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
OrganisationsDATEN

Dec 4th, 2012
EDDI2012 – Bergen, Norway

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

Contents

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

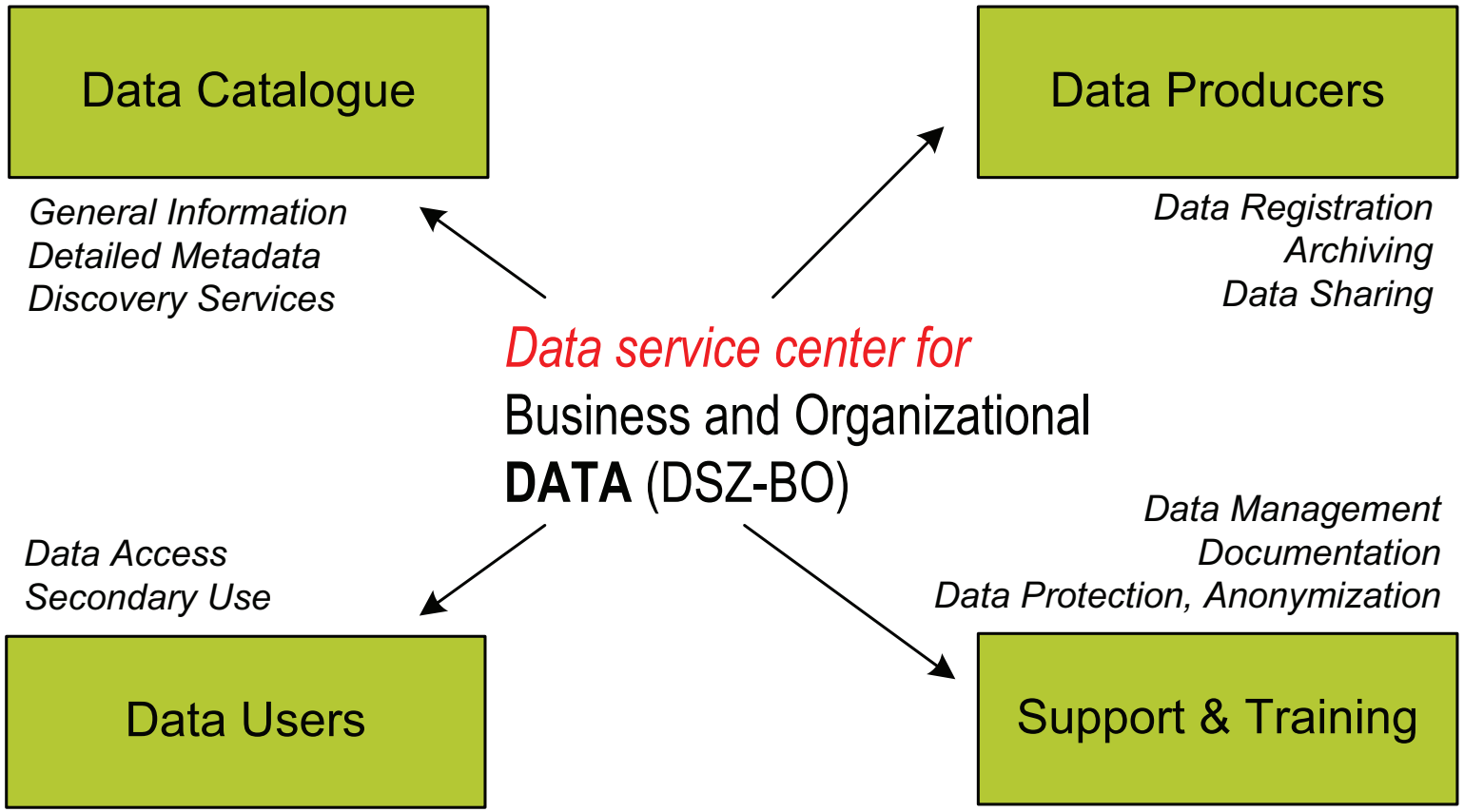
Contents

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).

