DDI Dagstuhl Sprint 2016 Week 2, 23. – 28. Oktober 2016,

DDI Moving Forward: Improvement and Refinement of Selected Areas

**High Level Production Model – Controlled Vocabularies**

The High Level Production Model of the DDI Controlled Vocabularies Group (DDI CVG) (see *Image1* below) displays the procedures currently carried out by the DDI CVG when producing a new Controlled Vocabulary (CV). This includes three phases: the Initiation, the Production Process and the Publishing Process



*Image1*: High Level Production Model of the DDI Controlled Vocabularies Group (detailed)

**Initiation**



*Image2*: Initiation of a CV

Initiation: There are several ways to initiate a CV. Some of the CVs stem from the list of classes that support CVs in DDI Lifecycle and from (major) shifts within the DDI specification. Others stem from the needs and demands of the users community, including CESSDA countries. There is also the possibility to contact the DDI CVG in order to express a need for a CV via email. However, there is no formal way of expressing such a need at the moment.

Prioritization: Currently there are no set rules for prioritization. Decisions on the priority of a CV request have been done on a case-to-case basis. Often the emphasis has been on producing a CV for elements for which many organizations had an internal CV, thus indicating a wide-spread demand.

**Production Process**



*Image3*: Production Process of a CV

Define coverage: DDI CVG looks at definition of the element in question in the DDI specification to see what area the element covers. If needed, the CVG expands or refines the definition to better describe the scope of the CV.

Review existing lists: The CVG examines the internal CVs of group member organizations and tries to get hold of the lists of other organizations and documentation standards. Additionally, Google searches are made.

Establish a base list of entries: Next, the existing lists are evaluated against coverage and common terms, and terms may be created, edited, or added. The terms in the CV should be mutually exclusive, with no overlapping. The list in itself should be as exhaustive as possible to cover many different types of data.

Review of list: The amended list is re-evaluated and tested by documenting real data with it.

Write definitions: Draft definitions are made and reviewed within the group.

Changes to list: Based on testing, reviewing and definition creation, the terms in the list are often amended. This loops the working process back to the writing of definitions: the terms and definitions are reviewed iteratively until there are no more changes to be made, which then triggers the output list.

**Publishing Process**



*Image4*: Publishing Process of a CV

Publish: In order to publish a CV, the final list is captured as an XML as well as an Excel document (produced manually). Both documents are reviewed by the DDI CVG for any kind of errors and inconsistencies (down to punctuation). The ICPSR web support runs a stylesheet on the XML file in order to turn it into HTML. The end result is a package of the CV in three different formats (XML, Excel, HTML). The overview table of the DDI CVs on the DDI webpage gets updated and the newly produced CV can be retrieved from the web. It is possible to download individual files or a zip package containing all files. Available are the latest version of a CV as well as all the previous ones. If several CVs are published at one time, the group may send an email to the DDI users list to inform them about the new release.

Revision: Revision is triggered by feedback from users. User reports are validated and decisions on the corrections needed are made. If necessary, the process loops back to the step “changes to list” (within the production process phase), in which terms and definitions are revised in an iterative process.

Currently, DDI CVs are published in American English (variants of them exist only in local organizations) though there are a few other language versions for some vocabularies.