Carleton UNIVERSITY

Canada's Capital University

Introduction to DDI: for the uninitiated

NADDI 2019
Jane Fry
April 24, 2019

Agenda

- Introductions
- Brief background of DDI
- Exercise 1
- Getting started
- Examples
- Integration into a data lifecycle workflow
- Exercise 2
- Wrap-up

Introductions

- **Name**
- **Where you work (name, city, country)**
- **In 20 words or less**
  - What you do

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

Definition

- **DDI**
  - Data Documentation Initiative
  - *“An effort to develop a specification for documenting data files in XML. The DDI Alliance is the organization that created the specification, ... More information can be found on the [DDI website](#).”*

Source: https://www.icpsr.umich.edu/icpsrweb/ICPSR/support/glossary

Why Use DDI?

- *"DDI encourages **comprehensive description** of data for discovery and analysis and supports **effective data sharing**. Because DDI is a **structured** standard, it facilitates **machine-actionability and interoperability** and it can actually be used to **drive systems**. Another feature of DDI is its focus on **metadata reuse**; "enter once, use often" means you can reuse metadata over the course of the data life cycle to avoid costly duplication of effort.*
- *DDI has advantages for several different audiences:*
  - *Librarians*
  - *Managers*
  - *Repositories*
  - *Researchers*
  - *Developers"*

Source: http://www.ddialliance.org/training/why-use-ddi

DDI: Librarians

- *"Standards are important to the effective functioning of libraries. Using a standard vocabulary to document research data leads to consistency and improved interoperability.*
- *DDI is designed to make research data independently understandable. DDI provides a standard structure for all of the metadata that accompanies a dataset and helps users of that dataset to interpret its contents. This is useful when assisting patrons and data analysts.*
- *DDI is an open, non-proprietary standard and anyone can use it."*

Source: http://www.ddialliance.org/training/why-use-ddi

DDI: Managers

- *“Metadata are expensive to produce, so reusing structured, standardized metadata makes good business sense.*
- *DDI promotes interoperability and thus supports partnerships with others that involve data and metadata exchange.*
- *DDI’s structure can enable effective search and discovery, subsetting, generation of syntax files, and flexibility in display, resulting in many efficiencies.”*

Source: http://www.ddialliance.org/training/why-use-ddi

DDI: Repositories

- *"Codebooks have long been used to interpret data files, but PDF and Word codebooks are not "intelligent." In contrast, DDI codebooks are structured and can be interactive, enabling users to navigate through a collection.*
- *DDI can serve as a foundation for data catalogs as it provides a standard structure for searching at both the study and variable levels to enable users to discover data of interest.*
- *Using DDI throughout the archiving life cycle can streamline the repository's workflow, leading to efficient ingest, management, and preservation of data."*

Source: http://www.ddialliance.org/training/why-use-ddi

DDI: Researchers

- *"Recent open access mandates from funders require that data be shared in order to validate results and to encourage new discoveries. This means that data must be well-documented, which is DDI's strength.*
- *Complex, longitudinal data projects require additional levels of data management. DDI can support this and can enable creation of reports, displays, and tools that leverage the richness of the data. Some examples are question banks, concordances, and interactive codebooks.*
- *The structure of DDI can support data comparison and harmonization."*

Source: http://www.ddialliance.org/training/why-use-ddi

DDI: Developers

- *"Using a structured standard optimizes machine-actionability and makes programming against the structure possible.*
- *DDI can actually drive process, leading to greater efficiencies.*
- *DDI can be used with relational databases to increase flexibility."*

Source: http://www.ddialliance.org/training/why-use-ddi

Machine-actionable

- **Definition**
  - *“This term refers to information that is structured in a consistent way so that **machines**, or computers, can be programmed against the structure. DDI provides **machine-actionable** metadata.”*

Source: https://www.ddialliance.org/taxonomy/term/198

Other Standards: Archival

- **Dublin Core**
  - Basic bibliographic citation information
  - Basic holdings and format information
- **METS**
  - Upper level descriptive information for managing digital objects
  - Provides specified structures for domain specific metadata
- **OAIS**
  - Reference model for the archival lifecycle
- **PREMIS**
  - Supports and documents the digital preservation process

Reference: Schloss Dagstuhl, 2014

Other Standards: Non-Archival

- **ISO 19115 – Geography**
  - Metadata structure for describing geographic feature files such as shape, boundary, or map image files and their associated attributes
- **ISO/IEC 11179**
  - International standard for representing metadata in a Metadata Registry
  - Consists of a hierarchy of “concepts” with associated properties for each concept
- **ISO 17369 SDMX**
  - Exchange of statistical information (time series/indicators)
  - Supports metadata capture as well as implementation of registries

Reference: Schloss Dagstuhl, 2014

More about DDI

- **Creates a standard format**
  - Used to markup codebooks
  - 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*

