

# STARDAT<sup>+</sup>

## DATA ARCHIVING SUITE

Monika Linne, Alexander Mühlbauer, Wolfgang Zenk-Möltgen

## Overview

✦ Initial Situation – Different Archiving Tools

Main Tools – DBKEdit/ DSDM/ CBE

Intention – Integration of Different Archiving Tools

Proceeding – Requirements Analysis

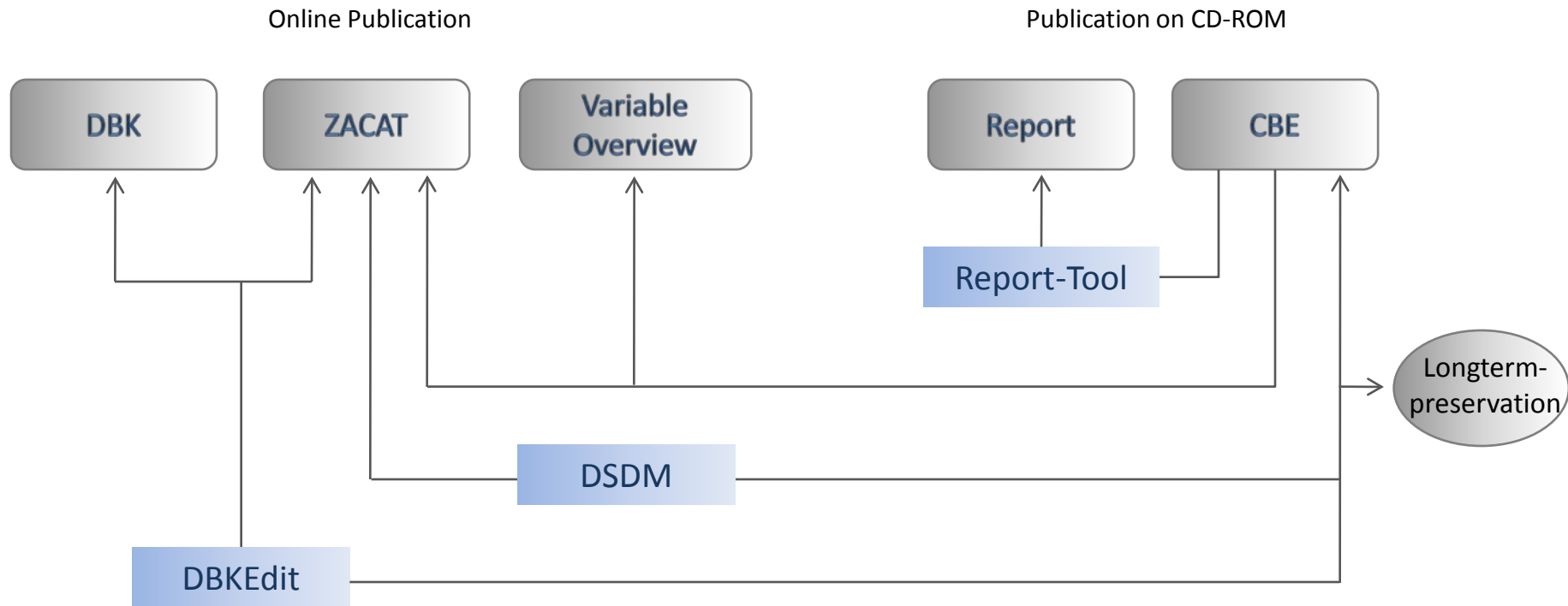
DDI Formats Currently Used

Requirements Concerning DDI Formats

Getting Started with STARDAT Development

Challenges of Going into DDI3

## Initial Situation – Different Archiving Tools



DBK  
ZACAT  
DBKEdit

Data Catalogue  
Online Study Catalogue  
Data Catalogue Edit-Tool

DSDM  
CBE

Dataset Documentation Manager  
CodebookExplorer

## Overview

Initial Situation – Different Archiving Tools

✦ Main Tools – DBKEdit/ DSDM/ CBE

Intention – Integration of Different Archiving Tools

Proceeding – Requirements Analysis

DDI Formats Currently Used

Requirements Concerning DDI Formats

Getting Started with STARDAT Development

Challenges of Going into DDI3

## Main Tools

### 1. Data Catalogue Edit-Tool (DBKEdit)

Documentation and management of metadata on study level of all archived studies

- A. *Inhouse-tool*
- B. *Web-based*
- C. *DDI-supporting*

#### Information provided

For each study/dataset (partly bi-lingual):

- Title & study number
- Primary investigators & fieldwork agencies
- Universe & sampling
- Abstracts
- Related publications
- Access categories
- Questionnaires, etc.
- Notes and comments for internal use

#### Publication via

- ZACAT (Nesstar)
- CBE (CodebookExplorer)
- DBK (Data Catalogue)

## DBKEdit – Management of Metadata on Study-Level

The screenshot displays the DBKEdit web application interface. On the left, there is a navigation menu with sections like 'Dienstleistungen' and 'Datenbestandskatalog'. The main content area shows a table of study records with columns for 'Titel', 'Erhebungsart', 'Finanzierung', 'Bemerkungen', 'System', 'Codebuch', 'Bemerkungen', 'Akquisition', 'Status', 'Vergabe', and 'Zugang'. A modal window titled 'Studienbeschreibung 3781, Sprache E' is open, showing a form for editing the study record. The form includes a 'Titel' field with the text 'EVS - European Values Study 1999/2000 (release 2, May 2006) - Spain' and a 'Rechtschreibung...' button. The interface also features a search bar and a 'Suchbegriffe eingeben...' field.

