Goal

CFDE Ontology Working Group (OWG) - Agenda & Notes

File location: *https://drive.google.com/drive/u/0/folders/1DTEgsecHc1RfzZMUkFEkOmnpVIVEhnqx*

Google Drive CFDE OWG folder:
https://drive.google.com/drive/folders/1DTEgsecHc1RfzZMUkFEkOmnpVIVEhnqx

What: *This working group will focus on establishing ontologies to be used for capture of metadata in the CFDE.*

When: *3rd Wednesday of the month, 11 am Eastern*

Where:
https://uab.zoom.us/j/88996931898?pwd=cHM0Z2FiU1VmcDJsNEdkSzAzUkh6dz09

Who: *Please contact Swathi Thaker (snthaker@uab.edu) if you do not have a calendar invite, or the OWG slack channel with any additional questions.*

Important Docs/Links:

Charter:
https://docs.google.com/document/d/1Pa1imd3wUIsmd03qCkf0-zzPt4kvJe2AAN9veN5XdhU/edit?usp=sharing

Finalized RFCs:
https://drive.google.com/drive/folders/1Yy9oP2SbZ9y1hnbN5h9r_tiWJAvmx_W4?usp=drive_link

General RFC Directory (those that are under comment):
https://drive.google.com/drive/folders/1BXI4DvygDyQyuUckiuVMnn0eP0Kv4pVN?usp=drive_link

Link to ontology-wg on Slack:
https://cfdeworkspace.slack.com/archives/C01GP14DLJX

Meeting Agendas and Notes ---Newest meeting on top---

TEMPLATE (FOR COPYING FOR EACH MEETING) <DATE>

Objective	see agenda	Time	Monthly, on the third Wednesday of the month, 11 am Eastern
Leader(s):	Before April 2024: Michelle Giglio May 2024 onwards: DRC team	Where	https://uab.zoom.us/j/88996931898?pwd=cHM0Z2FiU1VmcDJsNEdkSzAzUkh6dz09 Passcode: 276611
Participants: SIGN IN: Name & Affiliation	; ;		

Agenda Item	Action Items	Owner
•		
•	•	

Notes:

Agenda Topics for Future Meetings

Date	**Agenda Topic**	**Who**
	•	
	•	

Meeting Agendas and Notes:

Nov 20, 2024

Objective	see agenda	Time	Monthly on the third Wednesday of the month, 11am Eastern
Leader(s):	DRC team	Where	https://uab.zoom.us/j/88996931898?pwd=cHM0Z2FiU1VmcDJsNEdkSzAzUkh6dz09 Passcode: 276611
Participants: SIGN IN: Name & Affiliation	; ;		

Agenda Item	Action Items	Owner
● December metadata submission: Dec 1 - Dec 16	Announcement	DRC/Mano
● Requesting new entity/metadata to C2M2	Reminder	OWG/Srini
● RFC draft - Post-translational modification of Proteins	Discuss RFC	GlyGen/SPARC /OWG
● 4DN: Interest in harmonizing information on microscopy metadata	Discuss the proposed metadata utility	4DN and HuBMAP

Notes:

December metadata submission:

Important: Please do not include the folder 'external_CV_reference_files' and its contents in your C2M2 package submission. It takes ~9GB space and is of no use to us during ingestion.

Important reminder:

Before requesting new entity/metadata inclusion in C2M2, please read the OWG charter document to understand its purpose and scope. In summary,

1. Purpose & Scope: Utilize common metadata elements to link datasets across DCCs; unifying all metadata is beyond the scope.
2. Come up with Use case(s) that explain how including the new entity/metadata will help data discovery across CFDE.
3. Overall goal: Maximize data discovery using common metadata and minimize the burden of data deposition.

PTM Discussion:

4DN-led discussion on microscopy data/instruments:

- 4DN + HuBMAP: Interest in harmonizing information on microscopy metadata including instrument hardware configuration, acquisition settings, and quality control and data structure
- This is based on the 4DN-BINA-OME-QUAREP Microscopy Metadata Specifications
    - Published here: https://doi.org/10.1038/s41592-021-01327-9
- The specifications augment the OME-Data Model and are being maintained in consensus with BioImaging North America. QUREP-LiMi is adopted by 4DN, HuBMAP and SenNet for Instrument Hardware.
- The model is represented as XSD (NBO_MicroscopyMetadataSpecifications_ALL.xsd)
- Metadata is currently captured as a JSON file (with JSON schema for validation)

How many microscopy techniques/methods should be included? For ex: Light sheet microscopy, confocal, etc.,

Meeting Agendas and Notes:

Oct 22, 2024

Objective	see agenda	Time	3:30 - 5:00 pm (EST)
Lead:	DRC team	Where	Wisconsin Hall, The Bethesdan Hotel
Participants: SIGN IN: Name & Affiliation	; ;		

Agenda Item	Action Items	Owner
● **Update on PTM of proteins RFC** ● **Update on Biofluids RFC**	● Finalize the draft RFC ● Progress on Biofluids	GlyGen, SPARC OWG
● **B2AI Onboarding** ● *A brief overview of C2M2 and other Assets* ● *Submission process* ● *Understanding the needs of each B2AI center*	● Presentation by the DRC team ● Discussion with B2AI ● Discussion with A2CPS	DRC B2AI & OWG
● **Add items here**		

Notes:

Session 1 (3:30 - 4:10 pm @ Wisconsin Hall):

1. **PTM RFC:**
   a. We received inputs from the GlyGen team on the RFC document.
   b. Do we have a consensus on the required and optional fields for the PTM?
   c. Are the proposed 'field_names', their description, data_type, and value_sets standard terms in the community? For example:
      i. Field_names: [site_one; aa_site_one; site_two; aa_site_two; site_type; ptm_type; ptm_subtype; domain_type].
      ii. What would be the field name for cellular localization of proteins - SPARC's use case?
      iii. Value_sets for 'site_type': [defined, range, alternative and connect]. How about 'unknown'? unknown
      iv. Similarly, are the 'data_types' standard representation?
      v. How to handle missing values like unknown, not applicable, not recorded?
   d. What ontologies/controlled vocabularies to encode them?
   e. Will PTM be a separate table associated with 'protein.tsv' and 'collection_PTM.tsv' with the following attributes?

Field	Field Description	Required?	Field Value Type	Extra Info
accession	ID	Required	Valid Term	URI
site_one	Position of the site or first of two positions involved in the site.	Required	Positive integers and ‘unknown’	
aa_site_one	Amino acid at the first site	Required	Single letter code of AA	
site_two	Second position involved in the site	Optional	Positive integers and ‘unknown’	
aa_site_two	Amino acid at the second site	Optional	Single letter code of AA	
site_type	*defined*: event happens on position specified by position_one *range:* event happens on the peptide between position_one (start) and position_two (end) *alternative:* event happens on position_one OR position_two *connect:* event happens on position_one AND position_two. *unknown:*	Required	{defined;range;alternative;connect;unknown}	
PTM type	Major type of PTM	Required (XOR with domain_type)		
PTM subtype	Subtype of PTM	Optional		
Domain location	Subcellular localization	Optional		
Domain type	Domain type	Required (XOR with ptm_type)		
Disease/phe		Optional		

notype association				

2. Progress on Biofluid RFC

a. The Biofluid RFC is open for comments until October 31. Do the DCCs have any questions about the RFC?

b. We are updating the schema, preparation scripts and other codebases to accommodate biofluid related changes.

Session 2 (4:20 - 5:00 pm @ Wisconsin Hall)

Goals for B2AI-OWG meeting:

1. Onboarding B2AI centers to C2M2 (presentation by OWG/DRC)
    a. Brief overview of C2M2 - Schema, Scope, and Purpose
    b. Data submission process overview
2. *Make B2AI assets discoverable in the CFDE portal*
    a. Is the current C2M2 schema sufficient to enable the discovery of your datasets?
    b. If not, what Entities (Tables), Fields (Metadata) and Terminologies/CVs are required to encode their values? *[The goal here is not to accommodate all the B2AI metadata in C2M2 but to enable discovery with common metadata elements of participating DCCs].*
    It would be very helpful if each B2AI center could provide a synthetic metadata sample and encoding terminologies.
    c. B2AI uses clinical terminologies like SNOMED, RxNORM, and LOINC. What is the translational loss in X-mapping to DO, HPO, UBERON, PubChem, etc.?
    d. Strategies to tackle the above issue:
        i. Using slim terms in DO/HPO/UBERON…so the search will subsume all the concepts under the slim term?
        ii. Requesting these terms in DO/HPO/UBERON… It will be unmanageable if there are too many terms to request.
        iii. Use InterLex?
3. Processed data
    a. What are the processed datasets that B2AI centers are willing to submit to the CFDE?
4. What are the requirements/restrictions for accessing B2AI data?
    a. For example, PPH requires user registration to access non-identifying and non-sensitive data.
    b. Are there any issues in making your data discoverable via C2M2 without registration?
5. [Please add any other items for discussion here]

6.

Please see below for some helpful resource links about the agenda items:
CFDE Workbench Documentation: https://info.cfde.cloud/documentation
C2M2 Documentation: https://info.cfde.cloud/documentation/C2M2
Detailed schema: https://info.cfde.cloud/documentation/C2M2#c2m2-tables
C2M2 Submission Script: https://info.cfde.cloud/documentation/C2M2#submission-prep-script
Data Submission to the portal details: https://data.cfde.cloud/submit
Submission portal: https://data.cfde.cloud/submit/form

Agenda documents of DRC- Bridge2AI center meetings

- **Precision Public Health: Voice**
- **Salutogenesis: AI-READI**
- **Functional Genomics: CM4AI**
- **AI/ML for Clinical care: CHoRUS**

Discussion with A2CPS:

Ari Kahn: Interested in exploring some clinical metadata related aspects in conjunction with their potential role as biomarkers

Meeting Agendas and Notes ---Newest meeting on top---

Sep 18, 2024

Objective	see agenda	Time	Monthly on the third Wednesday of the month, 11am Eastern
Leader(s):	DRC team	Where	https://uab.zoom.us/j/88996931898?pwd=cHM0Z2FiU1VmcDJsNEdkSzAzUkh6dz09 Passcode: 276611
Participants: SIGN IN: Name & Affiliation	Srinivasan Ramachandran (MW, DRC); Sherry Jenkins (LINCS, DRC); Heesu Kim (LINCS, DRC); Mano Maurya (MW, DRC); ; ; Chris Kinsinger; ; ; Natalie Vineyard (NIH); ; Jimmy Zhen (MoTrPAC); ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;David Chen (ERCC) ; ; ; ; ; ; ; ; ; ; Swathi Thaker (ICC); ; ; ; ; Raja Mazumder (GlyGen); Matt Roth (ERCC); ; ; ; ; Jeremy Yang (IDG) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;		

Agenda Item	Action Items	Owner
● RFC draft - Post-translational modification of Proteins	Discuss RFC	GlyGen/SPARC /OWG
● RFC draft - Biofluid representation in biosamples	Discuss RFC	ERCC/OWG
● (If anyone has an item to discuss, please list it here)		

Notes:

PTM Discussion:

Rene:
How to represent an unknown position: How about leaving the field empty (optional).

Henning: Using the N-terminal-to-C-terminal range as a stand-n for "unknown" will make it very difficult to get entries with unknown range back. Better use an explicit data model.
Several such issues discussed in HUPO PSI-MI work group.
Data model from the HUPO PSI-MI work group:
https://psidev.sourceforge.net/molecular_interactions//xml/doc/MIF.html#element_location_Link02BD4A18

If you allow more than one ontology to be used, you might need to explicitly include WHICH ontology has been used, it might not automatically be clear from the accession number.

To detect sequence changes, you could use the sequence checksum provided by UniProt. But again, it might be overthinking.
Just to throw another spanner into the works, a range may well be discontinuous, for example a disulphide bridge. **Raja:** We could model this by having two positions and a type. Type would say “range”, “alternate”, “connection”. Connection would be your case but we probably need a better term.

Rene: Amino acid: what if it is a range, then it is not a fixed AA.
FALDO: can handle OR (e.g., position 21 or 25); specifies grammar but no instantiation; so, it will require additional columns for complete specification.
How to capture mutation: already captured by the fields (jointly) Ref AA and Altered AA

Version control or record the release version/date, etc for various ontologies used.

Biofluid discussion:

Both ‘anatomy’ and ‘biofluid’ columns will be optional.

Use InterLex if no Uberon ID available.

