DDI: What is it? How is it used?

Ontario Data Community Meeting

December 6, 2018

Jane Fry & Michael Steelworthy

Agenda

- What is DDI
  - Brief Background
  - Brief History
  - DDI Development
- DDI Case Studies #1
- DDI Case Study #2

Brief Background

- DDI
    - Data Documentation Initiative
        - http://www.ddi-alliance.org/
    - An international specification
    - Formats documentation for a social science data file
        - More useful than a word or text file
    - Supports the entire research data lifecycle

- This is one of many standards used
    - Dublin Core, METS, OAIS, ISO 19115, ISO 17369 SDMX

Brief Background (cont'd)

- Creates a standard format
  - Used to markup codebooks
  - Meaningful and consistent
  - Metadata is both human and meaningful

- Gives codebook level details such as
  - dataset contents
  - variable labels
  - frequencies
  - question text for each variable
  - …

Caveat!

- DDI is powerful provided that
  - *the information is entered into the appropriate fields when marking up the document*

Brief Background (cont'd)

- Remember
  - DDI facilitates the creation of metadata

- Expressed in XML
  - The XML schema is a way of tagging text for meaning, not appearance
  - Defines
    - Which tags are available
    - The order the tags will appear in a document
    - Whether the tags are required or optional
    - Whether the tags are repeatable or not

Definitions

markup

The characters and codes that change a text document into an XML or other Markup Language document. This includes the < and > characters as well as the elements and attributes of a document.

tags

Fragments of text used to organize content, usually delimited in a set format. Example of XML tags:

```
<book> <chapter> <title>The Beginning</title> <intro>blah blah blah...</intro>
</chapter> </book>
```

In the example above, `book`, `chapter`, `title`, and `intro` are tags. They do not convey content, but rather the context of the content. The < and > are used to signify what is a tag and what is content.

Reference: https://www.icpsr.umich.edu/icpsrweb/ICPSR/support/glossary#T

Example of DDI Compliant Codebook

```xml
<?xml version='1.0' encoding='UTF-8'?>

<codeBook version="1.2.2" ID="ctums_82M0020XCB_E_2004_ann_person-file">

<docDscr>

 <citation>

  <titlStmt>

   <titl>

     Canadian Tobacco Use Monitoring Survey, 2004: Annual, Person File

   </titl>

   <subTitl>

    Annual, Person File

   </subTitl>

   <altTitl>

    CTUMS 2004: Annual, Person File

   </altTitl>

    <IDNo>

    ctums_82M0020XCB_E_2004_ann_person-file

   </IDNo>
```

Benefits

- Interoperability
- Rich content
  - Granular
  - Expansive
- Increased search capability
  - Precision in searching
- International community

Challenges

- Complexity

- Level of researcher buy-in
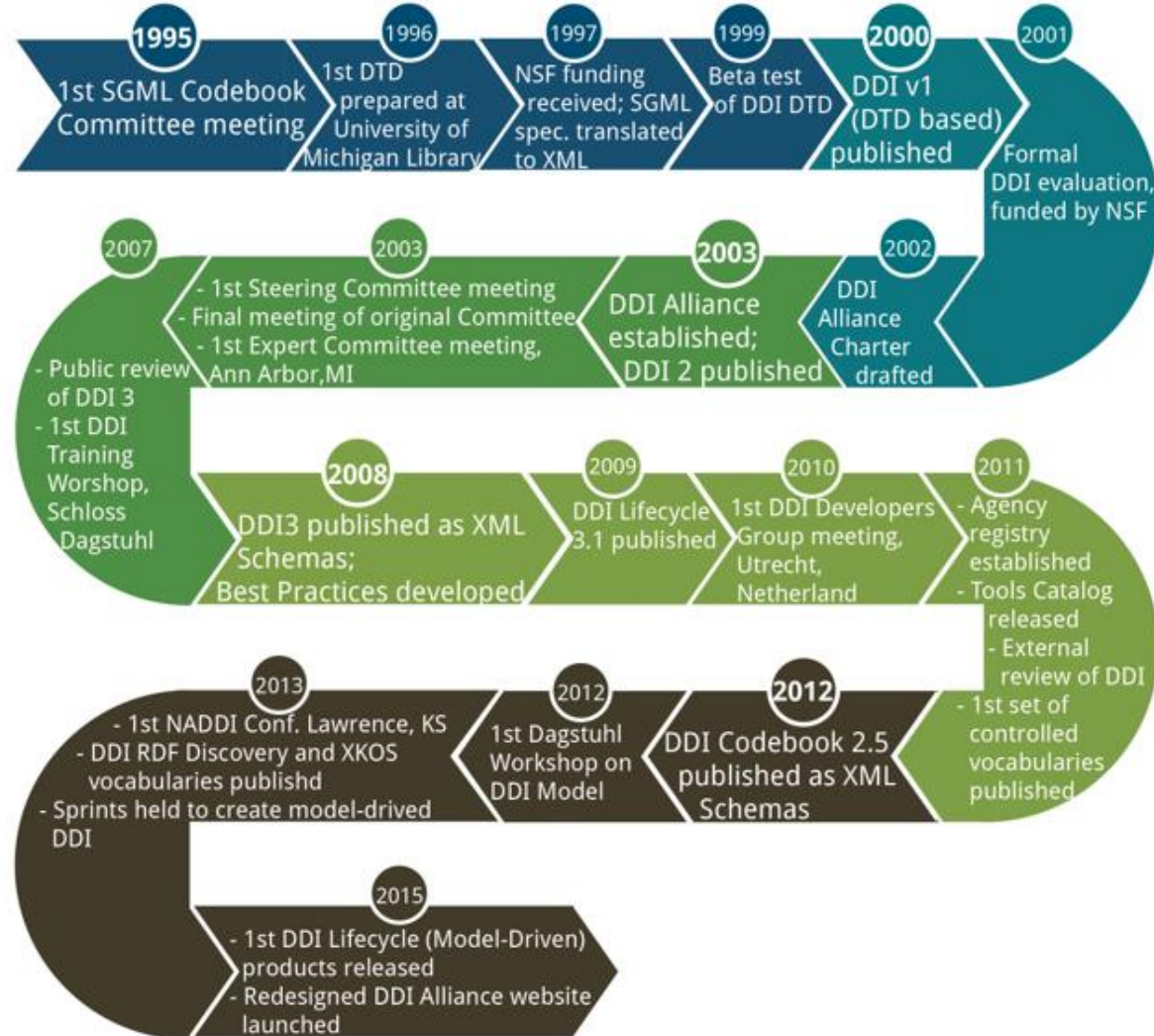
- Need for tools