Definition

- **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.”*

Source: https://www.icpsr.umich.edu/icpsrweb/ICPSR/support/glossary

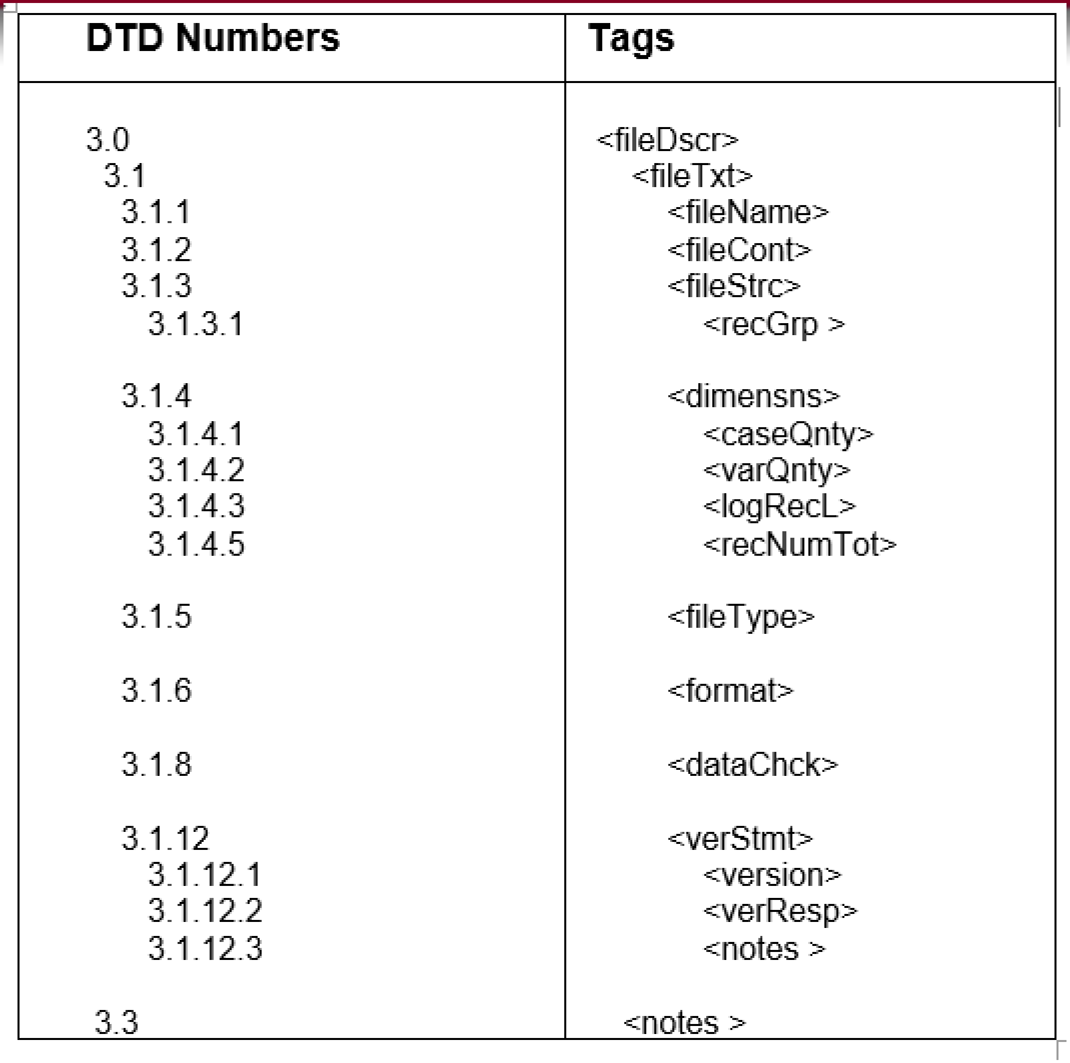
ntro> </chapter> </book>

Definition

- **Tags: "***Fragments of text used to organize content, usually delimited in a set format."*

Source: https://www.icpsr.umich.edu/icpsrweb/ICPSR/support/glossary

Example of Tag Names



DTD Numbers	Tags
3.0	<fileDscr>
3.1	<fileTxt>
3.1.1	<fileName>
3.1.2	<fileCont>
3.1.3	<fileStrc>
3.1.3.1	<recGrp >
3.1.4	<dimensns>
3.1.4.1	<caseQnty>
3.1.4.2	<varQnty>
3.1.4.3	<logRecL>
3.1.4.5	<recNumTot>
3.1.5	<fileType>
3.1.6	<format>
3.1.8	<dataChck>
3.1.12	<verStmt>
3.1.12.1	<version>
3.1.12.2	<verResp>
3.1.12.3	<notes >
3.3	<notes >

Source: http://bit.ly/2bo01MR

Examples of DDI Tags

- **<titl>**Canadian Community Health Survey, 2012: Annual Component **</titl>**

  <labl>Questionnaire (.pdf)**</labl>**

- **<dataDscr><notes>**The variables in this study are identical to earlier waves. **</notes></dataDscr>**

- **<titl>**Canadian Gallup Poll, May 2000**</titl>**

  <dataChck>Quality checks were performed by Carleton University Data Centre. **</dataChck>**

- **<titl>**Survey of Household Spending, 2001 [Canada]**</titl>**

  <varQnty>255**</varQnty>**

- **<titl>**Canadian Gallup Poll, May 1949, #186**</titl>**

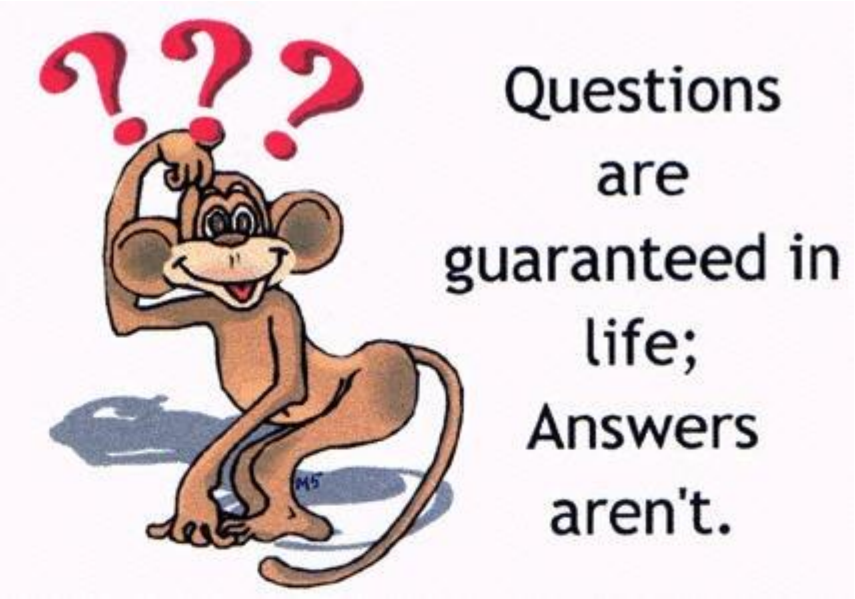
  <copyright>Copyright Gallup Canada Inc., 1950**</copyright>**