Agenda Topics for Future Meetings

Date	Agenda Topic	Who
	●	
	●	

August 21, 2024

Objective	see agenda	Time	Monthly on the third Wednesday of the month, 11am Eastern
Leader(s):	DRC team	Where	https://uab.zoom.us/j/88996931898?pwd=cHM0Z2FiU1VmcDJsNEdkSzAzUkh6dz09
Participants: SIGN IN: Name & Affiliation	Mano Maurya (MW, DRC); ; ;Shankar Subramaniam (DRC, MW), Swathi Thaker (ICC-Admin) ; ; ; ; ; ; ; ; ; Matt Roth (ERCC); ; ; ; William Khan (CHOP); ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Rahi Navelkar (4DN) ;David Chen (ERCC) ;Natalie Vineyard (NIH) ; Jimmy Zhen (MoTrPAC); Heesu Kim (LINCS, DRC); Anna Byrd (LINCS, DRC); ; ; ; ; ; ; ; Jeet Vora (GlyGen); ; ; ; ; ; Raja Mazumder (GlyGen); ;		

Agenda Item	Action Items	Owner
• **September 2024 C2M2 submission period** Open date: September 1, 2024 Close date: September 15, 2024		DRC
• Updated details related to C2M2: https://info.cfde.cloud/documentation/C2M2		DRC
• Post-translational modifications: PTM_Ontology_Requirements	Discuss >> Decide the ontology >>> RFC	GlyGen SPARC
• Biofluids ontology: Biofluids_Ontology_Requirements	Discuss >> Decide the ontology >>> RFC	ExRNA SPARC
• Ontology for gene-disease relationship: SIO DisGeNET ontology: https://www.disgenet.org/rdf Please see notes from KGWG on May 21, 2024 (https://docs.google.com/document/d/1WvpkLxWPW0XxZsam6jEJeEUQr2sQ0EWC/edit)		
• Data-set types/ data types, which can also capture the protocol/platform to generate the data		DD/KGWG: Deanne
•		
Slot for WG discussions at the fall meeting	• Discuss the metadata needs of Bridge2AI	Swathi

Notes:

Metadata submission:

- The general procedure to prepare the C2M2 files is exactly the same as before.
- Details on C2M2: https://info.cfde.cloud/documentation/C2M2
- Procedure for metadata submission: https://info.cfde.cloud/documentation/C2M2#datapackage-submission
- Submission portal: https://data.cfde.cloud/submit/form
- Resources:
    - JSON schema: https://osf.io/c63aw/
    - Ontology files and preparation script: https://osf.io/bq6k9/ : The file for provisional OBI terms may get updated before September 1 [Term Tracking].
- **DRC submission system helpdesk** slack channel: https://app.slack.com/client/TJ6R830D8/C0713HD4H2B

Expanding C2M2:
PTM:

- SPARC provided input; UniProt appears to be enough for their needs of cellular localization of the protein domains (e.g., intra/extra-cellular or transmembrane).

Biofluids:

- ERCC: David Chen: Current C2M2 doesn't allow specifying biofluids.
- Interlex: has record of other ontology IDs too, e.g., Uberon.
- Plan to add another field/column to the biosample table (or a closely related table) besides the existing 'anatomy' column.

Shankar: Bridge to AI data

- 4 types of data, including protein interaction, pharmacological perturbation data (may be compound treatment column may already take care of this), voice data (e.g., for lung infection)
- OWG will figure out what additional metadata elements and ontologies will be needed
- Already asked them to provide what type of metadata and ontologies they use internally.

Raja: Documentation

- Better to copy/move the ontology details (see https://docs.google.com/document/d/1MeUXtpKaNHITaVnUNeP62NIj29FXrBZH3q55jzaLmSM/edit and related documents) to the DRC/Information portal (perhaps on this page: https://info.cfde.cloud/documentation/C2M2).

July 17, 2024

Objective	see agenda	Time	Monthly on the third Wednesday of the month, 11am Eastern
Leader(s):	DRC team	Where	https://uab.zoom.us/j/88996931898?pwd=cHM0Z2FiU1VmcDJsNEdkSzAzUkh6dz09
Participants: SIGN IN: Name & Affiliation	; ; ; ; Mano Maurya (MW); ; ; ; ; ; ; ;Jeet Vora (GlyGen) ; ; ; ;Avi Ma'ayan (LINCS, DRC) ;Sherry Jenkins (LINCS, DRC) ;Anna Byrd (LINCS, DRC) ;John Erol Evangelista (LINCS, DRC) ; ;Daniel Clarke (LINCS, DRC) ; ; ; ;George Papanicolaou (NIH) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;Henning Hermjakob (EMBL-EBI / ICC-SC) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Matt Roth (ERCC); ; ;David Chen (ERCC) ; ; ; ; ; ; ; ; ;Christy Kano ;Natalie Vineyard (NIH, SPEC) Swathi Thaker (ICC-Admin); ; ; Suvarna Nadendla (HMP); ; ; ; Jimmy Zhen (MoTrPAC); ; ; ; ; ; ; ; ; ; ; Andy Schroeder (4DN) ; Srini Ramachandran (MW) ;		

Agenda Item	Action Items	Owner
• Post-translational modifications: PTM_Ontology_Requirements	Collect requirements > Discuss >> Decide the ontology >>> RFC	GlyGen SPARC
• Biofluids ontology: Biofluids_Ontology_Requirements	Collect requirements > Discuss >> Decide the ontology >>> RFC	ExRNA SPARC
• Ontology for gene-disease relationship: SIO DisGeNET ontology: https://www.disgenet.org/rdf Please see notes from KGWG on May 21, 2024 (https://docs.google.com/document/d/1WvpkLxWPW0XxZsam6jEJeEUQr2sQ0EWC/edit)		
• Data-set types/ data types, which can also capture the protocol/platform to generate the data		DD/KGWG: Deanne
• Semantic types		DD/KGWG: Deanne
Slot for WG discussions at the fall meeting		Swathi

Notes:

PTM:

- In the protein-scape: Both Uniprot and faldo needed to satisfy the needs of GlyGen; Input from SPARC needed.
- In the gene-space, another ontology might be needed: DisGeNET, Sequence Ontology, or HSCLO (genome-version(37/38):chromosome (indexed windows, e.g., 100kb, or

10kb):position/range:variable-resolution/hierarchy)? Deanne: will look at it further [Ben and Taha working on some aspects of it]. HSCLO used in the data-distillery (KG).

Biofluids:

- ERCC: David Chen: Current C2M2 doesn't allow specify biofluids.
- Interlex: has record of other ontology IDs too, e.g., Uberon.
- Plan to add another field/column to the biosample table (or a closely related table) besides the existing 'anatomy' column.

From the KGWG discussion

- Semantic types: July 16 discussion: https://docs.google.com/document/d/1WvpkLxWPW0XxZsam6jEJeEUQr2sQ0EWC/edit#heading=h.gjdgxs
- To do: What was the input data (type of ontological terms), what operation/transformation was carried out; seems like 'workflow'?
- Find all major data-types in a query.
- A new ontology: data-set types. C2m2.file.data_type may not be fine-grained enough, e.g., to capture specific protocols (chemistry) or platforms for single-cell experiments.
- Srini: Reach out to HubMAP, GlyGen and others such as MW who have protocol/platform information (e.g., mass spec types for MW).
- Deanne: Maybe start a document within DD/KGWG, refine it and later pass it on to the OWG. Useful to explore: https://www.ncbi.nlm.nih.gov/geo/summary/ .
- Suvvi: [in a different project] sub/specimen type: cell for single-cell, RNA for single-nuclei; needs to be explored further; find which existing c2m2 table/column it is related to if any.
- Capture assay/techniques (somewhat similar to process, but specific differences may exist) as well.

SIO DisGeNET Ontology: Notes from KGWG May 21 discussion

- https://www.disgenet.org/rdf
- More "location" oriented for sequence: http://www.sequenceontology.org/browser
- Could help with post-translational modification locations
- Has been (semi-)adopted for bulk processing for gene transfer format (gff - describes features on a sequence and a common bulk dump format from databases)
- Also a GA4GH working group thinking about standardizing certain file formats and controlled vocabularies within the columns of those files
- Could also look at linking the Sequence Ontology and HSCLO Ontology [https://github.com/TaylorResearchLab , https://www.biorxiv.org/content/10.1101/2024.02.15.580505v1]

May 15, 2024

Objective	see agenda	Time	Monthly on the third Wednesday of the month, 11am Eastern
Leader(s):	DRC team	Where	https://uab.zoom.us/j/88996931898?pwd=cHM0Z2FiU1VmcDJsNEdkSzAzUkh6dz09
Participants: SIGN IN: Name & Affiliation	; Peipei Ping (ICC-SC/UCLA); Wei Wang (ICC-SC/UCLA); Henning Hermjakob (ICC-SC/UCLA/EBI) ; Dean Wang (PM, ICC-SC/UCLA) ; ; ; ; ; ; ; ; ;John Erol Evangelista (LINCS/DRC) ;Heesu Kim (LINCS/DRC); Sherry Jenkins (LINCS/DRC); Sherry Xie (LINCS/DRC); Daniel Clarke (LINCS/DRC) ;Srini Ramachandran (MW/DRC) ; Mano Maurya (DRC/MW); ; ; ; ; ; ; Stephanie Olaiya (LINCS/DRC); Rahi Navelkar (4DN); ; ; ; Bernard de Bono (SPARC) ; ; ; ;Giacomo Marino (LINCS/DRC) ; ;Michelle Giglio ; ; ; ; ; ; ; ; ; Suvarna Nadendla (HMP); ; ;Andy Schroeder (4DN) ; ; ; ; Jimmy Zhen (MoTrPAC); ; ; ; David Chen (ERCC); ; ;Raja (GlyGen) ; ; ;Nasheath Ahmed (LINCS) ;Matt Roth (ERCC) ; ; ;Sean Davis ; ; ; ;Swathi Thaker (ICC): : ; Avi Ma'ayan (LINCS/DRC) : ;Jeet Vora (GlyGen); Jake Chen (ICC)		

Agenda Item	Action Items	Owner
• **June 2024 C2M2 submission period** Open date: June 1, 2024 Close date: June 15, 2024		DRC
• Updated details related to C2M2: https://info.cfde.cloud/documentation/C2M2		
• Adding support for post translational modifications (PTMs) to the C2M2		GlyGen, SPARC
• Adding DRS support to C2M2 assets		LINCS/DRC

- ○ https://www.disgenet.org/rdf
- ○ Mano/Srini to look into its utility for Ontology working group
- ○

Notes:

Metadata submission:

- The general procedure to prepare the C2M2 files is exactly the same as before.
- Details on C2M2: https://info.cfde.cloud/documentation/C2M2
- Procedure for metadata submission: https://info.cfde.cloud/documentation/C2M2#datapackage-submission
- Submission portal: https://data.cfde.cloud/submit/form
- Resources:
  - JSON schema: https://osf.io/c63aw/
  - Ontology files and preparation script: https://osf.io/bq6k9/
- **DRC submission system helpdesk** slack channel: https://app.slack.com/client/TJ6R830D8/C0713HD4H2B
- **access_url** column in file table (and possibly others): handled internally based on information in the persistent_id column.
- Daniel: DRS support: related to the above (**access_url** and persisent_id). access_url will be in DRS/https compatible format. Andy (4DN): great idea. This will help access/download/fetch the actual files,e.g., into a workspace.

Extending C2M2:

- GlyGen: post-translational modification, e.g., glycosylation, phosphorylation, and others
  - The site concept needs to be added to the C2M2 model.
  - Choice of ontology: in the process of brain-storming by the GlyGen team.
  - For protein: UNIPROT; site is an accession. Is this enough or do we need finer granularity?
  - BdB: range of substring not just site; for SPARC: for neuroscience, location of PTM for plasma membrane proteins; part of the substring is intracellular and part is extracellular.
  - Site e.g., 51, can be written as 51-51 in the range framework.
  - Criteria for selecting an ontology.
  - MG: make proteins a core entity as opposed to CV.
  - Henning: lessons from mass-spec PTM-related data submission. PSI-MOD may not be active.
  - Action plan: Collect requirements from the concerned DCCs (GlyGen, SPARC, others), Discuss and Choose an Ontology that satisfies most needs (Raja and team)
- George: How to prioritize requests for change in C2M2 to accommodate additional or change in metadata from different DCCs
  - Should it depend on how many DCCs request it.
  - MG: In the past it was first-come-first-serve basis.
- David (ERCC) and BdB (SPARC): biofluids related metadata/ontologies; how to represent absorption.
  - Any potential ontologies: Raja: https://ontology.buffalo.edu/smith/articles/BFLO_ICBO_2011.pdf
  - Is Uberon enough: Shankar: https://www.ebi.ac.uk/ols4/ontologies/uberon/classes/http%253A%252F%252Fpurl.obolibrary.org%252Fobo%252FUBERON_0006314?lang=en
  - InterLex for post-coordination of terms: Michelle
  - SNOMED CT: David & Srini

- Chris Nemarich: frictionless validator no longer maintained?
    - Daniel: We can fork it and maintain it. It is self-validating as well. Can develop CLI to submit to the new portal.
    - Avi: If something doesn’t work during the submission, we will still handle it, though it can take some additional time.
    - Avi: FAIR assessment: in plans; not for June 2024 submission.
    - Chris to contact DRC team to discuss what issues their team faced.
- MG: Utility of InterLex for intermediate and combinatorial terms

For PTM we can do the following.

