What can DDI do for you? An introduction to the DDI

Monday November 30, 2021

Adrian Dusa - University of Bucharest
Jane Fry – Carleton University

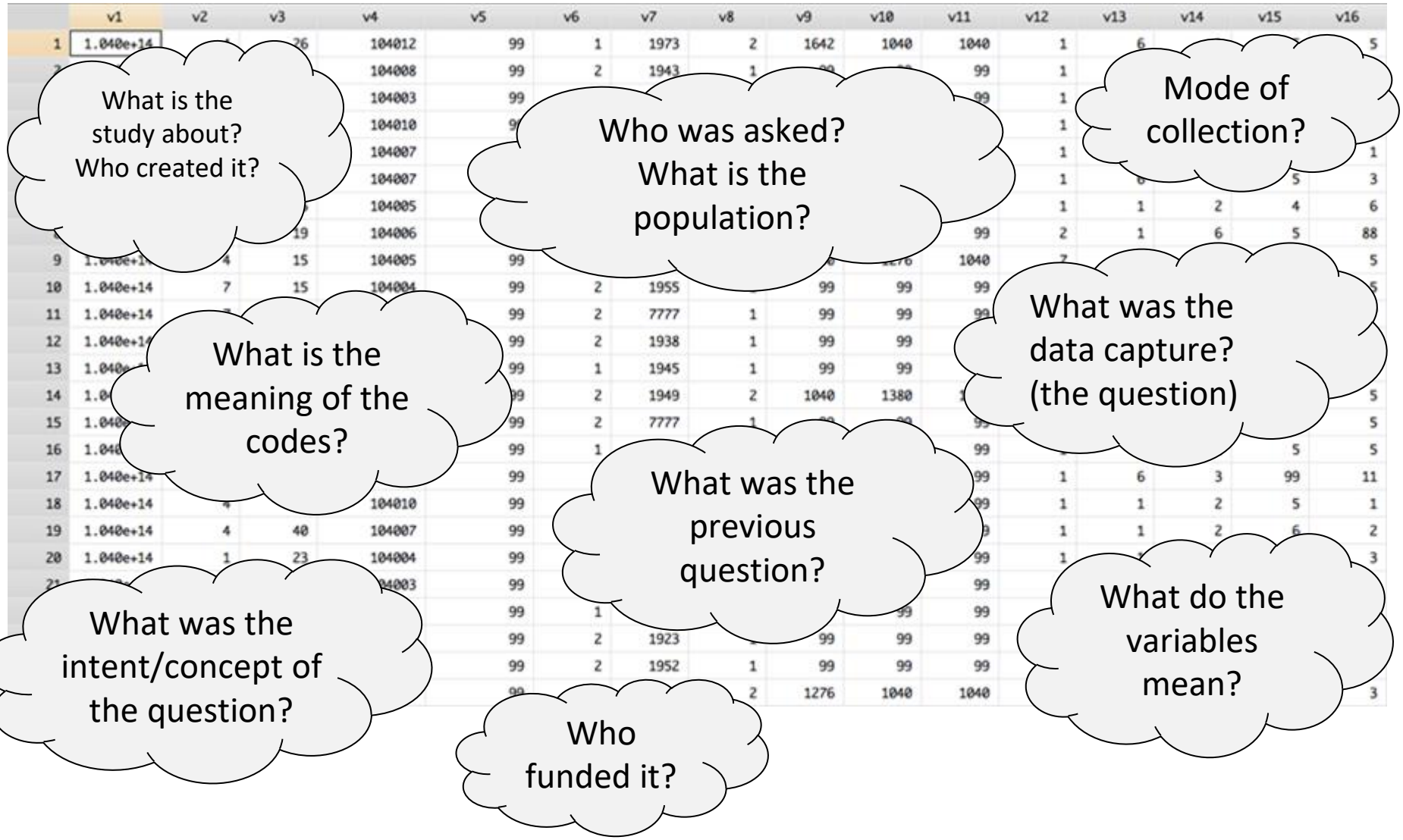
Outline

- The importance of Metadata
- What is DDI?
- What is the DDI Codebook?
- DDI and FAIR
- Live Demonstration

Metadata


Unico
COLA
COLA
355 mL
CHUNKY
CLASSICO
pizza
sauce
à pizza
Lindt
HAZELNUT · NOISETTE
Flocons
de jambon
MAPLE LEAF
156 g

