

# CCDH Internal Monthly Report - 2020-11-11

## 2020-11-11

CCDH Internal Meeting was not held, in observance of the US Holiday honoring Veterans of the US Armed Forces.

---

## 2020-11-04

- **Data Model Harmonization Workstream Update.**

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

- i. Presented updates of ongoing data model harmonization and CDA MVP work during the CRDC All-Hands meeting on 10/26/2020
    - ii. Completed first iteration of MVP deliverables for end-to-end harmonization efforts with CDA
      - 1. Biolinkml representation of the model now in a CancerDHC repo: <https://github.com/cancerDHC/ccdhmodel>
    - iii. Provided initial set of MVP deliverables to CDA during meeting on 10/22/2020 and held a follow-up deeper technical discussion with the CDA data modeling team on 10/28/2020

- b. **Status (compared to our timeline):**

- i. Making progress towards our [Phase 2 milestone](#) and the work with CDA provides an opportunity for us to get critical feedback on the CDM and an impetus to move to a more automated and computable modeling workflow

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

- i. Continue to refine work on the MVP deliverables based on feedback from CCDH and CDA
      - 1. Biweekly meetings with 1) broader CDA team and 2) technical team
    - ii. Meet with Melissa Cook and others in the CRDC Data Submission Working Group to propose a set of minimum metadata elements for submission Data submission working group has consumed the ADM and expanded the scope to include IDC and CTDC to CRDC nodes, with a focus on the needs of the CDS
    - iii.
      - 1. They have some bandwidth in the coming weeks to help support this effort to define a minimum set of metadata
      - 2. One constraint is that the deeper harmonization that CCDH undertook in creating the CDM has been limited to biospecimen and administrative data subdomains; clinical metadata is not yet covered and so recommendations on minimum metadata would

either need to be based off more superficial harmonization (i.e. ADM-level harmonization) or be scoped to include only subdomains that have been more deeply harmonized and refactored.

- iv. Review the Data Submission Working Group's alignment of IDC and CTDC data elements with existing ADM data elements as a first step in incorporating these node models into the CCDH pipeline

---

## 2020-10-28

### 1. (15 minutes) **Tools & Data Quality Workstream Update.**

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

- i. Presented work on data harmonization with caDSR and Ptolemy.V to IDC and NCI and (later) to the CRDC F2F meeting.
  - 1. They like the work, but don't have an immediate need for harmonized data: as discussed at the CRDC F2F yesterday, they are hoping for a centralized node responsible for clinical data harmonization and storage. At the least, they're not going to harmonize the datasets in TCIA until their users ask for it.
  - 2. Gaurav has committed to providing feedback to NCI that they could use to improve Ptolemy.V.
  - 3. Slides with examples for *Harmonizing Clinical Data with caDSR and Ptolemy.V*, starting [here](#).
- ii. Started work on identifying minimum metadata that might be useful across the nodes.
  - 1. Met with Melissa Cook and Smita Hastak to discuss the CRDC Data Submission Working Group; Laura is now part of that group and meets with them regularly.
  - 2. Received a [preliminary list](#) from CDS; we're now looking at [some example datasets](#) suggested by CDS.
- iii. Laura beginning work with Davera on end-to-end focus group.
- iv. Started focused meetings with Terminologies team and demonstrated SNOMED in DICOM mapping procedure to them.
  - 1. [Updated DICOM mappings](#) based on David Clunie's feedback.

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

- i. Need to finish two things by Friday to stay [on timeline](#):
  - 1. [Compare PFB and CEDAR instance representations](#) of the harmonized data from the IDC Radiomics and ISPY1 datasets.
  - 2. [Develop strategy and plan aligned with CRDC needs](#), which we have been doing around the end-to-end information flow effort.

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

- i. Aim to complete our milestone for [2020 Q4](#) by December 23.

- ii. Complete work on SNOMED in DICOM matching by:
  - 1. [Integrating BioPortal](#) into the mapping process.
  - 2. [Integrating RadLex and AGROVOC vocabularies](#) into the mapping process.
  - 3. [Grouping terms](#) based on the NCI hierarchy.

---

**2020-10-21**

During this week, the team prepared a joint presentation for the CRDC All Hands Meeting held Oct 26-27. The presentation can be found [here](#).