

Data Management, DDI-based Documentation and Visualization of Business and Organizational Research Data at the DSZ-BO

Johanna Vompras

University Library Bielefeld

Datenservicezentrum

Betriebs- und
Organisations**DATEN**

Dec 4th, 2012

EDDI2012 – Bergen, Norway

Session B1: *Infrastructure for Data Collection, Research, and Archiving*

Contents

- 1 DSZ-BO
- 2 Data Infrastructure
- 3 Technical Solutions and Tools
- 4 Summary

You have never heard about 'Bielefeld' ???



Population: 323.000

Location:
in the east of North Rhine-Westphalia
(Ost-Westfalen)

City naturally divided by the
Teutoburg Forest

Famous companies:
Dr. Oetker, Seidensticker, Schüco

University of Bielefeld



Contents

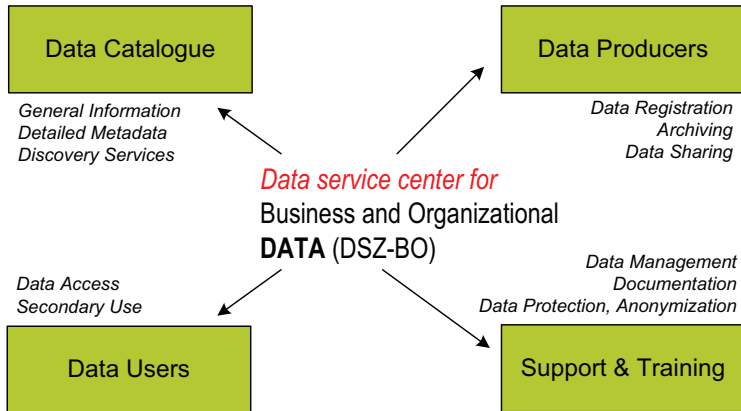
- 1 DSZ-BO
- 2 Data Infrastructure
- 3 Technical Solutions and Tools
- 4 Summary

Task and Scope of the DSZ-BO

Collect, archive, distribute and maintaining a catalogue of
Business and Organizational Data from the Social Sciences,
like ...

- Surveys with multiple organizations, e.g. interviews with human resource managers of different firms,
- Qualitative case studies and mixed methods,
- Process generated numbers, e.g. average time of patients in different hospitals, business catalogues,
- Observations, e.g. informal processes in one local office,
- Linked employer employee data (LEE).

DSZ-BO Services

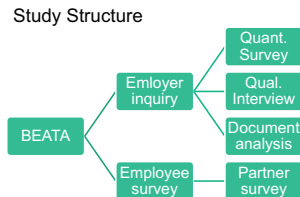


Contents

- 1 DSZ-BO
- 2 Data Infrastructure
- 3 Technical Solutions and Tools
- 4 Summary

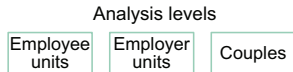
Specific Requirements

- Standardized Documentation: Mapping of "complex" study structure into DDI:

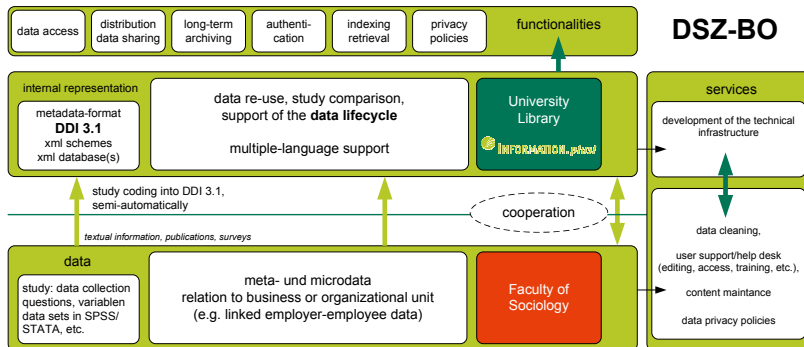


- Data Catalogue:

