

LIM 522: Schema in 10 points

Sarah Tribelhorn

Data Documentation Initiative (DDI) described in 10 points:

1. DDI is a free standard that can be used to describe data from surveys

DDI is used by researchers in the social, behavioral, economic, and health sciences to describe qualitative and quantitative data from surveys, statistical data files, and questionnaires (DDI, 2019). It was initially developed in 1995 (Wikipedia, 2020) to create a standard schema for data description. It can be used to conceptualize, collect, process, distribute, discover, and archive data during the research data lifecycle (DDI, 2019); thus, its description as **D**ocument, **D**iscover, and **I**nteroperate. DDI is governed by the [DDI Alliance](#) (DDI, 2019).

2. DDI can be used to standardize metadata from survey data

DDI can be used by various stakeholders, including “data producers, data distributors, data libraries, data archives, data users, researchers, software developers, and vendors” (DDI, 2019, para. 3) so that the data can be documented in a standardized format to increase discoverability, accessibility, and interoperability.

3. DDI uses XML and UML

DDI mostly uses XML so that it is machine- readable and -actionable, and web-friendly (Wikipedia, 2020). It uses tags and attributes, and does not rely on a particular platform. It defines which tags are available; order of tags; required tags; and repeatable tags. In addition, DDI also uses Unified Modeling Language (UML), which is a standardized language generally used in software engineering to visualize a system design in a

standard way (Wikipedia, 2021). Therefore, to use DDI successfully, it is necessary to have some understanding of both XML and UML. However, specific tools, such as Nesstar, Colectica, and Dataverse are required to interpret the XML format.

4. Specific uses of DDI

DDI can be used to “generate interactive codebooks, implement data catalogs, build question banks, create concordance mappings, harmonize and compare data, and manage longitudinal data sets” (DDI, 2019, para. 2). It enhances data understanding and interpretation, and is highly functional, since it can be used by people, software systems, and computer networks (DDI, 2019).

5. DDI Standards

DDI used three different standards, including DDI Codebook (DDI-C); DDI Lifecycle (DDI-L); and DDI Cross-Domain Integration (DDI-CDI) (DDI, 2019).

DDI-C can be used to document single survey or simple data, and is based on XML. The XML schema documentation for DDI-C is [here](#). DDI-L is also based on XML, and can be used to document the different states of a dataset lifecycle, and for longitudinal datasets. In addition, it can also be used to describe multiple datasets. The XML schema documentation for DDI-L is [here](#). DDI-CDI can be used to integrate data from different disciplines and domains. It is used to describe the data and data processes, and can be used for different types of data structures. It is based on UML, but can be expressed using XML. The relationships between different standards are shown in Figure 1 below from DDI (2019):

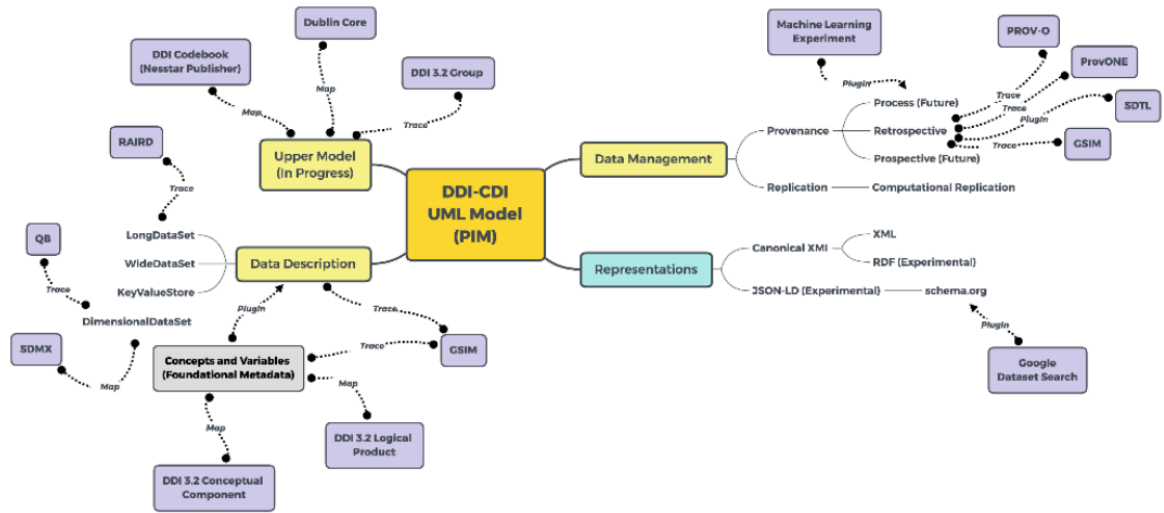


Figure 1. Relationships between DDI and different standards (DDI, 2019).