## DBK – Research and Information Provided for the User

The screenshot shows a web browser window displaying the GESIS Data Catalogue search results. The search criteria are 'EVS' in the title and 'Religion' in the abstract. The results list 34 entries, with the first entry selected. A detailed view of the selected entry, 'EVS - European Values Study 1999/2000 (release 2, May 2006) - Spain', is shown on the right.

**Data Catalogue Search Results:**

Choice	Study No.	Title
<input type="checkbox"/>	3776	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3777	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3778	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3779	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3780	EVS - European Values Study 1999/2000 (release 2, M
<input checked="" type="checkbox"/>	3781	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3782	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3783	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3784	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3785	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3786	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3787	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3788	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3789	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3790	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3791	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3792	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3793	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3794	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3795	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3796	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3797	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3798	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3799	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3800	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3801	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3802	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3803	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3804	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3805	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3806	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3807	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3808	EVS - European Values Study 1999/2000 (release 2, M
<input type="checkbox"/>	3811	EVS - European Values Study 1999/2000 (release 2, M

**Study Description 3781: EVS - European Values Study 1999/2000 (release 2, May 2006) - Spain**

Study No.	ZA3781
Title	EVS - European Values Study 1999/2000 (release 2, May 2006) - Spain
Current Version	<ul style="list-style-type: none"> <li>1.0.0, 13.04.2010, <b>doi:10.4232/1.3781</b></li> <li>Version number created automatically (Implementation of a uniform versioning policy)</li> </ul>
Date of Collection	1999
Principal Investigator/Authoring Entity Institution	<ul style="list-style-type: none"> <li>J. Elzo</li> <li>Universidad de Deusto, Bilbao</li> <li>F.A. Orizo</li> <li>DATA, Madrid</li> </ul>
Data Collector	DATA, Madrid
Abstract	<p>Moral, religious, societal, political, economic and social moral concepts of the Europeans.</p> <p>Topics:</p> <p>The question program is divided in three parts: besides a common question program for all countries there is a voluntary supplement program and further country-specific questions.</p> <p>1. Here only the common question program is presented:</p> <p>leisure time:</p> <ul style="list-style-type: none"> <li>importance of areas of life;</li> <li>feeling of happiness;</li> <li>memberships and honorary activities in clubs, parties, organizations, citizens' initiatives and occupation organizations;</li> <li>interactions in leisure time;</li> <li>tolerance regarding minorities, those of other beliefs and foreigners;</li> <li>interhuman trust;</li> <li>self-effectiveness (scale);</li> <li>general contentment with life (scale).</li> </ul> <p>working world:</p> <ul style="list-style-type: none"> <li>importance of selected characteristics of occupational work (scale);</li> <li>personal employment;</li> <li>general work satisfaction (scale);</li> <li>self-determination at work (scale);</li> <li>work ethic (scale);</li> <li>attitude to achievement-based pay and following work instructions without criticism;</li> <li>priority of nationals over foreigners as well as men over women with shortage of jobs;</li> <li>assumed priority of individual or social reasons for the situation of economic need of individuals;</li> <li>freedom of the unemployed to reject a job offer (scale).</li> </ul> <p>politics:</p> <ul style="list-style-type: none"> <li>party preference;</li> <li>attitude to foreign workers in one's country;</li> <li>fear of the future;</li> <li>assimilation and integration of immigrants;</li> </ul>

## Main Tools

### 2. Dataset Documentation Manager (DSDM)

Documentation of metadata on variable-level for special datasets and data collections

- A. *Inhouse-tool & external projects*
- B. *ACCESS-based desktop-application*
- C. *DDI-supporting*

#### Publication via

- ZACAT (Nesstar)
- CodebookExplorer

#### Information provided

For each Variable:

- Exact question and answer texts
- Multilingual documentation
- Documentation of deviations in questionnaires fielded in different waves or countries
- Interviewer instructions
- Forward/backward instructions for filter questions
- Documentation of additional archival notes (e.g. recoding, deviations)
- Generating variable categories



## DSDM – Management of Metadata on Variable-Level

The screenshot displays the Dataset Documentation Manager (DSDM) interface. The main window shows a list of variables in the dataset catalogue, with Q6A selected. The 'Selection in the question catalogue' pane shows the question text for Q6A: 'Le voy a preguntar con qué frecuencia hace Vd. ciertas cosas. Para cada actividad ¿diría Vd. que la hace al menos cada semana o casi cada semana; una vez o dos al mes; sólo unas pocas veces al año; o nunca en absoluto?'. The 'Question text preview' pane shows the question text in Spanish: 'Pasará tiempo con gente en clubes y asociaciones voluntarias (deportivas, culturales, comunales)'. The 'Selection in the dataset catalogue' pane shows the variable label: 'Study - Variable - Label: ZA3781 - v51 - how often spend time in clubs+voluntary associations (Q6A)'. The 'Selection in the question catalogue' pane shows the question label: 'Question: Q286: F286: ZA3811-V48-how often spend time with friends (Q6A)'. The 'Question text preview' pane shows the question text in Spanish: 'Pasará tiempo con gente en clubes y asociaciones voluntarias (deportivas, culturales, comunales)'. The 'Question text preview' pane also shows a list of response options: -3 no aplicable, -2 no contesta, -1 No sabe, 1 cada semana, 2 Una o dos veces al mes, 3 Unas pocas veces al año, 4 Nunca en absoluto.

