Collaboration in Data Documentation: Developing „STARDAT - The Data Archiving Suite“

Wolfgang Zenk-Möltgen

IASSIST 2011 - Data Science Professionals: a Global Community of Sharing
May 30 – June 3, 2011, Vancouver, Canada
Overview

✦ Old and new demands on survey documentation

Current applications and workflow

The STARDAT project

Acknowledgements to the STARDAT Team

• Martin Friedrichs
• Monika Linne
• Brigitte Mathiak

• Alexander Mühlbauer
• Johann Schaible
• Andias Wira-Alam
Old and new demands on survey documentation

• Accuracy
• Completeness
• Inter-operability
• Multi-linguality
• Availability
• Identification
• Clarity
• Integration
• Long-term preservation
Survey documentation in the research data life cycle

Data and information research

Survey planning and design

Data analysis

Data registration and archiving

Data collection and preparation
Means of achieving these demands

• Define the documentation workflow
• Define quality levels of documentation
• Define interfaces between different pieces of documentation
• Use structured instead of unstructured documentation
• Use standards, especially DDI
• Use controlled vocabularies where possible
• Build software that helps to use standards
• Check usability of the software and standards used
Overview

Old and new demands on survey documentation

Current applications and workflow

The STARDAT project
Current applications and workflow

Online Publication

da|ra  →  DBK  →  ZACAT  →  Variable Overview  →  Study Overview  →  SDEdit  →  Report-Tool  →  CBE

DBKEdit  →  DSDM

Offline Publication

Report  →  Report-Tool  →  Longterm-preservation

DBK  Data Catalogue
ZACAT  Online Study Catalogue
DBKEdit  Data Catalogue Edit-Tool
SDEdit  Editing-Tool for Study Method Reports
DSDM  Dataset Documentation Manager
CBE  CodebookExplorer
da|ra  Registration Agency
Main Tools

• Data Catalog (DBK) for study descriptions
• DSDM for variable descriptions
• CodebookExplorer for relations and structure of variables (and offline usage)
• Variable Overviews
• Study Overviews for methodology information
• da|ra for persistent identifiers & metadata
• ZACAT for retrieval and downloads
Main Tools

Quicksearch on metadata

1. European Values Study 2008: Lithuania (EVS 2008)  
   GESIS  
   DOI: doi:10.4232/1.10344  
   Versions 1.0.0  
   Principal Investigators: Juknevicus, Stanislavas (Institute of Culture, Philosophy and Art, Lithuania)  
   Erhebungstermin: 21.07.2008 to 15.08.2008  
   Publikationsagentur: GESIS - Leibniz-Institut für Sozialwissenschaften  

2. European Values Study 2008: Malta (EVS 2008)  
   GESIS  
   DOI: doi:10.4232/1.10364  
   Versions 1.0.0  
   Principal Investigators: Treidl, Joseph (European Centre for Gerontology, University of Malta, Malta)  
   Erhebungstermin: 16-09-2008 to 23-09-2008  
   Publikationsagentur: GESIS - Leibniz-Institut für Sozialwissenschaften  

   GESIS  
   DOI: doi:10.4232/1.10307  
   Versions 1.0.0  
   Principal Investigators: Tcherima, Jean Francois (Tcherima Etudes Consoll, Paris, France)  
   Erhebungstermin: 07-09-2008 to 04-09-2008  
   Publikationsagentur: GESIS - Leibniz-Institut für Sozialwissenschaften  

4. European Values Study 2008: Spain (EVS 2008)  
   GESIS  
   DOI: doi:10.4232/1.10302  
   Versions 1.0.0  
   Principal Investigators: Calvoca, Maria Silvestre (University of Deusto, Faculty of Political Sciences and Sociology, Bilbao, Spain)  
   Erhebungstermin: 28-05-2008 to 13-07-2008  
   Publikationsagentur: GESIS - Leibniz-Institut für Sozialwissenschaften  

5. European Values Study 2008: Austria (EVS 2008)  
   GESIS  
   DOI: doi:10.4232/1.10309  
   Versions 1.0.0  
   Principal Investigators: Zulehner, Paul M. (Institut für Pastoraltheologie, Katholisch-Theologische Fakultät, University of Vienna, Austria)  
   Erhebungstermin: 21-07-2008 to 22-10-2008  
   Publikationsagentur: GESIS - Leibniz-Institut für Sozialwissenschaften  

   GESIS  
   DOI: doi:10.4232/1.10230  
   Versions 1.0.0  
   Principal Investigators:  
   Erhebungstermin:  
   Publikationsagentur: GESIS - Leibniz-Institut für Sozialwissenschaften
Overview

Old and new demands on survey documentation

Current applications and workflow

춘 The STARDAT project
Intention – Integration of Different Archiving Tools

- Integrated management system for metadata
- Transfer of the features of DBKEdit, DSDM, CBE and further tools
- Interoperability with standards like DDI 2, DDI 3, ISO 20252
- Multi-language documentation on study and variable-level
- Web based module for structured metadata capture, management and dissemination (Web Based Data Ingest)
- Controlled vocabularies (Thesauri)
- Related publications, continuity guides, scales, trends and additional metadata
- Long-term preservation with DDI
- Export in different portals like ZACAT, Cessda Data Portal, Sowiport, da|ra
Future applications and workflow

Online Publication

- da|ra
- DBK
- ZACAT
- Variate Overview
- Study Overview

Offline Publication

- Report
- CBE

STARDAT

Longterm-preservation

DBK Data Catalogue
ZACAT Online Study Catalogue
STARDAT The Data Archiving Suite
CBE CodebookExplorer
da|ra Registration Agency
Requirements and Challenges concerning DDI

- Export to DDI 2.1 still needed
  - for publication on ZACAT (Nesstar) server
  - for data exchange with portals like da|ra, sowiport
- Export to DDI 3.1 needed
  - for long-term archiving
  - for Enhanced Publication Editor (linking publications to datasets), etc.
- Import from DDI all versions needed
  - for data exchange with primary researchers/projects
- Future DDI versions need to be supported
- Usage of Resource Packages needed for re-using elements
  - for own elements and elements from other institutions
- Concept for long-term archiving of re-used elements needed
- DDI Migration issues need to be solved
- Performance issues: XML – RDB transformation
Requirements Analysis, Architecture and Prototyping

- Structured interviews with the main stakeholders
- Requirements specification document
- System Architecture defined
- Current state: paper prototype
- Build working prototype
- Do evaluation with users
- Implement full functionality
Technologies for STARDAT

- Architectural requirements
  - Browser-based web application
  - Transaction-based multi-user access
  - Distributed undo-/redo-mechanism
  - Multi-threading
  - Automation of metadata management tasks
  - Extensible framework for further needs of metadata documentation
  - Keeping in mind: REST and SOA

- Grails, GORM, Java based development
  - Controllers are generated from templates to enable scripting for CRUD automatically
  - Auto-complete and duplicate detection integrated in view templates
  - Domain classes based on DDI 3.1, including validation
  - Flexible data storage: Database or XML
Build a virtual research environment that replaces some of the publication tools.

But that will take some time...
Thank you!

Questions, remarks, and ideas welcome!