towards a common framework: less burden from the users perspective, does not necessitate immediate changes in the accreditation rules and procedures at country level
    - Consultation of data and service providers will be conducted in 2013: information required for safe researcher, safe institution, safe project...(also taking into account EU regulation guidelines and forms)
    - Results could be presented at the next WGSC November 2013
  - Accreditation rules and procedures: identify clusters where it could be possible to move towards agreements in case of multiple accreditations needed for comparative projects
    - For countries requiring institution accreditation, look for coordination with the EU regulation process

European Remote Access Network (ERAN)

- Survey of current remote access solutions
  - Shows broad similarities in principles yet differences and first outputs
  - Report available on DwB website
  - Will be updated
- Towards a European Remote Access Network
  - 2 steps
    - A pilot for a RDC-in-RDC approach
    - Extension of the RDC-in-RDC approach
- To be presented and discussed at the NTTS conference back-to-back DwB workshop
Pilot for a RDC-in-RDC approach

- On the road
- Based on IAB (Germany), SDS/UKDA (UK) and CASD/GENES (France)
- Used for testing contracts, security, organizational issues...
  - Technical issues
    - Depend heavily on organizational and legal restrictions
    - Virtualization could help
  - Organizational issues
    - Single user account management
    - Responsibility for the network
    - Structure of the workflow (user-contracts, support)
    - Output control
    - Cost
    - Etc
  - Legal issues
    - What is possible on a transnational base
  - Virtual research environment

Extension of the RDC-in-RDC approach

- Pilot as a point of departure:
  - Reaching IAB data from France and UK
  - Not a two-ways approach
    - Reaching French and UK from IAB
    - Not needed for French data: SdBox (CASD/GENES) can be located in any university
  - Does not allow to combine data from different countries
- Feasibility of an ERAN
  - First discussions identified main issues
    - Single point of access needed
    - Access point may be safe centre, institution, anywhere
    - Security
    - Data storage: data has to be stored where it has to be stored due to legal issues, yet constraints may be more or less high depending on countries (anywhere, given country, data provider)
      - Permanent central repository may appear ideal yet most probably not the solution: a lot of data, where?
      - Temporary space for research projects: bi-lateral, multilateral decentralised solutions requiring agreements and contracts, more flexible to different national constraints
4. Next developments in 2013

- DwB workshop to be organized back-to-back to the NTTS conference, March 4, 2013, Bruxelles: European Remote Access Network
- Consultation on accreditation to start in March,
- DwB Users' Conference, 21-22 March 2013, Mannheim
- 2nd DwB Training course on European microdata and confidentiality, 15-17 February 2013, Bucharest
- 1st DwB Regional Workshop, Ljubljana, March 2013: Cooperation between NSIs and Data Archives
- Staff visits in RDCs offering remote access to start in 2013

Thanks for Listening

Contact: roxane.silberman@ens.fr

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