Benefits of DDI

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

Challenges of DDI

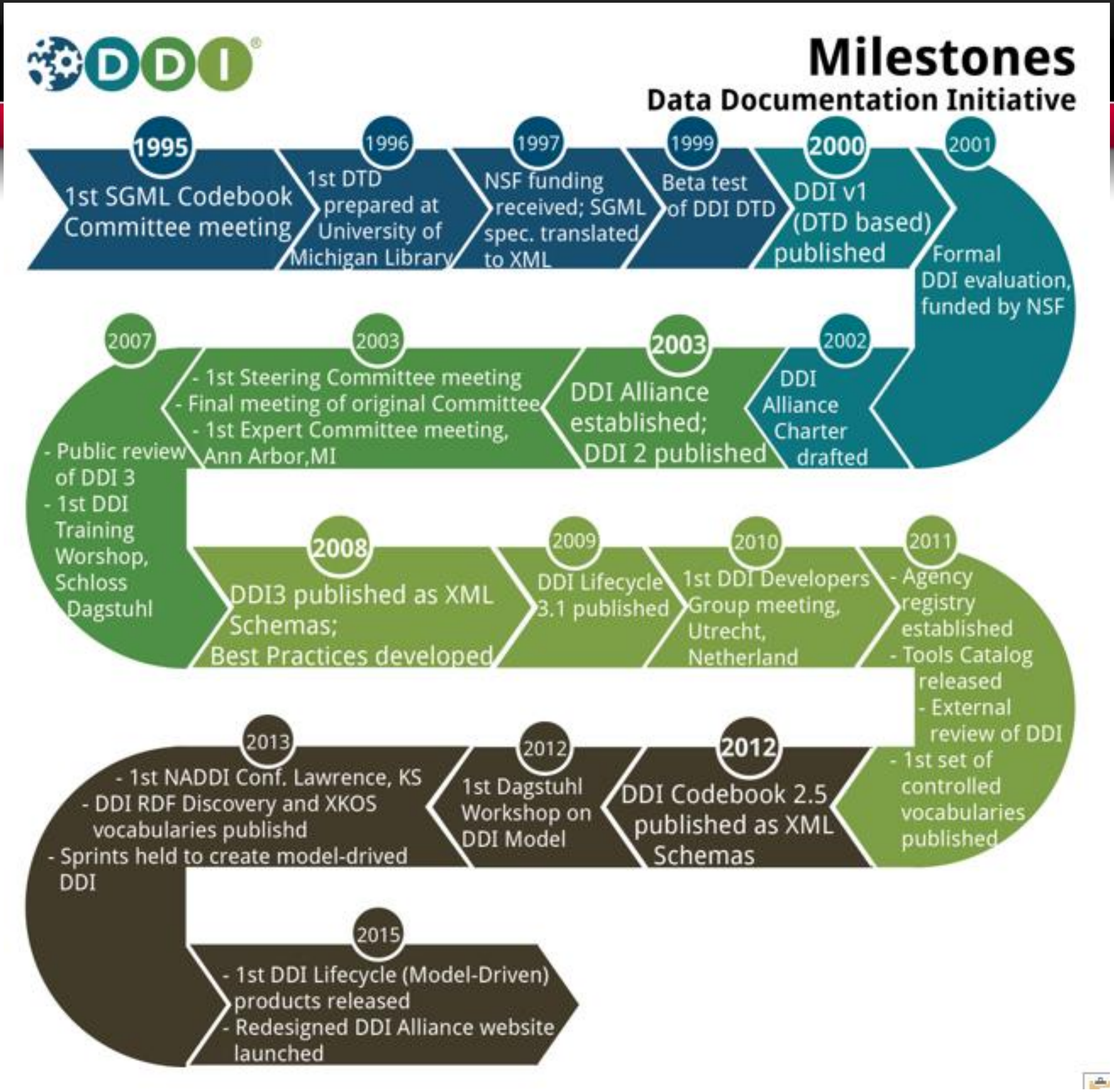
- Complexity
- Level of researcher buy-in
- Need for tools



http://taitegallery.net/wp-content/uploads/2012/02/unanswered-questions.jpg

Brief History of DDI

- **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**
  - Started in 2013



Milestones

Data Documentation Initiative

1995 1st SGML Codebook Committee meeting

1996 1st DTD prepared at University of Michigan Library

1997 NSF funding received; SGML spec. translated to XML

1999 Beta test of DDI DTD

2000 DDI v1 (DTD based) published

2001 Formal DDI evaluation, funded by NSF

2002 DDI Alliance Charter drafted

2003 DDI Alliance established; DDI 2 published

2003
- 1st Steering Committee meeting
- Final meeting of original Committee
- 1st Expert Committee meeting, Ann Arbor,MI

2007
- Public review of DDI 3
- 1st DDI Training Worshop, Schloss Dagstuhl

2008 DDI3 published as XML Schemas; Best Practices developed

2009 DDI Lifecycle 3.1 published

2010 1st DDI Developers Group meeting, Utrecht, Netherland

2011
- Agency registry established
- Tools Catalog released
- External review of DDI
- 1st set of controlled vocabularies published

2012 DDI Codebook 2.5 published as XML Schemas

2012 1st Dagstuhl Workshop on DDI Model

2013
- 1st NADDI Conf. Lawrence, KS
- DDI RDF Discovery and XKOS vocabularies publishd
- Sprints held to create model-drived DDI

2015
- 1st DDI Lifecycle (Model-Driven) products released
- Redesigned DDI Alliance website launched

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

DDI Development

- **DDI now branched into 2 separate development lines or metadata standards**
- **DDI Codebook (2003)**
  - aka DDI C
  - Formerly DDI 2
  - Built to emulate a physical codebook
  - Latest version is 2.5
  - Sections
    - *Document Description • Study Description • Data Files Description • Variable Description • Other Study Related Materials*

Source: http://www.ddialliance.org/explore-documentation

DDI Development (cont'd)

- **DDI Lifecycle (2008)**
  - aka DDI L
  - Formerly DDI 3
  - Supports the research data lifecycle
  - The one most new users are learning
  - Latest version is 3.2
  - Sections
    - *Study Concept • Data Collection • Data Processing • Data Distribution • Data Archiving • Data Discovery • Data Analysis • Repurposing*

Source: http://www.ddialliance.org/explore-documentation

Comparison

DDI 1 and 2 (DDI C)

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

DDI 3 (DDI L)

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

Reference: Jim Jacobs, 2006

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?