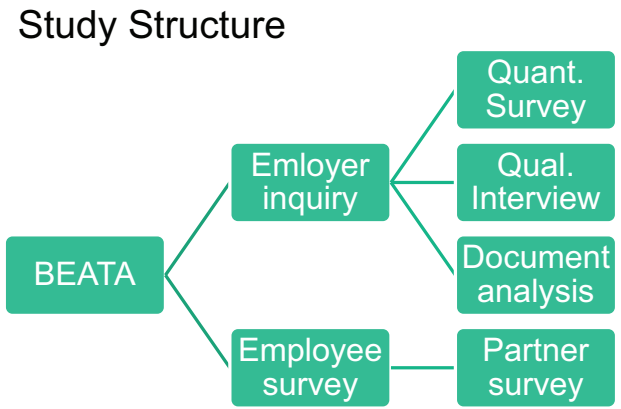
DSZ-BO Services



Contents

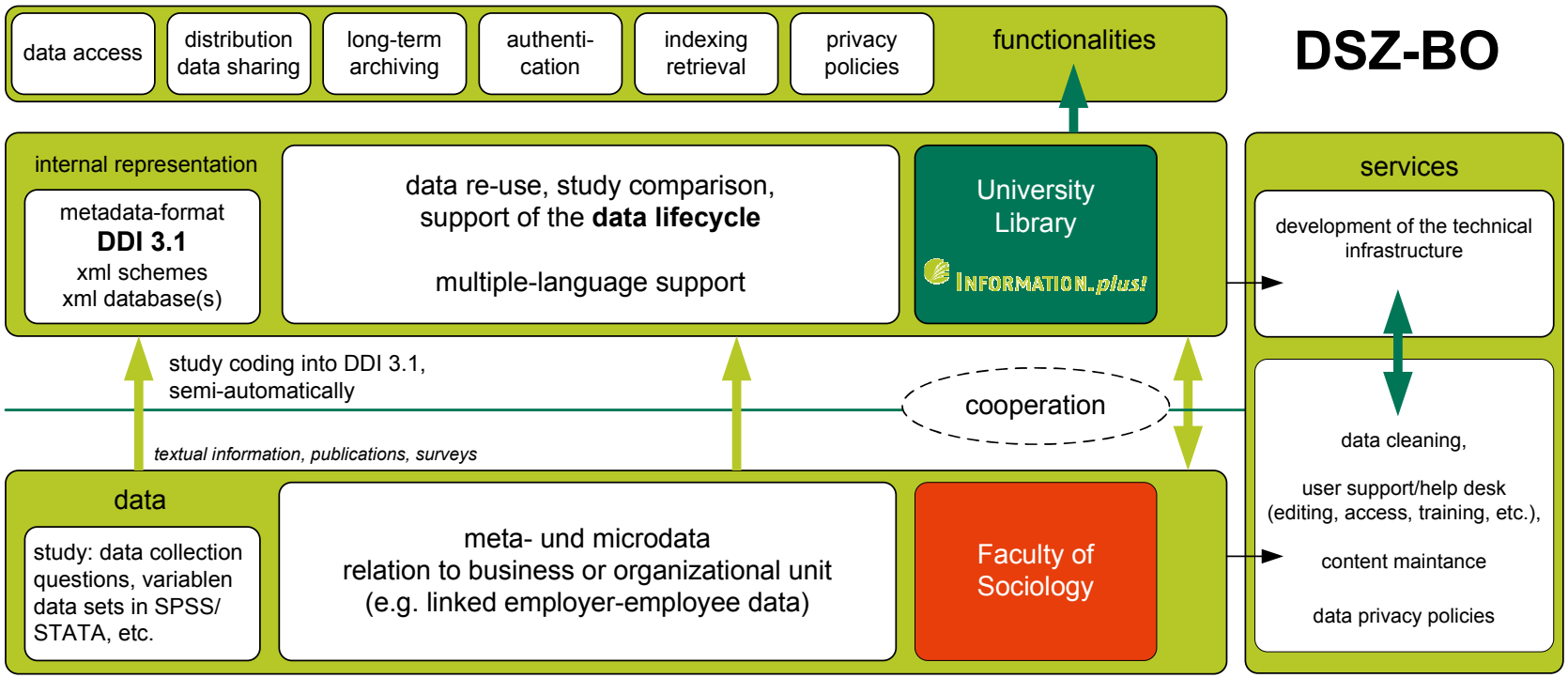
Specific Requirements

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

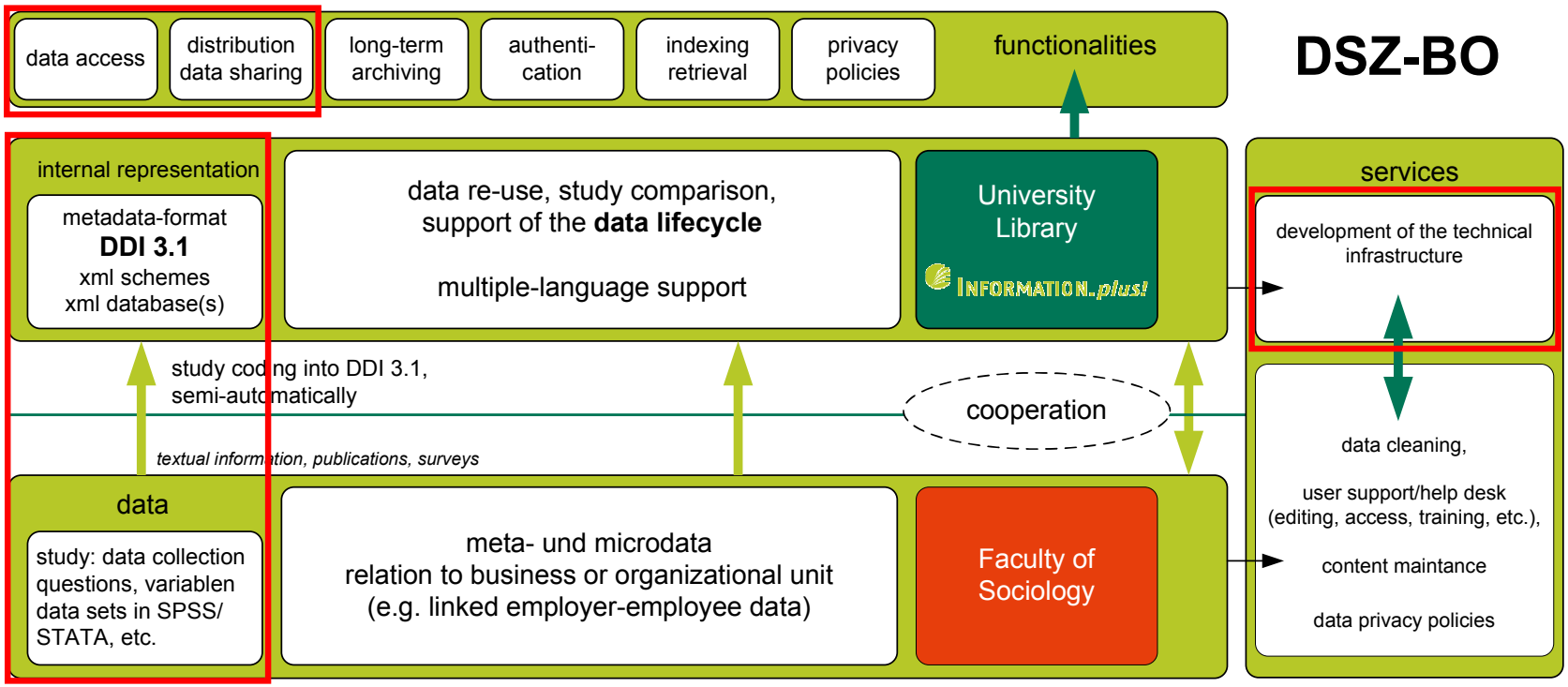


- Data Catalogue:
  - Data with certain methods,
  - Data which contains variables that operationalize certain research questions,
  - Datasets with certain levels of analysis,
  - Examples for good practice in a certain field.

Co-Operation DSZ-BO and Library Services



Co-Operation DSZ-BO and Library Services

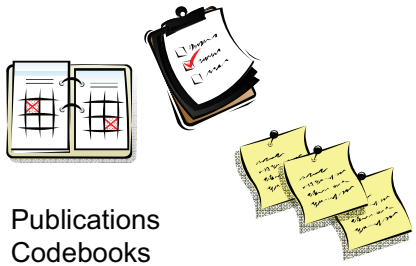


Contents

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

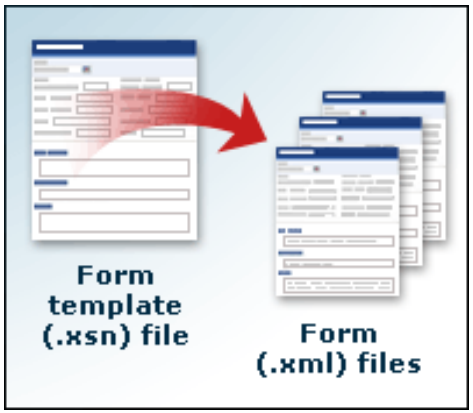
Examples:

