Challenge to implement DDI in a social science research institute

Report on first DDI experience and plans for a restart

Göteborg, 5th November 2011
Dr. Karsten Stephan & Daniel Buck
Overview

• HIS GmbH and HIS-HF
• Tools used and developed by HIS-HF
• First steps and problems in DDI 2009 to 2011
• Planning the „restart“ in 2012
• Discussion
HIS – Higher Education Information System

• HIS is a central service provider for institutions of higher education in Germany
  – Universities, national and federal ministries
• Founded in 1969
• Consists of four divisions, each of them focused on one field of activity
• HIS in total employs 432 people
Challenge to implement DDI in a social science research institute
HIS-Institute for Research on Higher Education (HIS-HF)

• Currently HIS-HF has 82 employees and additionally students as supporting staff
  – 63 social scientists, 4 administrative clerks,  
  – about 50 students supporting the conduction of surveys  
  – 15 computer scientists maintaining and developing software and information infrastructure
  • to support social scientists within the institute
  • to support federal ministries with statistical information systems
Project-Center „Methods and Services for Surveys“

- Our project-center maintains & develops software and supports many social science research projects in conducting studies.

- From the perspective of the whole research institute and the other departments our most important tasks is
  - the support of the current survey projects

- Therefore from our perspective we have to handle the current survey projects and our further goals:
  - Improving our information infrastructure in long term perspective
  - Finding resources to develop it
Information Infrastructure development in our project-center

• Efficiency in projects important
  – It is difficult to get more institutional resources to finance delays or „software development experiments“

• DDI or other new technologies are challenging in this context
  – New technologies have to be learned and used relatively fast
  – For bigger projects we need additional external funding

• Opportunities
  – Our institute does the whole data lifecycle including study planning, collecting & analysing data, archiving, disseminating
  – DDI seems to be perfect for us
Research topics and empirical studies

• Most important fields of research are:
  – Students, Graduates, Lifelong Learning
  – Steering, Funding and Evaluation in Higher Education

• Focus on quantitative methods and data, but often additional use of qualitative methods (expert interviews, biographic narrative interviews)

• 2011: 86 social science projects and additionally IT projects
Research topics and empirical studies

• For example in 2011:

• HIS-HF carried out 26 projects with cross sectional studies and waves of longitudinal cross sectional panels

• We conduct studies for NEPS, the National Educational Panel Study, a longitudinal study with 60,000 participants

• We run two online-access panels with 45,000 participants

• Furthermore in 2011 55 projects reusing official statistics, data collected by HIS-HF and universities
Some examples of empirical studies

<table>
<thead>
<tr>
<th>Study</th>
<th>~ N respondents</th>
<th>Number of waves per cohort and cohorts</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Semester Students Survey*</td>
<td>9.000</td>
<td>2 waves; every second cohort</td>
<td>Quantitative, paper&amp;pencil, online, longitudinal cross sectional</td>
</tr>
<tr>
<td>Graduates Survey*</td>
<td>10.000</td>
<td>3 up to 4 waves; every fourth cohort</td>
<td>Quantitative, paper&amp;pencil, longitudinal cross sectional panel</td>
</tr>
<tr>
<td>NEPS*</td>
<td>projected figure of online survey respondents of about 17.800, actual figure not yet published</td>
<td>Multi cohort model</td>
<td>Quantitative, Cati, online, longitudinal cross sectional panel, multi cohort model</td>
</tr>
</tbody>
</table>

* Data representative for Germany, not only for single universities or regions
IT Systems developed and maintained by HIS-HF

### Tools supporting conduction of surveys

<table>
<thead>
<tr>
<th>Tool</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIOB</td>
<td>Online Survey Tool, prototype used for NEPS survey</td>
</tr>
<tr>
<td>HIS-Code</td>
<td>coding of open ended questions</td>
</tr>
<tr>
<td>HIS-Plausi</td>
<td>plausibilisation of coded data</td>
</tr>
<tr>
<td>HIS-Dokudat</td>
<td>Generating cross tabulations</td>
</tr>
</tbody>
</table>

### Statistical Information Systems (SIS)

<table>
<thead>
<tr>
<th>SIS</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICE</td>
<td>Information System with official statistics on education topics</td>
</tr>
<tr>
<td>Dastat / Fostat</td>
<td>Information &amp; Reporting System on national research fundings, used by the Federal Ministry of Education and Research</td>
</tr>
</tbody>
</table>
Modernisation concept: situation 2009