Brief History

- Started in 1995
  - By ICPSR
  - International cttee
  - In Quebec city

- First version published in 2000
  - DDI 1

- New versions published
  - 2003 – DDI 2
  - 2008 – DDI 3

- Sprints

Reference: http://www.ddialliance.org/what/history.html

Who uses DDI

- Norwegian Social Science Data Services
- Harvard University
- American University
- DLI (Statistics Canada)
- Health Canada
- Bureau of the Census
- University of Michigan
- Bureau of Labor Statistics
- ICPSR

- Yale University
- ESRC Data Archive (UK)
- University of California, Berkeley
- University of Southern Denmark
- The Roper Center
- Zentralarchiv für Empirische Sozialforschung (GESIS)

Who else?

- CESSDA Data Portal (European quantitative social science datasets)
- Australian Social Science Data Archive
- DAMES Project (UK)
- DataFirst (at University of Cape Town)
- Israel Social Science Data Center
- Philippines National Statistics Office
- Statistics New Zealand
- Vision of Britain (historical view between 1801 and 2001)
- World Bank (International Household Survey Network)
- ODESI

DDI Development

- DDI now branched into 2 separate development lines or metadata standards
  - DDI Codebook
    - aka DDI C
    - Formerly DDI 2
    - This was the original one
    - Built to emulate a physical codebook

  - Sections
    - Document Description • Study Description • Data Files Description • Variable Description • Other Study Related Materials

DDI Development (cont'd)

- DDI Lifecycle
  - aka DDI L
  - formerly DDI 3
  - Supports the research data lifecycle
  - The one most new users are learning

- Sections
  - Study Concept • Data Collection • Data Processing • Data Distribution • Data Archiving • Data Discovery • Data Analysis • Repurposing

DDI Alliance

- 2003
- Membership organization
- Self sustaining
- Members have a voice in DDI development
- Membership documents on line
  - Charter, by-laws, forms, …
- Publications all on line
- http://www.ddialliance.org/alliance

Which DDI do I use?

- DDI C

  - Relatively straight forward
  - If you want to catalog a dataset
  - If you are describing a single study

- DDI L

  - If you are focusing on a lifecycle model
  - Broken down into different functions
  - Are you documenting questionnaires?
  - Are you documenting data?
  - Are you doing both?

18

Final Challenge

- How to go from DDI C to DDI L?

- What are the challenges?
    - Too much work to convert to DDI L
    - Insufficient resources (people, $)
    - Nesstar uses DDI C

- DDI 4 has an integrative vision

More information

- DDI website
  - http://www.ddialliance.org/
  - Excellent resource
    - FAQ
    - TOOLS
    - Markup examples
    - Metadata resources
    - Training resources
    - Publications

Reference: http://www.ddialliance.org/community/join

Case Study #1

Case Study #2

Thank you!

Jane Fry

Data Services Librarian

MacOdrum Library

Carleton University

613.520.2600 x1121

jane.fry@Carleton.ca

Michael Steelworthy

Coordinator, Research Data Services

Wilfrid Laurier University Library

Wilfrid Laurier University

519.884.0710 x3574

msteelworthy@wlu.ca