- Use UniProtKB accession and position (single position or range).
- Use PSI-MOD (https://bioportal.bioontology.org/ontologies/PSIMOD). Might not be actively maintained. Others: FALDO, Glycoconjugate Ontology, Sequence Ontology
- Use GlycoCoO and Protein Ontology (PRO) (https://bioportal.bioontology.org/ontologies/PR) to connect protein site with Glycan.

OWG google folder:
https://drive.google.com/drive/folders/1DTEgsecHc1RfzZMUkFEkOmnpVIVEhnqx

Charter:
https://docs.google.com/document/d/1Pa1imd3wUIsmd03qCkf0-zzPt4kvJe2AAN9veN5XdhU/edit?usp=sharing

RFCs:
https://drive.google.com/drive/folders/1BXI4DvygDyQyuUckiuVMnn0eP0Kv4pVN

Start Here Document
https://docs.google.com/document/d/1MeUXtpKaNHITaVnUNeP62NIj29FXrBZH3q55jzaLmSM/edit?usp=sharing

November 15, 2023

Objective	see agenda	Time	Monthly, on the third Wednesday of the month, 11 am Eastern
wLeader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09

Participants: SIGN IN: Name & Affiliation	;;Bob Carter (CFDE-CC); ; Mano Maurya (MW); ; Suvarna Nadendla (CFDE-CC); ; ; ; ; ; Jeremy Yang (IDG); ; ; ; ; ; ; ; ; ; ; ; ;David Chen (ERCC); ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Jimmy Zhen (MoTrPAC); ; ; ; ; ; ; ; ; ; ; ; ; ; ;;; ; ; ; ; ; ; ; ; ; ; ; ; ;Jonathan Silverstein (HuBMAP/SenNet) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;John Erol Evangelista (LINCS) ;

Agenda Item	Action Items	Owner
● **December C2M2 submissions**		Bob
● anything else anyone wants to announce or discuss	●	
Canceling OWG call on Dec 20th. The next call will be on Jan 17th, 2024.		

Notes:

- Slims that are used in CFDE are posted here: https://github.com/nih-cfde/cfde-deriva/tree/master/cfde_deriva/configs/portal_prep/
  assay_type_slim.tsv
  data_type_slim.tsv
  disease_slim.tsv
  file_format_slim.tsv
  ncbi_taxonomy_slim.tsv.gz

- Canceling Dec 20th OWG call. See you next year.

October 2023 meeting canceled due to CFDE PI/PM meeting

September 20, 2023

Objective	see agenda	Time	Monthly on the third Wednesday of the month, 11am Eastern

wLeader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;; ;Suvarna Nadendla (CFDE-CC); ; Michelle Giglio; ; ; ; ;George ; ; ; ; ; ; ; ; Jeremy Yang (IDG); ; ; ; ; ; ; ; ; ; ; Sherry Xie (LINCS); Daniel Clarke (LINCS); ; ; ; ; John Erol Evangelista (LINCS); ;Chris Nemarich (KF); William Khan (KF) ; ; ; Jimmy Zhen (MoTrPAC); ; ; ; ; ;Srini Ramachandran (MW) ; ; Mano Maurya (MW); ; ; ; ; ; ; ; ; ; ; ; ; ; David Chen (exRNA); ; ; ; ; ; ; ; ; ; ; ; ; Jonathan Silverstein (HuBMAP/SenNet); ; ; ; ;Matt Roth (exRNA) ; ; ; ; ; ; ; ; ; ; ;		

Agenda Item	Action Items	Owner
● **check-in on C2M2 Sept submissions**		
● **C2M2/KG integration work as part of OWG mtgs**	●	

Notes:

- Next C2M2 submission is Dec 15 but a possibility of moving it to Jan 15 after discussing with CFDE-CC.
- C2M2/KG - Working Doc
- Jeremy Yang: How governance should take place with OWG, KG, Data Distillery, and few more groups? Should some of the governance details be included in RFCs.
- Only clinical data is not dealt with in OWG all other main goals have been accomplished.
- General consensus: Keeping the groups separate and not merging them as each group has its own complexity and depth and have their own uses.More discussion will happen in CFDE PI/PM meeting next month.
- exRNA (David Chen) - uses biofluids, is there a possibility of integrating that into C2M2? Michelle will touch base with David for discussing the possibilities.

August 2023 meeting canceled.

July 19, 2023

Objective	see agenda	Time	Monthly on the third Wednesday of the month, 11am Eastern
wLeader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	; Mano Maurya (MW);Suvarna Nadendla (CFDE-CC) ; Michelle Giglio; ; ; ; ; Jeremy Yang (IDG); ; ; ; ; ; ;John Erol Evangelista (LINCS) ; Deanne Taylor (CHOP/KF); ; ; ; ; Jimmy Zhen (MoTrPAC); ; ; ; ; ; ; William Khan (KF); ; ; Taha M. Ahooyi (KF); ; ; ; ; ; ; Sherry Xie (LINCS); ; ; ; ; ; ; ;Jeffrey Grethe (SPARC) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;;Srini Ramachandran (MW) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;Christy Kano ; ; ; ; ; ; ; ; ; ; ; ; ; David Chen(ERCC-exRNA); ; Keyang Yu (ERCC); ; Matt Roth (ERCC); ;		

Agenda Item	Action Items	Owner
• **Update on KG/C2M2 connections**		
•	•	

Notes:

Overview of current state of work on KG/C2M2 connections

Next steps
-json in's and out's of what we need
-will allow parallelization of the work
-Distillery has some resources, CFDC-CC has some resources, will see

Question on storage of evidence of how got from node-edge-node.
Yes,it would be good to be able to do that. but don't have that info in the resources now - as they are summarized and agreggated - so not on the immediate list.
First will do the more straightforward cases and hopefully move on to this more complex application later.

Discussion of scope and focus of Ontology WG - will evolve as we enter this 4th year.

June, 2023 - meeting canceled

May 17, 2023

Objective	see agenda	Time	Monthly on the third Wednesday of the month, 11am Eastern
wLeader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;; ; ; ; ; ; ; ; ; ; ; ;John Erol Evangelista (LINCS) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Jeremy Yang (IDG); Srini Ramachandran (MW); Mano Maurya (MW); William Khan(KF); ;Christy Kano ; ; ; ; ; ;; ; ; Michelle Giglio ;; ; ; ; ; ; ; ; ; George Papanicolaou NIH; ; ;		

Agenda Item	Action Items	Owner
• HPO slim update		
• Reference file updates for the June submissions	•	
• Timeline for June submissions		

__Notes:__

HPO slim update:
Srini took the lead on this work. He contacted HPO to see if they had any existing relevant slims. They didn't have anything that met our needs. So, Srini extracted the first level children of 'phenotypic abnormality' and we decided to use them as the slim set. Suvvi will use the ROBOT tool to map all of the HPO terms to their corresponding slim term.

New file updates:

- Disease Ontology

- Human Phenotype Ontology
- Mammalian Phenotype
- Uberon
- UniProt is in progress

Updated slim mappings to come (but are not needed for DCCs to do submissions.)

Submission schedule:
Submission window will open no later than June 1, possibly earlier.
June 15th is our deadline for DCC submissions.

April 19, 2023

Objective	see agenda	Time	Monthly on the third Wednesday of the month, 11am Eastern
wLeader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;;Suvarna Nadendla (HMPand CFDE-CC) ; ; ; ; ;; ; ; Jeremy Yang (IDG); ; ; ; ;; ;; ;Srini Ramachandran (MW) ; Mano Maurya (MW); ; Chris Kinsinger (NIH); ; ; ; ; Sherry Xie (LINCS); Daniel Clarke (LINCS); John Erol Evangelista (LINCS); ; ; ; ; ; ; ; Dan Lyman (GlyGen); ; ; ; ; ; ; ; ; ; Taha M. Ahooyi (KF); William Khan (KF); Deanne Taylor (KF); ; ; ; ; ; David Chen (ERCC); ; ; Matt Roth (ERCC); ; ;Andy Schroeder (4DN) ; Julia Markowski(4DN) ; ; ; ; ; Jimmy Zhen (MoTrPAC); ; ;Bob Carter (CFDE-CC) ; ;		

Agenda Item	Action Items	Owner
● Report from Bethesda mtg breakout		
● Slimming HPO and MP	●	
● GlyTouCan and PubChem		
● submission file updates for Anatomy/Disease		
● timeline for June submissions		
● New Portal release coming		

Notes:

- Report from Bethesda mtg: Breakout presentation available here. Good conversion on semantic harmonization,went back to Tim Berners-Lee and Semantic Web Tech and it's role in KG. Discussions on Clinical metadata (OMOP and UMLS integration).
- HPO and MP: For June submissions, submitters can start using either HPO and MP terms in phenotype field. Chris Mungall, Melissa Haendel, Peter Robinson might already be involved in MP and HPO harmonizing efforts (mapped). We will be slimming HPO and MP terms separately. Will not try to merge now, but will explore that later looking at work of upheno and Cynthia Smith from KF who are doing mappings.
  Srini Ramachandran volunteered to help with slimming of HPO and MP.
  Jeremy suggested use of auto slimming tools. Deanne thinks that there is a slim already.
- GlyTouCan and PubChem: Right now, the compound field accepts PubChem terms and those GlyTouCan terms that are not already represented in PubChem. We are considering treating these the same as we do HPO and MP - wanted to give the group a heads-up on that and we will finalize plan after we have some time to think about it and play around with test data.
- Anatomy and Disease updates: We will be updating disease and anatomy reference files will be made by Mid May. We will also update the slims before that deadline.
- June submissions: Deadline for submissions is June 15th. All the updates to ontologies and slims will happen by mid-May. DCCs can start submitting as of June 1 or possibly earlier - we will send an announcement when submissions can start. It will be no later than June 1.
- Portal graph changes: The main data graph on the portal home page will not be a three dimensional graph anymore after the portal release due soon, instead it will just be a two dimensional graph with X and Y axis.
- **Next meeting is May 17th**

March 8, 2023

Objective	see agenda	Time	Monthly on the third Wednesday of the month, 11am Eastern
Leader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;;Suvarna Nadendla (HMP); Mano Maurya (MW); ; ; ; ; Sherry Xie (LINCS); Daniel Clarke (LINCS); ;John Erol Evangelista (LINCS) ; ; ; ;Daniel Lyman (GlyGen) ;Srini Ramachandran (MW) ;Taha M. Ahooyi (KF) ;William Khan (KF) ; Chris Nemarich (KF); Leslie Duffy (KF); ; ;David Chen (ERCC) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Matt Roth (ERCC); ; Andy Schroeder (4DN); ; ; ; ; Michelle Giglio; ;

Agenda Item	Action Items	Owner
• **selection of which Wednesday to meet**		
• **update on KG/C2M2 translation layer work**		
• **use of both HPO and MP**	•	
• **Additional slim work**		

Notes:

- 20 attendees
- Selection of which Wednesday to meet : Shifting the monthly meeting to the third Wednesday. The next meeting will be April 19th, 2023.
- Update on KG/C2M2 translation layer work: A group of 6 people met for the first time a few days back. Meeting notes here - C2M2/KG layering/translation
- use of both HPO and MP: We propose the use of both HPO and MP in C2M2. We will not make an additional field for MP but will allow use of MP terms in the same field as HP terms for phenotypes. Having more than two kinds of ids in the same field is not ideal. But it is an easy way to use terms from both the ontologies. This change will be in place for June submission but not for March submission.
- Additional slim work: Phenotypes are something that can be slimmed.

  NOTE: Shifting the monthly meeting to the third Wednesday. The next meeting will be April 19th, 2023.