Final Challenge

- **How to go from DDI C to DDI L?**
  - What are the challenges?
    - *Very, very, very much work to convert to DDI L*
    - *Insufficient resources (people, $)*
    - *Nesstar uses DDI C*
  - Solution
    - *Crosswalks, other resources*

DDI 4

- **In production mode**
  - To meet the needs of the user community
  - Has an integrative vision
    - *DDI C (DDI 2) and DDI L (DDI 3)*
  - Aimed at machine-actionable processing at the beginning, that is, will be auto-generated
  - Still undergoing development
  - Check out the FAQ
    - http://www.ddialliance.org/faq-about-ddi-4

DDI Alliance

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

More information

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

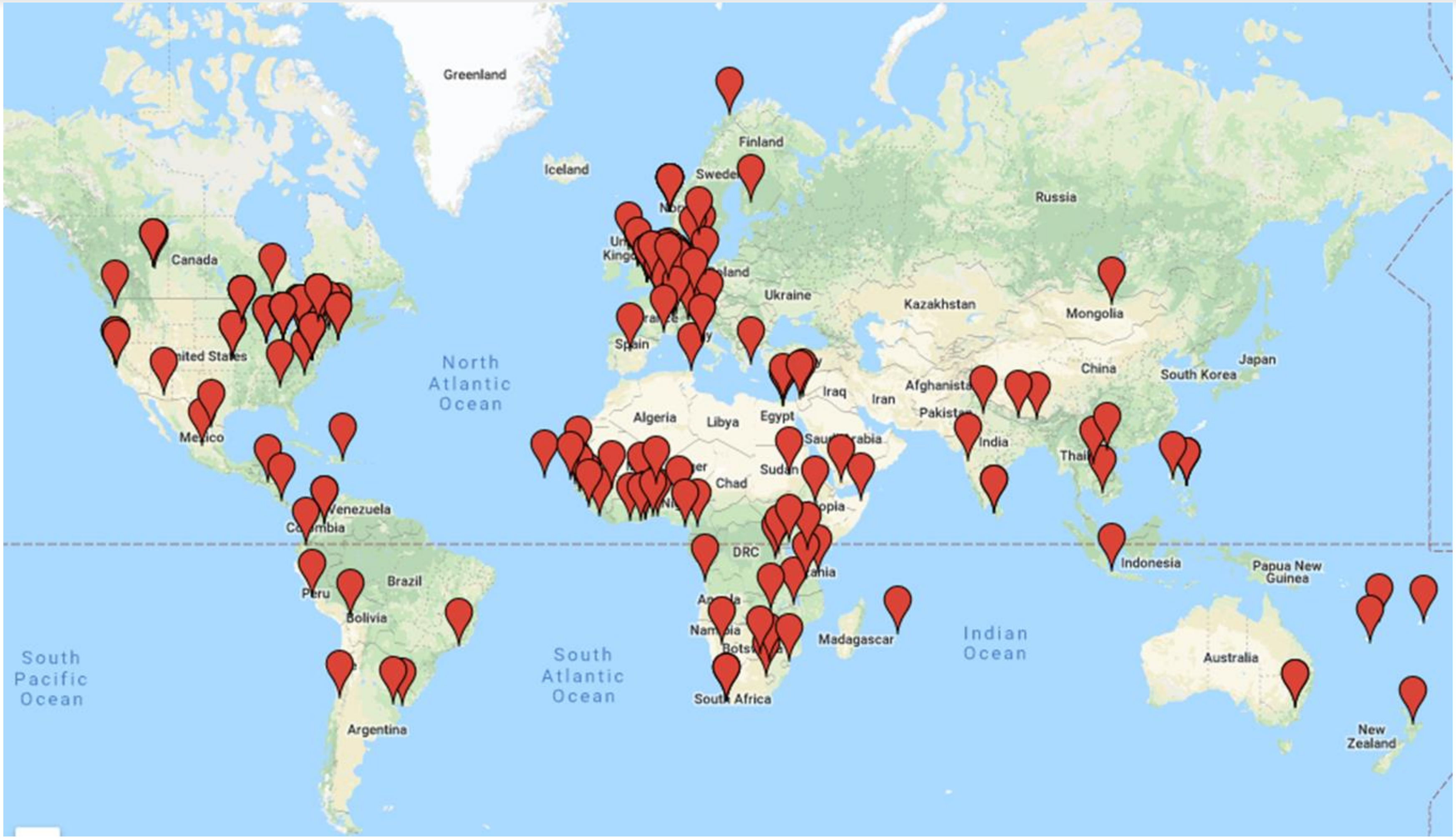
Who uses DDI

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

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

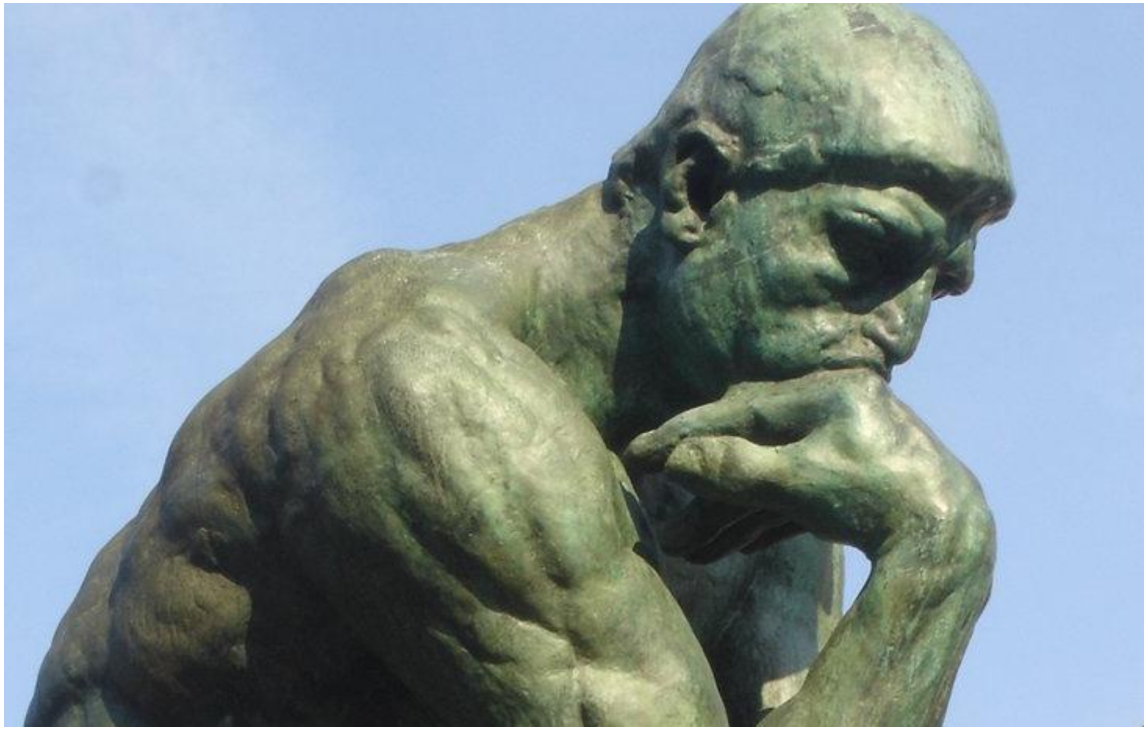
What projects use DDI?

- 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
- ICPSR Data Catalog
- Vision of Britain (historical view between 1801 and 2001)
- World Bank (International Household Survey Network)
- ODESI (Ontario Data Portal)
- …



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

Time to think!



Source: http://deacondance.com/wp-content/uploads/2012/05/Thinking-Man-Rodin.jpg

Exercise 1

- As a researcher, what metadata do you *absolutely* need?

- How do you want it to *streamline* your research?

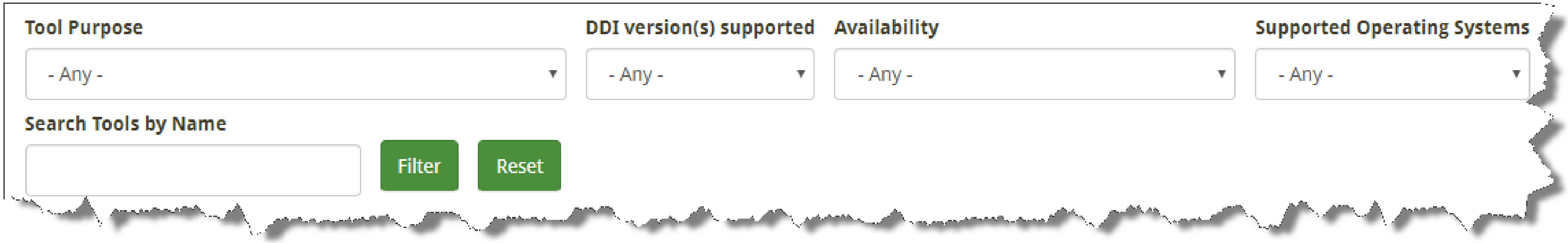
- What metadata would you like to have, if it is available, but it is *not integral* to your research?

10 minutes!

Getting Started with DDI