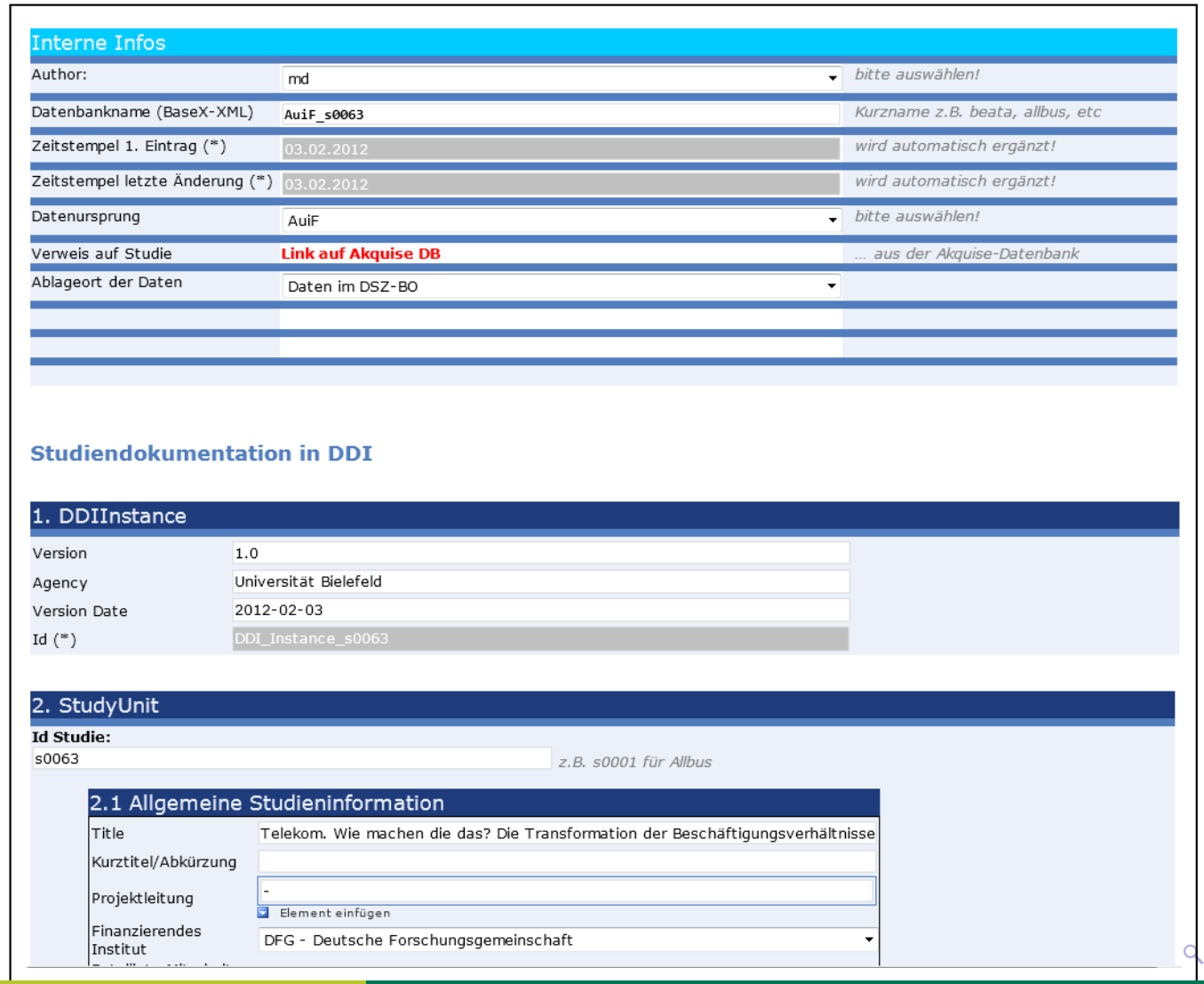
- Conversion of SPSS to DDI for variable level documentation
- DDI-XML-Templates for A-Standard documentation
- XML Forms for C-Standard documentation

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.