Database file: C:\Dokumente und Einstellungen\Ninne\Desktop\3781\_Spain\_EVS\_3W\_Pub\_07\_10\_22 (Version 2-3-1).mdb

## DSDM → Publication via ZACAT

**GESIS: ZACAT - Mozilla Firefox**

GESIS: ZACAT - Mozilla Firefox

http://zacat.gesis.org/webview/index.jsp

GESIS: ZACAT Extended Variable Overview - Europea...

Leibniz-Institute for the Social Sciences

Data for the social sciences.

Contact us Help FAQ ZACAT Home

**Description** Table Analysis

**Literal Question**  
Q1  
<SHOW CARD 1>  
Please say, for each of the following, how important it is in your life.

Q1.B Family

-5 other missing  
-4 question not asked  
-3 nap  
-2 na  
-1 dk  
1 very important  
2 quite important  
3 not important  
4 not at all important

Comparability:  
Trend question (EVS 2008=EVS 1999).  
Comparable questions: WVS 1995 (v4) and WVS

**Trend Variables 1999-2000-2008 (controlled ti**  
V2 - Integrated Dataset - EVS 1999-2000

**Comparable Variables 1981-2008**  
V2 - Integrated Dataset - EVS 1999-2000  
Q117B - Integrated Dataset - EVS 1990

**Original Language Documentation (all lang**  
v2 - France - EVS 2008  
v2 - Germany - EVS 2008  
v2 - Austria - EVS 2008  
V2 - Spain - EVS 2008  
V2 - Portugal - EVS 2008  
V2 - The Netherlands - EVS 2008  
v2 - Denmark - EVS 2008  
v2 - Northern Ireland - EVS 2008  
v2 - Ireland - EVS 2008  
v2 - Estonia (Estonian) - EVS 2008  
v2 - Estonia (Russian) - EVS 2008  
v2 - Latvia - EVS 2008  
v2 - Lithuania - EVS 2008  
v2 - Poland - EVS 2008  
v2 - Czech Republic - EVS 2008  
v2 - Slovak Republic - EVS 2008  
v2 - Hungary - EVS 2008  
v2 - Romania - EVS 2008

**Display this Variable in Extended Variable Overview**  
Variable v2: how important in your life: family (Q1B) how important in your life: family (Q1B)

NOUS COMMENÇONS AVEC QUELQUES QUESTIONS SUR LA VIE EN GÉNÉRAL, LES LOISIRS

Q1  
<MONTERER CARTE 1>  
Pour chacune des choses suivantes, pouvez-vous me dire si, dans votre vie, cela est très important, assez important, peu important ou pas important du tout ?

Q1.A Travail

1 Très important  
2 Assez important  
3 Pas très important  
4 Pas important du tout  
8 NSP  
9 SR

Note:  
Missing values from field questionnaire recoded into -1 to -5 values in IDS and NDS.

**CARTE 1**

Sehr wichtig  
Ziemlich wichtig  
Nicht sehr wichtig  
Überhaupt nicht

Très important  
Assez important  
Pas très important  
Pas important du tout

print this page

You can resize the window to increase/decrease the showcard picture reset this window

Fertig Fertig

Copyright © 2004 NESSTAR - All Rights Reserved

POWERED BY nesstor

## Main Tools

### 3. CodebookExplorer (CBE)

Tool on context-level which gives easy retrieval possibilities to end-users for databases of specific topics

- A. *Tool for end-users or members of external projects*
- B. *Inhouse-tool supporting the publication of context information (e.g. comparable questions, show cards)*
- C. *Not DDI-supporting*

#### Publication via

- CodebookExplorer (CD-ROM/Internet)
- ZACAT

#### Information provided

For each study/data set

- Browsing studies or trends, topics, scales etc. (if the database provides them)
- Structuring of variables according to content or methodical aspects
- Compare question texts
- Make simple analysis like frequencies or crosstabs
- Show questionnaires
- Access to SPSS datasets

## CBE → Publication and Management via CD-ROM and Internet

Extended Variable Overview - European Values Study - ZA4800 - v1 - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

http://info1.gesis.org/ejs/variables/compview.asp?db=QEV52008&id=ZA4800&var=v1&lang=en&id2=ZA3811&var2=v1&lang2=en

GESIS: ZACAT



**European Values Study**

[Home](#)  
**Extended Variable Overview for Comparable Variables and Original Languages of the European Values Study 1981, 1990, 1999/2000, and 2008 Waves.**

[EVS 2008 Integrated Dataset \(ZA4800\)](#) - 
 [EVS 1999-2000 Integrated Dataset \(ZA3811\)](#) - 
 [EVS 1990 Integrated Dataset \(ZA4460\)](#) - 
 [EVS 1981 Integrated Dataset \(ZA4438\)](#)