- **Daunting at first**
  - Process is broken down into steps
- **Lots of help available**
  - DDI Alliance
  - http://www.ddialliance.org/training/getting-started
  - Colleagues
  - Other researchers
- **DDI List-serv**
- **DDI Best Practices**
  - Work in progress
    - *Feedback always welcome*

- **Tools to help you get started**
  - *http://www.ddialliance.org/resources/tools*
    - *Drop down menu*
    - *Browse a list*
  - Choose the one that will suit your purposes
  - For different versions of DDI



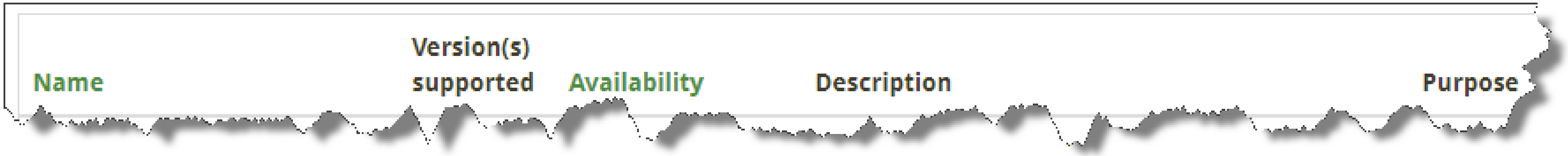
Tool Purpose: - Any -

DDI version(s) supported: - Any -

Availability: - Any -

Supported Operating Systems: - Any -

Search Tools by Name:



Name	Version(s) supported	Availability	Description	Purpose

Getting Started with DDI (cont'd)

- **One tool: Nesstar Publisher**
  - Norwegian Social Science Data Services
  - Data management program
  - Freeware
  - Data and metadata conversion and editing tools
    - *Enhance datasets*
      - Combine catalogue and contextual information
  - Merge DDI documents with markup for different sections of the DDI for the same study
      - Merge variable descriptions from SPSS/SAS with DDI

Getting Started with DDI (cont'd)

- **Nesstar Webview**
  - Metadata
  - Any associated documentation
  - Variable groups
  - Conduct basic analysis
    - *Subsetting*
    - *Crosstabs*
  - Bonus

Getting Started with DDI (cont'd)

- **Nesstar Webview**
  - Downloading
    - *Documentation*
      - PDF format
      - Export files with study descriptions and question text
    - *Data exported in format of choice*
      - SPSS, SAS, Stata, ASCII, …

Getting Started with DDI (cont'd)