February 8, 2023

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;; Mano Maurya (MW); ; ; ; Sherry Xie (LINCS); ; Jeremy Yang (IDG); ; ; ; ; ; ; ; ; Matt Roth (ERCC); ; ;Keyang Yu (ERCC) ; ; ;David Chen (ERCC) ; Suvarna Nadendla (HMP); ; ; ; ; ; ; ; ; ; ;Rahi Navelkar (4DN) ; ; ; Srini Ramachandran (MW); ; ; ; Dan Lyman (GlyGen); ; ; ; Andy Schroeder (4DN) ; ;William Khan (KF) ; Daniel Clarke (LINCS) ;John Erol Evangelista (LINCS) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;		

Agenda Item	Action Items	Owner
● Announcements/Updates ○ Asiyah slacked the group about proposed changes to the federal register for terms on race ○ work on layered knowledge graph/C2M2 idea discussed in KGWG		
● Last chance to comment on the draft documentation		
● Discussion of **essential** C2M2 additions/expansions - how much can be handled with the KG/C2M2 layers effort?		

Notes:

- Mano (Metabolomics) : Need for Cell Ontology and Cell Line Ontology terms. Cell Ontology terms link into anatomy ontology.
- Will Mammalian Phenotype Ontology be more useful- in C2M2 instead of HPO?
  GlyGen -Viruses, Drosophila
  IDG - Mouse
  4DN - Fly, Zebrafish, mouse
  Kids First - Mammalian
  No objections to shifting to MPO - human info would map to less granular terms but could still be at the granular level in the KG layer. Need to think about way to link in organisms outside of mammals.
- May have to focus on clinical metadata in future (ICD)
- KG/C2M2 layers subgroup will work on requirements for this system (Deanne Taylor, Jonathan Silverstein, Jeremy Yang, Michelle Giglio) - Others are welcome - contact Michelle if you'd like to be part of that
- People's thoughts on having this call monthly instead of biweekly - No objections, so it will be on the **second Wednesday monthly** at 11AM (EST) moving forward.
  Next call March 8th
  Jeremy Walter will cancel old invite and send a new invite for the monthly calls. We will encourage the use of slack for between-meeting discussions/work.

January 25, 2023 - No Meeting

January 11, 2023

Objective	see agenda	Time	Every other Wednesday 11am Eastern

wLeader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;; ; Mano Maurya (MW); ;Suvarna Nadendla (HMP) ; ; ; ; ; ; ; George Papanicolaou (NIH); ;Srini Ramachandran (MW) ; Chris Nemarich (KF); Deanne Taylor (KF); ; ; ; ; ; ; Mark Musen (HuBMAP); Vince Metzger (IDG) ; ; ; Daniel Clarke (LINCS); ; ; ; ; ; ; ; ; ; ; ;; ; ; ; ; ; ; ; ; ; ; ; Sherry Xie (LINCS); ; ; ; ; ; Jeremy Yang (IDG); ; ; ; ; ; ; ; ; ; Dan Lyman (GltGen); ; ; ; ; Taha M. Ahooyi (KF); ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;Matt Roth (ERCC) ; ; ; ;		

Agenda Item	Action Items	Owner
● **OWG Documentation draft** **Start Here - a guide to Ontology WG products**		
● **Slim RFC (still) in comment phase**		
● **Revisit use of Human Chromosomal Location Ontology - these *can* be linked to genes**		
● **Further discussion of source/OWG/UMLS layers idea raised before the holidays**	●	

Notes:

- OWG Documentation draft : A guide to OWG products as well as documentation on ontologies and RFCs. Please provide comments and suggestions.

- Building Ontology Slims RFC is out for comments. If you have any comments please add them to the doc.

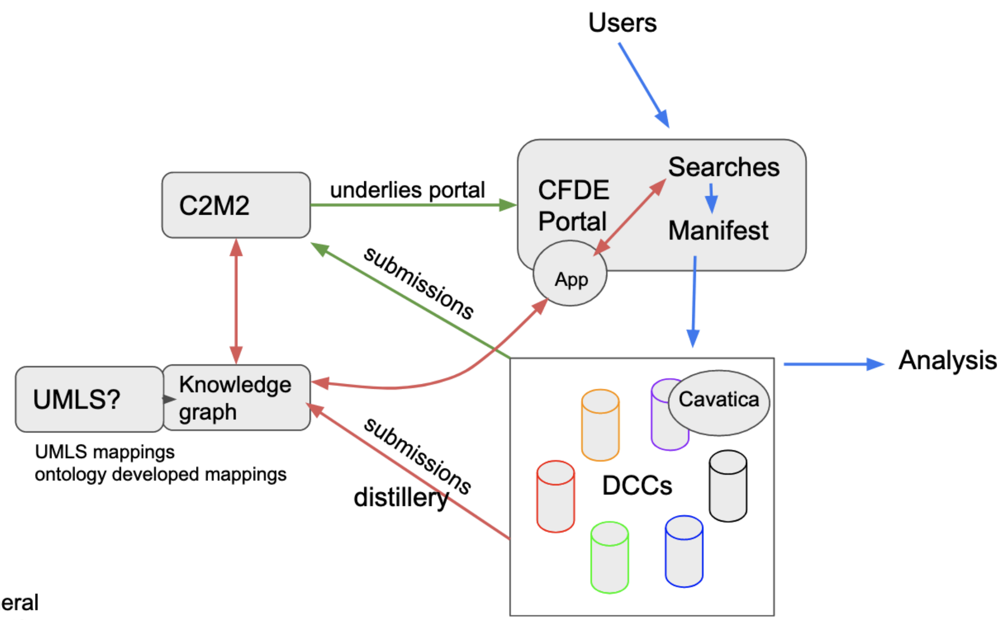
- Human Chromosomal Location Ontology - these *can* be linked to genes via association tables. Helpful for regional search of a chromosome. Need to think about this possible expansion of C2M2 in light of other C2M2 expansion areas. Other areas that have been wished for in the past are: variants attached to genes, expansion of disease/phenotype/symptom linkages, expansion of chemical entities.
  Homework: everyone think of things that they feel are essential to have in C2M2 to achieve project goals - keeping in mind that C2M2 can't capture everything and shouldn't try to capture everything. It's job is to be a guidepost. Once we have the full list, then we can prioritize.

- Continuation of discussion that started in the December meeting of the idea of structuring things around source/OWG/UMLS layers of information.
  The Knowledge Graph effort will be capturing information coming in from the Distillery project that is not, and will not, be captured in C2M2. Also, the Knowledge Graph effort will allow linkages to vocabularies well beyond those chosen for harmonization within C2M2. Question for us is how to make use of the KG info within the context of the Portal/C2M2.
  -First proposed use case that was explored: translation between vocabularies.
  An API could be built that would include within it all of the C2M2 associations as well as all of the KG associations where associations are linked to each other through UMLS or ontology-provided mappings (which might be equivalence or other types of mappings).
  If a user would like to search at the Portal with a term that is not part of a vocabulary used in C2M2, the App would be able to look up what C2M2 terms are linked to the query term via the mappings/linkages in the KG

and return this information to the user. The App could also provide some “roll up” or “roll down” abilities to provide more general or more specific information as relevant.

Subgroup of volunteers to spec this out: Michelle, Jonathan, Jeremy, Deanne
Homework: Michelle will contact subgroup.

Outline of idea - (**NOT** a strict data flow diagram, more of an interaction diagram)
-green is current path of information entering the C2M2 and being used in the portal
-blue is current user path for searching portal and use of results
-red is proposed new connections for the proposed App that would allow translation



use cases
specific to general
general to specific
translation - allow entry by any vocab - precalc/preingest or transit graph with back-and-forth API calls in an app attached to the portal; can do equivalence or limited roll ups

December 14, 2022

Objective	see agenda	Time	Every other Wednesday 11am Eastern

wLeader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;; Suvarna Nadendla (HMP); Mano Maurya (MW); ; ; ; ; ; ; ; ; ;John Erol Evangelista (LINCS) ; ; ; ; ; ; ; ; ; Srini Ramachandran (MW); ; ;Tony Kirilusha ; ; ; ; ; Sherry Xie (LINCS); Daniel Clarke (LINCS); ; ; ; ; ; ; ; ; ; ; ; ; ; Dan Lyman/GlyGen; ; ; ; ; ; ; Chris Kinsinger; ; ; ; ; ; ; ; ;Jeremy Yang (IDG) ; ; ; ; ; ; ; ;Taha Mohseni Ayooyi (KF) ; ; ; ; ; ; ;Bob Carter ; ; ; ; ; ; ; ; ; ; ;Matt Roth (ERC) ; ; Keyang Yu (ERCC); Jimmy Zhen (MoTrPAC); ;		

Agenda Item	Action Items	Owner
1. Check on status of using the new fields and CVs 2. Touch base on chemical terminology subgroup 3. Phenotype needs – HPO, EFO 4. Use of non-OWG ontologies in other CFDE contexts – is there anything we can or should do to help with harmonization. Leads into #5… 5. Discussion of possible need for OWG to provide mapping files between ontologies 6. Priorities for the new year: -Develop guidelines for new DCCs -Consistency checks -PubChem slim -Mapping files (maybe) -Phenotype terminology work? -Other things???		

Notes:

Review of C2M2 changes:

- Biosample.assay_type is removed but instead you have a new field - biosample.sample_prep_method.
- Sex: “transexual (M to F)” and “(F to M)” is removed. Neither of these have been used by DCC so no harm removing them. Sex will be furthur explored keeping in mind the federal standards.
- Race : c2m2 has an outdated list.
    - The current “Asian or Pacific Islander” will be split into two as “Asian” and “Native Hawaiian or Pacific Islander”.
    - “American Indian or Alaskan Native” will have a name change to “American Indian or Alaska Native”
    - “Black” will be changed to “Black or African American”.
    - “White” will have no change

Chemical Terminology: GlyGen has some chemical entities which were not in PubChem. We have added GlyTouCan terms alongside PubChem for GlyGen.Similar case, proposal is to add RefMet terms along with PubChem in C2M2.
Phenotype needs : Srini suggests/prefers SNOMED for phenotypes. Deanne - One should use two HPO terms to describe laterality (left, right, proximal, distal etc), same is true with UBERON. Eg: Right hand. SNOMED is a one-stop shop for clinical terms - radiology, diagnosis, procedure, observations etc. Jeremy - collect the ontologies and terms used by DCCs and integrate in CFDE. But this will make searches difficult in CFDE. Jeremy also suggests using CUIs from UMLS.
Srini suggests use of HPO/SNOMED as sources and map to chosen set of ontologies and in the background will be mapped to CUIs from UMLS. This led to the below discussion.

Discussion of granular source term assignments and what should be part of C2M2: Proposal for new layered system
We revisited the discussion of what should be the scope/mandate for C2M2.
Discussion coalesced around an idea that would have three layers of metadata information:
-Layer 1 - Source annotations: source entries that would require ontology:term structure for entries. DCCs could submit as much as they want as long as they conform to the format. everything would be taken and could be linked to any of the entities.
-Layer 2 - OWG standard ontology annotations: source annotations would be mapped (as needed) to the selected OWG ontologies (DO, HPO, Uberon, etc). This layer would form the basis for data summary views in portal. This layer would be searchable.
-Layer 3 - UMLS CUIs (concept unique identifiers) - work done by the KG and Data Distillery would provide mappings from the source or OWG standard ontologies to UMLS CUIs.
<u>Questions/issues:</u>
-would we want searches on the portal to be against UMLS too?
-I'm assuming we wouldn't be doing searches on the portal against the source layer - but is that what everyone else thought?
-how would we provide the source info to users? Maybe when people use the OWG standard ontologies (layer 2) to get search results, we then provide an optional download of the source metadata annotations? Thoughts?

Priorities for the new year:

-Further discuss the above layers proposal and reach a decision/plan
-Develop guidelines for new DCCs, put all OWG info onto a web page for people to easily access
-Consistency checks
-PubChem slim
-Chemical and Phenotype terminology work

Human Chromosomal Location Ontology
Taha (KF) - New Ontology - Human Chromosomal Location Ontology, a custom ontology used to capture the physical location of entities on human chromosomes at different resolutions. This

ontology might be useful for data generated by ATAC-seq, Hi-C assays etc. Suggest adding this ontology to CFDE. Genes, regulatory elements are attached to an ontology term.
Michelle: No provision in current C2M2 to attach such ontology terms to genes. Genes would need to become entities.

The next OWG call is on Jan 11, 2023

November 16th and November 30th meetings were canceled.