**Variables of Dataset Integrated Dataset - EVS 2008**

- StudyNo (GESIS Study Number)
- StudyNoC (GESIS Study Number (National datasets))
- Version (GESIS Archive Version)
- VersionC (GESIS Archive Version (National datasets))
- id\_cocas (unified respondent number)
- caseno (original respondent number)
- intrno (interviewer number)
- wave (wave)
- year (survey year)
- country (country code)
- country1 (country code [with split ups])
- c\_abrv (country abbreviation)
- c\_abrv1 (country abbreviation [with split ups])
- weight (weight)
- weight\_c (weight for country specific characteristics)
- cntry\_y (country\_year)
- cntry1\_y (country\_[with split ups]-year)
- v1 (how important in your life: work (Q1A))
- v2 (how important in your life: family (Q1B))
- v3 (how important in your life: friends and acquaintances (Q1C))
- v4 (how important in your life: leisure time (Q1D))
- v5 (how important in your life: politics (Q1E))
- v6 (how important in your life: religion (Q1F))
- v7 (how often discuss politics with friends (Q2))
- v8 (taking all things together)

**Dataset ZA4800:** Integrated Dataset - EVS 2008  
**Variable v1:** how important in your life: work (Q1A)

WE START WITH SOME QUESTIONS ABOUT LIFE IN GENERAL, LEISURE TIME ACTIVITIES AND WORK

Q1  
 <SHOW CARD 1>  
 Please say, for each of the following, how important it is in your life.

Q1.A Work

-5 other missing  
 -4 question not asked  
 -3 nap  
 -2 na  
 -1 dk  
 1 very important  
 2 quite important  
 3 not important  
 4 not at all important

Comparability:  
 Trend question (EVS 2008=EVS 1999).  
 Comparable questions: WVS 1995 (v8) and WVS 2000 (v8).

CARD 1

very important  
 quite important  
 not important  
 not at all important

**Trend Variables: 3rd and 4th EVS wave (the year of the waves is mentioned in the title)**

- V1 (Integrated Dataset - EVS 1999-2000)

**Comparable Variables: all four EVS waves**

- Q116A (Integrated Dataset - EVS 1990)
- V1 (Integrated Dataset - EVS 1999-2000)

**All Original Languages (all languages for all countries)**

- v1 (Albania - EVS 2008)
- v1 (Armenia - EVS 2008)
- v1 (Austria - EVS 2008)
- v1 (Azerbaijan - EVS 2008)
- v1 (Belarus - EVS 2008)
- v1 (Belgium - EVS 2008)
- v1 (Bosnia-Herzegovina - EVS 2008)
- v1 (Bulgaria - EVS 2008)
- v1 (Cyprus - EVS 2008)
- v1 (Czech Republic - EVS 2008)
- v1 (Denmark - EVS 2008)
- v1 (Estonia (Estonian) - EVS 2008)
- v1 (Estonia (Russian) - EVS 2008)
- v1 (Finland - EVS 2008)
- v1 (France - EVS 2008)
- v1 (Georgia - EVS 2008)
- v1 (Germany - EVS 2008)
- v1 (Greece - EVS 2008)
- v1 (Hungary - EVS 2008)
- v1 (Ireland - EVS 2008)
- v1 (Kosovo (Albanian) - EVS 2008)
- v1 (Kosovo (Serbian) - EVS 2008)
- v1 (Latvia - EVS 2008)
- v1 (Lithuania - EVS 2008)

**Dataset ZA3811:** Integrated Dataset - EVS 1999-2000  
**Variable V1:** how important in your life: work (Q1A)

Q1  
 <Show card 1>  
 Please say, for each of the following, how important it is in your life.

Q1.A Work

-5 missing source unknown  
 -4 not asked in survey  
 -3 not applicable  
 -2 no answer  
 -1 don't know  
 1 very important  
 2 quite important  
 3 not important  
 4 not at all important

CARD 1

1- Very important  
 2- Quite important  
 3- Not important  
 4- Not at all important

## Overview

Initial Situation – Different Archiving Tools

Main Tools – DBKEdit/ DSDM/ CBE

✦ Intention – Integration of Different Archiving Tools

Proceeding – Requirements Analysis

DDI Formats Currently Used

Requirements Concerning DDI Formats

Getting Started with STARDAT Development

Challenges of Going into DDI3

## 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 3, ISO 20252
- Multi-language documentation on study and variable-level
- Web based modul for structured metadata capture, management and dissemination (Web Based Data Ingest)
- Controlled vocabularys (Thesauri)
- Related publications, continuity guides, scales, trends and additional metadata
- Longterm-preservation with DDI
- Export in different portals like ZACAT, Cessda Data Portal, Sowiport

## Overview

Initial Situation – Different Archiving Tools

Main Tools – DBKEdit/ DSDM/ CBE

Intention – Integration of Different Archiving Tools

✦ Proceeding – Requirements Analysis

DDI Formats Currently Used

Requirements Concerning DDI Formats

Getting Started with STARDAT Development

Challenges of Going into DDI3

## Proceeding

### Requirements Analysis

- Structured interviews with the main stakeholders:
  - Investigation of usage of the tools and workflows
  - Identification of problems
  - Feedback from stakeholders about necessary/desired changes→ Result: Description of all functionalities, also structured in diagrams
  
- Requirements specification document
  - Detailed description of requirements
  - Transfer into BPMN-Diagrams
  - Coverage of not yet existing functionalities
  
- Next steps:
  - Building software-modules
  - Evaluation of the new tool by users



## Overview

Initial Situation – Different Archiving Tools

Main Tools – DBKEdit/ DSDM/ CBE

Intention – Integration of Different Archiving Tools

Proceeding – Requirements Analysis

✦ DDI Formats Currently Used

Requirements Concerning DDI Formats

Getting Started with STARDAT Development

Challenges of Going into DDI3

## DDI Formats Currently Used

### DBK

- Export to DDI 2.0
  - for publication on ZACAT (Nesstar) server
  - for long-term archiving
  - for data exchange with portals like da|ra, sowiport

