DDI Alliance Executive Board Meeting
4 May 2020

Present: Bill Block, Cathy Fitch, Maggie Levenstein, Jared Lyle, Steve McEachern, Barry Radler, Joachim Wackerow

Memorandum of Understanding

Jared's meeting with a UM lawyer to review the Memorandum of Understanding with the international organization was postponed. They will meet this Wednesday instead.

State of the Alliance and Strategic Directions

Steve reviewed a draft State of the Alliance and Strategic Directions presentation he will give at the upcoming annual Meeting of Members. It included a quick summary of activities over the last 12 months, current issues, and strategic directions - aligned to the three new work products established over the last 12 months. The later sections tried to align the submitted budget proposals with each of the three work products.

The Board discussed the funding proposals. It was asked whether the Board can put a moratorium on Alliance travel, which would open up funding toward promotion and training.

One of the funding proposals is a new position to help coordinate Alliance activities, including assisting with marketing and promotion. It was proposed that this position be staffed at ICPSR and report to the Executive Director. Maggie and Jared will discuss staffing for this position.

DDI-CDI

Arofan Gregory, moderator of the group developing DDI-Cross Domain Integration (DDI-CDI), presented about DDI-CDI, including about the background, groups and events, functionality, and current status and timelines (see Appendix 1).

Appendix 1

DDI – Cross Domain Integration (DDI-CDI)

Feature Summary and Status

4 May, 2020

Outline

- Background: MRT and DDI 4 Core
- Group and Events
- DDI-CDI Functionality
- Current Status/Timelines

MRT and DDI 4 Core

- In the margins of the 2018 EDDI meeting (Berlin) it was agreed that a “core” of the DDI 4 work should be brought to market
- A 1-year timeframe was proposed
- The Modelling, Representation and Testing (MRT) group was formed in early 2019
- The working process was to base models on implementations, tested against real-world use cases
  - ALPHA Network
  - DDI R Libraries
  - Others (BLS for time series, etc.)

Group and Events

- Small group (9 members) meeting weekly (and more) for over a year
- No turn-over – members have been extremely focused and disciplined
- Toronto Sprint in margins of NADDI 2019
- Dagstuhl Sprint in October 2019
- Public Review Release April 2020
- Communications with management, technical committee work, marketing, and training have been emphasized

Evolution in Purpose

- DDI-CDI was expected to be the “core” of a model-driven DDI
  - A “next generation” after DDI-Lifecycle
- Implementation cases showed that something else was needed: a focus on data provenance and data integration
- DDI-CDI has emerged as a *companion* to DDI-Codebook and DDI-Lifecycle, not a replacement for them
- The SBE community needs better data integration tools
  - So do other domains!

DDI-CDI Functionality

- Describe data formats:
  - Rectangular/unit-record
  - Long/event
  - No-SQL/”big data”
  - Multi-dimensional
- Describe data provenance/process
  - Procedural process
  - Declarative process
- Describe “foundational” metadata
  - Codes/categories/classifications
  - Concepts, variables, etc.

Design Goals

- Produce a useful, implementable product based on real use cases
- Produce a standard which would be useful across technology platforms (model-driven)
- Produce a standard which is more approachable and easier to understand
  - W3C specifications used as a model
  - Lots of examples at different levels

Data Description

- Focus is on the role played by individual datums across different types of structures
  - The same data point performs different functions (measure, descriptor, identifier) in different data sets
- We can describe 4 types of data structure
  - The model can easily extend to describe others
- Data transformation tools perform this kind of thing all the time
  - DDI-CDI can express the relevant metadata for tracking datums across different structures
  - No other standard has this capability

Process and Provenance

- DDI has never attempted to describe the processes which are combined to actually produce data
  - Focus has always been on low-level data processing (stats packages/SDTL)
- DDI-CDI describes processes at a higher level, and connects them with low-level processing descriptions
- Directly implements common models for provenance and process (PROV, BPMN)
- Supports “black box” parallel processing as well as stepwise “flow” processing
  - New feature of DDI
  - Becoming common in the real world

Foundational Metadata

- Building on years of work in DDI 4
- Sophisticated model for variables, conceptual underpinnings/application
- Works flexibly with different ontologies/concept systems/thesauri
- Well-aligned with DDI-Lifecycle

Current Status/Timeline

- Public review period ongoing through July 2020
- Series of webinars to recruit meaningful review from other domains
  - CODATA is supporting this activity
- Revised review version released in September 2020
- Focused review at intensive Dagstughl workshop or virtual equivalent
  - CODATA has offered to convene a working group of reviewers from external domains to feed requirements into MRT
- First production release early 2021

Questions?