November 2, 2022

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio Suvvi Nadendla	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	; Mano Maurya (MW); Suvarna Nadendla (HMP); Daniel Clarke (LINCS); ; ; ; John Erol Evangelista (LINCS); ; Vince Metzger (IDG) ;Rahi Navelkar(4DN) ; Srini Ramachandran (MW); ; ; ;Michelle Giglio ; ; ; ; Taha M. Ahooyi (KF); Chris Nemarich (KF); ; ; ; ; ; ; ; ; ; ; ; ; Keyang Yu (ERCC); ; ; ; ; ; ; ;Matt Roth ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;Christy Kano (NIH) ;		

Agenda Item	Action Items	Owner
• updates on CVs for sex and race		
• timeline for data submission for next portal release		
• presentation by Eoin Fahy on RefMet	• tutorial on RefMet	

Notes:

24 attendees

C2M2 changes:

- Biosample.assay_type is removed but instead you have a new field - biosample.sample_prep_method.
- Sex: “transexual (M to F)” and “(F to M)” will be removed. Neither of these have been used by DCC so no harm removing them. Sex will be furthur explored keeping in mind the federal standards. Possible shift to field name of “sex for clinical use”.
  George offered to help in this area as we move forward.
- Race : c2m2 has an outdated list.
  - The current “Asian or Pacific Islander” will be split into two as “Asian” and “Native Hawaiian or Pacific Islander”.
  - “American Indian or Alaskan Native” will have a name change to “American Indian or Alaska Native”
  - “Black” will be changed to “Black or African American”.
  - “White” will have no change

  More than one race can be assigned to the subject in association tables.
- Timeline:
  - Dec 1st DCCs can start submitting
  - Jan 9th deadline for all data to be submitted
  - Jan 17th new data release in the portal

RefMet:
RefMet tutorial: tutorial on RefMet
Slides from today’s presentation:
https://drive.google.com/file/d/1WY3d93Sbi_mtJ1QQ7IZ_kuqYwtBq51qw/view?usp=sharing
RefMet contains metabolites at four levels of knowledge ranging from known full structure as the most specific to only class level. Mappings to PubChem and CheBI are made whenever possible.
Discussion:
We can combine mappings across PubChem, GlyTouCan, RefMet, and CheBI into a knowledge graph that can be used to translate between the vocabularies.
Could we create a merged vocabulary with a primary vocab and then additions of unique terms from the other vocabs?
We decided to form a subgroup to look at this and come up with a proposal. So far to include: Eoin, Mano, folks from GlyCan, Michelle, Suvvi, Jonathan, Jeremy Yang.
Michelle will reach out to get the ball rolling.

Note: the Nov 16 Meeting is Cancelled - Michelle will be away on vacation and no others volunteered to run the meeting

October 19, 2022

Meeting canceled (Michelle was teaching)

October 5, 2022

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio Suvvi Nadendla	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;; Suvarna Nadendla (HMP); ; ; ; ; ; George Papanicolaou; Tony Kirilusha (NIH) ; ; ; ; ; ; Andy Schroeder (4DN); ; ; Michelle Giglio; ; ; ; ; ; ; ; Jeremy Yang (IDG) ; ; ; ; ; ; ; ; ;; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Taha M. Ahooyi (KF); ; ; ; ; ; ; ;Jeffrey Grethe (SPARC) ; ; ; ; Daniel Clarke (LINCS); John Erol Evangelista (LINCS); ; ; ;Eric Wenger (KF) ; ; Chris Nemarich (KF); ; ; ;Keyang Yu (ERCC) ; ; Mano Maurya (MW); ; ; ; ; ; ; ; Sherry Xie (LINCS); ; ;Dan Lyman (GlyGen) ;Rahi Navelkar (4DN) ; ; ;		

Agenda Item	Action Items	Owner
● updates to C2M2 biosample table		
● OBI ‘planned process’ branch		
● assay_type vs analysis_type in file table		
● use of InterLex terms	●	

Notes:

As of submissions starting December 1st:

1. Previously, there was a field in the biosample table called assay_type. Its name and definition were incongruous with each other. Therefore, the values people put in the field were inconsistent across DCCs and a bit of a mess. We are keeping the field but renaming it to be consistent with its definition. Therefore it is now called “sample_prep_method”. Terms here should describe the process by which the biosample was generated.
2. GlyGen’s new ‘synaptosome preparation process’ term, OBI:0003384, will go in the “sample_prep_method” field. Terms for populating this field should come from the ‘planned process’ branch of OBI, with the exclusion of things from the ‘assay’ branch.

Assays result in data, not in biosamples (as defined by OBI). Some nodes of interest in 'planned process' are: 'material processing', 'specimen collection process', and 'material component separation'.

3. assay_type vs. analysis_type in the file table:
   assay_type is used to capture the assay that was done on the biosample that led to the creation of the file in question. So RNA-seq might be an assay that led to a fastq file of sequence. When entering info into the file table for the fastq file, you would put OBI:0001271 ('RNA-seq assay') into the assay_type field in the file table.
   analysis_type is used to capture the kind of analysis that was done on a file to create the file in question. So sequence alignment might be the analysis type that leads to a BAM file. When entering info into the file table for the BAM file, you could put OBI:0002567 ('sequence alignment') into the analysis_type field in the file table. Terms for populating the "analysis_type" field could come from the "data transformation" node in OBI.
   (see discussion notes below for additional info)
4. Use of InterLex terms in submissions. InterLex terms can now be used in submissions. They are to be used for data_type and file_format fields only (as of now, since those are the only two types of InterLex terms we currently have).

You can start making new submission files now, but new files with these changes can not actually be submitted until December 1st. (It may end up being sooner than Dec. 1, that is a "worst case" date. Michelle will let you know if submissions can start earlier.)

Notes on discussion of above items:

There was discussion of using both assay_type and analysis_type for the same file. Michelle had said that it should usually be one or the other but not both. Andrew commented that it might be good to use both to indicate the experiment that a derived analysis file was linked to. If there was a "file derived_from file" linking table that could solve that issue. But there currently isn't. In that absence using both fields makes sense. Michelle will follow up on this issue with the C2M2 team.

There was also a suggestion for adding links to the relevant areas of OBI to the ingest documentation.

Suggestions for future meeting topics/presentations

- Mano's suggesting a presentation on use of RefMet terms/ids for metabolites names. Currently PubChem doesn't have a lot of the metabolites that Metabolomics Workbench needs. And getting terms for them in PubChem is difficult because they require a structure. Mano wil email Michelle to set up a time for that presentation.
- Assay might be defined differently in chemical studies than perhaps what is in OBI, future topic is to explore the needs of CFDE in the context of definitions of assay in other ontologies.
  Update to this post meeting: Michelle checked again on the definition of 'assay' in OBI and it turns out that she misspoke - the definition is: "A planned process with the objective to produce information about the material entity that is the evaluant, by physically examining it or its proxies." So, an assay is something that produces information about a material entity - that material entity could be a biosample or something else. Michelle will follow up with Jeremy Yang to discuss this.

September 7, 2022

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio Suvvi Nadendla	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;;Suvarna Nadendla (HMP) ; ; Sherry Xie (LINCS); ; Michelle Giglio; ; ; ; ; Christophe Lambert (IDG) ; Jeremy Yang (IDG); Vincent Metzger (IDG) ; Srinivasan Ramachandran (MW); ; ; ; ; ; ; ; ; Owen; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Haluk Resat; ;; ; ; ; ; ; ; ; ; ; ; Chris Kinsinger; Daniel Clarke (LINCS); John Erol Evangelista (LINCS); ; Mano Maurya (MW); ; ; ; ;Taha M. Ahooyi (KF) ; ; ; Daniel Lyman (GltGen); ; ; ; ; ; ; ; ; ; ; ;		

Agenda Item	Action Items	Owner
• Discussion of Bioregistry paper - led by Jeremy Yang		
•	•	

<u>Notes:</u> (20 attendees)

Discussion of Bioregistry paper:
https://www.biorxiv.org/content/10.1101/2022.07.08.499378v2.full.pdf

Looking at this paper in the context of the scope of the OWG as we move into this next stage of the CFDE.

Possible relevance to the workflow partnership project group, a CFDE partnership.
Relevance to Knowledge Graph working group.

Raja - Way to take CURIEs and distill info for use in Distillery project and others?
Jeremy - phenotypes will be complex and Bioregistry can help with this. Also with disease.
Perhaps we could align with Bioregistry and make c2m2 use it as reference.
What was proposed with use of PubChem.

Questions on interpretation of what they mean by cross-registry mappings. Any time two registries were linked at all? Not sure.

Several DCCs are in the registry - MW, 4DN, LINCS - may be more…

Possible use of Bioregistry to do the harmonization challenges that we would face with phenotype, clinical info, etc.

Jeremy will report back on more technical details.

There will be no meeting on Sept. 21 - next meeting in 4 weeks

---

August 24, 2022 - Meeting Canceled

Homework: read the Biorepository paper

August 10, 2022 - Meeting Canceled

July 27, 2022

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio Suvvi Nadendla	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;; ; Suvarna Nadendla (OMP); Srini Ramachandran (MW); ; ; ; ; ; ;Michelle Giglio ; ; ; ; ; ; ;; ; Mano Maurya (MW); ; ; ; ; ; ; ;John Erol Evangelista (LINCS) ; ; ; ; ; ; ; ; ; Jeremy Yang (IDG) ; ; ; ; Vincent Metzger (IDG); ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Taha M. Ahooyi (KF); ; ;Keyang Yu (ExRNA) ; ; ; ; ; ; ; ; ; ; ; ;		

Agenda Item		Action Items	Owner
• review of decisions from last meeting			
• ICBO submission		• https://docs.	

		google.com/document/d/10Sw9WGTj6RFF95yQwMWKxqpC5G6QOMFmWSkIvTGxblQ/edit?usp=sharing	
● status updates			

Notes:

Reminder of what we covered at our last meeting a month ago:

1. Phenotypes**:** No one on the call reported an immediate need for phenotype ontologies beyond HPO. Therefore, we will take more time to work through how best to expand this and into what areas. If you have phenotype information in-hand but can't currently submit it to C2M2 because HPO is the only ontology supported, let Michelle know.

   Notes from today: Jeremy (IDG): Needs terms from EFO to be able to capture gwas traits. We will follow up on this with later conversations.

2. Consistency Guidelines: Thanks to those who have already submitted guidelines. Everyone else, please submit your internal DCC consistency guidelines to the folder or if you don't have them, email Michelle to that effect.

   Update from today: waiting for new version of corrected term use report. Encouraged more DCCs to share internal guidelines

3. Use of GO Cellular Component terms in anatomy: No one on the call reported a need for using cellular level anatomy terms. Thus we will deprioritize backend changes in C2M2 and the submission system for this. Let Michelle know if a need for use of cellular anatomy terms arises.
   Update from today: no one expressed a need.

4. The use of InterLex to make provisional EDAM terms is underway: Let Michelle/Suvvi know if you need terms.
   Update from today: no one expressed a need.

