Data without Boundaries
Objectives, 1st year achievements and 1st lessons

Coordination : Roxane Silberman

CNRS/Réseau Quetelet

PARIS, OECD, CSTAT Expert group on microdata Access , May 31, 2012

Introduction

- **Access to government microdata now « on the table » everywhere**
- **With specific focus on:**
  - ➢ **highly detailed (confidential, sensitive …) microdata**
  - ➢ **Across borders**
- **Pressure comes from the researchers, yet researchers are not the only actors claiming for more access (economic actors, local authorities, international organizations, non governmental organizations …)**
  - ➢ **Frontiers and reasons for frontiers not always clear**
- **Increasing number of initiatives/projects**
  - ➢ **different objectives (aggregate data, PUF, SUF, confidential microdata…) and perimeters (kind of users, national/European/international …)**
  - ➢ **increasing need for information and coordination**
- **Aim for this presentation**
  - ➢ **Objectives and perimeter of *Data without Boundaries* project**
  - ➢ **1st year achievements and some first lessons**

Outline:

1) **DwB project in a nutshell**

2) **Context**

3) **Main objectives**

4) **Perimeter**

5) **Facilitate data discovery. 1st year achievements and some first lessons**

6) **Facilitate transnational accreditation. 1st year achievements and some first lessons**

7) **Improve transnational access. 1st year achievements and some first lessons**

8) **Enlarge discussions to the whole communities**

   Conclusions

1. DwB project in a nutshell

- **A four-year EC funded FP7-13 project (2011-2015)**

- **Linking the capacity of the research community with the important resources of the government microdata in Europe**

  - **Surveys and administrative datasets, combined files**

  - **Focus on confidential (highly detailed) data**

  - **Focus on crossing national boundaries within the ERA**

- **An essential step towards a European Research Infrastructure for the social sciences**

  - ➢ **Mechanism = Coordination of existing infrastructures**

  - ➢ **CESSDA (Data Archives) and the ESS (NSIs, NSAs, CB coordinated by Eurostat)**

  - ➢ **Based on volunteers : 27 partners of which 1/3 are NSIs/NSAs (ONS, CBS, INSEE, SORS, IAB, SCB, DESTATIS, CSIC, CNPS-INS), 1/3 Archives and 1/3 Universities involved in methodology about SDC**

  - ➢ **A method : Propose standards, implement pilots with the objective of enlarging cooperation to the whole communities**

SEVENTH FRAMEWORK
PROGRAMME

DwB

2. Context : general improvement …

- **At national level**
  - ➤ **Researchers access to government microdata, including highly detailed ones, has improved (particularly within the ERA), though still uneven in some countries**
  - ➤ **Progress in the " sharing data" culture and mutual understanding and trust**
  - ➤ **Strong cooperation in some countries between Data Archives and NSIs**
  - ➤ **Changes in the legal frameworks in many countries**
  - ➤ **New technologies and development of RDCs (remote access), some providing access to foreign researchers**

- **At European level**
  - ➤ **The ESFRI roadmap and the CESSDA ERIC process to build a RI for data access including government microdata : DwB project as a contribution to that process**
  - ➤ **Other projects and initiatives within the framework of the European Statistical System (Discussions about a new EC Regulation about researchers access to the European (integrated) microdata expected about 2012/2013; several ESSnet)**

- **At international level**
  - ➤ **Other initiatives (WDA, OECD …) yet less structured**

Yet …a lot of issues unsolved for trans-national access

➢ **Due to differences**

- ▪ **different providers (NSIs, Archives),**
- ▪ **different metadata standards,**
- ▪ **different legal frameworks, criteria and institutional arrangements for accreditation,**
- ▪ **different terminology (PUF, SUF, anonymized, highly detailed …)**
- ▪ **different modes of access (no access, safe centres, remote execution, remote access)**
- ▪ **different views about security, anonymization, output checking,**
- ▪ **different languages**

➢ **Some hard issues :**

- ▪ **who can sue in case of breach in another country ?**
- ▪ **what about running a single analysis on data from different countries**

3. Main objectives of DwB

- **Facilitate discovery of data**
    - **Building a central European point of access where a researcher can discover what are the available OS microdata (national and European), where and how they can be accessed**
    - **Proposing a metadata schema compatible with SDMX (NSI current standard) and DDI (CESSDA Archives standard) and usable for harvesting NSIs metadata**
    - **Servicing the use of OS data: providing tools (format, routines for harmonization), training the researchers for the use of the data, ensuring feedbacks to the producers**

- **Facilitate transnational accreditation particularly for comparative projects,**

    - **Proposing a standard for a European accreditation across borders and suggesting changes in the legal frameworks**