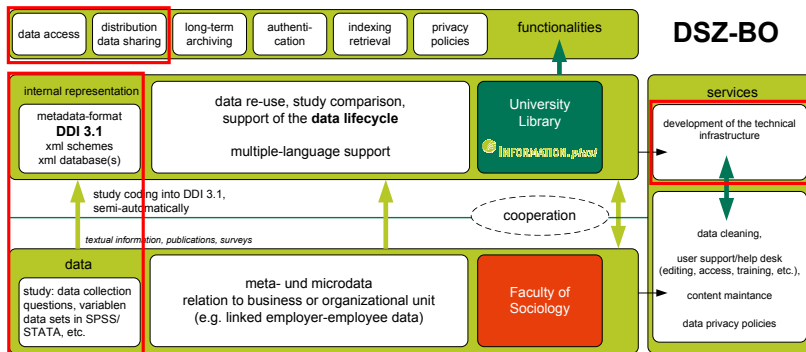
- Data with certain methods,
- Data which contains variables that operationalize certain research questions,
- Datasets with certain levels of analysis,
- Examples for good practice in a certain field.



Co-Operation DSZ-BO and Library Services



Co-Operation DSZ-BO and Library Services



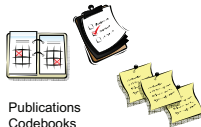
Contents

- 1 DSZ-BO
- 2 Data Infrastructure
- 3 Technical Solutions and Tools**
- 4 Summary

Documentation or putting "research" into a DDI ...

Study Investigation

- Expert Knowledge
- Systematic Literature Search and Selection
- Acquisition of Data from Data Producers



Publications
Codebooks
Documentations
(machine and non-
machine readable)
Questionnaires,
References
Statistical Data Files

Study Classification

Standards: A, B, C

- A: Own Research / Detailed Documentation
- B: Normal Documentation
- C: Minimal Documentation for Cataloguing and Registration purposes



Additional parameters for classification:

Kind of Data,
Data and Documentation Quality,
Traceability of Research Work,
Expert Knowledge,
Processing Effort

Study Documentation

using DDI v3



Specific Data Processing and Documentation Workflows



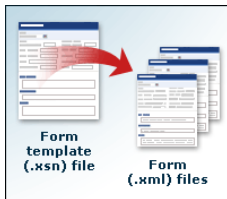
Examples:

- Conversion of SPSS to DDI for variable level documentation
- DDI-XML-Templates for A-Standard documentation
- XML Forms for C-Standard documentation

Tool 1: Editor for C-Standard Studies

Form template (.xsn)
provides a view on XML
data → fill it out!

- Form is compressed,
- XML Schemas, default XML data,
- XSLT files for view in the form,
- Script files and form definition files.



Interne Infos

Author:	md	bitte auswählen!
Datenbankname (BaseX-XML)	AuIF_s0063	Kurzname z.B. beata, albus, etc
Zeitstempel 1. Eintrag (*)	03.02.2012	wird automatisch ergänzt!
Zeitstempel letzte Änderung (*)	03.02.2012	wird automatisch ergänzt!
Datenursprung	AuIF	bitte auswählen!
Verweis auf Studie	Link auf Akquise DB	... aus der Akquise-Datenbank
Ablageort der Daten	Daten im DSZ-BO	

Studiendokumentation in DDI

1. DDIInstance

Version	1.0
Agency	Universität Bielefeld
Version Date	2012-02-03
Id (*)	DDI_Instance_s0063

2. StudyUnit

Id Studie:	s0063	z.B. s0001 für Allbus
------------	-------	-----------------------

2.1 Allgemeine Studieninformation

Title	Telekom. Wie machen die das? Die Transformation der Beschäftigungsverhältnisse
Kurztitel/Abkürzung	
Projektleitung	-
Finanzierendes Institut	DFG - Deutsche Forschungsgemeinschaft

Tool 2: Content Administration Backend

Inspired by researchers workflows for **Content Creation and Publishing**

Concentration of **all processing steps** needed to publish the 'DDI-Instance'

- 1 Upload and Archival of the DDI XML file,
- 2 Assignment of internal ID,
- 3 Storage of Metadata and XML Database Generation,
- 4 Preview and Publishing Options.