New Stuff

ICBO paper

Michelle/Suvvi were invited to be part of a workshop at the upcoming International Conference on Biomedical Ontology (ICBO) this September. Their process is for presenters to submit a “paper” that will then become part of the Proceedings of the conference that will be posted online at CEUR.
ICBO Conference - https://icbo-conference.github.io/icbo2022/
ICBO workshop:
https://docs.google.com/document/d/1jPnvE3ZN66QQE_hiyHVX6Y5NlU9B9mBhmP2iZCgxfyA/edit?usp=sharing
ICBO OWG paper draft:
https://docs.google.com/document/d/10Sw9WGTj6RFF95yQwMWKxqpC5G6QOMFmWSkIvTGxblQ/edit?usp=sharing
Affiliation sheet:
https://docs.google.com/spreadsheets/d/1JzpVP6OHl91TkYHdq-QFkROYxrcKyGaKG953QQmOQtM/edit?usp=sharing
Please make edits/suggestions by noon Eastern on Friday.
PLEASE USE SUGGESTING MODE

NTR updates:
All of the 4DN OBI terms are now live in the ontology
few more terms are still pending

New item:
Jeremy (IDG): Suggested reviewing Bioregistry paper -
https://www.biorxiv.org/content/10.1101/2022.07.08.499378v2

July 13, 2022

No meeting

June 29, 2022

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio Suvvi Nadendla	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;;Suvarna Nadendla (HMP) ; ; George Papanicolaou; Sherry Xie (LINCS); ; ; ; ; Michelle Giglio; ; ; ; ; ; ; ; ;Srinivasan Ramachandran (MW) ;John Erol Evangelista (LINCS) ; Daniel Clarke (LINCS); ; ; ; ; ; ; Vincent Metzger (IDG); ; ; ; ; ; ; ; ; ; ; ;Dan Lyman (GlyGen) ; ; Keyang Yu (ExRNA); ; ;Matt Roth (ExRNA) ; ; ; ; ; ; ; Taha M. Ahooyi (KF); ; ; ; ; ; ; ; ; ; ; ; Haluk Resat; ; ; ;		

Agenda Item	Action Items	Owner
● **Homework Review** ○ **phenotype thoughts** ○ **guidelines collection**		
● **GO CC in anatomy**		
● **New Terms/Interlex**	●	
● **NO CALL on JULY 13th**	**next call is July 27th**	

<u>Notes:</u>

- Homework from last call:
  - Phenotype ontologies in CFDE - Having more than one phenotype ontology in CFDE is kind of difficult. Want to know from DCCs if they are facing any difficulty in selecting phenotype terms since CFDE has only HPO to use.
    Did not hear any issues from members attending the call.
  - Consistency guidelines: Here is the link to HMP consistency guidelines - https://docs.google.com/spreadsheets/d/135O-gxwg1NVInGLGI8n7_R6MCKIvNeoP/edit?usp=sharing&ouid=109292309009829585903&rtpof=true&sd=true
    LINCS and Metabolomics also submitted their guidelines in the folder.
    Michelle and Suvvi will go over the guidelines from HMP, LINCS and Metabolomics to see if there is consistency in term usage for similar data. We encourage the other DCCs to post similar files for their internal guidelines. If they don't have such files, please email Michelle.
    NOTE: the current version of the term use report appears to have some errors in it - Michelle/Suvvi are going to follow up on this to get a corrected report.
- GO CC in anatomy: This came up in the context of the "synaptosome" term for GlyGen. GO CC will be incorporated into CFDE for anatomy term usage in the next submission if any DCC requires cellular level anatomy terms immediately.
  Did not hear any immediate need for GO CC terms, so we will be deprioritizing its incorporation into CFDE for now.
- New Term Request tracking - https://docs.google.com/spreadsheets/d/1zoHaSpJ4W5q_tVtiWMdLdMr9REHmRZPWb1-WmAKIquk/edit?usp=sharing
  This consists of new OBI terms requested by DCCs that have been already released by OBI and provisional ids for terms that are not released by OBI yet. It also contains the proposed EDAM terms that we will create in InterLex pending an EDAM release.
  Jeff made a community area in InterLex for CFDE. Michelle has made two terms in Interlex for adding new EDAM terms - https://scicrunch.org/CFDE/
  DCCs should review the EDAM terms in the new term tracking file (linked above) and let Michelle know if they see any issues they'd like to discuss.

Moving forward, DCCs will submit new EDAM terms to Michelle and Suvvi,we will work together to get them defined and identify appropriate parents. We will announce them to the OWG. Then Michelle/Suvvi will add them to Interlex and post the interlex ids to New Term Request Tracking google file. These ids can be used in future CFDE submissions.

NOTE: No Call on July 13th
Michelle will not be able to chair the call on July 13th, so no meeting that day. The next meeting will be on July 27th.

June 15, 2022

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio Suvarna Nadendla	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;; ; Suvarna Nadendla (HMP); ; ; ; ;Michelle Giglio ;Chris Kinsinger (NIH) ; ; ; ; ; ; ; ; ; ; ; ; ;; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;John Erol Evangelista (LINCS) ; Daniel Clarke (LINCS); ; ; ;Taha M. Ahooyi (KF) ; ; ; ; Sherry Xie (LINCS); ; ; ; ; ;Matt Roth (ERCC) ; ; ; ; ; ; ; ; ; Jeremy Yang (IDG); ; Vincent Metzger (IDG); Daniel Lyman (GlyGen); ; ; Mano Maurya (MW); ; ; ; ;		

Agenda Item	Action Items	Owner
- NTR updates	- if you need a term and haven't heard from us or aren't' in communication with us reach out to Michelle and Suvvi	
- sharing other harmonization efforts	Let Michelle know if you would like to present on internal or partnership harmonization work	
- Review of phenotype needs	- Think about your use of phenotype terms, be ready for future discussion	
- Guideline building	- Post your DCC internal rules/guidelines for term use to the folder linked below: https://drive.google.com/drive/folders/1sZuyEDw5pEWcYQjBZm6VzO-1WWGMk4a6?usp=sharing	

Notes:

NTR updates

- 4DN assay terms: Will be released in the upcoming OBI release scheduled soon.
- EDAM terms: Working with Jeff from SPARC to get EDAM terms into Interlex pipeline for provisional ids.
- GlyGen “synaptosome” term: We have a provisional id for the term. It will be included in OBI release after the upcoming one this month.

Phenotype Ontologies

- There is an overlap of concepts between these ontologies. Have to check with Arthur (C2M2 group) how it would affect the model if more than one ontology terms will be used for this metadata.

Based on the info people provided we would need:
-Human Phenotype Ontology
-Mammalian Phenotype Ontology
-Zebrafish Phenotype Ontology
-Ontology of Microbial Phenotypes

Questions:
-model phenotypes vs. human phenotypes they are modeling
-EFO ontology has relevant phenotype terms
-what is and isn’t a phenotype
-disease vs. phenotype vs. symptom
-keep what goes in the field kind of loose
-separate model phenotypes from human phenotypes?
this would be important for clinical directions
-don’t reinvent the wheel - work with what AGR has done, Monarch

Consistency Guidelines

A folder has been made in OWG space for DCCs to add any guidelines they have been following. Please put files with you guidelines here:
https://drive.google.com/drive/folders/1sZuyEDw5pEWcYQjBZm6VzO-1WWGMk4a6?usp=sharing

Useful reference for helping with consistency is the Term Use Report: (Ontology terms used by DCCs) - latest one is here (in the same folder as above):
https://docs.google.com/spreadsheets/d/1rlT69FtiBSj2Yi095F2oLigy_KEBICr0wpjIdccSzoU/edit?usp=sharing

May 18th, 2022

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio Suvarna Nadendla	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;;Suvarna Nadendla (HMP) ; ; ; ; ;Matt Roth ; Jared Nedzel (GTEx) ; Duyen Nguyen (GTEx) ; Vincent Metzger (IDG) ; ; ; ; ; ; Owen White; ; ; Sherry Xie (LINCS); Daniel Clarke (LINCS); Mano Maurya (MW); ; ; ; Chris Kinsinger (NIH); ; ; ; ; ; ; Asiyah Lin (NHGRI); ; ; ; ; ; ; ; ; ;Jeremy Yang (IDG); Dan Lyman (GlyGen); ; Andy Schroeder (4DN); ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Vincent Metzger (IDG) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;		

Agenda Item	Action Items	Owner
● update on biosample.assay_type field change		
● presentation from Keyang Yu on Variant WG plans	•	
● two presentations on harmonization of data between GTEx and KidsFirst from Duyen Nguyen and Meen Chul Kim		

Notes:

- ○ update on biosample.assay_type field change : We’ve run into an unexpected issue on timing. It turns out that the changes needed in the ingest and portal infrastructure are such that there isn’t time to officially change the field name for the next data/portal release in June. However, we suggest that you go ahead and use the biosample.assay_type field in the way that we intend to use the renamed biosample.sample_prep_method field.

Things for follow-up
-linking variant SEPIO capture of variant info to c2m2 in some way
-working with GTEx on use of uberon slim to help their query tool

May 4, 2022

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;; Suvarna Nadendla(HMP); ; ; ; George Papanicolaou; ; ; ; ; Owen White; Jeremy Yang (IDG); ; ; ; ; ; ; ; ; ;Matt Roth ; ; ; ; ; ; ; ; ; ; Daniel Clarke (LINCS); Sherry Xie (LINCS); ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Mano Maurya (MW); ; ; ; ; ; Sarah Reiff (4DN); Dan Lyman (GlyGen); ; ; ;John Erol Evangelista (LINCS) ; ; ; ; ; Arthur Brady ; ; ; ; ; ; Lynn Schriml (Human Disease Ontology)		

Agenda Item	Action Items	Owner
• Collection of HW results - species with phenotype info in hand		
• Discussion of synapse/synaptosome needs and larger issue of biosample preparation terms	•	

Notes:

- *biosample.assay_type* to be renamed to *biosample.sample_prep_method*
  - this will be a provenance field: “how did this biosample come into being?”
  - values will be drawn from OBI “planned process” subtree, excluding “assay” node
  - to express what sorts of assays were performed *on* biosamples, populate *file.assay_type* for files that were products of those assays
  - we will submit a term request to OBI for “synaptosome prep;” we expect them to rapidly issue a provisional ID, which can then be used in this field to identify synaptosome samples

- Future directions: Jeremy Y. has assigned himself homework to explore the portal; see how OWG activities have informed past CFDE/Portal work; think about the role of the OWG moving forward.

Chat:

Lynn: “Phenotype ontologies, unified representation at the Alliance of Genome Resources

For the bio sample data types, the metadata fields defined by the Genomic Standards Consortium could be reused. https://www.ncbi.nlm.nih.gov/biosample/docs/packages/
Includes sample preparation fields
Lynn Schriml to Everyone (11:10 AM)
https://github.com/GenomicsStandardsConsortium/mixs/tree/main/release/excel"

List of species that people have phenotype data for:

Mouse (MGI): Sue Bello and Cynthia Smith - Mammalian Phenotype Ontology– see Alliance of Genome Resources - phenotype alignment across organisms, including human, via Human Phenotype Ontology

LINCS – human only (although LINCS data mostly uses cell lines, so "phenotype" is not very relevant for us, as most of the cancer cell lines are already covered by the DO disease metadata)

MW: [Mano]: The species for subjects with disease/phenotype mentioned in internal MW system (assuming these are just animal models of a human phenotype, I used HPO to prepare the metadata table): Homo sapiens, Mus musculus, Rattus norvegicus [these three species cover most subjects with phenotype], Danio rerio, Felis catus, Macaca mulatta, Sus scrofa, Canis lupus familiaris, Bacteroides xylanisolvens, Ovis aries, Streptococcus mutans

IDG: Mouse (JAX/MGI, IMPC) employing Mammalian Phenotype Ontology.

4DN: no phenotype data currently (expecting human in future, maybe mouse - ontologies not decided on but potentially HPO)

exRNA: human (95% of data) and mouse (5%).

Mouse, rat, human: GlyGen

April 20, 2022

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;;Suvarna Nadendla (HMP) ; ; ;Allen Baron (DO) ; ; ; ; ; ; ; ; John Erol Evangelista (LINCS); ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Sherry Xie (LINCS); ; ; ; Jeremy Yang (IDG); Christophe Lambert (IDG) ; ; ; ; ; ; ; ; ; ; ;Owen White ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Mano		

	Maurya (MW); ; ; ; ; ; ; ; ; ; ; Daniel Lyman (GlyGen); ; ; ; ; ; ; ; ; Arthur Brady (CC); ; ; ; ; ; ; Haluk Resat; ; Keyang Yu (exRNA); ; ; ; ; ; Sarah Reiff (4DN); ; ; ; ; Andy Schroeder (4DN); ; ; ; ; ; ; ; ; ; ; ; ;

Agenda Item	Action Items	Owner
● NTR status		
● terms to avoid in submissions	●	
● use of GO in anatomy field		
● species phenotype info		
● presentations at upcoming meetings ○ variant WG standards https://docs.google.com/document/d/1p5wsaf2pLr6WZWQX8v2sqHCnu14S8riU/edit?usp=sharing&ouid=107880690455417800285&rtpof=true&sd=true ○ GTEx/Kids First internal harmonization for partnership		
● the OWG moving forward		

Notes:

For item #2 in the agenda:

1) Terms to avoid in submissions:
    - data (data:0006)
    - Format (format:1915)
    - assay (OBI:0000070)
    - anatomical system (UBERON:0000467)
    - anatomical entity (UBERON:0001062)
    - tandem mass spectrometry (OBI:0200198 - not under assay node in OBI) (will request this new OBI term - keeping for now until new term is ready)

- New term needs : Reach out to Michelle and Suvvi for new ontology terms.
- Terms to avoid in submissions: Decision: Leave the field blank if there is no specific ontology term to be used.

