News from Nuremberg

4th Workshop on Data Access
(WDA)
2012/03/26, Luxembourg

Stefan Bender, IAB

David Schiller, IAB

Outline

1. Introduction

2. Data Access ways

3. Output control

4. Data documentation

5. Research networks

6. Conclusion/Outlook

The RDC of the BA at the IAB

- Established 2004
- Survey and administrative data for
- non commercial empirical research

Some figures per year:
- 110 new projects
- 1700+ job submissions
- 60+ publications by data users

The RDC of the BA at the IAB

Cooperation and innovation

- Focus on international cooperation
- Projects to force developments on
  - available data and (new datasets)
  - infrastructure for data analysis (access, documentation etc.)
- Interest in sharing thoughts about the future of data access

Data access ways

The FDZ provides three data access ways:

- Scientific Use Files
- Guest stay
- Job submission

Progress is made on all three of them.

More secure Scientific Use Files

- Scientific Use Files (SUF) contain less detailed microdata
- They can be used by researchers at there own PC worldwide.
- Data can be downloaded from the FDZ-webpage
- Still some security restrictions are important:
  - Data has to be kept secure
  - Data must be send back after the project duration
  - The data provider doesn't now what happens to his data

More secure Scientific Use Files

The Cornell Restricted Access Data Center (CRADC) provides a environment to work with data in a secure way.

- Data is stored at Cornell, the researcher can log in
- Statistical tools are provided by CRADC
- No output check needed due to SUF data structure
- After project duration the user account is shut down
- SUF for north-America will be provided via CRADC

RDC-in-RDC approach (guest stay)

- Project PFiF, 10/2010-09/2013

- High detailed microdata must stay at the facilities of the BA
- Accessing data remotely (seeing it on the screen) is treaded as unsafe
- Result: researchers are forced to travel to and stay in Nuremberg (costly and not easy to schedule)

RDC-in-RDC approach (guest stay)

- Access from within a safe centre (comparable with the one in Nuremberg)
- via a safe internet connection
- Data stays in the BA, researchers have less to travel -> Bringing the data to the people

- Currently five additional locations:
  - Berlin, Bremen, Dresden, Düsseldorf, Ann Arbor (USA)
  - …and more in Europe, US to come.

Job submission (remote execution)

- 1700 job submissions per year
- Code files as email attachment
- FDZ staff runs programs on the server
- A lot of manual labour

- The FDZ implements the JoSuA (Job Submission Application) tool from IDSC at IZA in Bonn.

Job submission (remote execution)

- Researcher can log in to a restricted workspace
- They upload their jobs (code files)
- The programs run on the FDZ-server
- Outputs are stored within the workspace, where researcher can access them (including older submissions).

- This results in a more automated workflow and less administrative work in the background.

Output control

- Output has to be checked for disclosure risk (guest stay and job submission)

- The FDZ will use the JoSuA functionality to develop the output control procedure

- This is mainly for the 1700 job submission per year

Output control - new procedure

Distinction between external and internal output

- Internal output (95% of overall output):
  - For research work
  - More detailed output (factually anonymity)
  - Access able via the JoSuA Workspace but no download
- External output (5% of overall output):
  - For publications
  - Less detailed output (absolutely anonymity)
  - Access able via the Workspace, download possible

Output control - Advantages

- Less output is send to the researchers (less straying data with unknown usage)
- Faster internal output (because no intensive check in needed)
- Internal output is closer to the guest stay output (less modified data due to anonymization)
- External output only needed for publication

- **In short: faster and safer output for projects**

Data documentation

- IAB is member of the DDI Alliance
- Useful and standardized data documentation is crucial
- IAB has advertised a bid for a comprehensive data documentation solution
- Deadline was March 2012
- International cooperation are very welcome
- More on that topic at EDDI 2012 in Bergen

Research networks

- The infrastructure for scientific research was intensively expanded in the last years

- Resulting in sophisticated solutions for single data providers, i.e. countries and/or disciplines
  - Remote Access
  - Job submission
  - Data documentation
  - etc.

Research networks

- The next steps are predefined by new developments:
  - Growing amounts of data from different sources
  - Statistical and technical solutions to merge data
  - Internationalization
- Network solutions that bring single data providers together are needed
- For defined projects as well as for a future infrastructure for empirical research

Virtual working environment (VirtAug)

- For the SOEB (Reporting on socioeconomic development in Germany) different partners work together
- Data sharing and joint calculations were work intensive and inefficient

- A virtual working environment should improve the flow of work

Virtual working environment (VirtAug)

Some basic information:

- Data providers and universities are involved
- Different kinds of data result in different data security issues, e.g.
  - data has to stay at the data provider facility
  - data can be stored centralized
- Cloud and Grid solutions are involved

Remote Access Network

- EC FP7 project "Data without Boundaries"
- WP 4 – Improve access to OS microdata

- First step: Two-way connections between three European data providers

  - SDS, UK ⇔ IAB, Germany ⇔ CASC, France

Remote Access Network

- Second step: Pilot for a European Remote Access Network
- What has to be covered:
    - Data storage issues (where is the data "located")
    - Different Remote Access solutions (under what circumstances is RA allowed)
    - Workplace to access data (anywhere, institution (e.g. university), safe centre)
    - Comparability with existing solutions.

Conclusion / Outlook

- FDZ is heading to "international access".
- A lot of activities, which are complex and are pieces of a greater puzzle.
- Need for experience sharing, coordination and cooperation between RDCs, researchers, archives, NSIs...

Thank you for your attention

Stefan Bender

stefan.bender@iab.de

+49 911 179-3082

David Schiller

david.schiller@iab.de

+49 911 179-8381

http://fdz.iab.de