### DSDM

- Export to DDI 2.0
  - for publication on ZACAT (Nesstar) server
  - for long-term archiving
- Export to DDI 3.1
  - for Enhanced Publication Editor (linking publications to datasets)

### CBE

- No DDI currently used

## DDI Formats Currently Used

### Example: Study Description from DBK to DDI 2.0

contains

- Study number and title
- Principal investigators
- Abstract and classifications
- Research area and data collection information
- Further documents and bibliographic references
- Dataset information

Leit <?xml version="1.0" encoding="ISO-8859-1"?>  
 <!DOCTYPE codeBook SYSTEM 'http://info1.gesis.org/DDI/Version2-0.dtd'>  
 <codeBook>

DD

1

```

    <docDscr>
      <citation>
        <titlStmt>
          <titl>Dataset Title</titl>
          <subTitl/>
          <altTitl/>
          <parTitl/>
        </titlStmt>
      </citation>
    </docDscr>
    <stdyDscr>
      <citation>
        <titlStmt>
          <titl>Dataset Title</titl>
          <subTitl/>
          <altTitl/>
          <parTitl/>
          <IDNo>Study Number</IDNo>
        </titlStmt>
        <rspStmt>
          <AuthEnty>Principal Investigator, Affiliation: Institute</AuthEnty>*
        </rspStmt>
        <prodStmt>
          <producer></producer>
          <fundAg></fundAg>
        </prodStmt>
        <distStmt>
          <distrbtr abbr='GESIS' URI='http://www.gesis.org/'>
            GESIS - Leibniz-Institut für Sozialwissenschaften</distrbtr>
          <depositr/>
          <depDate/>
        </distStmt>
        <serStmt><serName/></serStmt>
        <verStmt><version/></verStmt>
      </citation>
  
```

```

<stdyInfo>
  <subject>
    <topcClas>Classification</topcClas>*
  </subject>
  <abstract><![CDATA[ Abstract ]]></abstract>
  <sumDscr>
    <collDate/>
    <universe><![CDATA[ Research area ]]></universe>
  </sumDscr>
</stdyInfo>
<method>
  <dataColl>
    <timeMeth><![CDATA[ Collection Dates ]]></timeMeth>
    <dataCollector> Data Collectors </dataCollector>
    <sampProc><![CDATA[ Sample Information ]]></sampProc>
    <collMode><![CDATA[ Data Collection Mode ]]></collMode>
    <weight/>
  </dataColl>
  <onlyInfo><respRate/></onlyInfo>
  <stdyClas/>
</method>
<dataAccs>
  <setAvail><avlStatus> Access Class </avlStatus></setAvail>
  <useStmt><contact/></useStmt>
</dataAccs>
<othrStdyMat>
  <relMat>
    <citation>
      <titlStmt>
        <titl>Related Materials</titl>
      </titlStmt>
    </citation>
  </relMat>

```

```

<relMat>
  <citation>
    <titlStmt>
      <titl><![CDATA[ Document Name and Type ]]></titl>
    </titlStmt>
    <holdings URI='Document URL'/>
  </citation>
</relMat> *
<relStdy/>
<relPubl>
  <citation>
    <titlStmt>
      <titl>Related Publications (only for first relPubl-Element)</titl>
    </titlStmt>
  </citation><![CDATA[ Bibliographic Reference ]]>
</relPubl> *
<othRefs>
  <citation>
    <titlStmt>
      <titl>Further Remarks</titl>
    </titlStmt>
  </citation><![CDATA[ Further Remarks ]]>
</othRefs>
</othrStdyMat>
</stdyDscr>
<fileDscr>
  <fileTxt>
    <dimensns>
      <caseQty>Number of Cases</caseQty>
      <varQty>Number of Variables</varQty>
      <recPrCas>Records per Case</recPrCas>
    </dimensns>
    <fileType>File Format</fileType>
  </fileTxt>
</fileDscr>
</codeBook>

```

## DDI Formats Currently Used

### Example: Variable Descriptions from DSDM to DDI 2.0

(for long-term archiving)

contains

- Study number and specific list (countries, waves, panels)
- SPSS variable names, labels, values and value labels
- Variable groups
- Question numbers, texts, interviewer instructions, filter instructions
- Sub-question numbers and texts
- Answer codes and categories, answer groups
- Several variable notes, e.g. for derivation

DD

1

```
<?xml version="1.0" encoding="UTF-16"?>
<!DOCTYPE codeBook SYSTEM "http://info1.gesis.org/ddi/DSDM/DSDM-LTA-200.dtd">
<!-- Selection of DDI 2.0 DTD Elements and Attributes used by DSDM for long-term archiving -->
<!-- The original DDI 2.0 DTD can be found at http://www.icpsr.umich.edu/DDI/Version2-0.dtd -->
<codebook ID="Archive Study ID">
  <docDscr>
    <citation>
      <titlStmt>
        <titl>Langfrstsicherungsdokumentation der Studie 'Archive Study ID'</titl>
      </titlStmt>
      <rspStmt>
        <AuthEnty>GESIS</AuthEnty></rspStmt>
      <prodStmt>
        <producer>GESIS</producer>
        <copyright>© 2007-2009 GESIS</copyright>
        <prodDate>Date (DD.MM.YYYY)</prodDate>
        <software>DSDM DDI Long-Term Archiving 2.0.0</software></prodStmt>
      </citation>
    </docDscr>
    <stdyDscr>
      <citation>
        <titlStmt>
          <titl></titl></titlStmt></citation>
      <stdyInfo>
        <sumDscr>
          <nation* ID="country identifier" abbr="short name of country">long name of country</nation>
        </sumDscr>
      </stdyInfo>
    </stdyDscr>

    <dataDscr>
      <varGrp* ID="identifier of variable group" var="list of variable references, coded under <var>">
        <labl>name of variable group</labl>
      </varGrp>
```



