

Research Infrastructure for Official Microdata in Europe: Data without Boundaries and Beyond

Roxane Silberman (CNRS-RQ) and Christof Wolf (GESIS)

Presented by Christof Wolf

*DwB Standing Committee member and Scientific Director at GESIS – Leibniz Institute for the Social Sciences
International Year of Statistics, Luxembourg, 26-27 November, 2013*



Initial situation (pre 2011)

- Access to official microdata including access to confidential data has made **progress at national level** in many European countries; however, big differences remain
- For a researcher access throughout Europe remains **dispersed, fragmented, burdensome for discovery as well as for access and use**
- **Transnational access to data from official statistics remains difficult**, particularly for confidential data
- **Merging confidential microdata across borders** to run a single comparative analysis is **highly problematic**

The answer: Data without Boundaries

- A project funded by the European Commission under the 7th Framework Programme
- 28 partners from 13 countries representing National Statistical Institutes (NSIs), social sciences data archives from the CESSDA network and universities
- Project funded for 4 years, from 05/2011 to 04/2015



Data without Boundaries – Objectives

- Enhance researchers' access to official microdata in Europe, with a particular focus on confidential data and transnational access
 - Both integrated European data and national data (including register data)
- Build trust and cooperation between the ESS (Eurostat + NSIs, NSAs, ECB, CBs) and CESSDA (now set up as a European Research Infrastructure), as an essential prerequisite to the construction of a data infrastructure that will provide services for OS data in the future

How are objectives reached?

The work and work packages of DwB are organized into three blocks:

- **Front office:** Building a single point of access and metadata issues
- **Access to OS microdata:** accreditation, Eu-RAN feasibility, Secure Data Centre issues, calls for transnational access
- **Communication and training:** promote discussions, increase knowledge and reach agreements

Communication and training

We aim at fostering trust and cooperation between European Statistical System, CESSDA members and the research community as well as increasing knowledge and competencies

- European Data Access Forum
- Specialised workshops, e.g. metadata, Eu-RAN, support for Eastern countries
- European User Conferences for Official Statistics
- Training sessions
- Staff visits
- Website



Access to OS microdata

We aim at improving researchers' access to OS microdata by

- Giving overview of accreditation procedures and developing best practices which eventually should lead to a European standard
- Testing the technical feasibility of a European Remote Access Network and building a pilot network
- Developing procedures and routines used in the production of anonymised microdata as well as output checking for confidential microdata in Secure Data Centres
- Supporting transnational access to confidential microdata involving 7 Research Data Centres in Europe

Front office

We aim at improving researchers' access to OS microdata by

- Identifying and describing standards used by NSIs and CESSDA in the production of data (e.g. metadata, classifications)
- Developing a common metadata model
- Building test application to harvest model-compliant metadata
- Collecting, structuring and publishing metadata for OS microdata from Member States
- Collecting, structuring and publishing metadata for integrated European data from Eurostat and European censuses
- Setting up a single point of access for information on OS data aligned with the CESSDA portal

Selected midterm results

- Overview of [accreditation criteria and procedures](#) in the ERA shows important similarities opening paths for more harmonization and possible delegation
- Overview of [legal frameworks](#) demonstrates in many cases the “silence of the law” on transnational access opening path for progress in the interpretation of security requirements and for bilateral/multilateral agreements
- An [architecture for a European Remote Access Network](#) has been proposed and intensively discussed both between partners and with other RDCs and stakeholders
- [Pilot implementation](#) with a RDC in RDC approach is on the road
- First outputs of the [survey on researchers needs for remote access](#) points to the [importance of Virtual Research Environment](#) that would allow the researchers to work together and to the need to [combine microdata](#) from different RDCs [across borders](#)

Selected midterm results, continued

- [State of the art report on metadata standards issues](#) provides a clear overview of current developments and the potential of GSIM for capturing the whole lifecycle, including dissemination, of OS microdata
- A [metadata model](#) fitting SDMX and DDI has been developed
- Conceptual foundation has been laid and web based editor constructed for structured, searchable documentation of OS data: [CIMES database](#) currently containing metadata for 1,796 datasets from 22 European countries
- [Support to 27 projects requiring transnational access](#) to confidential microdata has been provided
- Two [hands-on training courses](#) have been conducted; first time campus files

... and more results from the second half of project ...



But how do we guarantee sustainability?

