DDI Lifecycle: Moving Forward
Status of the Development of DDI 4

Joachim Wackerow
Technical Committee, DDI Alliance
Should I Wait for DDI 4?

• No!
  – DDI Lifecycle 4 is a long development process
• DDI Lifecycle 3.2 is a stable and comprehensive specification for the data lifecycle
• DDI Codebook 2.5 is suitable for single studies in an archive context
• RDF Vocabularies are available now for usage in Linked Data
• Migration paths to DDI 4 will be provided from the DDI specifications above
Why DDI 4

• Need a stable yet flexible specification that can adapt to new requirements
  – Address needs of new communities
  – Extend lifecycle coverage
  – Improve integration with other standards
    • For example: strong relationship to GSIM (General Statistical Information Model)
High-Level Goals

• Improve DDI usability and accessibility
• Enhance the documentation
• In general, develop a sustainable and transparent approach
Architecture Goals

- Information model as the DDI foundation (formal UML model), this will enable:
  - robust and persistent model
  - Improved communication with other disciplines and standards efforts
  - Flexibility in terms of technical expressions of the model
  - Streamlined development and maintenance
- Will simplify the implementation of DDI
- Model-driven specification
- Support both XML and RDF technologies
  - XML for exchange and preservation
  - RDF for discovery and linkage to other domains in Semantic Web
Architecture Goals

• Information model as the DDI foundation (formal UML model), this will enable:
  – robust and persistent model
  – Improved communication with other disciplines and standards efforts
  – Flexibility in terms of technical expressions of the model
  – Streamlined development and maintenance
• Will simplify the implementation of DDI
• Model-driven specification
• Support both XML and RDF technologies
  – XML for exchange and preservation
  – RDF for discovery and linkage to other domains in Semantic Web
Content Goals

• Abstraction of data capture/collection/source to handle different types of data
• New content on sampling, survey implementation, weighting, and paradata
• New content related to qualitative data
• Framework for access to data and metadata
• Process description across life cycle, including support for automation and replication
• Integration with GSBPM/GSIM, SDMX, CDISC, Triple-S
• Data management planning
• Fieldwork
Design Principles

1. Interoperability, 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
Model-Driven Approach

- Model represents understanding how the data lifecycle can formally be described in objects and their relationships
- Related XML Schemas, RDF vocabularies, and documentation can be generated from the model
- Changes and additions will be only made in the model
Functional Views

• Subsets of the model represent all objects related to a specific purpose like questionnaire or data file

• Users (software developers, instance editors) will only see the selected functional view in the chosen technical format (XML, RDF)
  – They have not to focus on large object library
In DDI 4 terms

Library

Package

Data sets
Physical data structures
Logical data structures

Study
Glue
Constructs
In DDI 4 terms
Efforts to Date

Sprints

• Dagstuhl October 2012, 5 days
• Dagstuhl October 2013, 5 days
• EDDI, Paris – December 2013, 2 days
• NADDI, Vancouver – April 2014, 5 days
• IASSIST, Toronto – June 2014, 5 days
• Dagstuhl – October 2014, 5 days
• EDDI, London – November 2014, 5 days
• Plus virtual meetings in between
First Draft Releases of Selected Parts in First Quarter of 2015

• Packages
  – Primitives
  – Complex Data Types
  – Identification
  – Process Core
  – Conceptual
  – Agents

• Functional Views
  – Agents
  – Discovery (without related packages)
Purpose of Draft Releases

• Draft releases of selected parts (packages and functional views)
• Iterative process: improvement and release of further drafts
• Goals:
  – Quality improvement by broader review of model, comments from community, and involvement of more perspectives
  – Test and review of technical production framework
DDI 4 Products

• Whole Model:
  – A PDF specification of the entirety of DDI 4. (Needs high-level documentation)
  – An XMI file containing DDI 4 (possibly in different flavours like Enterprise Architect)
  – HTML documentation for DDI 4 (Needs high-level documentation)

• For each Functional View:
  – PDF specification of the view (Needs high-level doc)
  – XSD for the View (with field-level documentation inline)
  – Clickable XSD documentation (created using the same tool as other versions of DDI-Lifecycle)
  – View OWL (in RDF XML) - field level doc as comments
  – View HTML documentation for OWL
  – View model documentation in HTML

• Tools:
  – The production framework itself as a toolkit for those making their own views

• Other deliverables are being considered (SQL, Jason, Java, C#, etc.)
Future Work

• Packages for next draft release
  – Data Capture
  – Data Description

• Some further topics in the works
  – Enhanced Data Citation
  – Codebook
  – Qualitative Data
  – Methodology
  – Controlled Vocabularies
Get Involved

• Commenting draft releases
• Contributing to working groups
• Participating in sprints
• Support from organizations by contributing:
  – project management, programmers time, meeting rooms

• Resources:
  – High level overview:
    • http://www.ddialliance.org/ddi-moving-forward-process-summary
  – Details at collaboration platform:
    • https://dditools.atlassian.net/wiki/display/DDI4/*Moving+Forward+Project
  – Development platform:
    • http://lion.ddialliance.org/
  – IASSIST 2014 presentation
    • http://www.library.yorku.ca/binaries/iassist2014/2G/2G-Wackerow.pptx