- In 2009 we decided to improve our workflow from technical perspective:
  
  - Different file formats were used, which are not compatible to each other (questionnaires were stored in text processing software as well as in desktop publishing and online survey systems) => interoperability could be better

  - The documentation (study descriptions, questionnaires, codebooks) was stored using text processing file formats => limited possibility of automatic content processing

  - The metadata was stored in different locations (local hard disc drive, file servers, online survey server) and in different versions without using a version control mechanism => which one is the last version?
Modernisation concept: objectives

- Improvement of interoperability
- Possibility to query and edit lifecycle metadata (e.g. search variable descriptions, question bank)
- Standardised documentation for SUF, PUF, CUF
Modernisation concept end of 2009

- We planned to implement a central repository, which holds all metadata in DDI 3 format
- We wanted to implement DDI 3 in the core of (new) software applications, which are developed by HIS-HF (online survey software / text coding)
- We planned to implement a web interface to add, edit and retrieve the information in the repository
- We decided to start the DDI implementation in the context of the development of a new online survey software: HIOB
HIOB: project overview

- HIOB is an online survey software
- It is developed to fulfil the requirements in the field of scientific research
- Main features are the ability to conduct complex surveys, data security, interoperability
- Three software developers are working on the project
- Development started in the beginning of 2010
- Prototype was released in May 2011
- The first major survey, which was conducted with HIOB was a panel wave of the German national education panel study (NEPS) in October 2011
- The first “public” version (version 1.1) will be released in 2012
- HIOB is open source, it's developed under the AGPL licence
- Programming language is Java/Java EE
HIOB: challenges implementing DDI 3

- We planned to use the DDI data model to manage datasets, measurement instruments and other information in the core of the HIOB software, BUT:
  - DDI 3 is very complex
  - There was too little information available for us to understand completely the semantics of instrument markup with DDI
  - Some parts of DDI seemed to be work in progress
  - Especially in the field of instrument documentation it was difficult for us to determine which properties of a HIOB questionnaire could be marked up and which not
  - Also, it's very difficult, to add additional XML elements with a non DDI namespace to the complex standard without full understanding of the DDI semantics
  - Restraints in development resources, internal resources are sometimes urgently needed for daily business in supporting the current social science projects, software development projects may be delayed
HIOB: challenges implementing DDI 3

- We received help from the DDI community but after some months we decided to stop the DDI activities in the context of the HIOB project due to temporal restraints.
- The HIOB project had to be completed within 2 years. We were responsible for providing a running online survey system with elaborate features until 2012 – under all circumstances.
Restart in 2012: new roadmap

- Our institute decided to improve its data management and sharing substantially. Therefore we want to implement a small scale, but technically well equipped data research centre.
- We want to improve the Documentation in terms of quality, completeness and common coordinated standards.
- We want to introduce processes to assure quality standards.
Restart in 2012

- In this context we would like to restart our DDI activities concerning our whole information infrastructure
- We will start with a basic approach and expand our DDI activities step by step
- This time we are not going to implement DDI 3 completely. Instead of that we are going to use a defined subset of DDI Elements at first
- We are going to limit the export Elements from HIOB to existing DDI Elements at least in a first productive system
- We plan not to use the DDI data model in the core of software applications. Instead of that we are going to implement import and export routines for the programs used/developed by HIS-HF
Summary / Discussion

• It seems to be difficult for a mid sized research institute to develop DDI 3, mostly due to its complexity. But we thank the DDI community for all of its work and support!
• An extensive implementation of DDI 3 in the whole infrastructure in HIS-HF fully financed by internal resources in one big “DDI 3 rollout project” was not possible in 2009, is still hardly possible now
• The extent and usability of documentation influences strongly our ability to implement DDI 3
Summary / Discussion

- The speed and scale of DDI 3 Implementation within HIS-HF depends strongly on external funding. If we get external funding, we are going to develop DDI 3 tools and support the DDI community this way.
- We are planning to implement DDI 3 but at this time starting in a less complex way and step by step.
- We are going to establish a data research centre including a basic DDI 3 infrastructure that can be expanded by additional tools.