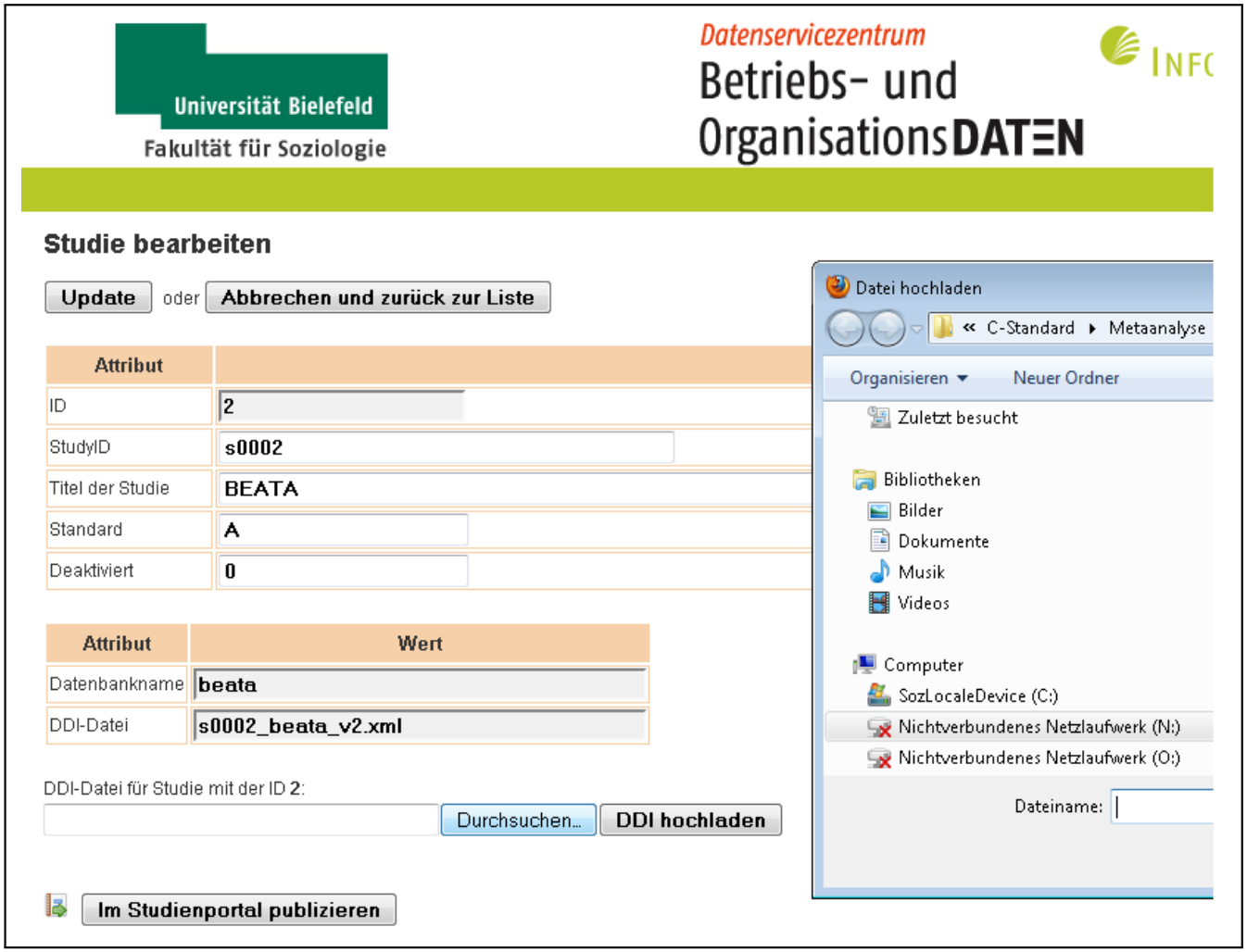


Tool 2: Content Administration Backend

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

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

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



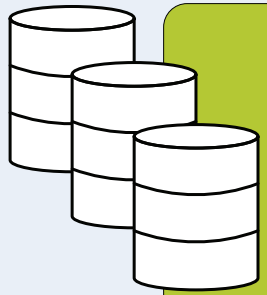
Technical Infrastructure: 3 Layers

Requirements:

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

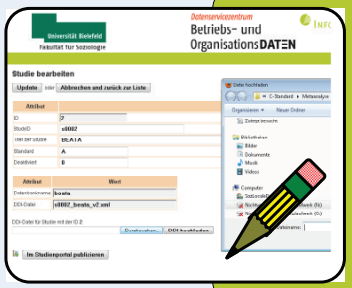
Information Portal

- Search
- Query Processing
- Visualization



Storage Layer

- XML/SQL Databases
- Other Ressources
(Tables, Statistical Files, etc.)