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16
1	1.040e+14	4	26	104012	99	1	1973	2	1642	1040	1040	1	6	1	5	5
2	1.040e+14	1	19	104008	99	2	1943	1	99	99	99	1	1	6	4	5
3	1.040e+14	4	98	104003	99	1	1990	1	99	99	99	1	6	3	99	11
4	1.040e+14	2	98	104010	99	1	1983	2	1756	1040	1040	1	6	3	99	11
5	1.040e+14	1	18	104007	99	2	1927	1	99	99	99	1	1	6	4	1
6	1.040e+14	2	19	104007	99	1	1983	1	99	99	99	1	6	2	5	3
7	1.040e+14	4	15	104005	99	2	1970	1	99	99	99	1	1	2	4	6
8	1.040e+14	4	19	104006	99	1	1942	1	99	99	99	2	1	6	5	88
9	1.040e+14	4	15	104005	99	1	1965	2	1040	1276	1040	7	4	1	5	5
10	1.040e+14	7	15	104004	99	2	1955	1	99	99	99	1	1	2	5	5
11	1.040e+14	7	15	104003	99	2	7777	1	99	99	99	1	1	6	77	77
12	1.040e+14	4	18	104005	99	2	1938	1	99	99	99	1	3	6	5	5
13	1.040e+14	7	17	104005	99	1	1945	1	99	99	99	1	1	6	5	4
14	1.040e+14	4	18	104005	99	2	1949	2	1040	1380	1040	1	4	6	5	5
15	1.040e+14	2	15	104003	99	2	7777	1	99	99	99	1	1	2	4	5
16	1.040e+14	4	32	104007	99	1	1974	1	99	99	99	1	6	2	5	5
17	1.040e+14	4	98	104006	99	1	7777	1	99	99	99	1	6	3	99	11
18	1.040e+14	4	25	104010	99	2	1968	1	99	99	99	1	1	2	5	1
19	1.040e+14	4	40	104007	99	1	1967	1	99	99	99	1	1	2	6	2
20	1.040e+14	1	23	104004	99	1	1932	1	99	99	99	1	1	6	4	3
21	1.040e+14	2	18	104003	99	2	1965	1	99	99	99	1	6	2	5	5
22	1.040e+14	1	27	104011	99	1	1956	1	99	99	99	2	1	1	5	2
23	1.040e+14	4	18	104004	99	2	1923	1	99	99	99	1	3	6	3	3
24	1.040e+14	1	19	104006	99	2	1952	1	99	99	99	1	4	2	5	3
25	1.040e+14	3	16	104004	99	2	1947	2	1276	1040	1040	1	1	6	3	3



What is DDI?

- DDI = Data Documentation Initiative
- An international metadata standard
  - Used primarily in the social and behavioural sciences, economics, health
  - An open standard designed for data sharing and reuse
- A structure for describing data and its related information
- Describes data from surveys and other observation-based data collection methods
  - currently moving towards covering new data types and data from new domains.

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.*

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

DDI as a standard

- The DDI standard structure means
  - all computers, even if they are using different applications, can work on the same data and related information (metadata/documentation)
  - the metadata is put into a structured format

- Uses generic XML technology as the basis for cross-platform use

- The format is not proprietary to any specific system

DDI Products

- The DDI standard has developed over time
  - https://ddialliance.org/products/overview-of-current-products
- Currently has two main products:
  - DDI-Codebook
  - DDI-Lifecycle
- Each designed for a different purpose
- Other products:
  - Controlled vocabularies
  - RDF (XKOS)
  - Standard Data Transformation Language (SDTL)

Developing Products

- The DDI standard continues to develop based on what users want
  - https://ddialliance.org/products/developing-products-of-the-alliance
- Developing products:
  - DDI-Cross Domain Integration (DDI-CDI)
  - RDF (Disco)

Source: Lyle, Jared. (2021 November). Introduction to the DDI Metadata Standard. Presented at EDDI Training Fair, 2021.