DD

2

```

<var* ID="variable identifier" name="SPSS variable name">
  <labl>SPSS variable label</labl>

  <qstn* ID="identifier of question"
        seqNo="Questionnaire No"
        sdatrefs="list of country references, coded under <nation>">
    <preQTxt?>text before question</preQTxt>
    <qstnLit*>literal question</qstnLit>
    <forward*>forward instruction</forward>
    <backward*>backward instruction</backward>
    <ivulnstr*>interviewer instruction</ivulnstr>
    <postQTxt?>text after question</postQTxt>
  </qstn>

    Or (if only reference):
  <qstn* qstn="question reference, coded under <qstn>" />

  <txt* ID="identifier of sub question"
        Level?="sub question no"
        sdatrefs="list of country references, coded under <nation>">sub question text</txt>
    Or (if only reference):
  <txt* sdatrefs="sub question reference, coded under <var><txt>" />

  <catgryGrp* ID="identifier of answer value group"
        catgry="list of answer value references, coded under <catgry>"
        <labl>name of answer value group</labl>
  </catgryGrp>

    Or (if only reference):
  <catgryGrp* catGrp="answer value group reference, coded under <catgryGrp>" />

```

DD

```
<catgry* ID?="identifier of answer value" only exists if tag is AnswerValue in DSDM
Missing?="SPSS value missing" as Y|N" only exists if tag is VariableValue in DSDM >
<catValu>answer value code and SPSS value only numeric!</catValu>
<labl?>SPSS value label</labl> only exists if tag is VariableValue in DSDM
<txt* ID="identifier of specific answer text">country specific value label</txt>
</catgry>
```

3

*Or (if reference):*

```
<catgry* sdatrefs="specific answer text reference" >
<labl?>SPSS value label</labl> only exists if tag is VariableValue in DSDM
</catgry>
```

```
<derivation>
<drvdscr>description of derivation</drvdscr>
</derivation>
```

```
<notes? type="NoteNote">userdefined notes to variable: Note</notes>
<notes? type="Text">userdefined notes to variable: Text</notes>
<notes? type="Note">userdefined notes to variable: Archive Note</notes>
<notes? type="Problem">userdefined notes to variable: Problem</notes>
</var>
</dataDscr>
</codeBook>
```

## DDI Formats Currently Used

### Example: Variable Descriptions from DSDM to DDI 3.1

(for Enhanced Publications)

contains

- Study number
- SPSS variable names, labels, and URNs
- Variable groups
- Complete question, answer texts and notes for each variable
- SPSS values and value labels

```

<?xml version="1.0" encoding="UTF-16" standalone="no"?>
<ddi:DDIInstance
  id="gesis_Archive Study ID" agency="de.gesis" version="Version (Format Major.Minor.Revision default 1.0.0)"
  versionDate="Date (ISO Format YYYY-MM-DD)" xmlns:ddi="ddi:instance:3_1" xmlns:s="ddi:studyunit:3_1"
  xmlns:pd="ddi:physicaldataprotuct:3_1" xmlns:pi="ddi:physicalinstance:3_1" xmlns:c="ddi:conceptualcomponent:3_1"
  xmlns:l="ddi:logicalproduct:3_1" xmlns:r="ddi:reusable:3_1" xmlns:dc="ddi:datacollection:3_1" xmlns:a="ddi:archive:3_1"
  xmlns:xhtml="http://www.w3.org/1999/xhtml" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="ddi:instance:3_1 http://info1.gesis.org/ddi/3_1/instance.xsd">
  <s:StudyUnit id="Archive Study ID_SU">
    <r:UserID type="DBK Study Number"> Archive Study ID </r:UserID>
    <r:Citation>
      <r:Title> Archive Study ID </r:Title>
    </r:Citation>
    <s:Abstract id="Archive Study ID_A">
      <r:Content>Study description not available.</r:Content>
    </s:Abstract>
    <r:UniverseReference>
      <r:ID>UNIVERSE_REF</r:ID>
    </r:UniverseReference>
    <s:Purpose id="Archive Study ID_P">
      <r:Content>Archive Study ID DSDM to EPE DDI 3.1 Export (1.4.0)</r:Content>
    </s:Purpose>
    <dc:DataCollection id="Archive Study ID _DatCol">
      <dc:QuestionScheme id="Archive Study ID _QueSch">
        <dc:QuestionItem id="Archive Study ID _QSPSS Variable Name">
          <dc:QuestionText xml:lang="en">
            <dc:LiteralText>
              <r:Description>Complete Question Text</r:Description>
              <dc:Text>Complete Question Text</dc:Text>
            </dc:LiteralText>
          </dc:QuestionText>
          <dc:CategoryDomain>
            <r:CategorySchemeReference>
              <r:ID> Archive Study ID _QSPSS Variable Name_CatSch</r:ID>
              </r:CategorySchemeReference>
            </dc:CategoryDomain>
          </dc:QuestionItem>
        </dc:QuestionScheme>
      </dc:DataCollection>
  </s:StudyUnit>
</ddi:DDIInstance>