6. DDI is unique for data

DDI is different from other schemas previously studied, in that it is used to describe data; whereas the other schemas, such as Dublin Core or MODS, are used to records, such as describe books, articles, images, videos, etc. However, it is universal, in that XML can be used to describe the metadata.

7. Examples of the markups for the different DDI Standards:

An example of a DDI-C XML markup is [here](#).

An example of a DDI-L XML markup is [here](#).

An example of a DDI-CDI workflow is shown in Figure 2 below (DDI, 2019), and the XML markup is linked [here](#).

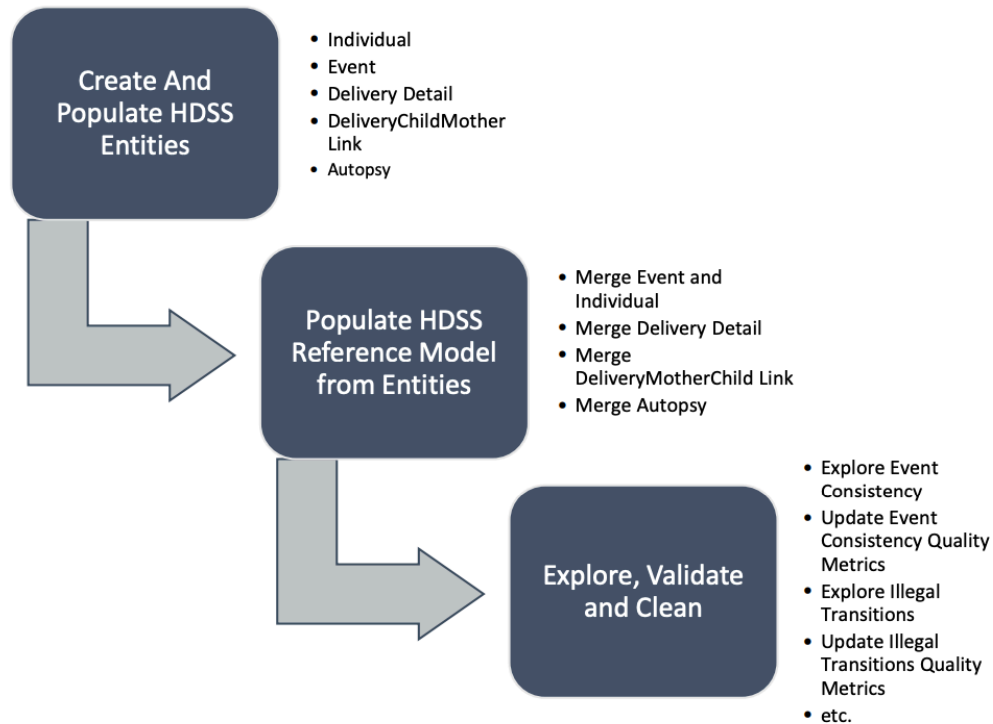


Figure 2. Example of a DDI-CDI workflow (DDI, 2019).

8. Benefits of DDI

The benefits of DDI include its relative interoperability across different systems. It has a rich content, which is both expansive and granular. It has precise and accurate searching capabilities, and it is used globally (DDI, 2019). Furthermore, DDI follows the FAIR principles when applied to metadata: Findable, Accessible, Interoperable, and Reusable (DDI, 2019).

9. Challenges of DDI

It is relatively complicated, and so requires training before use (DDI, 2019). Therefore, in a library setting, during the project planning phase, resources will need to be assessed and allocated for this if necessary.

10. DDI is used by many international organizations and projects

Global projects and organizations using DDI include (but are not limited to) the Norwegian Social Science Data Services; Harvard University; Statistics Canada; Census Bureau; Australian Social Science Data Archive; and DataFirst (University of Cape Town) (DDI, 2019). In addition, software packages, such as [EpiData](#), have been developed to allow for data definitions, documentation, entry, and analysis based on Dublin Core Elements, which are then implemented as DDI standards (DDI, 2019). Furthermore, several open source web applications have also been developed using DDI, such as the Dataverse Network, which was developed to “share, preserve, cite, explore, and analyze research data”, while making it accessible (DDI, 2019). Several dataverses can be hosted in a Dataverse repository. For example, the Harvard Library and Harvard University of Information Technology collaborated with the Institute for Qualitative Social Science to make the [Harvard Dataverse](#) available to data collectors and researchers globally to deposit data (DDI, 2019).

References

DDI. (1999). *Document, discover and interoperate*. <https://ddialliance.org/>

Wikipedia. (2020, May 30). *Data Documentation Initiative*.

https://en.wikipedia.org/wiki/Data_Documentation_Initiative

Wikipedia. (2021, October 27). *Unified modeling language*.

https://en.wikipedia.org/wiki/Unified_Modeling_Language