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

DDI users

Agencies

- Norwegian Social Science Data Services
- Harvard University
- DLI (Statistics Canada)
- Health Canada
- Bureau of the Census
- ICPSR
- Bureau of Labor Statistics
- ESRC Data Archive (UK)
- Zentralarchiv für Empirische Sozialforschung (GESIS)
- …

Projects

- CESSDA Data Portal
- Australian Social Science Data Archive
- DAMES Project (UK)
- DataFirst (at University of Cape Town)
- Israel Social Science Data Center
- ICPSR Data Catalog
- ODESI (Canada)
- Statistics New Zealand
- …

DDI Users in 2020

Source: https://ddialliance.org/community/join

DDI Audiences

- Librarians
- Managers
- Repositories
- Researchers
- Developers
- For more information, check out the DDI Alliance website

DDI Alliance …

- A self-sustaining member organization
  - Created in 2003
- Members have a voice in DDI development
- Executive Director
  - Jared Lyle
- All information is on-line
  - Membership, Charter, by-laws, forms, …
  - Publications, conferences, working groups, …
- http://www.ddialliance.org/about/about-the-alliance

DDI Alliance

- Organization consists of
  - **Executive Board -** The policymaking and oversight body of the Alliance.
  - **Scientific Board -** Responsible for the work to develop the standard.
  - **Technical Committee -** To maintain the various DDI products, in collaboration with the different working groups of the DDI Alliance.
  - **Working Groups -** Convened to work on different activities and topics within the work areas of the DDI Alliance.

Benefits of Membership

- Vote on Alliance products, including additions and modifications to specifications.
- Be elected to Alliance leadership, including the Executive Board and the Scientific Board.
- Have a seat at annual meetings, including the Annual Meeting of Members and the Annual Meeting of the Scientific Community.
- Participate on Working Groups which develop DDI products and advise DDI leadership.
- Display the Alliance trademarks on promotional material and publicize participation in the Alliance.
- Send participants to Alliance-sponsored events at reduced or no cost.
- Request access to Members-only information.

Source: Lyle, Jared. (2021 November). Introduction to the DDI Metadata Standard. Presented at EDDI Training Fair, 2021.

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

Interpretation of DDI

- Written in XML format
- Need a tool to interpret it
- Popular ones are
  - Nesstar
  - Colectica
  - Dataverse

More information: https://ddialliance.org/resources/tools

…

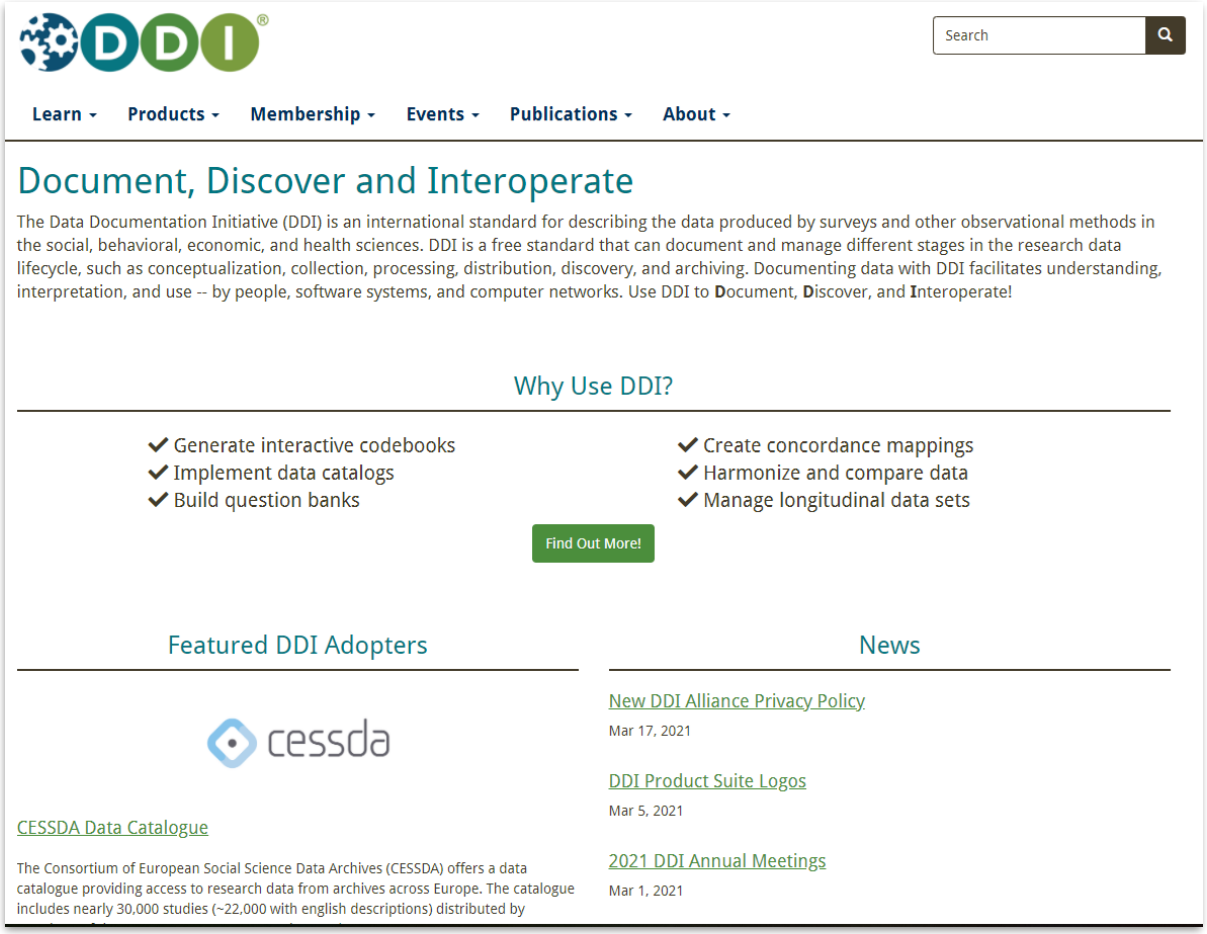
- Expressed in XML
  - The XML schema is a way of tagging text for meaning, not appearance
  - Defines
    - Which tags are available
    - The order the tags will appear in a document
    - Whether the tags are required or optional
    - Whether the tags are repeatable or not

