Design Principles

20 October 2015

# Participants

David Barraclough, Gary Berg-Cross, Michel Dumontier, Michelle Edwards, Jon Johnson, Wendy Thomas, and Mary Vardigan

# Goal / question

The goal of the group was to review the design principles for the new model-based DDI and compare DDI and FAIR principles for overlap.

We also sought to evaluate the principles against these criteria:

* clarity of principle
* self contained -> does not require additional lookup to understand the meaning
* implementability of principle
* relevance
* broadly applicable
* justification provided; related efforts

# Summary

The result of the review was a more concise well organized set of principles that will be reviewed by the Modelling Team and the Technical Committee. The principles were reduced in number and better organized for application. They were written to provide a basis for both design and evaluation of the model.

# Review the DDI design principles

* Provide rationale for each
* Prioritize in terms of impact on the model
* can these be used for evaluation? how can they be implemented?
* examples

There may be a way to combine some principles by theme. There are principles relating to design, development, and capability. The principles also relate to lifecycle phases.

## Design Principles

A set of design principles has been identified during the course of the various DDI sprints. The list is shown below:

1. Interoperability and Standards – The model is optimized to facilitate interoperability with other relevant standards.

rationale: members are using a suite of standards to meet their needs; the model needs to accommodate a plurality of standards.

seamless transformation between overlapping models; recognizing where the models abut and complement one another.

proposed: The model will be maximally interoperable with relevant standards

We will endeavor to understand the interplay of our member community-based standards and develop a mechanism to work towards interoperability (e.g., mapping).

See I1 in FAIR principles

1. Simplicity – The model is as simple as possible and easily understandable by different stakeholders.

rationale

worried about overengineering; complexity to accommodate all use cases can have a negative impact on understandability

proposed: The model should carefully balance complexity with functionality and understandability

1. User Driven – User perspectives inform the model to ensure that it meets the needs of the international DDI user community.

rationale

balance design with user need; be able to adapt to user needs;

proposed: The model will respond to user needs in an agile manner

1. Terminology – The model uses clear terminology and when possible, uses existing terms and definitions.

rationale

proposed: The model will be documented in such a way to facilitate understanding for multiple communities

through terminological correspondence

Vocabularies will be exposed in the use cases to ground the terminology.

the community will strive towards ready cognitive understanding of the model

1. Iterative Development – The model is developed iteratively, bringing in a range of views from the user community.

rationale

maintain stable core

addition of new aspects does not adversely affect the core

proposed: The model will respond to user needs in an agile manner with minimal disruption of previous commitments/developments

1. Documentation – The model includes and is supplemented by robust and accessible documentation.

proposed: The model is documented in a clear and accessible manner and provides justification for design decisions.

1. Lifecycle Orientation – The model supports the full research data lifecycle and the statistical production process, facilitating replication and the scientific method
2. Reuse and Exchange – The model supports the reuse, exchange, and sharing of data and metadata within and among institutions.
3. Modularity – The model is modular and these modules can be used independently.
4. Stability – The model is stable and new versions are developed in a controlled manner.
5. Extensibility – The model has a common core and is extensible.
6. Tool Independence – The model is not dependent on any specific IT setting or tool.
7. Innovation – The model supports both current and new ways of documenting, producing, and using data and leverages modern technologies.
8. Actionable Metadata – The model provides actionable metadata that can be used to drive production and data collection processes.

Additional lower-level principles have surfaced during initial DDI model development:

* Remodeling Discouraged – The model leverages existing structures in the specification whenever possible to avoid inefficiencies.
* Objects Represent Actual Things – The model includes objects that are functional and are used.
* Separation of Logical and Physical – The model supports a distinction between logical and physical representations.
* Names are Mutable – The model contains names and labels that may change to encourage accessibility.
* Common Expressions – The model will only have features that reflect the common expressive capabilities of supported syntaxes/technologies (e.g., no multiple inheritances)

These principles are needed to inform the design of DDI 4.0, and to assist in decision-making during development and maintenance of the standard in future.

# Review the FAIR principles

<http://micheldumontier.blogspot.de/2015/10/refining-fair-principles.html>

FAIR Principles (proposed)

Preamble

One of the grand challenges of data-intensive science is to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows. Here, we describe FAIR - a minimal set of guiding principles to make data Findable, Accessible, Interoperable, and Re-usable.

To be Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier

F2. data are described with rich metadata

F3. (meta)data are registered or indexed in a searchable resource

F4. metadata specify the data identifier

To be Accessible:

A1 (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 the protocol is open, free, and universally implementable

A1.2 the protocol allows for an authentication and authorization procedure, where necessary

A2 metadata are eternally accessible, even when the data are no longer available

To be Interoperable:

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

To be Re-usable:

R1. meta(data) have 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 their provenance

R1.3. (meta)data meet domain-relevant community standards

# Conclusion

## Agreed positions

## Non agreed positions

## Recommendations

We recommend the revision of the DDI principles to the following:

**DESIGN**

The model

\* is developed in an agile, modular and iterative manner

\* is responsive to community needs to produce actionable metadata

\* should balance complexity with functionality and understandability

\* is extensible and strives towards compatibility between different versions

\* is maximally interoperable with relevant community standards

\* supports a plurality of implementations

**DOCUMENTATION**

The documentation of the model

\* is clear, complete, and timely

\* is concise, comprehensible, accessible, and useable by multiple communities

\* provides justification for design decisions.

\* provides reference and functional perspectives

**CAPABILITY**

The model and its documentation

\* support the discovery, reuse, exchange, and sharing of (meta)data

\* support the capture, production, management, and analysis of (meta)data

\* support audit and reproducibility across the (meta)data lifecycle

*further actions*

-> evaluation of the principles against the evaluation criteria

-> determination of specific metrics for assessing the group’s conformance to the principles

# Still open / further discussion