- **Nesstar drawbacks**
  - For advanced statistical analysis -
    - *it is best to download the data and use a statistical analysis package*
  - Must have access to a server to publish the dataset
  - Not intuitive when starting to markup datasets
  - Not intuitive for first-time user in Webview
  - Downloading into SAS not user friendly

- **Not a drawback, just a consideration**
  - Uses DDI Codebook standard

- **Another tool: Colectica for Excel**
  - Tools to help with your metadata
  - https://www.colectica.com/software/
  - http://www.ddialliance.org/node/893
  - Documents variables and datasets directly from within Excel
  - Can be used to produce detailed (item-level) metadata for studies already completed
  - Creates metadata and documentation for surveys
  - DDI version 3.1, 3.2
  - Saves metadata directly in the Excel file
    - *When the file is shared, so is the metadata*

Getting Started with DDI (cont’d)

- **Colectica Reader**
  - Free tool
    - *To view the metadata*
    - *No specialized software is needed*
  - Generates documentation for variables and code lists
    - *PDF, Word, HTML*

Examples of using DDI

- **Check out how Nesstar Webview works**
  - Using the ODESI data repository
  - http://www.library.carleton.ca/help/odesi-how-to-use-odesi
    - *Navigating the ODESI repository*
    - *Searching for variables*
    - *Finding, subsetting and downloading*
    - *Creating a cross tabulation*
    - *Downloading a full dataset*

Examples of using DDI (cont'd)

- **Check out how Colectica works**
  - A number of videos to watch
  - http://www.youtube.com/user/Colectica/videos
  - Colectica Questionnaires
    - *Demo*
    - *Create a Survey and Add*
    - *View a Survey's Structure*
    - *Add metadata to a Survey*
    - …

Back in 15!

Source: http://logopond.com/logos/9a1bf9d159d18c327dc2f3b39ba12bad.png

Review

- **We have learned**
  - the history of DDI
  - who is using it
  - how it can streamline our research
- **We have seen**
  - 2 tools and examples
    - *Nesstar and Colectica for Excel*
- **Remember**
  - There are 2 different lines of DDI development
    - *DDI C and DDI L*

Data Lifecycle Workflow

- **Integrating Metadata and Research Data Management into a data lifecycle workflow**
  - What metadata
  - When or where to integrate it
  - What does research data management have to do with this

Data Lifecycle Workflow (cont'd)

- **DDI Metadata accompanies and enables**
  - Data conceptualization
  - Collection
  - Processing
  - Distribution
  - Discovery
  - Analysis
  - Repurposing
  - Archiving

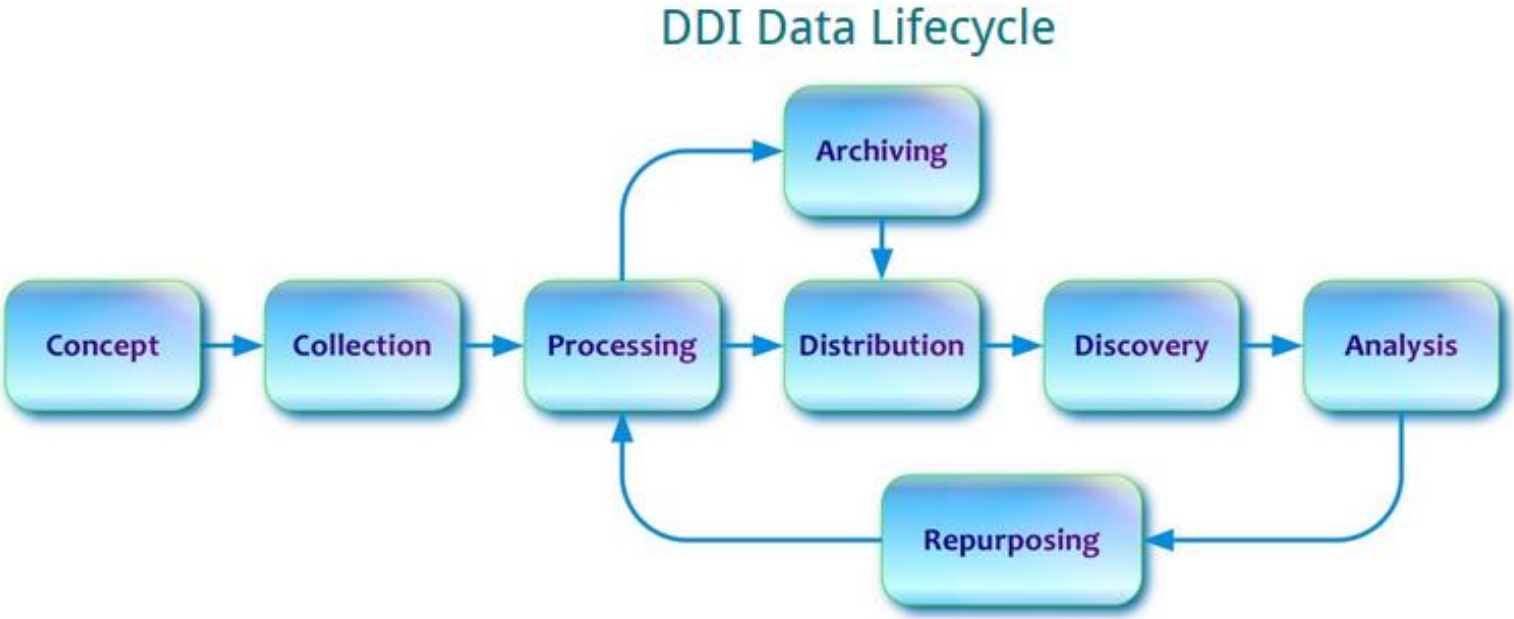
Exercise 2

- *When* or *where* would you like to see metadata integrated into a data lifecycle workflow?

- What does research data management have to do with this?

10 minutes!