Caveat!

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

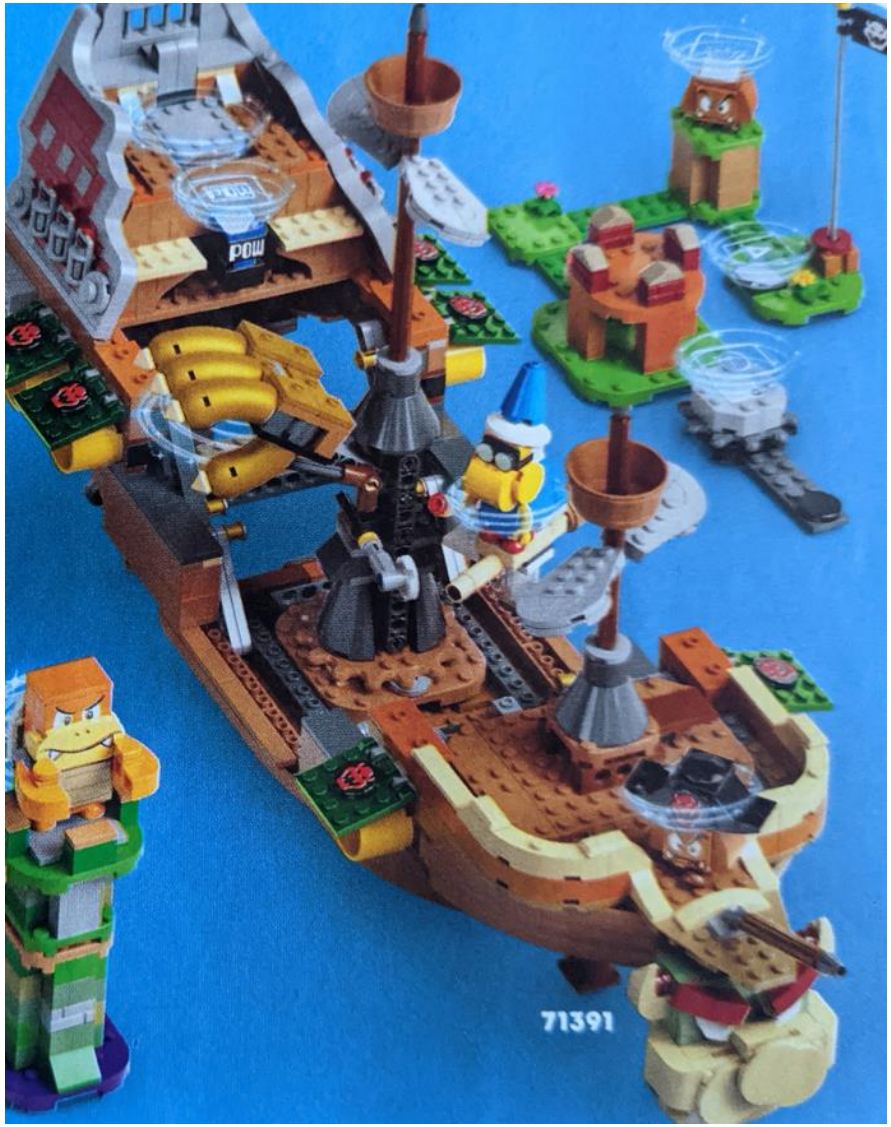
More information

- DDI website
  - http://www.ddialliance.org/
  - Excellent resource
    - Learn (training materials)
    - Products
    - Events
    - Publications
    - Collaborate
    - About