By establishing a

European Service Centre for Official Statistics

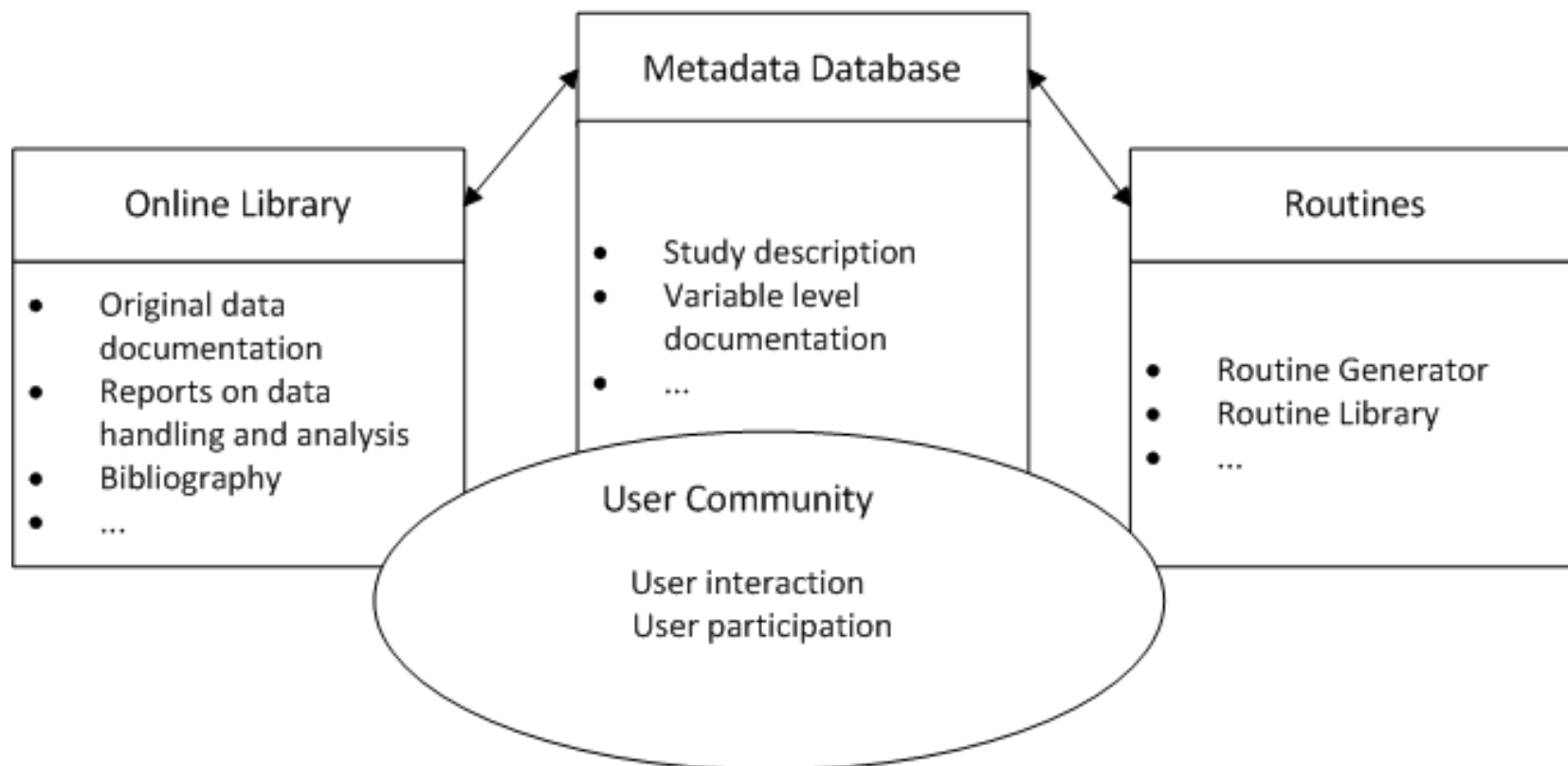


The Centre's Tasks

1. Establish a comprehensive web service for European OS-microdata & census data
2. Promote use of OS-microdata by hosting training courses, organizing user conferences and by incentivizing research
3. Improve data access
 - Provide information on access conditions, access sites and application procedures
 - Support expansion of transnational access networks and harmonization of conditions of use throughout Europe
 - Support researcher and the European Statistical System in accreditation
4. Support or manage (?) a European Remote Access Network



Task 1 – Web Service



Series by country

-  Austria
-  Belgium
-  Bulgaria
-  Cyprus
-  Czech Republic
-  Denmark
-  Estonia
-  Finland
-  France
- Training and vocational skills surveys
- Training and vocational skills survey - 1964
-  Germany
- The German Microcensus
-  Greece
-  Hungary
-  Iceland
-  Ireland
-  Italy
-  Latvia
-  Lithuania
-  Luxembourg
-  Malta
-  Netherlands
-  Norway
-  Poland
-  Portugal
-  Romania
-  Slovakia
-  Slovenia
-  Spain
-  Sweden
-  Switzerland
-  United Kingdom

Study

Training and vocational skills survey - 1964

Original Title :
Formation, Qualification Professionnelle - 1964

Original Alternative Title :
FQP 1964

Abstract :
The FQP 1964 survey (survey on training and vocational skills in France), conducted by INSEE between November 1963 and February 1964, is the first wave of a series which will result in the next 40 years, in a replication of 5 successive investigations: 1970, 1977, 1985, 1993 and 2003. It encompasses both labor force members, 15-65 years of age as of January 1st 1963, and out of the labor force individuals, aged 15-55 years at the same date. It provides information on respondents' occupational status (at the time of survey and 5 years before), on the schooling and vocational training history of individuals born in 1918 and after. The questionnaire consists of five parts: occupational position at the time of survey, changes in occupational status over the 5 years preceding the investigation, schooling and vocational education and training, estimation of occupational income. The survey was conducted on a representative sample of 25,598 individuals, with a count of 22 782 respondents.

