# CCDH Internal Monthly Report - 2021-07-14

2021-07-14

**Data Model Harmonization Workstream Update**

a. **Progress over the last month**:
   i. Ongoing collaboration with CDA
      1. Deep-dive [discussion](#) on data modality wrt. research activities and files. This was identified as a priority for query by CDA
      2. Clarifying remaining gaps in CDA's use case as identified in their ['experience'](#) document
   ii. IDC model integration
      1. Initial work will focus on priority attributes identified by CDA and IDC (David Pot), primarily including those exposed through the [IDC portal](#)
      2. [Frequency distributions](#) of priority attributes compiled from BigQuery source data
   iii. FHIR US Core IG research gap analysis
      1. Team from Samvit is reviewing the latest version of CRDCH and consulting with DMH team; their observations are helping to identify gaps in our documentation and has been a useful partnership on that front

b. **Timeline check**:
   i. Initial work on Phase III model extensions underway

c. **Work for the upcoming month**:
   i. Preparation for upcoming Homing in on Harmonization focusing on CRDCH
   ii. Expansion of CRDC-H model documentation including example instance data to demonstrate population of complex data structures (e.g. CancerStageObservationSet) and a visual representation of the model
   iii. Begin initial modeling of experimental subdomain
   iv. Continue incorporation of IDC model into CRDCH
      1. Understand the lineage of the instance data that is exposed in the BigQuery table to assist in our modeling

2021-07-07

**Tools & Data Quality Workstream Update**

a. **Progress over the last month**:
   i. LinkML model generation
      ● [Completed] Reorganized and reviewed changes made to LinkML model generation made in the run-up to the May 31 deadline.
      ● [Mostly complete] Development on LinkML to support CCDH goals.
   ii. NCIt+
      ● Meeting with Gilberto: CBIIT is developing a triplestore service for terminologies and value sets; we talked about the possibility of including our NCIt+ existential graph there.
      ● [Ongoing] Making plans to evaluate mappings between HPO and NCIt.
   iii. Example data workflow
      ● [Ongoing] Onboarding Mark Miller and other members of the LBNL team.
   iv. Workstream
      ● [Ongoing] Cleaning up issues and planning Phase III development
b. **Timeline check**:
   i. June: focus on LinkML generation
   ii. July: focus on Example Data Workflow
   iii. August: focus on Terminology and Data Model Harmonization needs for next major release of the CRDC-H model
c. **Work for the upcoming month**:
   i. Example Data Workflow: convert current Jupyter Notebook into an automated continuous integration workflow.
   ii. Work with Terminology on coming up with a workflow that integrates model generation and terminology enumeration lookups.

2. **Informal updates**
   a. Identifiers report
   b. LinkML to FHIR mapping

3. **Reminders**:
   a. Weekly reports are due tomorrow, Thursday at 10 AM (Pacific). http://bit.ly/ccdh-weekly-reports
   b. Fortnightly CCDH / CDA Tech Deep Dive will be held tomorrow, Thursday at 10 AM (Pacific). Agenda here.

**Notes:**
1. Tools and data quality updates:
   a. On the next xs-WS call, we need to figure out what the priorities with the nodes should be and whether we need to nag them for resources (e.g. data dictionaries, public APIs, etc.)

b. Would like a technical overview of NCIt+ for a future meeting: what a good alignment looks like, what a challenging one looks like.
   c. NIH SI would be more than happy to share resources on NCIt+ if that would be helpful.
   d. Working on *Biospecimen* is a good starting point as example of how NCIt Plus can be best utilized.
2. Identifiers report
   a. Along with these recommendations, we (CCDH) also made architecture recommendations earlier this year ([https://bit.ly/crdc-e2e](https://bit.ly/crdc-e2e)).
      i. Gilberto & Robinette could have a look at this document and help us understand what the best ways to move forward some of these recommendations.
3. LinkML to FHIR mapping
   a. Samvit + Warzel group are working on FHIR to US Core IG. As part of that, they are looking at gaps that may exist between CRDC-H and their FHIR IG. Started doing mappings from CDM to FHIR US Core, but are doing CRDC-H to FHIR US Core now.
   b. MH Would like to connect with this team to strategize together. We want to be on the same page about the strategy - and would like to have a regular check in. They have been attending DMH calls; we should invite them to join the X-WS call (next iteration is on 14-July).

---

2021-06-23

**Community Development Workstream Update**
   a. **Announcement**
         Kat to become Community Manager
   b. **Progress over the last month**:
      i. Website updates are ongoing
         ● Tickets can be found [here](here).
   c. **Timeline check**:
      i. Completed Phase 2 deliverables, currently planning and prioritizing for Phase 3
   d. **Work for the upcoming months**:
      i. Next [Homing in on Harmonization](Homing in on Harmonization) meeting **tomorrow** (June 24)
         ● Topic: Foundations of Data Harmonization (Part 2)
         ● Presenters: Sam and Chris C.
      ii. Upcoming HioH meeting on August 26
         ● Suggested topic: In-depth look at CRDC-H model
         ● Presenters TBD, want cross-workstream representation

- Present the APIs in more detail, along with potential queries and how they are used. Need to coordinate with the user testing (Gilberto)

iii. User testing to begin in Phase 3 - aiming to start within ~ 6 weeks
- Important to consider how this fits with HioH and Newsletter communication
- Developing testing suite with test queries
- Will be discussing with CDA tomorrow - we will ask about 'where' they would like to start testing.
  a. I just checked the [agenda](#), and it turns out that this is on the agenda for the *next* CDA-CCDH Tech Deep Dive meeting on July 8, not tomorrow. But soon!

iv. July Newsletter being prepared, will focus on a look ahead to ways we will engage with the CCDH community in the future
- Present at Comm Dev meeting next week and then to this group
- Aim to send out first week of July (in ~two weeks)