  Arthur: “note that null values by convention _always_ mean "no information available" in our environment, not "thing not present".
- A new anatomy request - “Synaptosome/synapse” is present in the GO cellular component. UBERON did not want to add it to their ontology as it is not a specific

anatomy term (instead it is a cellular level anatomy term). It results from a process/ experimental preparation.
One workaround is to include GO cellular compartment terms along with Uberon in c2m2 for anatomy. This is possible if there is no overlap between GO and uberon terms.
If there is an overlap between GO and uberon terms, then uberon terms will be used in c2m2 and GO terms will be synonyms for uberon terms.

Jonathan Silverstein :"I agree we probably have to be disciplined about use of an anatomy term as anatomy - so as long as the application/context retains that synapse is used as anatomy/location of something rather than what it is (prep) - we probably do need to get a "synaptosome prep" and many other "preps" as "things" if we want them in - for example, in HuBMAP we have "samples" and samples can be just blocks, sections, or suspensions at highest level so its tractable, with many subtypes".

Decision: We will think about this more and also explore the idea of "synaptosome" to be a sample (synaptosome preparation).

- Homework Assignment for DCCs to tell us what species you are collecting phenotype information for.
- OWG moving forward: Owen- What new ontology (use of new ontology already existing in the world) will make CFDE data more useful? Or in other words what use cases the users are trying to derive from the portal? This will inform us what new controlled vocabularies will add value.
Jeremy - Combine MONDO with DO in c2m2.

March 9, 23 and April 6

meetings were canceled due to Michelle having to teach and to lack of pressing agenda items

---

Feb. 23, 2022

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;;Suvarna Nadendla (HMP) ; ; ; ; Michelle Giglio; ; ; ; ; ; ; ; Tom Gillespie (SPARC); ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;Dan Lyman (GlyGen) ; ;Srinivasan Ramachandran (MW) ;Chris Kinsinger ; ; ; ; ; ; ; ; ; Daniel Clarke (LINCS); ; Arthur Brady; ; ; ; ; ; ; ; ; Sarah Reiff (4DN); ; Mano Maurya (MW); ;		

Agenda Item	Action Items	Owner
- C2M2 update - status and plans		
- CFDE Coordination center proposal for upcoming year (may 2022-April 2023)	- DRAFT_CFDE_20... - Draft as in "not yet approved as our funded activities", not draft as in "still being written". This was submitted to NIH as our proposal at the beginning of February 2022)	

Notes:

- Example of a markdown page : https://www.facebase.org/chaise/record/#1/isa:dataset/RID=1-X5F0
- CFDE-CC application for next funding period: https://drive.google.com/file/d/1vmvDcv715ZsA39LArdTZmKZ-I7c2QoUR/view (Draft as in "not yet approved as our funded activities", not draft as in "still being written". This was submitted to NIH as our proposal at the beginning of February 2022)
- Arthur's slides: https://docs.google.com/presentation/d/1EnXnhlEi18yHJFD58vPC2blCMkdJN-3B1bLQCSpbuSg/edit?usp=sharing

February 9, 2022

Objective	see agenda	Time	Every other Wednesday

			11am Eastern
wLeader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;; Suvarna Nadendla (HMP); ; ; ; ; ; ; ; ; ; ; ; Michelle Gilgio; ; ; ; ; ; ; ; ; ; Sarah Reiff (4DN); ; ; ; ; ; ; ; ; ; ; ; ; Jessica Binder(IDG); ; ; ; ; ; ; ; ; Sherry Xie (LINCS); ; ; ; ; ; ; ; ; ; ; ; ; ; ;; ; ; ; ; ; ; ; ; Jeremy Yang (IDG); ; Mano Maurya (MW); Daniel Clarke (LINCS); ; ; ; ; ;Christophe Lambert (IDG) ; ; ; ;Chris Kinsinger ; ; ; ; ;Dan Lyman (GlyGen) ; ; ; Raja Mazumder; ;Jared Nedzel ;		

Agenda Item	Action Items	Owner
● New term updates ○ EDAM term requests sent to Jeff for Interlex testing ○ GlyGen needs for new uberon terms		
● Draft slim RFC https://docs.google.com/document/d/1Tedisnul0gbaOPlGyLnC4q-OSvcHqGw0qMnD2EVr3e4/edit?usp=sharing	●	
● NCBI Taxon slim plans		
● What phenotype ontologies do we need?		
● Future meeting items ○ C2M2 updates/plans ○ drafting guideline documents		

Notes:

- CFDE portal has ~100 taxonomic organisms. It will be challenging to display them in the portal hence slim is necessary. We will be using the proteome slim terms and, as needed, the proteome clusters for creating taxonomy slim. Michelle and Suvvi will collect all portal taxonomy ids and map to the proteome clusters for creating the slim.
- Michelle delivered new EDAM terms to Jeff (SPARC) for incorporating in the interlex pipeline system. DCCs can provide their new CV terms to Michelle and parallelly we will also enter them in the interlex pipeline. Here is the template that DCCs can use to send their terms to Michelle - https://docs.google.com/spreadsheets/d/1zoHaSpJ4W5q_tVtiWMdLdMr9REHmRZPWb1-WmAKlquk/edit?usp=sharing
- Michelle and Suvvi sent 4DN and HMP new EDAM terms to Jeff to be included in the interlex pipeline. Michelle sent provisional term ids for new assay terms requested by 4DN.

- Slim RFC is ready for review by the group.
- What phenotype ontologies should be included in CFDE? One phenotype ontology will not be sufficient. Request DCCs to come up with some focussed phenotype ontologies. We should be inclusive of species for which phenotype data is available in CFDE. Some suggestions are Mammalian Phenotype Ontology, HPO, model organism phenotype ontologies. Next ontology working group call, we shall finalize a list. Including clinical phenotype information using ontologies can be tricky.
- In the next call, Arthur will update the current C2M2 model and future additions for the next quarter release. Will discuss clinical metadata also in the call. Some investigation needs to be done on a discussion during the steering committee meeting about FHIR metadata mapping?

January 26, 2022

No meeting - homework instead:
Everyone was asked to review the draft of the Slim RFC and provide comments/suggestions:
https://docs.google.com/document/d/1TedisnuI0gbaOPlGyLnC4q-OSvcHqGw0qMnD2EVr3e4/edit?usp=sharing

January 12, 2022

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;;Suvarna Nadendla (HMP) ; ; Michelle Giglio; ; ;owen white ; ; ; ; ; ;Steve Mathias (IDG) ; ; Jeremy Yang (IDG); ; ; ; ; ; ; Mano Maurya (MW) ; ; ; ; ; ; ; ; Sarah Reiff (4DN); ; ; ; ; ; Asiyah Lin (NHGRI); ; ; ; ; ;George Papanicolaou ; Arthur Brady; ; ; ; ; ; ; Daniel Clarke (LINCS) ;John Erol Evangelista (LINCS, IDG) ; ; ;Dan Lyman (GlyGen) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Sherry Xie (LINCS); ; ; ; ; ; ; Raja Mazumder (GlyGen); ;Philippe Rocca-Serra ; ; ; ; ; ; ;Jessica Binder(IDG) ; ; ; ; ; ; ;		

Agenda Item	Action Items	Owner
- **new term requests** ○ **updates** ○ **Interlex testing terms** ○ **term from Mano**		

● slim comments		
● RFC comments	●	
● Items for the new year: ○ documentation ■ guidelines docs ■ slim RFC ○ shift to Interlex ○ taxon slimming ○ ontologies ■ phenotype ■ data use ontology ■ others ○ update on current C2M2 and plans		

Notes:

- HuBMAP's new OBI assay terms have been released. 4DN assay terms are under review. Provisional OBI ids have been registered for 4DN terms. New Term Request Tracking document.
- 4DN - new format and data_type terms will be processed (given provisional ids) through Interlex.
- Mano (Metabolomics) - They have a new DO term request - 'cachexia'. Want to know if it's a disease or a phenotype. OWG didn't have an opinion, therefore, we'll submit the question as a DO github new term request.
- Ontology Slims: Sarah Reiff's(4DN) request for assay slim changes will be incorporated in the v2 version of assay slim.
  Mano - JSON is included under YAML and JSON slim nodes.
- All RFCs (EDAM, DO, OBI and chemical) have received good comments by the consortium so far. Comment period ends in a week.
- Plan is to generate one RFC for all slims describing the two methods of top down and bottom up approaches for making a slim. Then each slim will be listed and some information given on how it was generated.
- Discussion of guidelines documents and their importance. Should they be RFCs or not - not quite sure - they will be evolving documents, there will certainly be the need for comment from consortium. These are a high priority.

December 15, 2021

In lieu of this meeting, I’m asking everyone to instead use that hour to review the new slim drafts for data, format, and assay.

Suvvi and I have used a combination of top-down and bottom-up procedures to build these slims, with a focus on making sure to cover terms already in use by the DCCs in the current version of portal data.

We have files for each of the slims that are google sheets in this folder:
https://drive.google.com/drive/folders/1ghfrUZNidNlw891BgM1bKZB5NX7ByRuL?usp=sharing

Each slim file contains 3 tabs: the first tab is the list of just slim terms, the second tab is slim mappings for terms currently in the datasets in the portal submitted by DCCs, the third tab is slim mappings for all terms in the ontology (whether they are used by any of the DCCs or not). We have made sure that all terms used in the portal have a mapped slim term. For any terms that do not have a slim mapping in our slims, they will by default slim to themselves. (But again, all terms currently in use in the portal have a mapped slim term.)

We will revise the slims after the next portal data release to expand them to cover any additional terms that start being used by the DCCs (that weren’t used in the current release.)

Here are links to the files:

- data (EDAM): https://docs.google.com/spreadsheets/d/1OiFHlTUPNDJ7BWcrOb-L9aA7V1TdUnvfty6eunsLQNo/edit?usp=sharing
- format (EDAM): https://docs.google.com/spreadsheets/d/13DvJy9bFBS32txpSAUyqbfXBPdV3MB7ecuC7_5LG3fI/edit?usp=sharing
- assay (OBI): https://docs.google.com/spreadsheets/d/1ePntIjYkGgoQajl8WsZIp-NCwDqTBvzyu5GtGM7EuMk/edit?usp=sharing
- disease (already approved by OWG): https://docs.google.com/spreadsheets/d/1nL3aI_8I-pdbpai8pkahR3kCBy2I9DUKAA053Q_MFJw/edit?usp=sharing

If you have comments/questions please post them to the OWG slack channel for discussion.

Reminder of our remaining Action Items that we will work on in the new year:

- Guidelines documents: Formalize location, continue working on drafts
- Interlex
    - Send sample EDAM terms to Jeff for testing in InterLex
    - review status labels and expand as needed for use of InterLex
    - Jeff will create a CFDE community in InterLex
- work with Raja to develop a plan for using a combination of NCBI taxon tree and sequence similarity based clusters to slim taxa
- Discuss phenotype and data use ontology needs
- Identify any other areas where ontologies/CVs are needed

Thanks,
Michelle

December 1, 2021

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	Raja (GlyGen); Suvarna Nadendla (HMP); Eryk Kropiwnicki (IDG); Susanna Sansone (CDFE-CC); Michelle Giglio; Sarah Reiff (4DN); Jeffrey Grethe (SPARC); Steve Mathias (IDG) ; John Erol Evangelista (LINCS, IDG); Sherry Xie (LINCS); Daniel Clarke (LINCS); Jeremy Yang (IDG); Dan Lyman (GlyGen); Chris Kinsinger (NIH); Arthur Brady; Mano Maurya (MW); Srinivasan Ramachandran(MW); Haluk Resat; Michael tiemeyer (GlyGen, Glycoscience Group); Rayna Harris (CFDE), Jeremy Walter (CFDE)		

Agenda Item	Action Items	Owner
• Presentation from Raja on slimming taxonomy	work with c2m2 group on modeling the cluster slims along with the NCBI tree	
• Presentation from SPARC on Interlex	• Jeff create CFDE Community in InterLex • Jeff/Michelle/Others review status labels used so far, and ones in InterLex, expand as needed • Michelle/Suvvi, send Jeff EDAM term requests for use in testing	
• Slims ○ Disease ■ DO Slim final ■ DO_slim_mapping ○ Assay ■ assay_slim_mapping ○ EDAM file format	• Homework for all: ○ final review of DO slim ○ review proposed assay slim ○ (see links to files at left for these) ○ review EDAM	

		slims when they are sent out over the next week	

Notes:

- Raja - slimming taxonomy
    - 104 organisms
    - Option 1: we use the NCBI taxonomy tree
    - Option 2: we create a slim that takes in account of genomics and tax data
        - this is important for bacteria and viruses because the NCBI taxonomy doesn't always reflect actual relatedness
        - Representative genomes that can help define our tax slim
        - Scalable way to generate genome clusters, which takes in consideration reference organisms and emerging pathogens
    - ***Proposed solution***:
        - we could use cluster for bacteria and virus, and the NCBI tax tree for the rest
        - we will have both the NCBI tree and the cluster info in the data structure; will need to sort out UI issues
        - clusters for bacteria and viruses are already available and maintained by other groups (at SIB).