Keywords :
EDUCATION
LABOUR AND EMPLOYMENT: Working conditions

Geographic coverage : France.

Kind of data : Survey data.

Documents :
- Questionnaire - Questionnaire - French (<http://www.cmh.ens.fr/greco/questionnaires/FQP/III-0001q.pdf>).

Date of Data Collection :
1963-11 -> 1964-03

Sep 26, 2012 [15:03]

Dataset

Training and vocational skills survey - 1964 - Scientific Use File (SUF)

Original name :
Formation, qualification professionnelle - 1964 -

Disseminating Identification number :
III-0001

Access conditions :

	File Possession	Onsite Access	Remote Access	Remote Execution
Public	FREE	N/A	N/A	N/A
Students	FREE	N/A	N/A	N/A
PhD Students	FREE	N/A	N/A	N/A
Researchers	FREE	N/A	N/A	N/A



Study

Variables

Change variable

Pos.	Name
17	PB150
19	PB160
21	PB170
23	PB180
25	PB190
27	PB200
29	PB210
31	PB220A
33	PE010
35	PE020
37	PE030
39	PE040
41	PL030
43	PL035
45	PL015
47	PL020

Variable Details

Values / Frequencies

EU-SILC / 2009 / Cross-sectional / Personal Data / PE040



Variable name PE040 ?

Variable label ?

Classification ?

Reference Period ?

Description Target Variable ?

Country Specific Comments ?

Other Comments ?

Thematic Classification ?

Unit of Observation Household Person ?

Filter ?

Is ad-hoc module variable ?

Is derived variable type ?



Study

Variables

Change variable

Pos.	Name
9	PB110
11	PB120
13	PB130
15	PB140
17	PB150
19	PB160
21	PB170
23	PB180
25	PB190
27	PB200
29	PB210
31	PB220A
33	PE010
35	PE020
37	PE030
39	PE040
41	PL030
43	PL035
45	PL015

Variable Details

Values / Frequencies

EU-SILC / 2009 / Cross-sectional / Personal Data / PE040



PE040: Highest ISCED level attained

Current variable

Statistics

Total (Missing) 23832 (0)

Value	Value Label	Frequency	Total %	Valid %
1	primary education	445	1.9	1.9
2	lower secondary education	3067	12.9	12.9
3	(upper) secondary education	11207	47.0	47.0
4	post-secondary non-tertiary education	1505	6.3	6.3
5	1st & 2nd stage of tertiary education	7608	31.9	31.9
	Valid Total	23832	100.0	100
	Total	23832	100	

Implementation

- The European Service Centre could be a subunit of CESSDA
- As a part of CESSDA the European Service Centre can benefit from existing knowledge and connections of the CESSDA network
- A governing board should be implemented in which representatives from the European Statistical System, the research community and other stakeholders assist the development of the service centre

In a nutshell

- The European Service Centre for Official Statistics would:
 - ✓ Service researchers using official microdata within the ERA
 - ✓ Help share the burden of increasing demand for high quality data and metadata between the CESSDA archives and the ESS members
 - ✓ Resulting in advantages from both sides
 - Increasing use of official microdata for research and for public policies evaluation
 - Input for data quality
 - Less costs for European data infrastructures and more integration of services, less silos

Want to learn more?



DwB Deliverable 5.1:

[Report on concept for and components of virtual service centre](#)




26.11.2013

Year of Statistics, Luxembourg



Project N°: 262608



DwB
Data without Boundaries

ACRONYM: **Data without Boundaries**

DELIVERABLE D5.1

*Report on concept for and components of
European Service Centre for official statistics*

WORK PACKAGE 5

Servicing European researchers in the use of OS microdata

REPORTING PERIOD:	From: Month 1	To: Month 18
PROJECT START DATE:	1 st May 2011	DURATION: 48 Months
DATE OF ISSUE OF DELIVERABLE:	April 30 th 2012	
DOCUMENT PREPARED BY:	Partner Numbers 5, 25,1	Partners Names GESIS, CED, CNRS-RQ

Combination of CP & CSA project funded by the European Community
Under the programme "FP7 - SP4 Capacities"
Priority 1.1.3: European Social Science Data Archives and remote access to Official
Statistics



19

Thank you for your attention!

Contact: christof.wolf@geis.org

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

