

# The Need for Networks on Data Access

*Data Without Boundaries project  
and the Workshop on Data  
Access*

Roxane Silberman / CNRS-Réseau Quetelet

Stefan Bender / IAB

Anco Hundepool / CBS



*Tarragona, SDC-worksession, 26-28 Oct 2011*



# Outline

- 1) Main issues**
- 2) WDA initiative**
- 3) DwB Project overview**
- 4) A focus on Statistical disclosure control issues in DwB**
- 5) Future DwB Events 2012**

# 1. Main issues

- **Access for research purpose to anonymized official statistics more widespread even if still uneven**
- **Does not meet researchers needs**
- **New solutions for access to highly detailed (confidential) microdata : on site access, remote execution, remote access in RDCs**
- **Comparative research require access across borders to datasets located in different countries**
  - **Information fragmented**
  - **Access across borders not allowed in many countries**
  - **Multiple accreditation**
  - **No standard environment**
  - **Language issues**

## 2. WDA initiative – General idea I

The aims of the Workshop on Data Access (WDA) are

- to review access to confidential government data for research purposes
- to bring together researchers, share knowledge, outline a future research agenda and build up a network for future activities.

# WDA – General idea II

## “Bottom Up” Approach

- **Self-organized by (a group of) RDCs/Archives**
- **Topics are generated by the Archives/RDCs**
- **“Working base”, highly interactive**
- **Asia, Africa Australia/NZ, Europe, North America**

## WDA – History

- **1<sup>st</sup> August 2007 organized by IAB in Nuremberg**
- **2<sup>nd</sup> February 2009 organized by Felix Ritchie (ONS) in Newport.**
- **3<sup>rd</sup> May 2009 organized by ICPSR (Felicia LeClere), ISR (Maggie Levenstein) in Ann Arbor**
- **Now: WDA is part of DwB (WP4) and 4<sup>th</sup> Workshop will be spring 2012 in Nuremberg.**

# 3. Data without Boundaries (DwB) project

- **A four-year EU-funded FP7-13 project (2011-2015)**
- **Linking the capacity of the research community with the important resources of the official micro data in Europe**
- **Mechanism = Coordination of existing infrastructures**
  - **CESSDA Data Archives, and the ESS (NSIs coordinated by Eurostat, ECB)**
  - **Based on volunteers**
- **27 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, CIS (IPUMS) + MT (SME)**
  - **Coordination Roxane Silberman (CNRS/Réseau Quetelet)**

# 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
- **RDCs to RDCs :** how to build a European distributed remote access that meet researchers needs to access confidential microdata across borders : technical, standardization and methodological issues
- **Servicing the use of OS data:** provide tools (format, routines for harmonization), train the researchers for using European micro data
- **Immediately offer transnational access**



# 3 Blocks, 12 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)
- **RDCS to RDCs :** discuss and agree on security standards, architecture for a distributed remote access (WP3 and WP4)
- **Methodological issues:** (WP11)
- **Implement a pilot with 3 RDCs (IAB in Germany, SDS in UK, GENES/INSEE in France) (WP4)** building on current possibilities for national micro-data and 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 with **7 RDCs** (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 offering remote access**

## 4. A focus on Statistical disclosure control issues in DwB

### Work Package 11 (SDC)

#### Partners:

- CBS (Anco Hundepool, Jan Mol)
- URV (Josep Domingo-Ferrer)
- UPC (Jordi Castro)
- Destatis (Sarah Giessing, Chris. Gürke)
- ULL (JJ Salazar)
- ONS (Felix Ritchie)
- CISC (Vicenc Torra)
- SOTON (Nathalie Shlomo)
- SORS (Andrejka Smukavec)

# Work Package 11 (SDC) tasks

1. Selection/overview of statistical software tools
2. Statistical methods for dataset design to maximize utility whilst keeping disclosure risks low (micro data)  
*Synthetic data, Microaggregation, Risk/Utility*
3. Statistical software tools and methods for output design to remove residual disclosure risk in research results (tabular etc)  
*flex. tabulation, Cell suppression, CTA, Open Solvers*
4. Testing the results in real life situations  
*Record linkage techniques*
5. Producing standards and guidance (output checking)  
*Improved guidelines*

## 5. Future DwB events 2012

- **January 2012 : 1st call for transnational access to highly detailed microdata:**
  - France, Germany, Netherlands, UKDA RDCs
  - Give support for comparative research projects that required transnational access
  - Business cases for a European accreditation and a distributed remote access
- **March 27-28, 2012 in Bruxelles** 1<sup>st</sup> European Data Access Forum
  - “Legal framework and research accreditation in Europe”

# Thanks for Listening



Contact: [roxane.silberman@ens.fr](mailto:roxane.silberman@ens.fr)

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

**WDA**

Contact: [stefan.bender@iab.de](mailto:stefan.bender@iab.de)