- **Improve access to OS microdata:**

    - **Proposing an architecture for a European remote access network (decentralised)**
    - **Implementing a pilot building on current possibilities with 3 RDCs**
    - **Immediately supporting transnational access via open calls for researchers with 7 RDCs in 4 European countries**
    - **Improving methodologies for managing risk of access to data**

4. DwB perimeter

- **All kind of OS microdata:**
  - ➢ **surveys, administrative microdata (households, business data, fiscal data..)**
  - ➢ **From NSIs, NSAs, Central banks, other government agencies**
  - ➢ **Focus on highly detailed microdata and attention to combined files**
- **Europe**
  - ➢ **Yet building links with non European**
- **Focus on crossing borders for national microdata**
  - ➢ **Yet some issues include European microdata (providing improved metadata) and continuous coordination with Eurostat (and ESSnets)**
- **For the researchers**
  - ➢ **Part of the construction of a European data infrastructure for the social sciences (see the CESSDA ERIC process)**
  - ➢ **Yet questions about how to define a researcher, a research institution**

SEVENTH FRAMEWORK
PROGRAMME

DwB

5. Facilitate discovery of OS microdata
1st year achievements

What do we have so far :

- **A framework for a European Service Centre for OS microdata (ESC-OS), ideally as a unit of the future CESSDA ERIC, yet with a specific governance including the NSIs (report)**
  - ➢ **Can be seen as a framework for the future, ensuring the steps made during DwB won't be lost and that the different activities will continue.**
  - ➢ **To be discussed with main actors (Eurostat, NSIs, CESSDA …) and updated during the project time**
- **A metadata schema (with a hierarchical structure : series, studies, datasets, variables) to be used for collecting, structuring and coding the information on available OS microdata**
- **A first evaluation of the disparate body of metadata on Official Statistics (via a 1st questionnaire to NSIs)**
- **A first draft for an object model incorporating DDI and SDMX to be used to harvest NSIs metadata**

Next steps:

Collect and structure the information on main surveys and administrative data for each country (a 1st perimeter has been determined) and start structuring metadata and user friendly routines for Eurostat data

Discuss with NSIs how to access their metadata

5. Facilitate discovery of OS microdata
Some first lessons

- **About the metadata base on OS microdata:**

  ➢ **What degree of centralisation ? Avoid duplicating efforts (where detailed metadata are provided), yet in a context where available information is unequal**

  ➢ **Controlled vocabularies to use ?**

  ➢ **Avoid building silos : will also feed the CESSDA catalogue where researchers can discover all kinds of data**

  ➢ **Translation as a crucial issue**

- **About the object model for harvesting NSIs metadata**

  ➢ **Needs for continuous coordination with on-going developments about metadata standard issues**

  ➢ **Needs to follow up on NSIs projects (2nd survey to be launched)**

  ➢ **Needs for more interviews about users (researchers) needs**

SEVENTH FRAMEWORK
PROGRAMME

DwB

6. Facilitate transnational accreditation within the ERA
1st year achievements

What we have so far:

- **An almost complete database on existing practices in transnational accreditation within Europe**
  - **To be made available for external use (in coordination with the metadata base)**
  - **Will need to be updated (by NSIs ?)**
- **An analysis of similarities on which to build new solutions for transnational access complying with the legal frameworks for research data access**
- **A first analysis of inefficiencies based in the law for providing transnational access**

Next steps:

Design and consult NSIs on a transnational accreditation standard

SEVENTH FRAMEWORK PROGRAMME

DwB

Data without Boundaries: Objectives, 1st year achievements and 1st lessons
— Facilitate transnational accreditation within the ERA
— First Lessons

6. Facilitate transnational accreditation within the ERA
Some first lessons

- **Accreditation practices : Important and encouraging similarities in basic principles**
  - ➢ **Yet differences in implementation**
  - ➢ **Terminology heterogeneous**
  - ➢ **Some important pending issues:**
    - ➢ **Who is a researcher ? What is a research institution (for instance what about institutional organizations, private/public …) ? Different answers (see the EDAF)**
    - ➢ **Whether to continue relying on institutions to ensure a researcher is "safe"**

- **Legal frameworks : Positive interpretation for transnational access in some countries of the silence of the law**

- **How to move for simplifying transnational access for comparative research**
  - ➢ **Different paths: bilateral/multilateral agreements avoiding multiple accreditation once accreditation is given in one country vs a unified accreditation managed centrally**

SEVENTH FRAMEWORK PROGRAMME

DwB

7. Improve transnational access to OS microdata
a) Towards a European remote access network
1st year achievements

What we have so far :

- **A 1st detailed overview (via a questionnaire) about solutions used by RDCs providing remote access (will need to be updated)**

- **An analysis of similarities and differences to be used for the transition into a remote access network**

