

Implementation of DdiEditor in DDA

EDDI12 – 4th Annual European DDI User Conference
Bergen December 3rd-4th 2012

Anne Sofie Fink
Danish Data Archive



Outline of the presentation

- DDA – Danish Data Archive
- Introduction to the DdiEditor
- New work practices for data curation processes
- Experiences from on-going software development
- Further development

Danish Data Archive (DDA)

National data archive for structured research data for social science and health science

Kind of material:	Survey data
Activities:	Data retrieval Data processing and preservation Data dissemination
Funding:	Publicly funded
Staffing:	App. 20 employees, 2-3 software developers
Organisational affiliation:	Unit in the Danish National Archives (Statens Arkiver)

DdiEditor 2.0

Framework for data processing and processes

- Ingest of wiki text file and SPSS-file (incl. descriptive statistics)
- Linking of variables and questions
- Use of generic code
- Creation and use of resource packages
- Manipulation of metadata elements
- Integration of study description and metadata on variable level
- **Production of DDI-L data and documentation for long term preservation and dissemination**
 - Export to statistical software – SPSS Syntax file and csv data file
 - User documentation – style sheet-based codebook

Wiki syntax

- Data input in wiki syntax

Questions, variables, categories, concepts, universes, instrumentation (question flow)

- Syntax for repeated questions and categories
- Syntax for multiple questions
- Syntax for filtering (if then else/while/sequence)

- Naming of concepts

- Naming of sub universes

Section of wiki file

```
*V8 Q2A: Other, please write
''state'' Q3: What are your primary duties? Please tick the duties which you spend the most
time on. If you spend an equal amount of time on different duties, please tick all of them.
*V9 Q3A: Editorial management
**Not mentioned
**Mentioned
*V10 Q3B: Administrative management
**V9
*V11 Q3C: Editorial secretariat
**V9
*V12 Q3D: Locale news
**V9
*V13 Q3E: National news
**V9
*V14 Q3F: Other
**V9
''ifthenelse'' V14==1 V15 na Answer question 3G, Other
*V15 Q3G: What are your primary duties? Other: (Please specify)
===Deadlines and feedback===Deadlines and feedback valued by employees
''state'' Now you will be asked questions concerning deadlines and feedback
''interview'' You are doing god -about half way through the questionnaire :) NA
*V16 Q4: Do you have to meet a fixed deadline?
**No (Skip to questions 6 (V23))
**Yes, but not every week
**Yes, once a week
**Yes, several per week
**Yes once a day
**Yes, several times a day/almost all the time
```

SPSS file

- Variables – naming and labeling
- Codes
- Adding DDA-variables
(DDA no. and DDA serial number)
- Recoding of values for missing data (DDA standards: 9, 10, 11 - unknown, not relevant (na), not participating)
- Definition of measurement scale for variables (ordinal, nominal and scale)
- Production of descriptive statistics
(frequencies on ordinal and nominal variables)


Genericode and resource packages


- DdiEditor plug in for creation of controlled vocabularies
- Genericode for DDA CVs
 - Study state classes
 - Access restrictions
 - CESSDA Subjects categories
- Resource packages
 - International
 - E.g. occupation, education
 - National
 - E.g. family relationship, regions
 - Survey specific
 - Complex/complicated category structures e.g. health condition


DDI-XSLT Print


User documentation – Codebook style output - Variable

VAR V18 5B. DEADLINES CREATES MOTIVATION

 [Deadlines and feedback](#)

 [Answer questions 5A to 5F](#)

 Variable is filtered by: V16>1

 Q5: To which degree do the following apply to you? (Mark only once per line)

Q5B: Deadlines create motivation for the work

	% MD	% Number	Code	Category
7	7	144	1	To a very high degree
34	36	739	2	To a high degree
35	37	760	3	Moderately
11	12	246	4	To a low degree
4	4	90	5	To a very low degree/nor at all
3	3	57	6	Not relevant for me
2		44	9	Unknown
5		121	10	Irrelevant
100	100	2201		























Response Rate: 93

Scale: Ordinal

DDI-XSLT Print

User documentation – Codebook style output - Instrumentation

Main sequence

- ☐ Your current job. If you have more than one job, please answer the questions for your primary occupation i.e the job in which you work the most hours.
-  [V4](#): Q1: How long have you held your current position with this company? Number of years: (roughly)
-  [V5](#): Q1B: How long have you held your current position with this company, Number of months:
-  [V6](#): Q1C: Working hours per week in current job
-  Make sure only to tick one box.
-  [V7](#): Q2: What type of company do you work for?
-  If [V7==13](#) then [seqc-Q2: What type of company do you wor ...](#)
-  [V8](#): Q2A: Other, please write
- ☐ Q3: What are your primary duties? Please tick the duties which you spend the most time on. If you spend an equal amount of time on different duties, please tick all of them.
-  [V9](#): Q3A: Editorial management
-  [V10](#): Q3B: Administrative management
-  [V11](#): Q3C: Editorial secretariat
-  [V12](#): Q3D: Locale news
-  [V13](#): Q3E: National news
-  [V14](#): Q3F: Other
-  If [V14==1](#) then [seqc-Q3F: Other](#)
-  [V15](#): Q3G: What are your primary duties? Other: (Please specify)
- ☐ Now you will be asked questions concerning deadlines and feedback
-  You are doing good -about half way through the questionnaire :)
-  [V16](#): Q4: Do you have to meet a fixed deadline?
-  If [V16>1](#) then [Answer question 5A to 5F](#)
-  [V23](#): Q6: How would you rate your overall health?
-  [V24](#): Q7: Does your collective agreement or contract give you the right to be compensated for overtime (e.g. compensatory time-off or paid overtime)?
-  If [V24==1](#) then [Answer question 8 and 9](#) else
-  [V28](#): Q10: How often are your planned working hours changed in short notice?

Experiences from on-going software development and releases

- Between enthusiasm and resistance
- Bug tracking to replace zero-defects mindset
- SCRUM as software development methods
- Test, test, test, ...
- And corporation

Conclusion

- Support for new curation processes
- Production of dynamic data and documentation in DDI-L for preservation and dissemination
- Export to SPSS
- Creation of genericcode for DDA cv's
- Creation of resource packages

... and way forward

Further development

- Less dependence on wiki based input
- User friendly creation of resource packages
- Support for linking of metadata elements on study and variable level for longitudinal studies
- Support for automatic review processes
- Versioning of metadata elements
- Support for data exchange
 - Import from web survey tools to DdiEditor

What to know more?

- Come to the poster session
 - Try out the Ddieditor
 - Have a look at the DdiXslt (SND and DDA)
 - See the landing page solution
- Visit: <http://code.google.com/p/ddieditor>
- Questions(!) & thanks
- Contact information: asf@dda.dk