CONTENTS

• Prior experience with DDI
• Current use of DDI
• Moving forwards
• Accommodating legacy
• Experiments
• Anticipated benefits
PRIOR EXPERIENCE WITH DDI

- DDI Alliance – contributed via Technical Implementation Committee and Controlled Vocabulary subcommittee
- CESSDA Portal development - browse datasets from contributing CESSDA members
  - uses Nesstar architecture and ELSST thesaurus
- DExT development – worked with ODaF to build SPSS to DDI converter
  - import data, DDI metadata from SPSS files; convert to ASCII; re-export to SAS, Stata and SPSS
- More recent work with ODaF on use of the DDI for qualitative data
  - Assessment of DDI 3 for qualitative and mixed method collections, and development of complimentary Qudex schema
    - As covered in session B1 ‘Report back from the DDI Qualitative Working Group’ on Wednesday
CURRENT USE OF DDI

- Data ingest and dissemination
  - Depositor completes Data Deposit Form
  - Data goes in to a relational database
  - Data processors add metadata via bespoke tools
  - DDI 2.1 compliant XML record created and stored
  - Publish DDI 2.1++ catalogue record (levels 1-3 only)
  - Create Lucene Index from XML for website search
  - Run SPSS to DDI script to create variable level metadata
  - DDI 2.1 record imported into Nesstar Publisher and DDI variable level metadata added
  - Publish DDI 1.2 record in Nesstar catalogue
  - Make catalogue records available via OAI (DDI 2.1)
MOVING FORWARD

• Kicked off at internal workshop December 2010
  • Led by Arofan Gregory
• Approach is to
  • Build a parallel ingest/preservation/dissemination process
  • Migrate the metadata
  • Test, test, test
  • Shut down the old process
• All pretty straightforward?

We just need to tackle our legacy systems and metadata
Hatton flight, Grand Union Canal (21 locks, rise 146 ft)
ACCOMODATING LEGACY – STEP 1

• Question Bank - determine how best to recognise relationships between:
  • re-used questions
  • questions (child) and module (parent)
  • questions belonging to a recognised standard
    • e.g. ONS harmonised set
• Longitudinal Studies - determine how best to recognise:
  • continuities between data collections within the same study
    • including question continuity
    • methodological continuities
• Thesauri/Controlled Vocabulary/Ontologies – determine:
  • what have already
  • what else could be collectivised and reused
ACCOMODATING LEGACY – STEP 2

• Create DDI 3 profile via Process and Data Mapping:
  • Identify element set used from DDI 1.2 (Nesstar 3.5 variable level)
  • Identify element set used from DDI 2.1 (all other current DDI use)
  • Map to DDI 2.5
  • Map representative DDI 2.5 instance to DDI 3
ACCOMODATING LEGACY – STEP 3

- Re-architect from client server to web services
  - Break the dependency between UIs and existing databases
  - Allows for use of distributed, heterogeneous data sources
    - Relational Db, XML Db, flat files ....
- Increase use of Thesauri/CVs/Ontologies
  - e.g. Language, Country (fairly static)
  - e.g. Person, Organisation, Address etc (expandable)
- Manage evolution of terms in Thesauri/CVs/Ontologies
  - Machine actionable trail from past to current values
ACCOMODATING LEGACY – STEP 4

- Re-factor the data
  - Remove redundancy via use of CVs
  - Restructure to support new data model
    - concepts and relationships
  - Implement WS interfaces against DDI 3 profile
- Retool the ingest/preservation/dissemination process
  - Build new UIs to create and edit the metadata
- Handle DDI 3 conversion
  - to 2.1 for Nesstar v4, OAI
  - or to 1.2 for Nesstar v3.5
TOOL EVALUATION

- Retool the ingest/preservation/dissemintation process
  - Likely to be combination of own and third party tools
  - Expect to evaluate the following:
    - Colectica
    - Canadian RDC Network toolset
    - eXist
    - Nesstar API
    - Arisddi
    - Various format conversion utilities
EXPERIMENT: PRESERVATION METADATA

- Data provenance and lifecycle actions on data are important for preservation metadata
- Archive has always supported 'pre-ingest' part of the workflow
- As DDI 2 doesn’t take account of the full lifecycle, have been exploring METS and PREMIS standards
- Packaging anything which is not stored effectively with DDI 3 may be done with a METS wrapper
- Evaluate PREMIS vs DDI 3 to identify fit and overlap

- Can DDI 3 meet our needs?
EXPERIMENT: QUALITATIVE DATA

- DDI 2 fine for describing the study and an overview of a whole data collection

- Good down to the individual file level but cannot describe the content of files
  - e.g. the structure of an textual interview data or how files relate to each other

- Working on a standard called QuDex in conjunction with DDI 2 to ensure holistic and detailed description of complex data collections
  - QuDex has the power to relate data, parts of data and annotations to each other

- Can DDI 3 meet our needs?
ANTICIPATED BENEFITS

- Richer study level metadata => resource discovery++
- Much improved Survey Question Bank
- Better handling of Longitudinal studies
- Legacy systems tamed!

Before
http://blogs.sfweekly.com/thesnitch/SharkGW.jpg

After
http://www.seanomeallie.com/images/Shark-Fin-Pull-Toy.jpg