- **First discussions about an architecture for transition of separate Remote access centres into a remote access network (decentralised)**

- **Preliminary discussions for implementing a pilot with 3 RCDS (IAB in Germany, SDS in UK, GENES in France) that will produce first answers for the following topics :**

  - **Legal issues consequences**
  - **Technical compatibilities**
  - **Contracts needed ? (what is needed to cooperate?)**
  - **Organizational issues (how to structure the workflow?)**

Next steps:

All tasks to continue

Feedbacks from researchers (interviews)about needs and solutions

7. Improve transnational access to OS microdata
a) Towards a European remote access network
Some first lessons

- **It was necessary to agree about terminology :** remote access was defined as "the researcher can actually see the data"

- **Overview of remote access solutions shows that solutions differ only slightly yet building a remote access network is still not an easy task. Open issues:**
  - **Technical solutions differ: reasons ?**
  - **3 different data access versions (result of legal restrictions)**
    - ❑ **From anywhere**
    - ❑ **From an institution (open question: how is it defined?)**
    - ❑ **From a safe centre (what if not in the premises of an NSI ?)**

- **Transition of separate Remote Access Centres into a Remote Access Network**
  - **A decentralised architecture vs a centralised one**
  - **How to allow combining data from different countries ? Virtual space where the projects could be hosted ?**
  - **A number of open issues :**
    - ❑ **legal issues for hosting projects in a virtual space (in Eurostat ?) while data storage and management remain in the countries**
    - ❑ **Workplace to access data (anywhere, institution (ie University), safe centre**
    - ❑ **Responsibility for the network …**

7. Improve transnational access to OS microdata
b) Immediately support transnational access
1st year achievements and some first lessons

7 RDCs from 4 countries (Germany, UK, France and Netherlands) offer transnational access to confidential (highly detailed) microdata via open calls and selection by a User Selection Panel

What we have so far:

    1st call received 24 applications/proposals

❑   **4 rejected and 10 invited to resubmit after RDC check**

❑   **Of 13, 10 passed to the USP, 10 successful**

Next steps:

Continuous calls twice a year

Some first lessons:

➢   **Important challenges for administration, communication and co-ordination between RDCs**

➢   **Demonstrate**

❑   **Needs for centralised and standardised information about data**

❑   **Important language issues`**

❑   **Multiple accreditations and training problematic**

❑   **Needs for combining data from different countries**

SEVENTH FRAMEWORK
PROGRAMME

DwB

7. Improve transnational access to OS microdata
c) Improved methodologies for managing risks of access to data
1st year achievements and some first lessons

What we have so far:

- **A number of presentations/papers/reports on different issues**
  - ➢ **Selection of risk assessment software**
  - ➢ **Microdata protection**
  - ➢ **Tabular data protection**
  - ➢ **Experimental risk assessment via record linkage**
  - ➢ **Guidelines for output checking (just starting)**

Next steps: **Work will continue**

Some first lessons

How far these results meet the researchers' needs ?

Need to involve more the users (researchers) in discussing the outputs.

8. Enlarge cooperation to the whole communities

- **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**

What we have so far :
- **A 1ˢᵗ European Data Access Forum (Luxembourg, March 2012)**
  - ➤ **Attendance and inputs from DwB, from European Archives, NSIs, Eurostat, ESAC, ESGAB, central banks, researchers, EC, non European NSIs/RDCs, researchers**
  - ➤ **Most NSIs attended the Forum, some willing to join the project**
  - ➤ **Non European were invited**
- **2 additional sessions have been organized**
  - ➤ **Metadata issues in Gothenburg (dec 2011)**
  - ➤ **Workshop on accreditation in the Eastern countries in Bucharest (jan. 2012)**
- **1ˢᵗ training session for the use of OS microdata July 2012 Mannheim**

Next steps
- **1ˢᵗ regional conference on data access (Eastern countries) in Lubiana (Spring 2013)**
- **1ˢᵗ Users conference Mannheim (Autumn 2013)**
- **Staff visits in RDCs to start in order to promote remote access solutions**

Some conclusions

- **Positive developments yet a lot of pending issues to be solved**

- **Need for continuous attention to on-going changes at national and European level**

  ➢ **See on-going changes at national level in legal frameworks, accreditation**

  ➢ **On-going changes in regulations at European level (European regulation about research access to European microdata still pending, new European regulation on data protection, future European regulations on metadata standards, on integration of social surveys…**

- **Need for continuous attention to parallel initiatives and projects (ESSnet, EC funded projects, international initiatives/projects)**

- **A European project facilitated by the European context yet a basis for building solutions to be useful in other contexts.**

Thanks for Listening

Contact: roxane.silberman@ens.fr

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