Universität Bielefeld
 Fakultät für Soziologie

Datenservicezentrum
Betriebs- und OrganisationsDATEN


Studie bearbeiten

oder

Attribut	
ID	2
StudyID	s0002
Titel der Studie	BEATA
Standard	A
Deaktiviert	0

Attribut	Wert
Datenbankname	beata
DDI-Datei	s0002_beata_v2.xml

DDI-Datei für Studie mit der ID 2:

Datei hochladen

C:\Standard > Metaanalyse

Organisieren ▾ Neuer Ordner

Zuletzt besucht

Bibliotheken

Bilder

Dokumente

Musik

Videos

Computer

SozLocaleDevice (C:)

Nichtverbundenes Netzlaufwerk (N:)

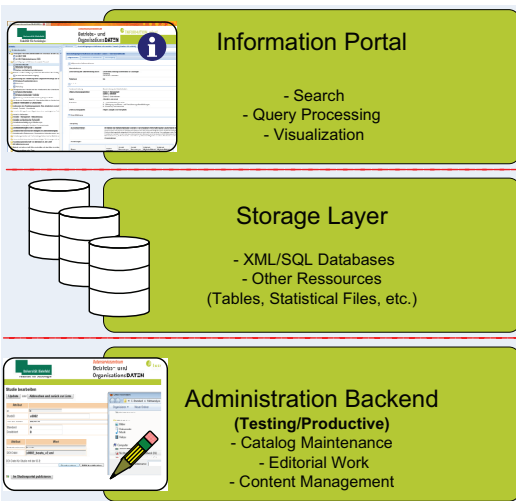
Nichtverbundenes Netzlaufwerk (O:)

Dateiname:

Technical Infrastructure: 3 Layers

Requirements:

- Easy to maintain (contents),
- Easy to extend (DDI model and queries)
- General approach for mapping into a visualization
- Data Encapsulation



Data Storage Layer

- Data storage:
DDI File: XML Database (BaseX)
Other Metadata: Relational Database
- Data Queries/Modification:
XQUERY/MySQL
- Data Visualization:
JavaScript Framework

PHP Script with XQUERY $\Rightarrow$ Results $\Rightarrow$ JSON $\Rightarrow$ Input for Components (e.g. GridPanel or DataView)

Data Retrieval and Visualization

- Each 'Study' is stored in a single XML database
- Queried by XQuery language, e.g. *'select all publications related to the study ALLBUS'*
- Results are returned as lists of items and transformed into JSON format

XQuery: find all publications related to ALLBUS study

```
FOR $node IN doc("allbus")//s:StudyUnit/r:OtherMaterial[@type='text']
RETURN
  ( CONCAT(
    data($node/r:Citation/r:Creator)," (" ,data($node/r:Citation/r:PublicationDate),"),",
    data($node/r:Citation/r:Title)),
    data($node/r:Citation/r:PublicationDate),
    data($node/r:ExternalURLReference)
  )
```

Visualization of structured JSON data

Components: Windows, (Tree)Panels, Tabs, (Grouping)Grid
Functions e.g. within GroupingGrid: *Sorting* and *Grouping*

The screenshot displays the website of the Datenservicezentrum (Data Service Center) at the University of Bielefeld, Faculty of Sociology. The header includes the university logo, the faculty name, and the center's name. A navigation bar contains tabs for 'Studien', 'Überblick', 'Publikationen', and 'TEST-GUI'. The 'Publikationen' tab is active, showing a list of publications under the heading 'Allbus Betriebsbefragung'. The list is organized by year (2009 and 2010) and includes details such as the title, authors, and a brief description of the research project. A 'Konsolle' (Console) window is visible at the bottom left, showing the text 'South'.

Universität Bielefeld Fakultät für Soziologie

Datenservicezentrum
Betriebs- und OrganisationsDATEN

INFORMATION.plus!