https://ddialliance.org/

DDI Codebook


POW
71391

DDI Codebook …

- A structure facilitating the production of machine-readable codebooks and data dictionaries.

- Built to emulate a physical codebook,
  - that is, to catalog a dataset, to describe a single study or a single round or wave in a repeated study.

- The latest version of DDI-Codebook in version 2.5.

DDI Codebook …

- Relatively straight forward
- Sections
  - *Document Description*
  - *Study Description*
  - *Data Files Description*
  - *Variable Description*
  - *Other Study Related Materials*

DDI Codebook - Section 1

- Document Description
  - *Describes the actual document that you are putting together to describe the survey or study*
  - *For example*
    - *Title*
    - *Date*
    - *Author (of document)*
    - *…*

DDI Codebook - Section 2

- Study Description
  - *Describes the actual survey or study that was conducted*
    - *For example: purpose, how, why, where, how long, funding of the study, ...*
  - *Does not describe the data file*
  - *Many of the metadata tags in this section are the same as the ones in the first section - 'Document Description'*

DDI Codebook - Section 3

- Data Files Description
  - *Exactly what it says!*
  - *Describes the datafiles that were created by the study*
  - *For example*
    - *Number of variables*
    - *Number of cases*
    - *Structure of the datafile*
      - *Hierarchical, rectangular (the default) or relational*

DDI Codebook - Section 4

- Variable Description
  - *Any and all information there is about each and every variable in the datafile*
  - *For example*
    - *the question text (including any pre-questions)*
    - *the variable label*
    - *all the values and associated labels for that particular variable*
    - *the population*
    - *any additional notes*
      - *for example, instructions for the interviewer*

DDI Codebook - Section 5

- Other Study Related Materials
  - *These are documents that are separate from this particular file you are creating*
  - *A link is provided for separate download*
  - *For example*
    - *Questionnaires*
    - *User Guides*
    - *Codebooks*
    - *…*

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>

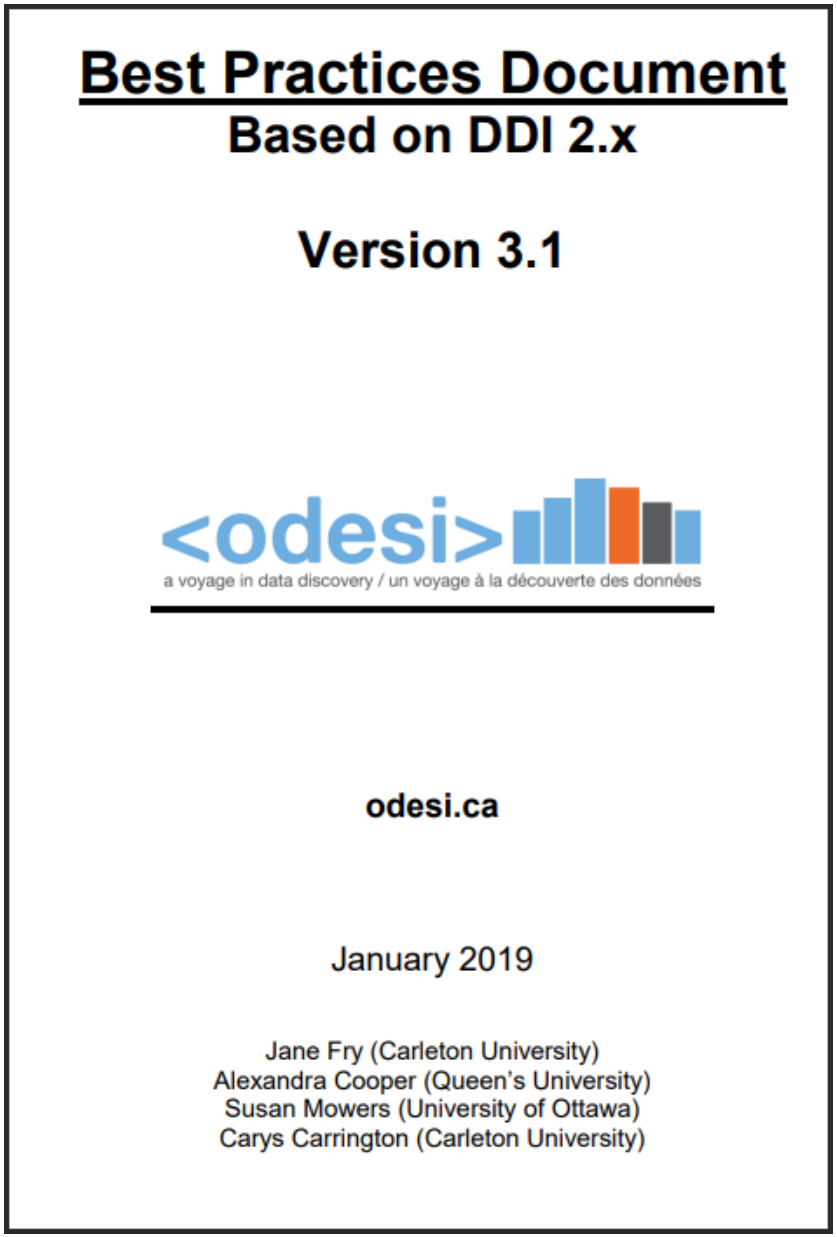
<titl>**Canadian Gallup Poll, May 2000**</titl>
<dataChck>**Quality checks were performed by Carleton University Data Centre.** </dataChck>