```

```

<l:LogicalProduct id="Archive Study ID_LogPrd">
  <l:DataRelationship id="DataRel">
    <l:LogicalRecord hasLocator="false" id="LogRec">
      <l:VariablesInRecord allVariablesInLogicalProduct="true">
        <l:VariableSchemeReference>
          <r:ID>Archive Study ID_VarSch</r:ID>
        </l:VariableSchemeReference>
      </l:VariablesInRecord>
    </l:LogicalRecord>
  </l:DataRelationship>

  <l:CategoryScheme id="Archive Study ID_QSPSS Variable Name_CatSch">*
    <l:Category id="Archive Study ID_QSPSS Variable Name_CatValue">
      <r:Label>SPSS Value Label</r:Label>
    </l:Category>*
  </l:CategoryScheme>

  <l:CodeScheme id="Archive Study ID_QSPSS Variable Name_CodSch">
    <l:Code>
      <l:CategoryReference>
        <r:ID>Archive Study ID_QSPSS Variable Name_CatValue</r:ID>
      </l:CategoryReference>
      <l:Value>SPSS Value</l:Value>
    </l:Code>*
  </l:CodeScheme>*

  <l:VariableScheme id="Archive Study ID_VarSch">
    <l:Variable id="SPSS Variable Name"
urn="urn:ddi:de.gesis:VariableScheme.Archive Study ID_VarSch.1.0.0:Variable.SPSS Variable Name.1.0.0">*
      <r:UserID type="SPSS Variable Name"> SPSS Variable Name </r:UserID>
      <r:Label maxLength="255" type="label"> SPSS Variable Label</r:Label>
      <l:QuestionReference>
        <r:ID>Archive Study ID_QSPSS Variable Name</r:ID>
      </l:QuestionReference>
    </l:Variable>
  </l:VariableScheme>

```

D

3

```

        <l:VariableGroup id="VariableGroup ID_VarGrp" >*
            <l:GroupTypeCoded codeListID="Group Type"
                codeListAgencyName="DDI">Subject</l:GroupTypeCoded>
            <r:Label> Variable Group Name </r:Label>
            <l:VariableReference>
                <r:ID>SPSS Variable Name</r:ID>
            </l:VariableReference>*
        </l:VariableGroup>
    </l:VariableScheme>
</l:LogicalProduct>
<a:Archive id="gesis_ARCHIVE">
    <a:ArchiveSpecific>
        <a:ArchiveOrganizationReference>
            <r:ID>ArchiveOrg</r:ID>
        </a:ArchiveOrganizationReference>
    </a:ArchiveSpecific>
    <a:OrganizationScheme id="gesis_OrgSch">
        <a:Organization id="ArchiveOrg">
            <a:OrganizationName>GESIS</a:OrganizationName>
        </a:Organization>
    </a:OrganizationScheme>
    <r:LifecycleInformation>
        <r:LifecycleEvent id="EVENT_1">
            <r>Date>
                <r:SimpleDate>Date (Format: YYYY-MM-DD)</r:SimpleDate>
            </r>Date>
            <r:AgencyOrganizationReference>
                <r:ID>ArchiveOrg</r:ID>
            </r:AgencyOrganizationReference>
            <r:Description>Export from DSDM to EPE DDI 3.1 (1.4.0)</r:Description>
        </r:LifecycleEvent>
    </r:LifecycleInformation>
</a:Archive>
</s:StudyUnit>
</ddi:DDIInstance>

```

## Overview

Initial Situation – Different Archiving Tools

Main Tools – DBKEdit/ DSDM/ CBE

Intention – Integration of Different Archiving Tools

Proceeding – Requirements Analysis

DDI Formats Currently Used

✦ Requirements Concerning DDI Formats

Getting Started with STARDAT Development

Challenges of Going into DDI3

## Requirements Concerning DDI Formats

- 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)
- 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
- Migration issues need to be solved



