

Proposing a Metadata Solution over Multiple RDCs in the German Context

EDDI 2013

03.12.2013, Paris

David Schiller, IAB

Ingo Barkow, DIPF

A wide range of data – the German context

<div> <div>Social</div> <div>Economy</div> <div>Education</div> <div>Health</div> <div>Psychology</div> <div>Other</div> </div>		
<div> <div>Social</div> <div>Economy</div> </div> <p>Research Data Center of the Federal Statistical Office (FDZ-Bund)</p>	<div> <div>Social</div> <div>Economy</div> </div> <p>Research Data Center of the State Statistical Offices (FDZ-Länder)</p>	<div> <div>Education</div> <div>Health</div> </div> <p>Research Data Center (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB)</p>
<div> <div>Health</div> </div> <p>Research Data Centre of the German Pension Insurance (FDZ-RV)</p>	<div> <div>Education</div> </div> <p>Research Data Centre of the Federal Institute for Vocational Education and Training (BIBB-FDZ)</p>	<div> <div>Education</div> </div> <p>Research Data Center (FDZ) of the Institute for Educational Quality Improvement (IQB)</p>
<div> <div>Social</div> <div>Economy</div> </div> <p>Research Data Center of the Socio-Economic Panel Study (FDZ-SOEP)</p>	<div> <div>Health</div> </div> <p>Research Data Centre ALLBUS at GESIS</p>	<div> <div>Other</div> </div> <p>Research Data Center International Survey Programmes at GESIS</p>
<div> <div>Health</div> </div> <p>Research Data Center Elections at GESIS</p>	<div> <div>Health</div> </div> <p>Research Data Center of the Survey of Health, Ageing and Retirement in Europe</p>	<div> <div>Education</div> <div>Health</div> <div>Psychology</div> <div>Economy</div> </div> <p>Research Data Centre of the German Centre of Gerontology (FDZ-DZA)</p>
<div> <div>Psychology</div> </div> <p>Research Data Center PsychData of the Leibniz-Institute for Psychology Information (ZPID)</p>	<div> <div>Health</div> </div> <p>Research Data Center of the German Family Panel (FDZ pairfam)</p>	<div> <div>Health</div> </div> <p>Research Data Centre Ruhr at the RWI (FDZ Ruhr)</p>
<div> <div>Economy</div> </div> <p>LMU-ifo Economics & Business Data Center (EBDC)</p>	<div> <div>Health</div> </div> <p>'Health Monitoring' Research Data Centre at the Robert Koch Institute (RKI)</p>	<div> <div>Education</div> </div> <p>Research Data Center German Microdata Lab (GML)</p>
<div> <div>Health</div> </div> <p>International Data Service Center (IDSC) at the Institute for the Study of Labor (IZA)</p>	<div> <div>Economy</div> </div> <p>Data Service Center for Business and Organizational Data (DSZ-BO)</p>	<div> <div>Education</div> </div> <p>Research Data Centre Education at the German Institute for International Educational Research (DIPF)</p>
<div> <div>Health</div> </div>	<div> <div>Social</div> <div>Economy</div> </div>	<div> <div>Economy</div> </div>

Research Data Centre in Germany

- Due to legislation, data has to be made available for scientific research
- The German Data Forum (for social sciences) accredits Research Data Centre
- Currently there are 26 Research Data Centre
- They offer a wide range of data
 - Social, Education, Economy, Health, Psychology, etc.
 - Survey data and administrative data

The good and the evil

- Data is there:
 - 26 organizations offer data for scientific research
 - They create or extract it and prepare it for research
 - According to that they are the experts for their data
- But what data is exactly where and how can it be reached?
 - 26 website offer information to the researcher
 - They have different structures and explain different things
 - They are not connected in a structured way

Researcher needs

- Researchers have to be supported in the way they work.
- First an idea for a research topic originates
- Then they look for appropriate data
 - E.g. content, “power” of data; how to work with it
- If they cannot find the information easily, they will likely skip the planned project
- A waste of the 26 data sources

Needs of data owners and funding bodies

- Data owners are normally experts in creating or extracting their data
- They are not all experts in data documentation
- Therefore they at least need tools to guide them through the documentation process
- Funding bodies are not interested in sponsoring 26 more or less equivalent infrastructures
- They are interested in cheap and efficient approaches

Single point for information

- The best thing would be one single point for all the needed information about research data
- Within this point the researcher can search in a structured way for:
 - Available data by topics
 - Detailed information about content and quality on variable level
 - Possibilities to combine data sources
 - Circumstances for accreditation and accessing data

Standard documentation

- Such a single point can only be established by using a standardized way of data documentation
 - First for the researchers to easily understand it
 - Second for the data owner to easily create it
 - Third for different tools to easily “work” with the documentation
- Thereby researchers and data owner shouldn't have to learn the documentation structure in the background, the tools should help them in an enjoyable way.

Obstacles

- No commonly used standard (DDI 20%); needs are
 - Standard, i.e. not moving to upgrades to often
 - Manageable coverage (profiles)
 - Practical exchange and storage format
- No central platform or access point in place
- No mechanism to pool (harvesting etc.) the documentations
- No organizational structure to node the data owners to a central platform

Ways to go

- DDI must proof that it is a practical standard
 - reliable standard, manageable, exchange and storage, connectable to other standards, supported by harmonized tools;
- Connectable tools have to be in place and easy to use
 - For people that are not developers the tool creates the standard by offering a comfortable working interface; the standards work invisible for the user in the background.
- Strong will to work within an organizational structure
 - Give services to others, by agreements and concentrate on the own area of expert knowledge (data creation)

Summary

- The German context
- Data is there
- A easy to use infrastructure to discover data is missing
- DDI has to position itself as the standard of choice by being practical
- Tools have to be harmonized
- I guess the discussed issue is not only a German one

Thank you for your attention

david.schiller@iab.de

barkow@dipf.de

<http://fdz.iab.de>

<http://www.dipf.de>