

CCDH Internal Monthly Report - 2020-09-16

2020-09-16

1. Tools Workstream Update

a. Progress over the last month:

- i. Created plans until 2021 Q1 as [document](#) and [Github milestones](#).
- ii. Shared results of worked example ([IDC NSCLC Radiomics](#), taking [input data](#) and [column and field mappings](#) to create [mapped data](#)) with IDC; who [responded positively](#). Currently discussing best steps forward with IDC.
- iii. Updated DICOM SNOMED to NCI mapping to produce [a set of mappings](#) in the SSSOM format.
- iv. Helped create the [resources list](#) for CCDH website.
- v. Have participated in CCDH data model specification format discussions.

b. Status (e.g. against a timeline):

- i. Making progress towards 2020 Q3 goals:
<https://github.com/cancerDHC/tools/milestone/2>

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

- i. Continue working on [IDC example data harmonization](#), with a goal of:
 1. Completely map a second example file.
 2. Choose a format for storing and sharing column mapping information.
 3. Choose a format for storing and sharing harmonized data.
 4. Determine what the next steps for IDC should be.
- ii. Continue working on DICOM SNOMED to NCI mapping, with a goal of:
 1. Develop [tools for looking up unmapped terms](#) in ontologies and on the Ontology Lookup Service.
 2. Choose a [format for storing mappings](#) between terms across different resources.
- iii. Develop [strategy and plan](#) aligned with CRDC needs.
- iv. Implement [boomer workflow](#) for data model harmonization.

2020-09-09

1. Terminologies Workstream Update

- a. Quarter 3 milestone: <https://github.com/cancerDHC/Terminology/milestone/1>
- b. Value set harmonization <https://github.com/cancerDHC/Terminology/issues/8>:
 - i. Assembled an example value sets (Sample) from GDC, PDC and ICDC in tabular format
 - ii. **ToDo**: Process HTAN value sets
 - iii. **ToDo**: Develop a workflow for streamlining the assembly and manual curation

- c. Value set modeling in BiolinkML (TCCM)
<https://github.com/cancerDHC/Terminology/issues/13>:
 - i. Created core data models in BiolinkML
 - ii. Value set data model in BiolinkML
 - iii. Created REST API definition for value sets service in openapi
 - iv. **ToDo**: Complete the modeling and service
 - v. **ToDo**: Prepare a TCCM presentation for the HOT community meeting in October
- d. Represent value sets as FHIR value sets
<https://github.com/cancerDHC/Terminology/issues/9>:
 - i. Completed the FHIR service based on caDSR and NCIt data. Example:
<https://fhir.hotecosystem.org/terminology/cadsr/ValueSet/5432508>
 - ii. Created a draft implementation guide using FHIR shorthand
 - iii. Made modification to the service using FHIR extension.
- e. BiolinkML support
 - i. Presented in the data harmonization workstream meeting
 - ii. Added components for future TCCM work
- f. Standards and Tooling page of the web portal:
<https://ccdportaldev.pedscommons.org/standardsTooling>

Notes:

- From Bill Duncan:
 - See the children of our attribute value class:
<https://microbiomedata.github.io/nmdc-metadata/docs/AttributeValue.html>
 - E.g., Consoled term value:
<https://microbiomedata.github.io/nmdc-metadata/docs/ControlledTermValue.html>
 - ... or geolocation value, in which we have both lat and lon parts:
<https://microbiomedata.github.io/nmdc-metadata/docs/GeolocationValue.html>

2020-09-02

1. Community Development Workstream Update

- a. Quarter 3 Milestones:
<https://github.com/cancerDHC/community-development/issues/6>
- b. Discussion on CDA / CCDH collaboration:
 - Think it would be a good idea to coordinate the creation of a timeline with CDA team; Brian says they are planning to have an implementable model ready within a 3 month timeframe but need to ensure that this is the best timeline for both teams
 - Will CCDH be supporting the 'T' component of ETL? - at a minimum want to be able to provide mappings for source elements to target elements, but there is a question as to how formal these mappings need to be
 - Need further coordination / conversation about who is responsible for various elements of the project: ie nodes, CDA, CCDH, all of the above