DDI Research Lifecycle



Source: http://www.ddialliance.org/training/why-use-ddi

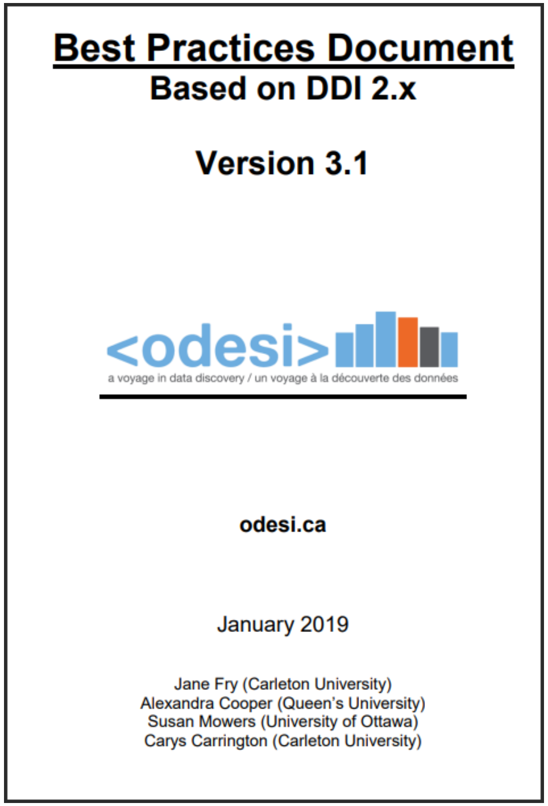
Data Lifecycle Workflow (cont'd)

- **Remember**
  - More efficient and easier to capture information about all stages of the research workflow ***at the time*** of its occurrence rather than after the fact

Best Practices

- **Do they matter?**
  - When
    - *During data collection, coding, ingest, …*
    - *At all times*
  - To whom do they matter
    - *You*
    - *Your research team*
    - *Future researchers*

Example



Best Practices Document
Based on DDI 2.x

Version 3.1

<odesi>
a voyage in data discovery / un voyage à la découverte des données

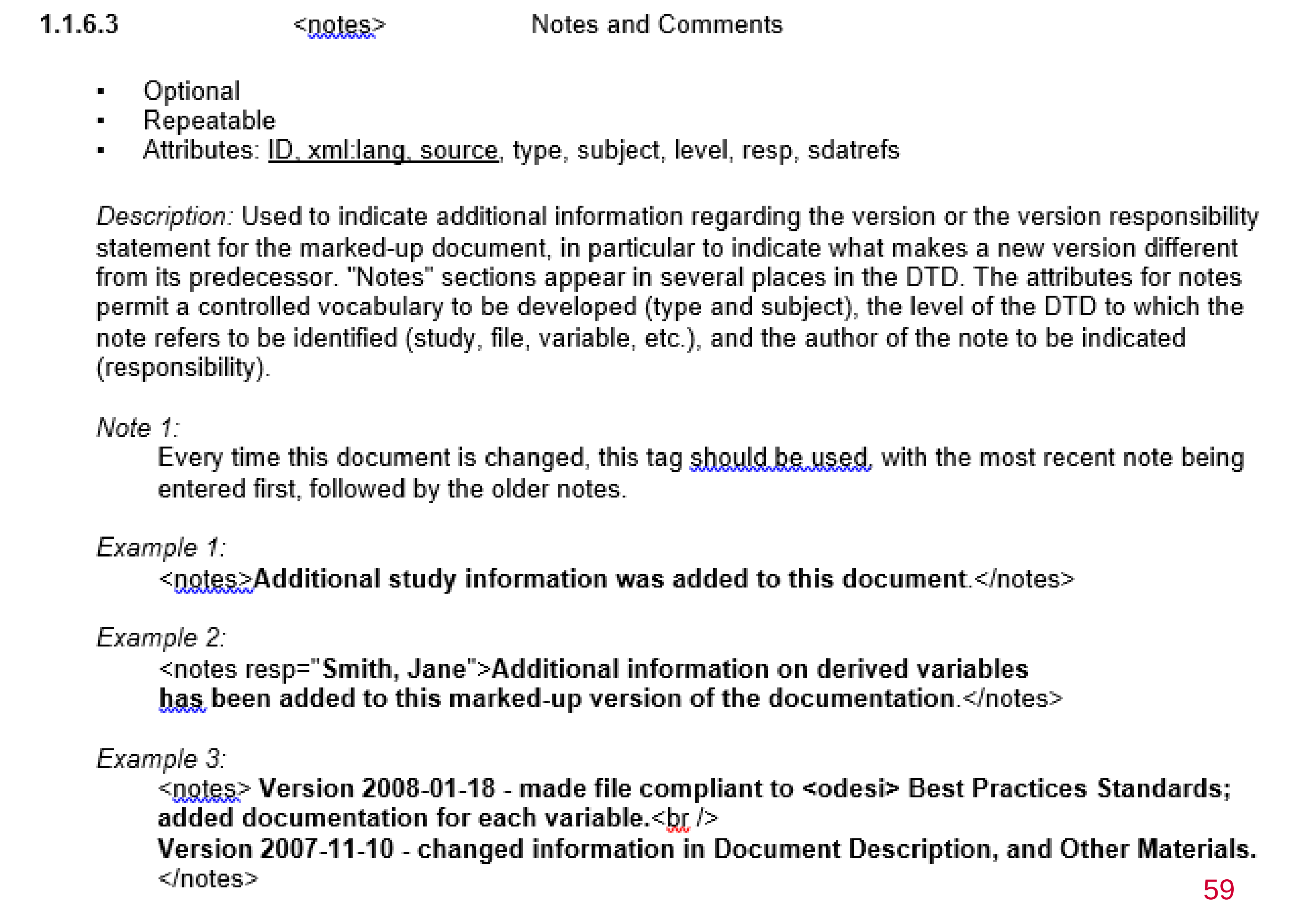
odesi.ca

January 2019

Jane Fry (Carleton University)
Alexandra Cooper (Queen's University)
Susan Mowers (University of Ottawa)
Carys Carrington (Carleton University)

Reference: https://library.carleton.ca/sites/default/files/help/data-centre/BPDv3-1-2019-01-28.pdf

ODESI BPD Example



1.1.6.3 <notes> Notes and Comments

- Optional
- Repeatable
- Attributes: ID, xml:lang, source, type, subject, level, resp, sdatrefs

Description: Used to indicate additional information regarding the version or the version responsibility statement for the marked-up document, in particular to indicate what makes a new version different from its predecessor. "Notes" sections appear in several places in the DTD. The attributes for notes permit a controlled vocabulary to be developed (type and subject), the level of the DTD to which the note refers to be identified (study, file, variable, etc.), and the author of the note to be indicated (responsibility).

