Survey and Analysis: DDI Relationship to Other Standards/Interoperability

# Overview

This paper provides a summary of discussions on the topic of DDI’s alignment with other standards over the recent past and describes various approaches to standards alignment which have been proposed. It is intended to serve as input for further discussion of this topic within MRT, and to frame up the issues which will need to be addressed by the group in the production of the DDI 4 Core.

Th summary is based on the discussions of this topic that have taken place since 2013. It lays out different types of relationships with other standards, provides a list of standards we have discussed aligning with, and suggests which type of alignment is appropriate for each of the relevant standards. Potential issues with various approaches are also identified.

# Problem Statement

DDI is the dominant standard in the social sciences data space. In many cases it needs to interact closely with other standards to support integration of data from several fields, work with standards that support high level searching for content (i.e. Dublin Core, DCAT, DataCite), or to transmit metadata to and from DDI to related standards. In addition, DDI needs to provide a clear transition path between different versions of the standard and to provide an implementation model for several conceptual models common in the social sciences data space.

The purpose of building relationships to is clearly define the commonalities between DDI and other standards, identifying exact matches, similarities, and the means of mapping data from one structure to another by providing clear metadata linkages. With the expansion of DDI to multiple expressions (UML, XML, RDF, etc.) there is also a need support the movement of metadata content between these different expressions without loss.

# Types of Relationships

The selection of approach is dependent upon how the information is intended to be used within DDI and as an interface with non-DDI systems/users.

### 1 - Informed modeling

This is the use of the external standard to inform the general content and relationships within related sections of DDI.

*Example:* ISO/IEC 11179-3 (Agency, ID, Version), ISO/IEC 11179-5 (Name, Label, Description), GSIM modeling of the ISO/IEC11179 Data Element

### 2 – Selective replication of content

This is the practice of using explicit structures that describe a set of information in a way that it is used by its community. The structure is replicated with a DDI namespace set of objects and results in a clean DDI to external mapping. These may be simplifications, such as limiting options for content. These are objects that must be handled in a consistent way within both standards and therefore clean mapping is a requirement.

*Example:* ISO 19115 Bounding Box, Spatial Object; OWL-S InParameter, OutParameter, Binding

### 3 – Inclusion of native objects from an external standard

This is an incorporation of all or part of an external standard within DDI using the external namespace. This is used where a clear and strong interface occurs between DDI and external systems.

*Example:* Dublin Core, XML types (xs:date, xs:string, etc.), XHTML

### 4 – Plug-in points that can be used to define specialized substitution content from an external namespace

This is an abstract class in DDI that allows for the specification of content from an external namespace

*Example:* DDI-Lifecycle PhysicalDataStructure/RecordLayout, DDI-Lifecycle and DDI4 Capture

# General Comments Regarding Approaches to Creating Relationships

* Version specific relationships and mapping – version changes in either standard can change the level of similarity in the relationship, and therefore relationships must be version specific.
* Maintaining mapping with new standards – as new standards emerge it may be possible to build on mappings by chaining across several standards. However, the validity of such mappings must be verified.
* The use of profiles may be sufficient in many situations for conveying relationships between standards. This may be especially true for less complex high-level standards that focus on discovery or focused publication.

### Issues related specifically to the inclusion of native objects from an external standard

1. Version changes in the external namespace can cause DDI to become out of sync with the published namespace. For example, Dublin Core has updated by expanding specializations of terms which required a new version of the DDI package as it directly contains the Dublin Core dc.xsd and dcterms.xsd. Inclusion of these files was done to support validation and of DDI where connectivity is an issue. The reference to Dublin Core from the DDI schema is not version specific resulting in the conflict between the version of Dublin Core available remotely and that available within the DDI published package.
2. There is a value in the incorporation of native elements when the purpose of the metadata within DDI is to facilitate a tight interface between DDI and the systems using the external standard. For example, the use of XML primitives and attribute types provides consistency of understanding and processing by the systems that use these objects.
3. DDI may wish to be able to extend these objects for internal management and/or to facilitate linking them to other DDI objects. For example, the addition of a DDI typing mechanism for a Dublin Core Identifier to differentiate between multiple identifiers used by the systems with which the DDI user is interacting.
4. DDI may wish to limit what is included either in terms of the number or complexity of the external objects.
5. DDI is a model which is expressed in a minimum of three languages (UML, XML and RDF). If the external standard is expression specific there may be objects whose structure does not translate well into other expressions. For example, UML directly supports only a subset of XML primitives and other content or structures may be XML or RDF centric causing difficulties in expressing the content correctly in each language.

# Suggested Approaches to Identified Standards

These are recommendations for possible approaches to providing relationships to standards that currently have relationships to DDI or have been noted as standards of interest.

## Considerations

* Standards that should inform DDI frequently are expressed primarily as conceptual models in the social science domain or a major sector of that domain such as those used by Statistical agencies.
* Selected replication is relevant when DDI intends to support the implementation of a conceptual standard or ensure close, well-defined transfer of content between the two standards.
* Direct inclusion has proved problematic in the past and should only be used with very stable standards if at all.
* Standards focusing on high-level discovery or focused publication have been limited to mappings or profiles. There are a number of these standards covering similar content at greater or lesser detail which are used in a number of domain spaces which interact with social science data. Selecting a single standard for deeper incorporation would be difficult given the developing landscape.

## Suggested Approaches Table

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| STANDARD | Informed Modeling | Selective Replication | Namespace Inclusion | Plug-in | Mapping/ Profile |
| GSBPM | X |  |  |  |  |
| GSIM | X | X |  |  | X |
| ISO 19115 | X | X |  |  | X |
| Dublin Core |  | X | ? |  | X |
| Text (ex. XHTML) |  |  | ? |  |  |
| DCAT |  |  |  |  | X |
| DataVerse |  |  |  |  | X |
| DataCite |  |  |  |  | X |
| PROV-O | X |  |  |  | X |
| PREMIS | X |  |  |  | X |
| CDISC | X |  |  | X | X |
| FHIR |  |  |  |  | X |
| EML |  |  |  | X | X |
| SDMX/DataCube | X | X |  | X | X |
| ISO 11179 | X | X |  |  |  |
| DDI Lifecycle | X | X |  |  | X |
| DDI Codebook |  | X |  |  | X |
| SKOS/XKOS |  | X |  |  | X |
| DISCO |  | X |  |  | X |