Exploring DDI3 in an Archive

Mari Kleemola
Information Services Manager

Matti Heinonen
Senior Systems Analyst

Finnish Social Science Data Archive

EDDI, Bonn, 4.12.2009

- Motivation
- Steps to take
- Test case ISSP 2006
- Building a database

A Step Aside

- KDK-PAS
  - = The National Digital Library, Long-term preservation project
  - – aiming for a national solution for preservation of digital objects
  - – outcome undefined at this point
- Expections of our customers
  - – whatsthatthingkidsaredoingnowdays? Web2.0?
- Need for more integrated production environment at FSD

Motivation to DDI3

- Reworking our production environment
- Need to produce various kinds of metadata
  - – metadata for long-term preservation
  - – multilingual metadata
  - – DDI3 for Cessda (no decisions yet)
  - – KDK-PAS, web20
- Opportunities for thematic subsets of our collection & their markup
- CESSDA PPP WP8

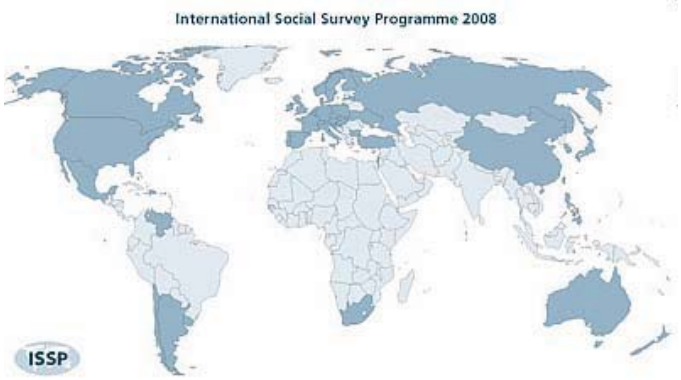
Steps to Take

- Map existing DDI2.1 documentation into DDI3, create DDI3 mark-up
- Identify future requirements (functionalities needed, what DDI3 supports)
- Take a look at FSD's current technical environment
- What have other organisations done / are doing?
- Investigate existing tools
- Explore possibilities for producing new tools

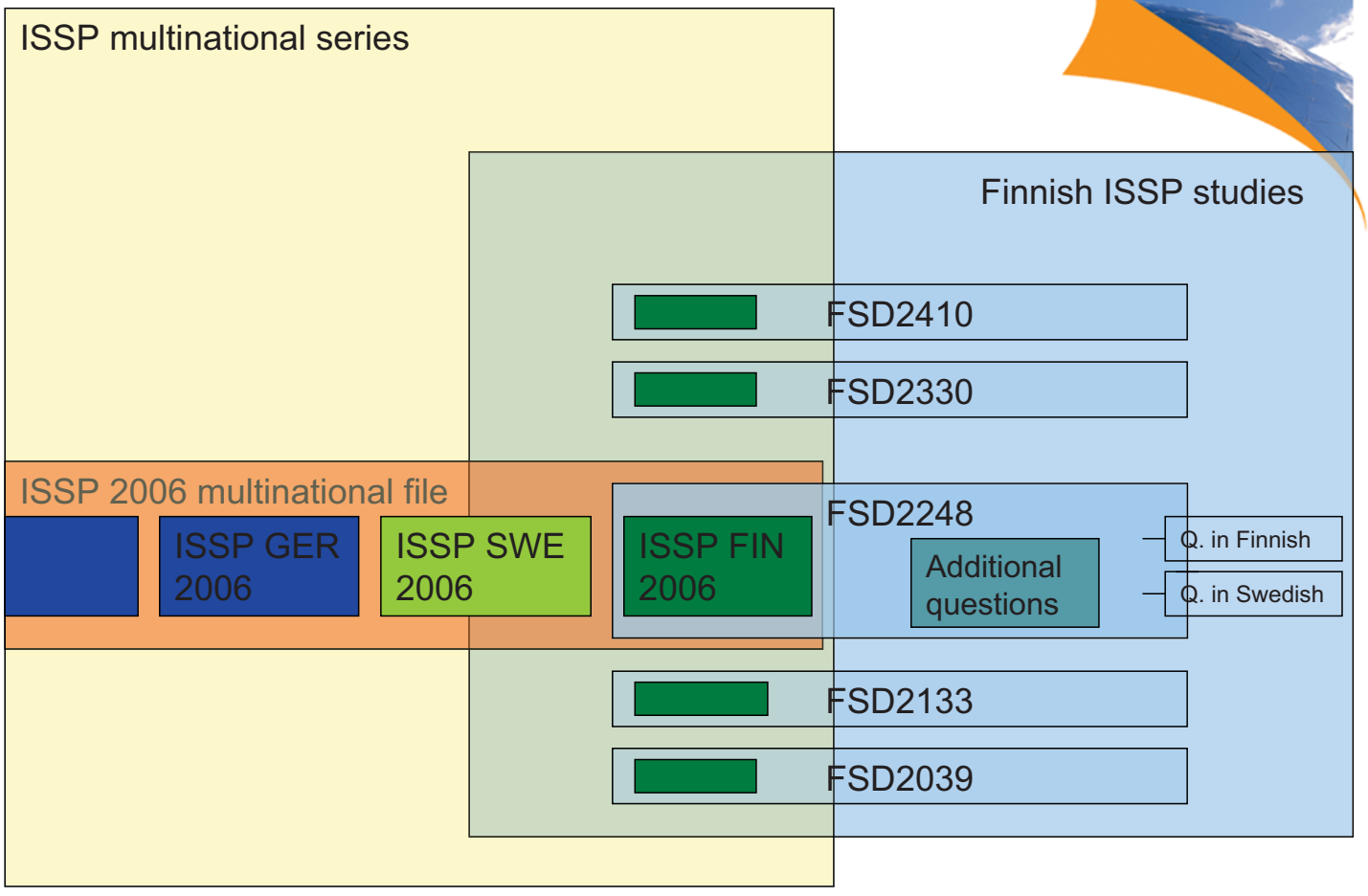
Test Case
ISSP 2006 Finnish Data

International Social Survey Programme

www.issp.org



- Cross-national & annual
- FSD one of Finnish ISSP partners
- Data collected by Statistics Finland
  - two questionnaires: fi and sv-FI
  - documentation in Finnish and in English
- A single study belongs to several groups



Results So Far

- Mapping of FSD's DDI2.1 metadata to DDI3.1
- A test DDI3.1 xml file
- Investigate further:
  - grouping, inheritance
  - comparison
  - map FSD metadata in operational db to DDI3.1
  - which studies (not all?) to "totally DDI3"
  - what customers want/expect/need?
  - European co-operation (CESSDA)

Test Case
Building a Database

Method

- Build DDI 2.1 by hand
  - been doing that for 10 years…
- Build DDI 3.x by hand
  - previous test case
- Build a test database from DDI 2.1, produce DDI 2.1 and DDI3.x from it
  - slowly working on this
  - will never be a production system

Results So Far

- FSD's use of DDI 2.1 well defined
  - good definitions of what we store in each element
  - no CDATA, phewww…
- Our usage of DDI2.1 maps to DDI3.x fairly well
  - but manual labour required

Problems

- Matching language versions
  - they should be identical, but are not
- Ends up with ”monolithic” DDI 3.x
- Database reflects DDI 2.1, not DDI 3.x or our needs
- Mapping elements
  - database may require redundant data to accomodate both 2.1 and 3.x

To Do

- Make more complete
  - cover more elements
  - refactor until satisfied with the data structure
- Incorporate data from our operational database
- Try producing other metadata formats

Thank you!