Note 1:
Every time this document is changed, this tag should be used, with the most recent note being entered first, followed by the older notes.

Example 1:
<notes>**Additional study information was added to this document.**</notes>

Example 2:
<notes resp="**Smith, Jane**">**Additional information on derived variables has been added to this marked-up version of the documentation.**</notes>

Example 3:
<notes> **Version 2008-01-18 - made file compliant to <odesi> Best Practices Standards; added documentation for each variable.**<br />
Version 2007-11-10 - changed information in Document Description, and Other Materials.
</notes>

Examples of DDI Tags

- **<titl>**Canadian Community Health Survey, 2012: Annual Component **</titl>**

  <labl>Questionnaire (.pdf)**</labl>**

- **<dataDscr><notes>**The variables in this study are identical to earlier waves. **</notes></dataDscr>**

- **<titl>**Canadian Gallup Poll, May 2000**</titl>**

  <dataChck>Quality checks were performed by Carleton University Data Centre. **</dataChck>**

- **<titl>**Survey of Household Spending, 2001 [Canada]**</titl>**

  <varQnty>255**</varQnty>**

- **<titl>**Canadian Gallup Poll, May 1949, #186**</titl>**

  <copyright>Copyright Gallup Canada Inc., 1950**</copyright>**

Resources

- **Fry, J., Cooper, A., Mowers, S. and Carrington, C. "Best Practices Document: Based on DDI 2.x, Version 3.1. January 2019.**

  https://bit.ly/2ZctTBb
- **ICPSR Glossary of Social Science Terms**

  https://www.icpsr.umich.edu/icpsrweb/ICPSR/support/glossary
- **XML Schema Tag Library – Version 2.1**

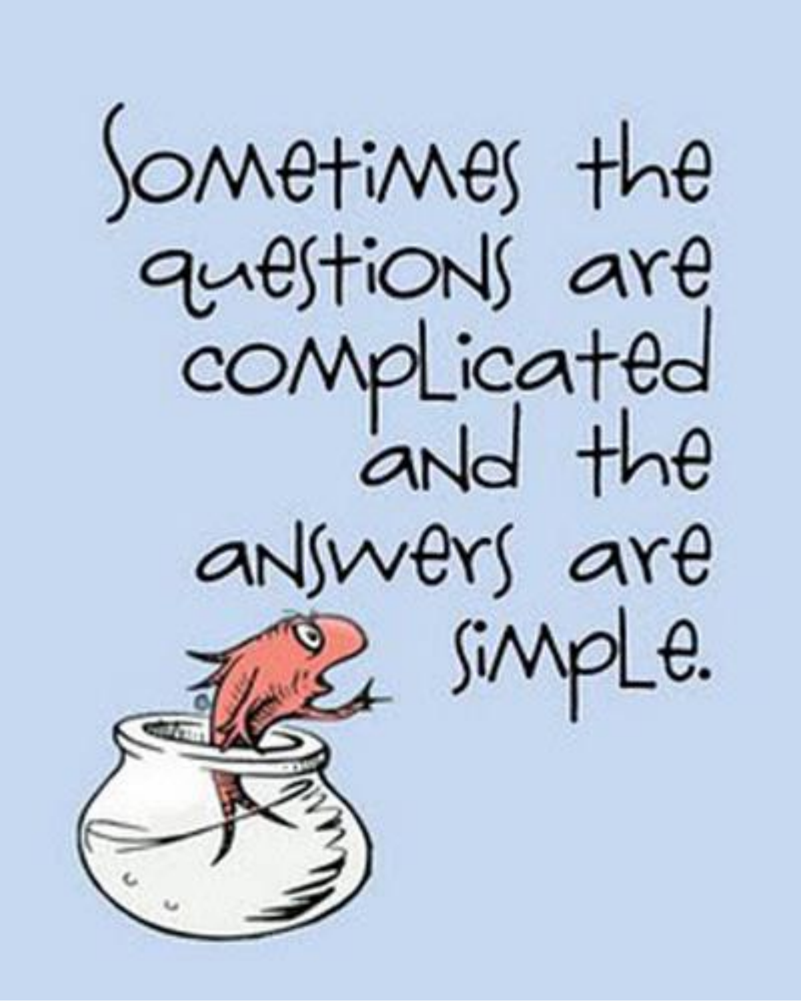
  http://bit.ly/2bo01MR

Workshop Materials

- **To access materials from this workshop**
  - They will be on the NADDI Conference website

References

- **DDI Alliance. <*ddi> Data Documentation Initiative*.**
  http://www.ddi-alliance.org/
- **Schloss Dagstuhl, October 2014. "DDI Basics".**
  https://bit.ly/2ZkdoTu
- **Iverson, J. & Stephenson, E. (2013). *DDI-Lifecycle and Colectica at the UCLA Social Science Data Archive*. Presentation at the North American Data Documentation Conference (NADDI) 2013.**
  http://hdl.handle.net/1808/11049
- **Jacobs. Jim (2006). "Evolution of Data Documentation". Workshop "A Gentle Introduction to DDI: What's in it for Me?" presented at IASSIST 2006.**
- **Vardigan, M. & Wackerow, J. (2013). DDI – A metadata standard for the community. Paper presented at the North American Data Documentation Initiative Conference (NADDI) 2013.**
  https://kuscholarworks.ku.edu/handle/1808/11056



http://www.quotationof.com/images/question-quotes-7.jpg

Contact Information

Jane Fry
Data Services Librarian
Rm 122, MacOdrum Library
Carleton University, Ottawa
613.520.2600 x1121
jane.fry@Carleton.ca