DDI - A Metadata Standard for the Community

Mary Vardigan
ICPSR – Inter-university Consortium for Political and Social Research

Joachim Wackerow
GESIS – Leibniz Institute for the Social Sciences

1st North American DDI User Conference
April 2-3, 2013 – Lawrence, Kansas
Today’s Presentation -- “State of the DDI“

• What is DDI? Benefits and Principles
• Barriers and Challenges
• DDI of the Future
• How You Can Help
What is DDI?

Application using parts of DDI
A Project of and for the Community

- Developed by a self-sustaining membership Alliance recently reconfigured with new Bylaws
- Two major development lines
  - DDI Codebook
  - DDI Lifecycle
- Metadata for both human and machine consumption
- Additional specifications:
  - Controlled vocabularies
  - RDF vocabularies for use with Linked Data
Primary Benefits

• Rich content (currently over 800 items)
• Metadata reuse across the life cycle
• Machine-actionability
• Data management and curation
• Support for longitudinal data and comparison
• Support for preservation and platform-independent software
• Support for a global network
Metadata Reuse

Data Documentation Initiative
Data Documentation Initiative

Machine-Actionability

DDI as backbone for structured metadata

Data/other documents are referenced from DDI

Custom Tools (e.g. Forms-based)
CAI Tools Colectica, MQDS etc.
Information extracted from SPSS etc.

Repurposing

Data Producer Archive

Codebook
Web information systems
Nesstar etc.

Search engines.
Statistical packages

Distribution
Discovery
Analysis
Longitudinal Data and Comparison


"Generic Longitudinal Business Process Model." http://dx.doi.org/10.3886/DDILongitudinal05
Paradigm Shift

• Using structured metadata in standardized form -- a paradigm shift for social science research infrastructure

• We might compare this to the change in the 1990s when people began to understand the value of linked documents
Distributed Approach

• Shift from centralized maintenance and control of metadata by the researcher or archive to a distributed approach

• More efficient and easier to capture information about an event at the time of its occurrence rather than after the fact
Barriers and Challenges

• Complexity
• Level of researcher buy-in
• Need for tools
• Changes in workflow
• Access to metadata
Measuring Success

• Adoption around the world
  – Projects range from the General Social Survey in the U.S., to the Research Data Centres in Canada, to Statistics New Zealand
  – Through IHSN, DDI is now being used in over 70 countries
  – DDI is a centerpiece of proposed CESSDA-ERIC (European Research Infrastructure Consortium)
Data Documentation Initiative

Global DDI
Moving Forward: RDF Vocabularies for Semantic Web

• DDI-RDF Discovery Vocabulary
  For publishing metadata about datasets into the Web of Linked Data
  Based on DDI Codebook and DDI Lifecycle

• XKOS
  RDF vocabulary for describing statistical classifications, which is an extension of the popular SKOS vocabulary
  Publication expected in second half of 2013
DDI of the Future

• Robust and persistent data model (for the metadata), with extension possibilities, variety of technical expressions
• Complete data life cycle coverage
• Broadened focus for new research domains
• Simpler specification that is easier to understand and use including better documentation
Why Now?

• Experiences in developing and using the initial DDI Lifecycle structure

• Lack of a data model proved limiting in:
  • Further improvement of the development line 3.*
  • Development of sustainable software based on different versions of DDI

• Pressure for changes from several directions at once
  • New content from substantive working groups
  • Being approached by new communities
Specific Changes Envisioned I

- Abstraction of data capture/collection/source to handle different types of data
- New content on sampling, survey implementation, weighting, and paradata
- New content developed by the Qualitative Working Group
- Framework for data and metadata quality
- Framework for access to data and metadata
Specific Changes Envisioned II

- Process (work flow) description across the data life cycle, including support for automation and replication
- Integration with GSBPM/GSIM, SDMX, CDISC, Triple-S
- Disclosure review and remediation
- Data management planning
- Development of standard queries and/or interface specifications (such as REST), needed to enable interoperable services
High-level Design Goals I

Coming out of Dagstuhl 2012 workshop:

1. Interoperability and Standards
2. Simplicity
3. User Driven
4. Terminology
5. Iterative
6. Documentation
7. Lifecycle Orientation
8. Reuse and Exchange
9. Modularity
10. Stability
11. Extensibility
12. Tool Independence
13. Innovation
14. Actionable Metadata
Data Documentation Initiative

DDI in Relation to GSIM and Other Standards

REFERENCE STANDARDS/MODELS
- SDMX
- CORE
- ISO 19115
- ISO/IEC 11179
- Neuchâtel
- Others*
- GSBPM

IMPLEMENTATION SYNTAXES/TOOLS
- Google DSPL
- SDMX
- Software Tools
  - PC Axis
  - Blaise
  - Others
- RDF Open Data Vocabularies
- XBRL
- CORE Prototype
- Others*

Key:
- Not widely implemented
- Widely implemented
Data Documentation Initiative

Alignment with GSIM
DDI Development Lines

• DDI Codebook
  – DDI 2.1 (DTD)
  – DDI 2.5 (XML Schema)

• DDI Lifecycle
  – DDI 3.1, 3.2 (XML Schema)
  – DDI X (model-based)
    • Bindings as
      – XML Schema
      – RDF
    • Data as a Service, Service Oriented Architecture (SOA), Web services, REST
    • Recommendations for schemas for database management systems
Time Line Moving Forward

• New work initiated at Dagstuhl workshops on “Moving Forward” in Autumn 2012
• Planned to continue in October 2013
• Ongoing virtual work involving the community in progress – roadmap developed yesterday
• User stories and core being defined
• Mapping among DDI, SDMX, GSIM under way
• First draft of model expected by end of 2014
You Can Contribute!

• Please join us on a working group
• Participate in technical work, including modeling
• Join the DDI Alliance!