Administration Backend

(Testing/Productive)
- Catalog Maintenance
- Editorial Work
- Content Management

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 ⇒ Results ⇒ JSON ⇒ 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)
        )
```

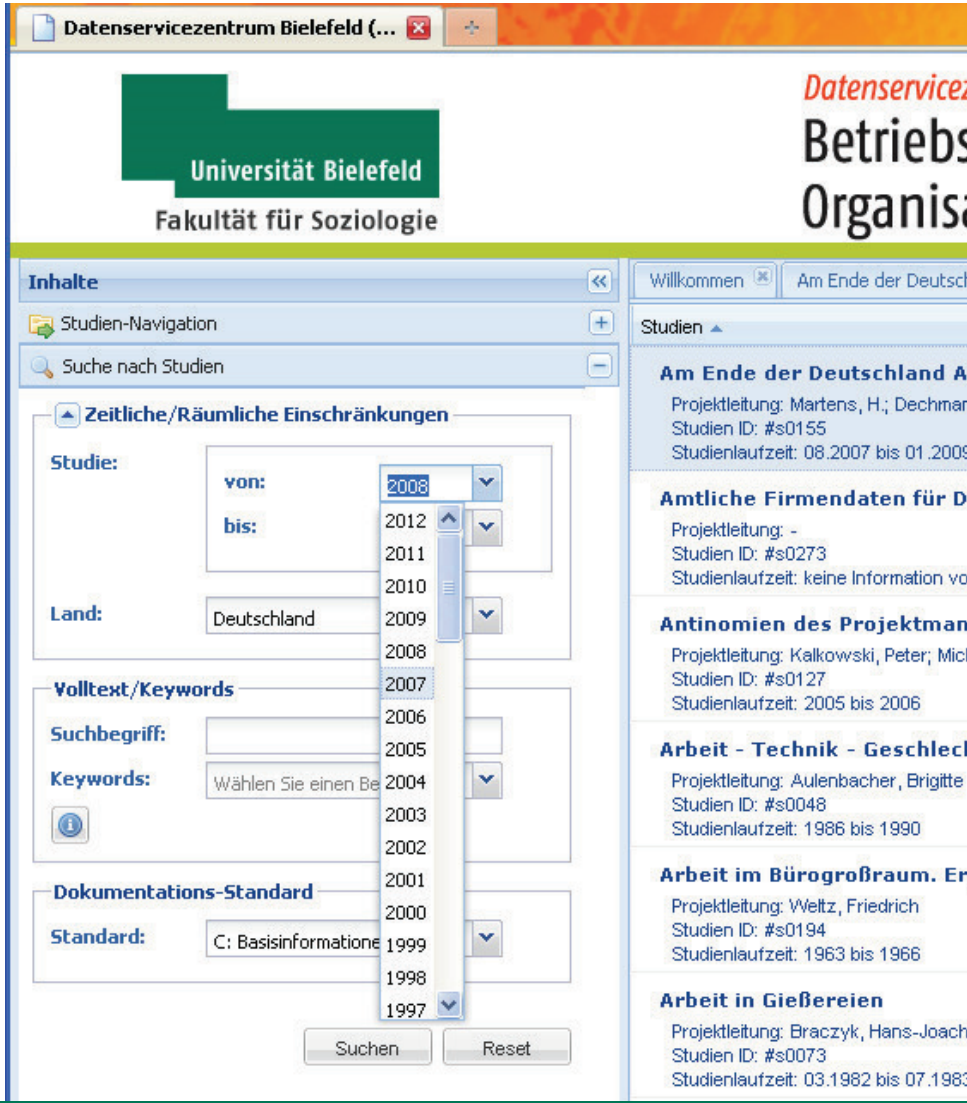
Visualization of structured JSON data

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

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)



Contents

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:

π(a:OrganisationName)

(a:Archive/a:OrganizationScheme/a:Organization $\bowtie_{@id==r:ID}$

r:FundingInformation/r:AgencyOrganizationReference)