HIS Hochschul Informations System GmbH

Using a DDI 3 Based Single Source Approach to Increase the Efficiency of Social Science Research Processes

Bonn, December 4., 2009
Karsten Stephan

Overview

- Higher Education Information System (HIS)
- Research Processes and Software Applications
- Limitations of Efficiency
- DDI Single Source Approach
- Examples
- Summary

Higher Education Information System (HIS)

- Central service provider for institutions of higher education as well as ministries and administrative bodies in the German scientific landscape

- Division “Higher Education IT” (software house for higher education administration)
- Division “Higher Education Development” (management consultation with the central topics of higher education organisation)
- Division “Research on Higher Education” (empirical research institute on higher education)

Division “Research on Higher Education”

- 60 Scientists
- More than 50 empirical studies annually
  (paper and pencil / online)
- 40 online studies with about 100.000 interviews in the year 2008

Research Processes

- construction of measurement instruments
- data collection
- data preparation
- data analysis
- reporting
- data archiving
- data dissemination

Software Applications

- measurement instruments => text processing, desktop publishing, online survey system
- data collection => online survey system
- data preparation => plausibility check, text coding, weighting
- data analysis => statistical analysis, text data analysis
- reporting => report generation system

- data archiving => saved as proprietary data format
- data dissemination => e.g. as scientific use file, terminal server