- Next step: interfacing with CDA to ensure both teams know what they should be doing moving forward
 1. Need a ½ page document to document this conversation (what tools, processes, etc are currently in use and who is responsible - this will be discussed in the DMH meeting tomorrow)
 2. Will help with the current in / out of scope conversation for CCDH
- c. Discussion on the CRDC org chart
 - Will help to understand who is doing what and eventually lead to the proper communication channels
 - Hope to be able to proactively plan who we are going to engage with
 - Sherri - think initially this should be done for the CCDH team to help complete our own work, rather than create a document for the entire CRDC (that may be out of scope)
 - How to gather information from various stakeholders? Will CRDC members be willing to share information about reporting structures, etc?
 - Continuously changing so it is difficult to capture
 - Start by taking current contacts and laying them out in a diagram format, will then fill in gaps in the chart
 - Ask questions to determine who technical teams and leadership is and then may organically grow into a document that is useful beyond the CCDH scope
 1. Ex. Erika could describe HTAN structure
 2. Will need to determine what information is important
 - Sam - think we should generate diagram and fill in as much information as possible to begin
 1. Use a tiered process - Kat will create initial document and share with the Community Development WS, then open to CCDH internal team, then open to node contacts
 2. Will eventually discuss opening this up to entire CRDC group, but want to check with Federal leads / internal team to reach agreement that it is appropriate to be completed more broadly

2020-08-19

1. Data Model Harmonization Workstream Update

a. Progress over the last month

- i. Discussed and documented processes, tasks, separation of duties, and staffing related to value set harmonization which will be a critical next step in evolving the CDM into an implementable model
- ii. Completed initial mapping of outstanding biospecimen and administrative entities and properties to BRIDG with assistance from Wendy
<https://github.com/cancerDHC/data-model-harmonization/issues/7>
- iii. Prepared for and conducted the first of our 'retrospectives' of DMH Phase 1 work <https://github.com/cancerDHC/data-model-harmonization/issues/13>

- iv. Conducted a review of data model change management procedures across the CRDC as a first step towards defining a process that keeps the CCDH data model artifacts in sync with changes to source models <https://github.com/cancerDHC/data-model-harmonization/issues/11>
 - v. Changed the format of Office Hours to focus on questions/topics submitted in advance by community members rather than an open question and answer format; no topics were submitted for the August 6, 2020 meeting so that was cancelled
 - vi. Began our regular cadence of bi-weekly meetings with CDA modelers on August 13, 2020. The groups are converging on a scope for the initial work that will include administrative and biospecimen subdomain entities, ideally from both GDC and HTAN. Provisionally set a deadline of 3 months for delivery of an implementable CRDC-H constrained to biospecimen and administrative subdomains. First steps will be scoped very narrowly to a small subset of properties as an end-to-end proof of concept. Next meeting will be August 27, 2020. Rolling minutes and agenda [here](#):
 - vii. Continued progress in modeling framework / language selection process
 - 1. Defining a set of evaluation criteria against which options will be assessed [here](#)
 - 2. Main candidates at this point are Biolink ML and FHIR (either leveraging existing resources or defining our own)
 - 3. To assist in our evaluation we will implement CDM.Specimen in BiolinkML in order to have a concrete representation of a modeling artifact to review
- b. Status (e.g. against a timeline):**
- i. On target for Phase II: <https://github.com/cancerDHC/data-model-harmonization/milestone/1>
- c. Next steps/work for the upcoming month in context of workstream or cross-cutting plans**
- i. Conduct additional retrospective sessions to review our work <https://github.com/cancerDHC/data-model-harmonization/issues/13>
 - ii. Produce a proposed data model change notification process based on our survey of existing processes, the information we would want to receive from an upstream model, and input from ICDC/CTDC stakeholders <https://github.com/cancerDHC/data-model-harmonization/issues/11>
 - iii. Finish the review of modeling language / framework options for CCDH and make a selection <https://github.com/cancerDHC/data-model-harmonization/issues/9>
 - iv. Presentation about phenopackets oncology use case(s) by Peter Robinson during office hours on August 20, 2020
- d. To do:** Schedule an in-depth technical call with the Clinical Trials Data Commons team (Resham to provide list of necessary participants; Kat to schedule)
- i. Chris willing to help lead this process