DDI High-Level Overview for Researchers, Data Producers, and Data Managers

# I. Challenges

The challenges are mostly overcoming the reluctance of researchers and data managers to take on extra work when they do not see the point. This requires an explanation of the benefits of good documentation, metadata management and reuse to them if these challenges are to be met.

Questions/issues from data producers & researchers:

* Why should I bother? It does nothing for ME.
* Ownership of data and reluctance to share: “This is my data – why should I let my competitors have it?”

Questions/issues from data managers:

* Documentation is additional work - existing systems already do it to a sufficient level.
* They are not considering the needs of end users/re-users for documentation to support discoverability or usability.

Common issues:

* Privacy and consent and associated risks make them reluctant to share.
* Concept of "upstream metadata" needs to be sold. What is the return on investment for upstream metadata?
* How does legacy metadata fit in?
* Where does the funding come from for this extra work?

# II. Approaches

In order to address these issues and answer these questions, we must introduce some basic concepts regarding structured documentation and metadata management.

- Present the concept of structured metadata vs. straight document-based codebooks.

- Present the idea of active metadata management. There are several levels of increasing sophistication, with attendant costs in terms of culture/process change and tooling:

Level I: Content guidelines/model for documentation (allows for future expansion)

Level II: Enhanced data dictionary in centralized data base (building on existing infrastructure)

Level III: Centralized metadata management using dedicated DDI tools (fully active metadata management)

- Describe what DDI does at each level. Introduce the distinction between study-level, dataset-level, and variable-level metadata.

- Introduce the concept of metadata re-use: “one source of truth.”

- Introduce the trends within research toward funding of larger collaborative projects

- Introduce requirements for data-sharing as a condition for publication (eg, *Nature)*

- Emergence of data citation as a way of enhancing researcher reputation

Once these ideas have been presented, the reasons for doing each type of metadata management can be presented.

1. Structured documentation

* Better support for providing metadata in different formats to different users
* Support for data discovery systems (portals, catalogues)
* Helps with communication across research projects with larger groups of collaborators (saves on learning curve), including passing data between analysis packages. Problems with institutional memory within projects when temporary researchers (students, 3-5 year contracts) leave. Single point of failure if everything is inside one person's head.
* Standardization provides a degree of consistency within communities and institutions which lowers costs for everyone over the long term (skills transfer between different studies, etc.). Increases in data quality/usability.

1. Centralized metadata in existing data management systems

* Recording metadata at each stage for use in future (questions in a centralized repository).
* Single source of truth supports metadata reuse

1. Dedicated DDI Metadata Management Tools

* For data managers - you can leverage your metadata across studies, and it will save you work. (Repeat data collection.)
* For researchers: Additionally, consistency because of reuse produces more comparable data.

1. General Benefits

* Data citation & data journals provide career credit
* Publishers require citation of data sources (DOIs) making your data more visible
* Funders require data sharing and reuse (data management plans) Germany - Domain Data Protocols - discoverability, too.

# III. Resources

Most existing overview material seemed too technical for this level of introduction. These slides are easy to write, and maybe we should simply do them from scratch.

- Old Dagstuhl training slides on use cases (overview diagrams) - where is this? Ask Wendy.

- Colectica slides that Chifundo has from EDDI 2014 training

- Presentation from Wendy Thomas and Marcel Hebing from EDDI 2013 (questionnaire example showing which parts of questionnaire goes into which field in DDI)

- “ Jane Fry: “Power of DDI Metadata” (researchers, what is needed? How they want to streamline research, etc. Tools screenshots to explain DDI creation. – in Training Libraries on DDI site.

- “Getting Started Guide – draft 1” (Word Document) – data catalogue and codebooks, explanation of importing Excel (Training Group Wiki under Archive Files)

- “Publishing Microdata <ODESI> Using Nesstar Publisher 4.x (Using DDI 2.x) “

- IHSN Microdata Toolkit Guidelines – “Quick Reference for Data Archivists”

- “Creative Codebook” by Amber Leahy (How to make a codebook with…) – under Training Group Archive Files