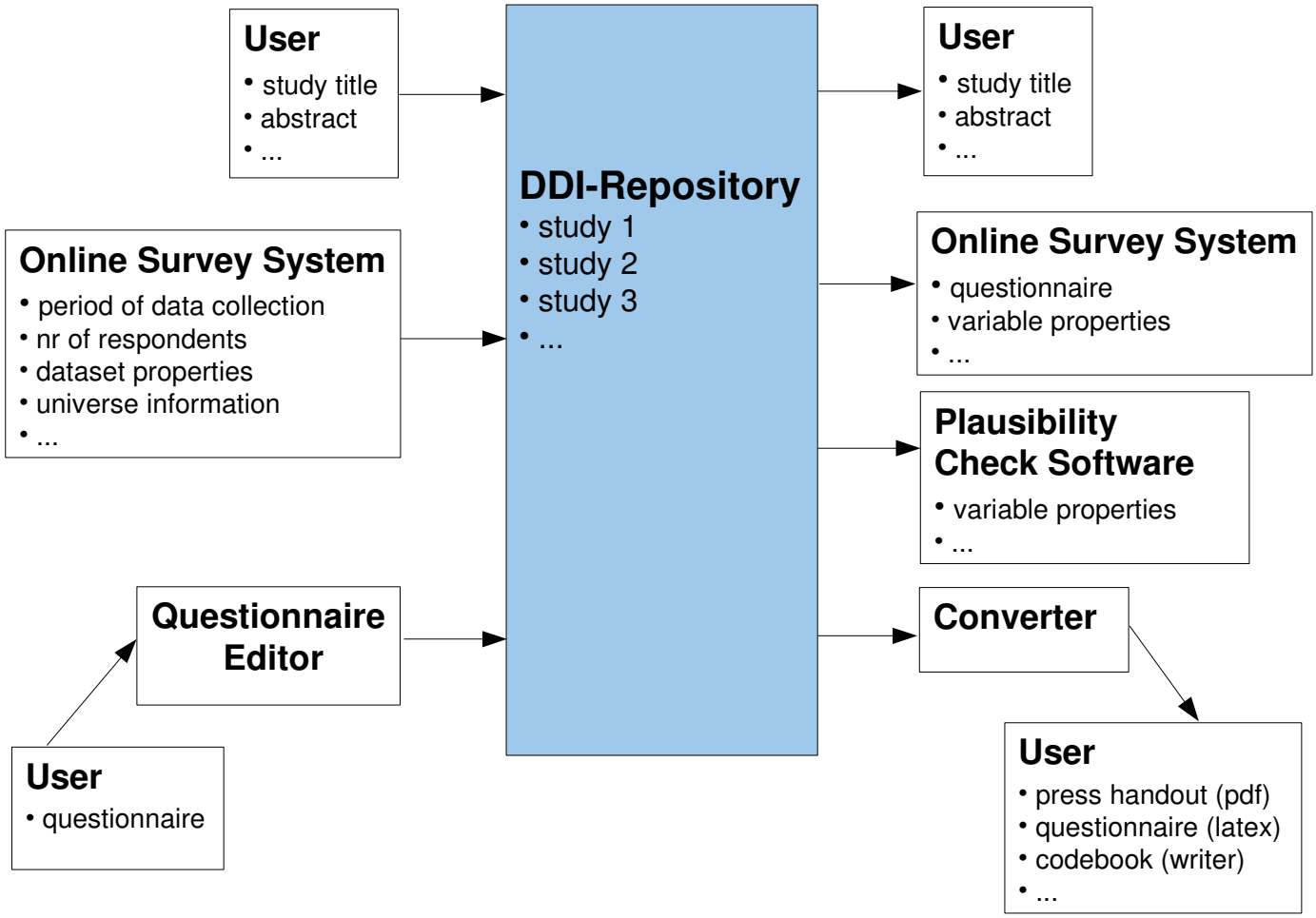
Limitations of Efficiency

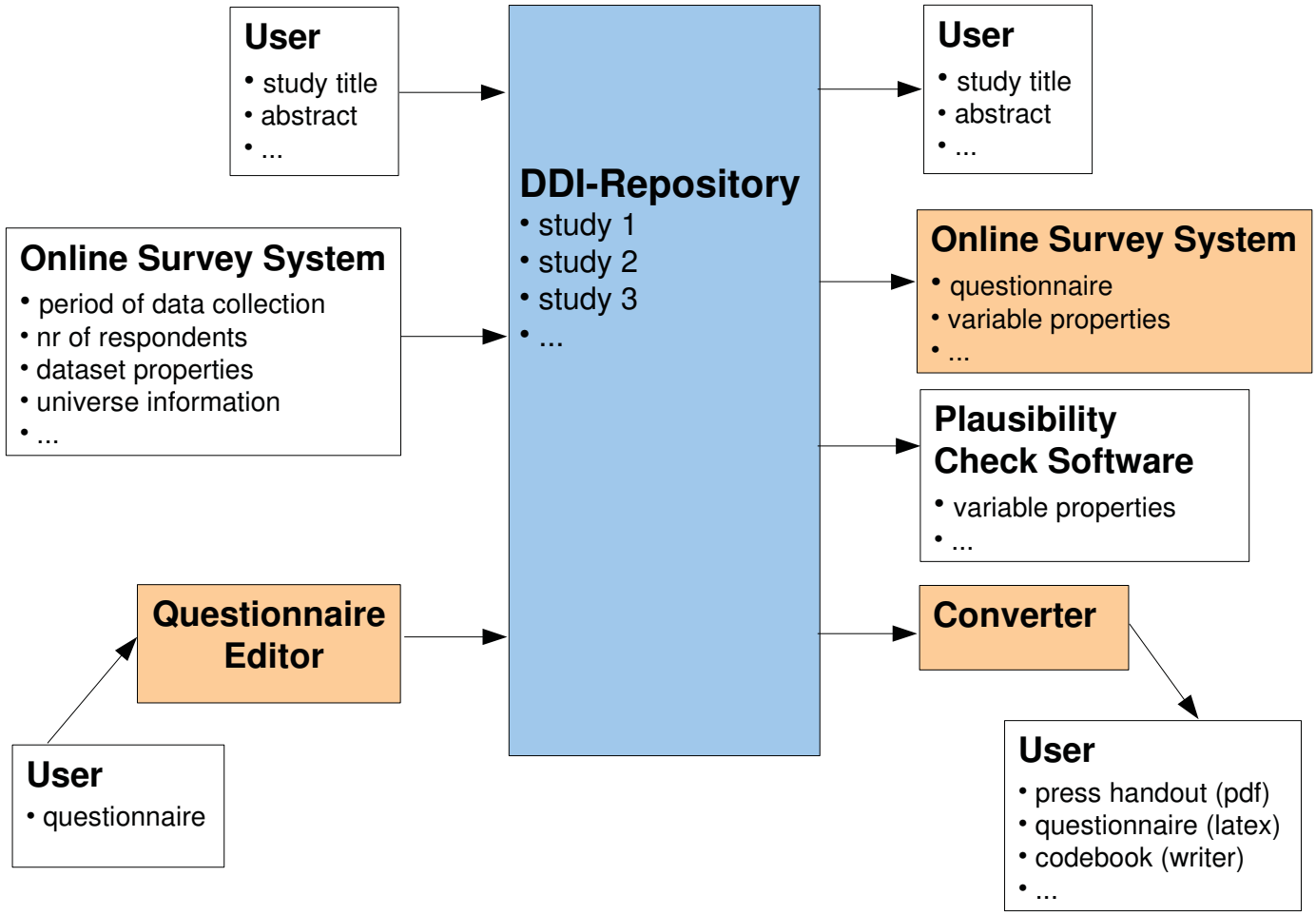
- different file formats are used wich are not compatible to each other. For example: questionnaires are stored in text processing, as well as desktop publishing or online survey system
  => the **interoperability** of the software systems **is suboptimal**.
- the documentation (study descriptions, questionnaires, codebooks and other information) is stored using text processing file formats.
  => There is a **limited possibility of automatic content processing**
- the metadata is stored in different locations (local hard disc drive, fileserver, online survey server) and in different versions without using a version control mechanism
  => **Which one is the last version?**