## Overview

Initial Situation – Different Archiving Tools

Main Tools – DBKEdit/ DSDM/ CBE

Intention – Integration of Different Archiving Tools

Proceeding – Requirements Analysis

DDI Formats Currently Used

Requirements Concerning DDI Formats

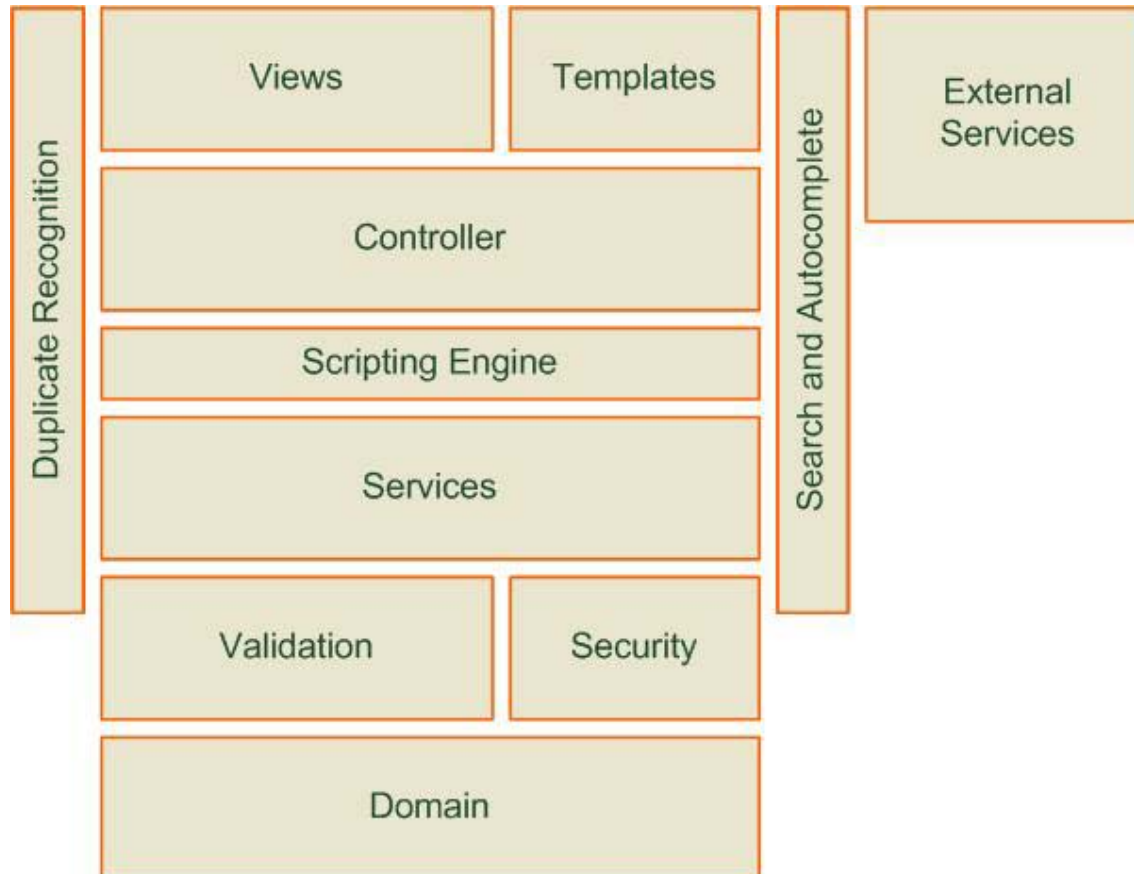
✦ Getting Started with STARDAT Development

Challenges of Going into DDI3

## Getting Started with STARDAT Development

- Architectural requirements
  - Browser-based web application
  - Transaction-based multi-user access
  - Distributed undo-/redo-mechanism
  - Multi-threading
  - Automation of metadata management tasks
    - e. g. long running tasks with progress reporting and cancelation support
  - Extensible framework for further needs of metadata documentation
  - Keeping in mind: REST and SOA
  
- Experimental prototyping
  - First milestone until May 2011
  - Implementation of only a small set of functional requirements
  - Check of feasibility of architectural and technical ideas
  - Iterative process to improve development and implementation

## Getting Started with STARDAT Development First Draft of Architecture



## Overview

Initial Situation – Different Archiving Tools

Main Tools – DBKEdit/ DSDM/ CBE

Intention – Integration of Different Archiving Tools

Proceeding – Requirements Analysis

DDI Formats Currently Used

Requirements Concerning DDI Formats

Getting Started with STARDAT Development

✦ Challenges of Going into DDI3

## Challenges of Going into DDI 3

### Really Internalize Lifecycle Orientation

- Managing documentation process of complex social science data
  - Apply adequate grouping approach
  - Identify a strategy to establish resource packages
  
- Migration issues
  - Find equivalent elements
  - Identify additional elements needed
  - Identify re-usable elements
  - Re-structure the elements
  
- Building software
  - Existing software tools are *static*
    - Only their combination supports lifecycle management
  - New software tool (!) shall be *dynamic*
    - Lifecycle management is inherently contained

## Challenges of Going into DDI 3 To Be ... or Not to Be

What does it mean when we talk about DDI 3 usage?

- supporting DDI 3
  - proprietary domain model
  - proprietary storage
  - I/O module with some squeezing mapping
  
- compatible with DDI 3
  - DDI 3 domain model, perhaps some proprietary extensions
  - storage in relational database
  - mapping between XML and relational database
  
- based on DDI 3
  - DDI 3 domain model, no proprietary extensions
  - storage in flat XML files or native XML databases
  - no mapping, full interoperability

## Challenges of Going into DDI 3

### What DDI 3.x? How to manage DDI version evolution?

- Extensions
  - no problem
  
- Modifications
  - not trivial
  - but assumed to be manageable
  
- Downgrade mismatch
  - unconditionally avoid nasty lock-in effect
  - but how?

## Challenges of Going into DDI 3

### Performance Issues

- The performance topic has been addressed
- Solutions are already discussed by the community
  - DDI light
  - XML – RDB transformation
 → We need to understand and adapt to our needs
- We need always keep in mind performance because of its extensive side effects on architecture
  - What usage of DDI is possible?
  - What storage technique do we use?
  - Where are the bottle necks?



## Challenges of Going into DDI 3

### Not at least Reuse, Reuse, Reuse of Software!

- More generalized thinking
- Stringent component orientation in architecture design
- Advanced documentation needs
- We would like to offer our results to the DDI3 community
- We are interested in exchange.
  - One first step done: We already started an exchange with DDA, we hope to reuse DDI3 Editor code.