

# CCDH Internal Monthly Report - 2020-12-16

---

**2020-12-15**

## **Data Model Harmonization Workstream Update**

### **1. Progress over the last month**

- a. Ongoing meetings with CDA technical teams
- b. Published an updated version of the CRDC-H biolinkml model [\[link\]](#)
- c. CDA MVP internal POC release generated some feedback for CCDH on the model and terminology deliverables [\[link\]](#)
- d. Review of CDS Minimum Metadata [\[link\]](#) and recommendation to align with node-specific models for modality-specific entities (e.g. sequencing entities align to GDC.ReadGroup)

### **2. Status (compared to our timeline)**

- a. On target to complete milestones for Phase 2 - 2020Q4 [\[link\]](#)

### **3. Next steps, work for the upcoming month in the context of workstream or cross-cutting plans**

- a. Update CRDC-H model documentation to reflect current release of model
- b. Continue development of CRDC-H model in those priority areas identified by the CDA as potential extensions for March [\[link\]](#). These include some diagnostic attributes and some treatment attributes
- c. Continue work on custom tooling to support a version-controlled data modeling workflow
  - i. Support for transformations to/from spreadsheet format, csv format, template-based documentation generation

---

**2020-12-09**

### **1. Tools & Data Quality Workstream Update (Jim, Gaurav, Laura)**

#### **a. Progress over the last month**

- i. file formats/models for harmonized data — to be presented below
- ii. kicking off E2E focus group work

#### **b. Status (compared to our timeline)**

- i. Would like to better sync with Terminologies WS to add granularity to current / new tickets.

#### **c. Next steps, work for the upcoming month in the context of workstream or cross-cutting plans**

- i. better coordination with Terminology group (e.g. terminology server, NCIT-plus pipeline)

### **2. CEDAR and PFB - Talk about when and what next steps should be. (Gaurav).**

- a. [Slides](#)

- **ToDo:** Need to investigate relationship between BiolinkML or CRDC-H model in BiolinkML and its relationship (how it can be realized) in CEDAR content.

---

**2020-12-02**

**Terminology & Ontology Ecosystem Workstream update.**

**1. Progress over the last month**

- Worked on value mapping of permissible values in the MVP model.  
[https://drive.google.com/drive/folders/1XQEXbKkIRjguRx3zrOQI\\_FzIFKXAKEL-](https://drive.google.com/drive/folders/1XQEXbKkIRjguRx3zrOQI_FzIFKXAKEL-)
  - Retrieved existing GDC permissible value to NCI mapping from EVS.
  - Algorithmically mapped some values based on common value meaning identifiers in caDSR and mappings in NCI
  - Manually mapped other values to NCI codes
  - Communicated with GDC, PDC and HTAN to verify the mappings.
  - Based on the mappings, did an exercise to create harmonized sets of permissible values for a harmonized attribute.
- Working on using a graph database and a REST API to manage the mappings, using a graph database. <https://github.com/HOT-Ecosystem/crdc-node-models>
  - Wrote scripts to import the data dictionaries into the graph database.
  - Created a graph model based on the iso-11179 model to represent the mappings.
  - Developing a REST API to retrieve and update mappings.
- Working on implementing an enumeration type in BiolinkML. It will be used to represent code sets in the harmonized model.  
<https://github.com/biolink/biolinkml/tree/enumerations>
  - Generic enumeration model -- from simple set of possible values for an attribute to embedded value set definition to reference to predefined value sets.
  - Includes CTS2/FHIR based binding parameters (versions and tags)
  - Serves as a gateway to TCCM / FHIR and other terminology services
- Started to gather requirements for the terminology service
  - Differentiate behaviors that already exist in servers (FHIR, NCI EVS, VSAC, ...) from value set and mapping specific behaviors that *depend* on existing servers but need a separate implementation
  - Differentiate authoring / browsing / discovery tools from run-time uses of these tools and determine the requirement “gaps”

**2. Status (compared to our timeline)**

- Closed milestone Q3 2020 and working on milestone Q4 2020.  
<https://github.com/cancerDHC/Terminology/milestones>

**3. Next steps, work for the upcoming month in the context of workstream or cross-cutting plans**

- Continue to work on harmonizing the permissible values in the MVP models

- b. Implementing the tool to manage the mappings
  - c. Complete the implementation of the enumeration type in BiolinkML
  - d. Collect the requirements for the terminology server and start implementation.
- 

**2020-11-18**

**Community Development Workstream Update.**

**1. Progress over the last month**

- a. Outreach to GDC, PDC re: terminologies mapping work
  - i. PDC responded on 11/11/20
  - ii. Sent reminder to GDC on 11/10/20
  - iii. Sent email to HTAN 11/16/20
- b. Closed tickets:  
<https://github.com/cancerDHC/community-development/issues?q=is%3Aissue+is%3Aclosed>
- c. Open tickets: <https://github.com/cancerDHC/community-development/issues>

**2. Status (compared to our timeline)**

- a. Working on open tickets:  
<https://github.com/cancerDHC/community-development/issues>

**3. Next steps, work for the upcoming month in the context of workstream or cross-cutting plans**

- a. Working on open tickets:  
<https://github.com/cancerDHC/community-development/issues>
- b. Working on next newsletter - early January 2021

**4. End-to-end analysis work has begun, will involve Community Development workgroup soon**

**a. Draft for End-to-End Analysis Plan will be submitted to SCC on Thursday .**

- i. We have been collaborating on this over the past two weeks and will have a draft to submit tomorrow at SCC. One of the reasons it took us a little longer to complete this document is that we wanted to get the end-to-end analysis started. The team has been meeting and started on the work.
- ii. This work is informing a draft of CCDH's vision for overall CRDC architecture - this will be the formal deliverable from this work
- iii. Anticipate this work over the next 8 to 10 weeks
- iv. Will need more involvement from the nodes as this work continues