=> High Transformation Costs

DDI Based Single Source Approach

- a cental repository holds all metadata in DDI 3 format
- researchers and applications can add metadata to the repository
  - researchers can add and edit metadata using a web interface or by using applications
  - software applications can add and edit metadata using software interfaces
- researchers and applications can receive metadata from the repository
  - researchers can read data using a web interface and transform data to various output formats using a converter
  - software applications can read the required data using software interfaces


User
• study title
• abstract
• ...
DDI-Repository
• study 1
• study 2
• study 3
• ...
User
• study title
• abstract
• ...
Online Survey System
• period of data collection
• nr of respondents
• dataset properties
• universe information
• ...
Online Survey System
• questionnaire
• variable properties
• ...
Plausibility Check Software
• variable properties
• ...
Questionnaire Editor
Converter
User
• questionnaire
User
• press handout (pdf)
• questionnaire (latex)
• codebook (writer)
• ...


User
• study title
• abstract
• ...
DDI-Repository
• study 1
• study 2
• study 3
• ...
User
• study title
• abstract
• ...
Online Survey System
• period of data collection
• nr of respondents
• dataset properties
• universe information
• ...
Online Survey System
• questionnaire
• variable properties
• ...
Plausibility Check Software
• variable properties
• ...
Questionnaire Editor
Converter
User
• questionnaire
User
• press handout (pdf)
• questionnaire (latex)
• codebook (writer)
• ...

Technical Approach

- a XML-database (eXist) is used to hold the DDI instances
- the web interface (DDI web editing framework) is implemented using the Java EE (JSF) programming language
- the DDI beans (apache XML-beans) are used to implement the Java DDI binding
- Java and XSLT are used to implement the converters
- the project is developed under the GPLv3 license

Summary

- the DDI based single source approach lowers the transformation costs
- it also offers the possibility of automatic content processing

=> the efficiency of the research process can be increased highly by changing the IT infrastructure to the drafted DDI 3 based single source approach

Thank you