<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>

Best Practices Document

- Started with the DDI Alliance technical document
  - Made it human readable
- Updated last year (2019)
  - Used when marking up in ODESI
  - Uses the Nesstar platform
- Uses lots of examples
- Great for training purposes


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)

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.**
Version 2007-11-10 - changed information in Document Description, and Other Materials.
</notes>

To whom is DDI useful?

- Data archivists, librarians
- Data scientists / statisticians
- Researchers of any kind
- Teachers, lecturers
- Research institutions
- Even the general public

DDI and FAIR

What exactly is FAIR?

Findable

Accessible

Interoperable

Reusable

Findable

https://www.go-fair.org/fair-principles/

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

Accessible

https://www.go-fair.org/fair-principles/

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
  - A1.1 The protocol is open, free, and universally implementable
  - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

Interoperable

https://www.go-fair.org/fair-principles/

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

Reusable

https://www.go-fair.org/fair-principles/

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
  - R1.1. (Meta)data are released with a clear and accessible data usage license
  - R1.2. (Meta)data are associated with detailed provenance
  - R1.3. (Meta)data meet domain-relevant community standards

How does DDI help?

DDI Codebook is essentially an XML text file, that is:

- machine readable
- has dedicated sections for:
  - itself as a document
  - the study
  - the data file
  - the variables, labels and other information in the data
  - other relevant materials

- data archivists, librarians
- data scientists / statisticians
- researchers of any kind
- teachers, lecturers
- research institutions
- even the general public

DDI Codebook advantages

- Fairly easy to produce, even by non-technical people
- Already part of free and / or open source software
  (Nesstar, Colectica, Dataverse, more recently StatConverter might help)
- The biggest effort is automatic (data description)
- Only the easy parts should be filled in manually (study, universe, weighting procedure etc.)
- Perfect for an "Archive in a box"

