

# Building Together a European Access to Confidential Microdata

## Main issues and recent initiatives

Roxane Silberman

Réseau Quetelet and CCDSHS

DwB coordinator

Kloster Johannisberg, June 17<sup>th</sup>-18<sup>th</sup> 2010

# Summary

- Context
- Main issues
- The DwB project

# 1. Context

- Increasing pressure on the researcher side for access to confidential microdata but also changes in needs
- More awareness on Eurostat and NSIs side and a global trend towards improvements
- Yet, far from a solution for the current and future needs of researchers

# Increasing pressure from researchers side but changes

- Highly anonymized data less and less fit researchers needs
  - Increasing level of anonymization
  - New statistical tools require detailed data
- Crossing borders even worse but crucial
  - Researchers increasingly want to compare data from different countries, European and non European
  - Researchers increasingly work together in different countries, within Europe and also outside Europe
- Different types of data
  - Less surveys, more administrative data, fiscal data ...
  - More longitudinal data, more complex datasets
- Combining data is an increasing need and practice
- Are we dealing with current and future needs or with past needs ?
  - Increasing number of researchers, not only surveys, not only within Europe

# More awareness on Eurostat and NSIs side

- Access to official statistics both anonymized and confidential still uneven in Europe at national level
- Access to Eurostat datasets even highly anonymized still burdensome
- But more awareness of researchers needs and a global trend towards improvements
  - Legal frameworks : changes at national levels in several countries, towards a change at European level
  - New technical solutions : several countries have set up or are setting up safe centres and increasingly remote access
- Discussions and initiatives for a distributed European access and possible cooperation with Data Archives (CESSDA)

## Yet, far from a solution for the current and future needs of researchers

- Changes in the legal frameworks still uneven across Europe
- Researcher accreditation highly dependent on the national legal framework but also on institutional arrangements and practices
- Standards for anonymization differ, level tends to increase
- Mode of access ranging from : no access, safe centres , remote execution to remote access
- Providers may be NSIs, statistical departments, government agencies, Data archives, other actors
- Metadata not always available
- Language issues

## 2. Crucial issues

- Confidential data : what are we talking about ?
  - Different views, different practices, lowest common denominator ?
  - Impact the level of the pressure for access to confidential data
- Solutions for a European distributed access
  - Different national context and arrangements
  - Need for a flexible system
- Checking outputs
  - Different views : systematic or not, methodological solutions, legal solutions
- Metadata and language issues
- Cooperation with Data Archives (CESSDA)
  - Sharing the burden, sharing knowledge
  - Single point of access : towards a European research infrastructure
- Building for the future : different types of data, increasing number of providers, non only Europe

# 3. Within this context : The DwB project

- Data without boundaries
- A FP7- 13 EC call
- Coordination of existing infrastructures
  - CESSDA, network of data archives across Europe providing access to data for researchers
  - NSIs and Eurostat in charge of official microdata
- Enhancing researchers access to official microdata in Europe and bridging different communities
- Focus on confidential data
- A major step toward a European research infrastructure (within the context of the ESFRI roadmap and the CESSDA ERIC process)



# Main issues

- Build a central point of access: what data are available ? how can I access them ?
- Legal issues and accreditation : Towards a European accreditation
- Metadata standards and interoperability: NSIs tend to use SDMX, CESSDA Archives use DDI
- Servicing the use of OS data : provide tools (format, routines for harmonization), train the researchers for using European microdata
- Data protection and methodological issues : anonymization, output checking
- Technical and standardization issues in developing a European distributed remote access : propose and implement a test case

# Structure of DwB

- Block 1 : Access facilities
- Block 2 : Front office : Portal and metadata standards
- Block 3 : Preparing data access

# Block 1 Access

- Legal issues and accreditation
  - Examine current national accreditation
  - Agree on best practices and common standards
  - Examine how they can be implemented within the existing legal framework
  - Propose and test an accreditation pilot for the DwB project
  - Suggest changes in the legal framework
- Technical issues in developing RA and SDC procedures
- Standardizing RA and SDC procedures
- Developing metadata and routines for OS databases
- Implement a test case for a distributed European remote access building
  - a) on current possibilities for national microdata
  - b) on the accreditation pilot,
  - 3) conditional to a change in 2012 on new possibilities for Eurostat microdata

## Block 2 Front office

- Provide a single point of access within the context of the CESSDA portal and the current CESSDA ERIC process
- Discuss standards and develop tools to harmonize metadata
- Devise techniques to harvest NSI metadata by CESSDA
- Improve metadata, formats and provide routines for OS and Eurostat microdata

# Networking

- Bridging different communities
- All WPs in cooperation
- Involve all non partners NSIs, Archives, Eurostat, researchers : annual European Forum, regional workshops or conferences, thematic workshops, users conferences

## Where we are ?

- Positive evaluation of the project
- Revision of the proposal and negotiation with the EC
- Partners
  - CESSDA partners
  - NSIs and other providers of official statistics: ONS, CBS, INSEE, SORS, SCB, IAB, DESTATIS ... more ?
  - Universities, IPUMS ...

## Conclusion

- Build trust and cooperation
- Build a pilot and provide a model
- Keep in mind other initiatives, possible changes over the period and build within this context