Studien Überblick Publikationen TEST-GUI

Allbus Betriebsbefragung

Titel

Jahr: 2009

Stefan Liebig, Christian Gerhards, Jennifer Elsner (2009): Datenhandbuch Projekt "Verknüpfte Personen-Betriebsdaten im Anschluss an den ALLBUS 2008" – ALLBUS Betriebsbefragung 2009

Stefan Liebig (2009): Interdisciplinary Longitudinal Surveys: Linking Individual Data to Organizational Data in Life-course Analysis (Working Paper No. 68)

Stefan Liebig (2009): Organizational Data (Working Paper No. 67)

Alexia Meyermann, Jennifer Elsner, Jürgen Schupp, Stefan Liebig (2009): Pilotstudie einer surveybasierten Verknüpfung von Personen- und Betriebsdaten. Durchführung sowie Generierung einer Betriebsstudie als nachgelagerte Organisationserhebung zur SOEP-Innovationsstichprobe 2007

Jahr: 2010

Christian Gerhards, Stefan Liebig (2010): Methodenbericht Projekt "Verknüpfte Personen-Betriebsdaten im Anschluss an den ALLBUS 2008" – ALLBUS Betriebsbefragung 2009

Konsolle
South

Functionalities of the Search

Search, Browse, Visualization

- Selection of studies, display of general study information
- Listings of data collections, questions, concepts, etc.
- Linking of data with questionnaires and publications, and other materials
- Search by keywords, or (thesaurus) concepts
- Filtering (e.g. by year, country, standard)

The screenshot shows the 'Datenservicezentrum Bielefeld' search portal. The header includes the university logo and 'Fakultät für Soziologie'. The main search area has a sidebar with filters: 'Zeitliche/Räumliche Einschränkungen' (Time/Space Restrictions) with a date range from 2006 to 2012 and a country filter set to 'Deutschland'; 'Volltext/Keywords' with a search bar and a dropdown for 'Keywords'; and 'Dokumentations-Standard' with a dropdown for 'Standard'. The main content area displays a list of studies, including 'Am Ende der Deutschland A', 'Antliche Firmendaten für D', 'Antinomien des Projektman', 'Arbeit - Technik - Geschlecht', 'Arbeit im Bürogroßraum. Er', and 'Arbeit in Gießereien'. Each study entry shows the project leader, study ID, and study period.

Contents

- 1 DSZ-BO
- 2 Data Infrastructure
- 3 Technical Solutions and Tools
- 4 Summary**

DSZ-BO: Experiences from technical point of view

To find a collection of *light-weighted* documentation tools tailored for specific scope of studies and data, which are *easy to learn* and *easy to operate*,

- Special requirements for afterwards documentation → decision about classification
- Adjustment to researchers' working workflows:
 - tools,
 - editors,
 - data processing, and
 - content authoring system.
- *WYSIWYG-like* information portal,
- Easy-to-use content management system, not only for technical staff.

Thank you! Questions?

Contact

Bielefeld University Library

Research Data Management Services and Infrastructure Projects

Johanna Vompras

johanna.vompras@uni-bielefeld.de

Data Service Center for Business and Organizational Data

dsz-bo@uni-bielefeld.de

Appendix: XQUERY Query Generation

Table *ddi_path*:

id	attribute	label	path	root	...
1	title	Titel	data(\$node/r:Citation/r:Title)	s:StudyUnit	
2	creator	Erstellt von	\$node/r:Citation/r:Creator	s:StudyUnit	
3	funding	Gefördert durch	for \$knoten in doc(\$database) ...		
...					
18	studyresults	Ergebnisse	NULL		
...					
144	realization.sampling_notes	Anmerkungen	\$node/r:Note/r:Content	d:DataCollection	

Example 'Funding':

```
FOR $knoten in doc($database)//a:Archive/a:OrganizationScheme/a:Organization
WHERE $knoten/@id = $node/r:FundingInformation/r:AgencyOrganizationReference/r:ID
RETURN
concat(data($knoten/a:OrganizationName),utilities:ifexistsPar(data($knoten/a:Nickname)))
```

corresponds to the relational statement:

$$\pi_{(a:OrganisationName)}(a:Archive/a:OrganizationScheme/a:Organization \bowtie_{@id==r:ID} r:FundingInformation/r:AgencyOrganizationReference)$$