DDI Codebook disadvantages

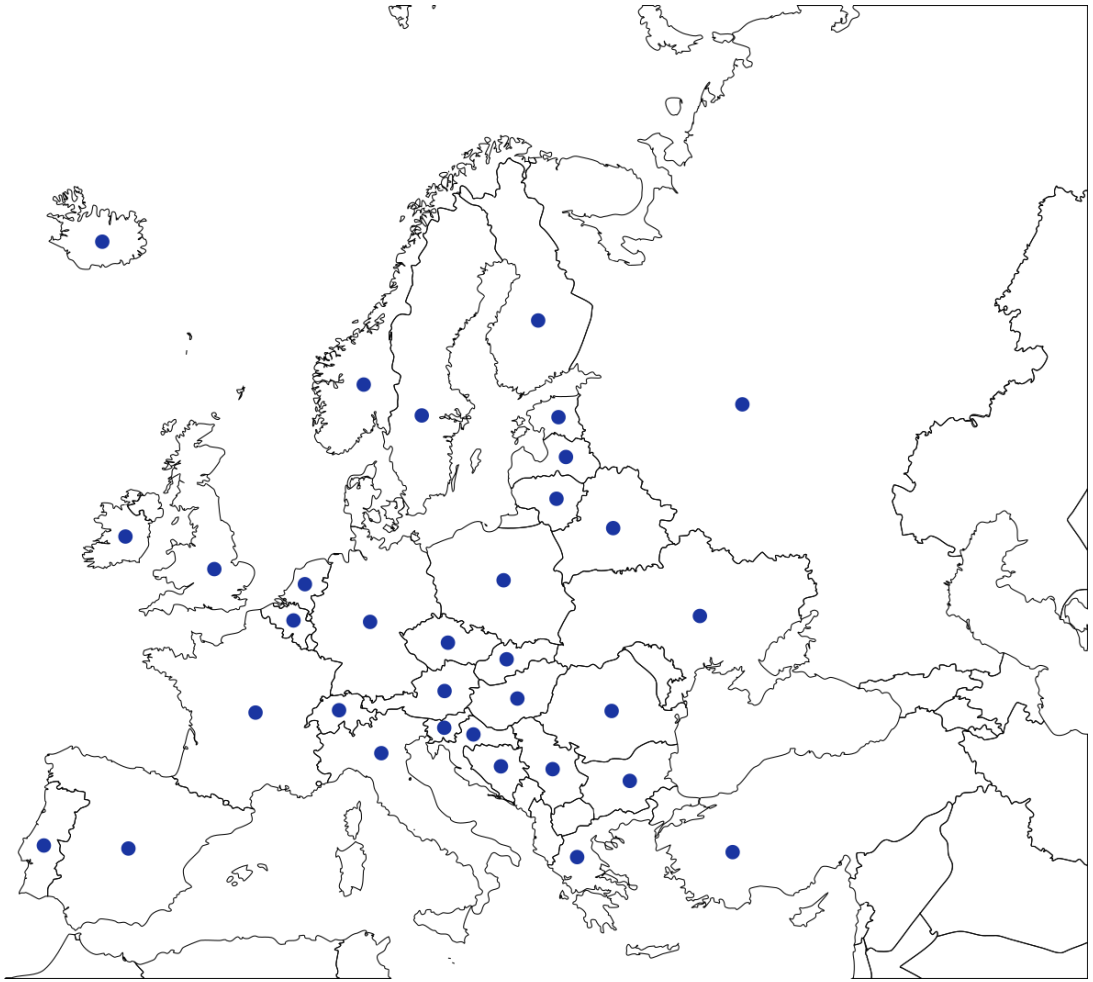
- It's a static, frozen document
- It can be automatically harvested, but it does not communicate with other Codebook files
- Individual datasets are often part of a study series (e.g. ESS rounds) and the Codebook version of DDI does not cover that
- Leads to a lot of duplicated information (repeated questions, or answer scales)
- Slightly more difficult to compare information between DDI Codebook files

Information reuse: DDI Lifecycle

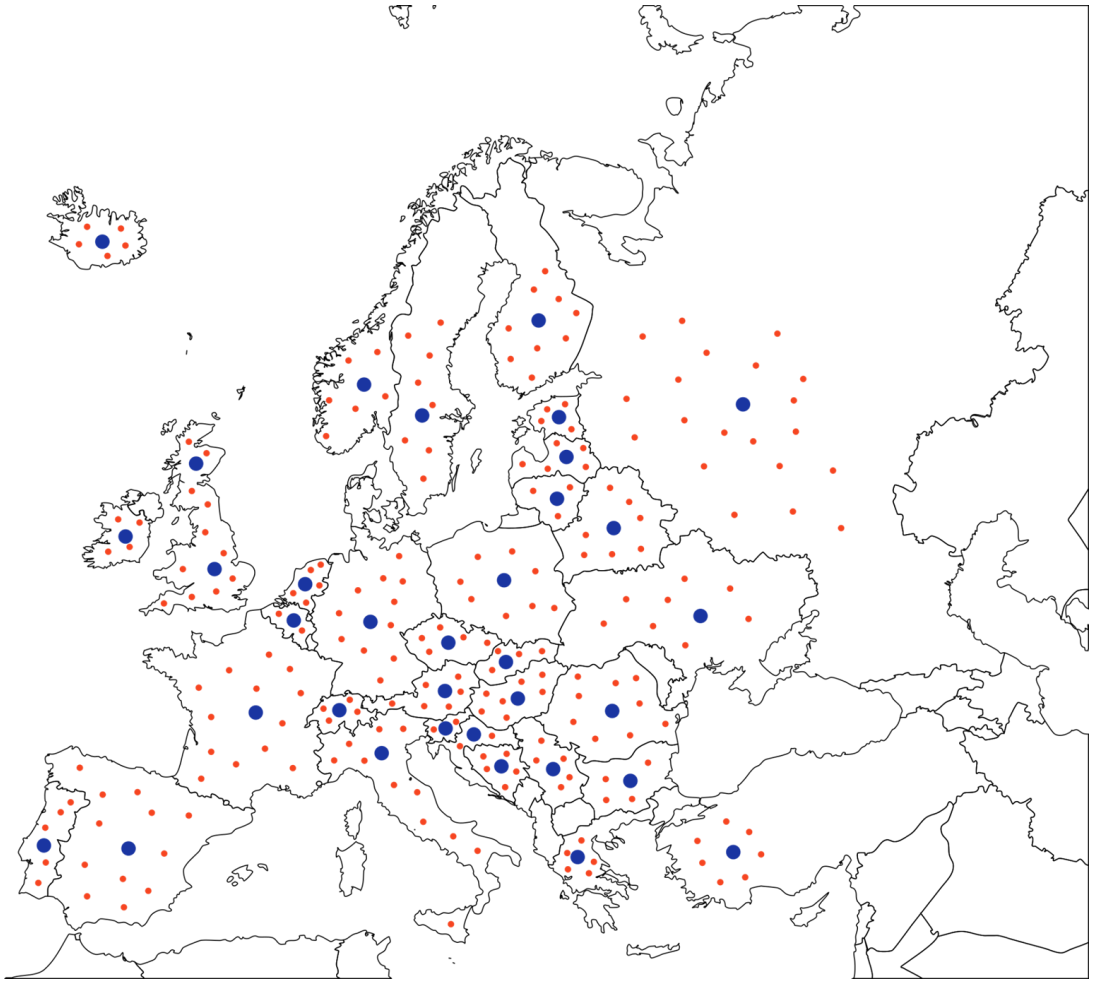
- End goal of research: compare and learn.
- Many examples of what can be compared, most often concepts.
- Concepts have operational models.
- Operational models have questions.
- Questions have answers of a specific type (e.g. a response scale)
- All of these can be reused, as well as any (re)combination of elements.

Increased search possibilities!

Formal network of data providers



Extended network of data providers



How do they all combine?

- Back to FAIR:
  - FAIR Implementation Profiles (FIPs)
  - https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/
  - FAIR Digital Objects (FDOs)
  - https://fairdo.org
  - FAIR Data Points (FDPs)
  - https://fairdatapoint.org

DDI in Europe

- Many, many institutions use DDI.
- I will mainly refer to the national data archives part of CESSDA - Council of European Social Science Data Archives:

  https://www.cessda.eu

- And also in real life research such as the ESS - European Social Survey:

  https://www.europeansocialsurvey.org

  - with its associated online data analysis tool:

    http://nesstar.ess.nsd.uib.no/webview/

Thank you!

Contact information

Adrian Dusa
dusa.adrian@unibuc.ro

Jane Fry
jane.fry@carleton.ca

References

- DDI Alliance. *Data Documentation Initiative*. http://www.ddi-alliance.org/
- Fry, J., Cooper, A., Mowers, S., & Carrington, C. (2019). "Best Practices Document: based on DDI 2.x, version 3.1". https://bit.ly/3mhLmmH
- Jacobs, J. (2006). "Evolution of Data Documentation". Workshop "A Gentle Introduction to DDI: What's in it for Me?" presented at IASSIST 2006.
- Orten, H., Beuster, B., & Jääskelainen, T. (2019). «What can DDI do for you? AN introduction to the DDI. Presented at EDDI 2019. DOI: 10.5281/zenodo.3597192
- Perry, A. & Fry, J. "Introduction to DDI: Basic Concepts and How to Develop Skills for Training Researchers" IASSIST 2019.
- Schloss Dagstuhl, October 2014. "DDI Basics". https://bit.ly/2ZkdoTu
- 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://bit.ly/2J7RDTQ