- Jeff - SPARC on Interlex / Term Management
    - Interlex Term Management - CFDE OWG
    - SPARC uses several terminologies for datasets, models, services; goal is to use a consistent and computable vocabulary for search, display, integration
    - Interlex tool: dynamic lexicon for biomedical terms
    - Term request pipeline
        - from request, review and engineering (editing and addition to the ontology, and use in the knowledge graph)
    - Term support pipeline - requirements:
        - Quick turnaround for PIDs
        - Term suggestion tracking via UI and API
        - Elicit suggestions, and support expert review
        - Prepare submission to external/community ontologies
    - Term management improvements, incl:
        - Community dashboard
        - Enhanced set of status tag
    - Discussion points:
        - Terminology review: who does what?

        - Submission to external/community ontologies - how to/who?
        - Extra status tags?
        - Integration with BioPortal?
    - ***Action Items***
        - create a CFDE Community space in Interlex
        - CFDE new term request tracking sheet has examples of tracking status for us to review
            - OBI is a good example both for the term management process and for the status tags
        - send EDAM term needs to Jeff for use in testing - EDAM will be an excellent test case for our use of InterLex

- Slim update and review
    - DO slim ready to go - had a few minor adjustments since last time
    - proposed OBI assay slim ready for review by OWG
    - EDAM format slim will soon be sent around

- Homework:
    - Action items as above
    - Final review of DO slim
    - Review and send comments on Assay slim

November 17, 2021

Meeting canceled due to 2-day CFDE Strategy Meeting

November 3, 2021

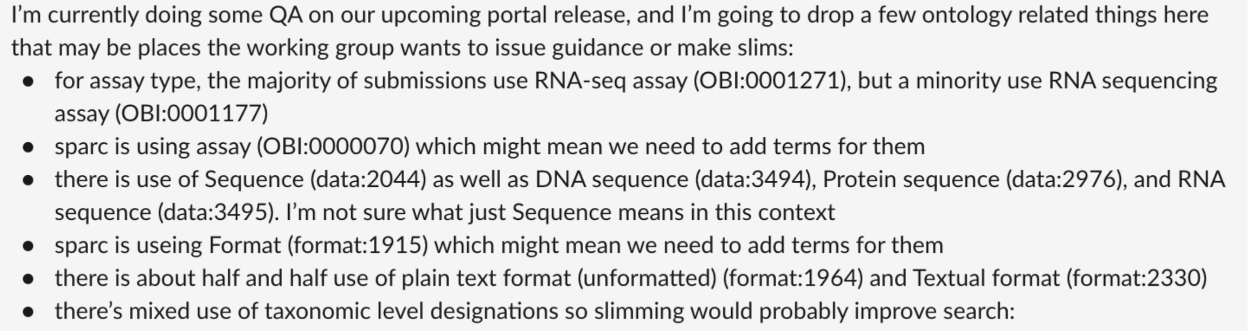
Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	Raja (GlyGen);;Suvarna Nadendla (HMP) ; Michelle Giglio; Owen White; ; ; Arthur Brady; ; ; ; ; ;Jeremy Yang (IDG) ; ; ; ; ; ; ; ;John Erol Evangelista (LINCS, IDG) ; ;Mano Maurya (MW) ; ; ; ; ; Asiyah Yu Lin (NHGRI); ; ; ; ; ; ; ; ;Chris Kinsinger ; ; ; ; ; ; ; ; ; ; ; ; ; Haluk Resat; ; ; ; ; ; ; Bernard de Bono; ; ;; ; ; ;Dan Lyman (GlyGen) ; ; ;Steve Mathias (IDG) ; ; ; ; ; ; ; ; ; ;Amanda Charbonneau ; ; ; ; ; ; ; ; Sherry Xie (LINCS); Daniel Clarke (LINCS); Eryk Kropiwnicki (IDG); Sarah Reiff (4DN); ; ; ; ; ;Philippe Rocca-Serra ;			

Agenda Item	Action Items	Owner
- Disease Ontology Slim https://docs.google.com/document/d/1SDpa0R7RUL83YIrLx24-ZDtwlW0AdVPAnFN4TBEEiXU/edit?usp=sharing	Michelle make final slim list and draft RFC	
- Question from Mano regarding use of DO for non-human subjects	- Follow up with C2M2 group on relationships between disease terms and what they are linked to	
- questions raised by Amanda in the OWG Slack ○ use of very general terms - do people need more specific terms? ○ in OBI: RNA-seq assay vs RNA Sequencing assay ○ use of data_type 'sequence' as opposed to more specific DNA/RNA/protein sequence ○ consistency with regard to text file formats ○ taxonomy slim (See details below) See info file here: https://docs.google.com/spreadsheets/d/1wkoN9fyUtVZWYnbK0S0AKJmiTS-yELCRKgkNqk870eA/edit?usp=sharing	- Add decisions below into nascent guidelines document - Work on taxonomy slim	

Notes:

- DO Slim working document:
    - Do we want to have "disease by infectious agent" as the slim term or have the sub categories as well? Raja: keep it simple and have the general term.
      Mano: Thinks it is useful to have granular slim terms as they have data related to bacterial infections.
      It is decided that we should keep it simple and complexity can be added if necessary.
      Michelle reminded everyone that curators should always annotate at the most granular level appropriate and that the slim is just for summary views/high-evel comparisons for the UI.
    - "disease of anatomical entity", general agreement that this structure is good as is from AGR slim
    - "disease of cellular proliferation" : organ specific cancer terms will stay as children of the overall 'cancer' term in the slim, thus terms sliming to the organ-specific terms will be counted twice in the slim - once to their organ-specific slim parent and once to the overall cancer term.
    - "genetic disease" : keeping it simple - using parent rather than AGR children terms.

    - Michelle will put together final slim based on this conversation and draft an RFC for it.
- Mano’s question: Metabolomics have mouse models.Is it appropriate to use DO ids for non-human models? It is appropriate to use DO for mouse datasets used as human models. We still need to be able to express the relationship between the disease term and the thing we are linking to - such as: subject ‘is_model_for’ DOterm . This needs to be built into the C2M2 structure.
- Amanda’s questions from slack and QC doc (link):



I’m currently doing some QA on our upcoming portal release, and I’m going to drop a few ontology related things here that may be places the working group wants to issue guidance or make slims:

- for assay type, the majority of submissions use RNA-seq assay (OBI:0001271), but a minority use RNA sequencing assay (OBI:0001177)
- sparc is using assay (OBI:0000070) which might mean we need to add terms for them
- there is use of Sequence (data:2044) as well as DNA sequence (data:3494), Protein sequence (data:2976), and RNA sequence (data:3495). I’m not sure what just Sequence means in this context
- sparc is useing Format (format:1915) which might mean we need to add terms for them
- there is about half and half use of plain text format (unformatted) (format:1964) and Textual format (format:2330)
- there’s mixed use of taxonomic level designations so slimming would probably improve search:

Sequence used as data_type. DCCs who used this term should look at the data to see if a more granular data type can be used. There is “nucleic acid sequence” in OBI that can be used if the file contains both DNA and RNA sequence.

For consistency HMP will change and use “mass spectrometry data” instead of “mass spectrum”.

Amanda asked about a guidelines document. Such a document has been started although it is not formatted or in any way that it is yet digestible. That is still on the to-do list.

There is a confusion between RNA sequencing assay and RNA-seq assay. There was a discussion at OBI whether to merge them or have more detailed definitions differentiating both. For regular RNA sequencing (RNA->cDNA->sequencing) use RNA-seq assay. This will be resolved at OBI soon.

There are DCCs using “pig” and “domestic pig”. DCC clarified that those two are the right terms as that is what was reported by the data submitters. Taxonomy has more terms now in the portal. Is order level slimming good? Raja - slimming should be data driven. Decision : will look at the species included in the portal and decide on the upper level taxonomic terms. Jonathan - We used very simply count at node and roll up to roughly equal counts in categories shown - and in a “perfect” interface showing the counts with each choice is super useful.

Amanda might have someone who can play around with the taxonomy.

October 20, 2021: Meeting Canceled

October 6, 2021

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	Michelle Giglio; Jeremy Walter; ; ;Suvarna Nadendla (HMP) ; ; Eryk Kropiwnicki (IDG); ; ; ; ; ; ; ; ; ; ; Susanna Sansone; ; ; ; ; ; ;Steve Mathias (IDG) ; ; ; ; ; ; ; Vince Metzger (IDG); ; ; ; ; ; ; ; ; ; Daniel Clarke (LINCS); John Erol Evangelista (LINCS, IDG); ; ; ; ; Jeremy Yang (IDG); ; ; ; ;Jonathan Silvestain ; Dan Lyman (Glygen); ; ; ; ;Arthur Brady ; ; ; ; ; Haluk Resat; ; ; ; ;Mano Maurya (MW) ; ; ; ;Owen ; ; ; Raja (GlyGen); ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;		

Agenda Item	Action Items	Owner
● **Review of Uberon RFC**		
● **Presentation from Jeremy: "PubMed Indexing, Entities and Ontologies"**	● **https://drive.google.com/file/d/15XlYW6Apj41PJmbFjNwl6I7_Qx2jgDDU/view?usp=sharing**	
● **Disease Ontology Slim - can we move forward with the AGR subset?**		
● **Uberon Slim - would anyone like to volunteer to suggest an alternative?**		

Notes:

- UBERON RFC
  - is ready for review by wider CFDE.
- Presentation from Jeremy: "PubMed Indexing, Entities and Ontologies"
- Entrez type of search and results will be good for CFDE (Eg: searching in the search bar which will give results for categories in the database - gene, diseases, phenotypes etc).
- Vince - autosearch and autofill functionality is very useful.
- Homework - Disease slim - DO_AGR_slim to be reviewed. Raja - cancer node is too broad. Explore a cancer slim. Raja ready to share a cancer slim with us.

- Suvvi has explored use of ROBOT tool (filter option) for generating mappings of granular terms to slim terms - thus we should be ready when the slim is finaized.
- Need to start working sessions to create a more formal and streamlined uberon slim. No one volunteered. We will go with what we have for now.
- Michelle will contact UI people to have a call to know the plans and what UI is able to do and what kind of help they need from us.

September 22, 2021

Objective	see agenda	Time	Every other Wednesday 11am Eastern
wLeader(s):	Michelle Giglio Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;Suvarna Nadendla (HMP) ; ; ; ; ; ; ; ;Amanda Charbonneau ; ; ; ; ; Sherry Xie (LINCS); ; ; ; ; ; ; ;Michelle Giglio ; ; ; ; ; ; ; ; Bernard de Bono ; ; ; ; Srinivasan Ramachandran (MW); ; ; ; ; ; ; Vincent Metzger (IDG);Jeremy Yang (IDG); ; ; ; ; ;Dan Lyman (GlyGen) ; ; ; ; ; ; Mark Musen; ; ; ; ; ; ; ; Mano Maurya MW); ; ; ; ; ; ; ; ;John Erol Evangelista (LINCS, IDG) ; Daniel Clarke (LINCS); ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Haluk Resat;Eryk Kropiwnicki (IDG) ; ; ; ; ; ; ; ; ; Arthur Brady; ;Chris Kinsinger ; ; ; ; Jonathan Silverstein (HuBMAP); ; ; ; ; ; ;Tom Gillespie (SPARC) ; ;Ellen Quardokus (HuBMAP) ;Raja Mazumder (GlyGen) ; ;		

Agenda Item	Action Items	Owner
- **Slims** ○ **uberon - new thoughts? Use cases? https://docs.google.com/spreadsheets/d/16x-6dTsmxWZH2nnyabSZF4-H3PIEiUPHJCIGwuhC1_k/edit?usp=sharing**		

○ **Disease Ontology - candidate is the existing DO_AGR_slim.owl https://github.com/DiseaseOntology/HumanDiseaseOntology/tree/main/src/ontology/subsets**		
● **Updates on C2M2 plans/work**	●	
● **Draft of uberon RFC** https://docs.google.com/document/d/1e14gCuo3K0qEEhitPzS8ybn6HBMTr-gqmggzYJ24Tpk/edit#		

Notes:

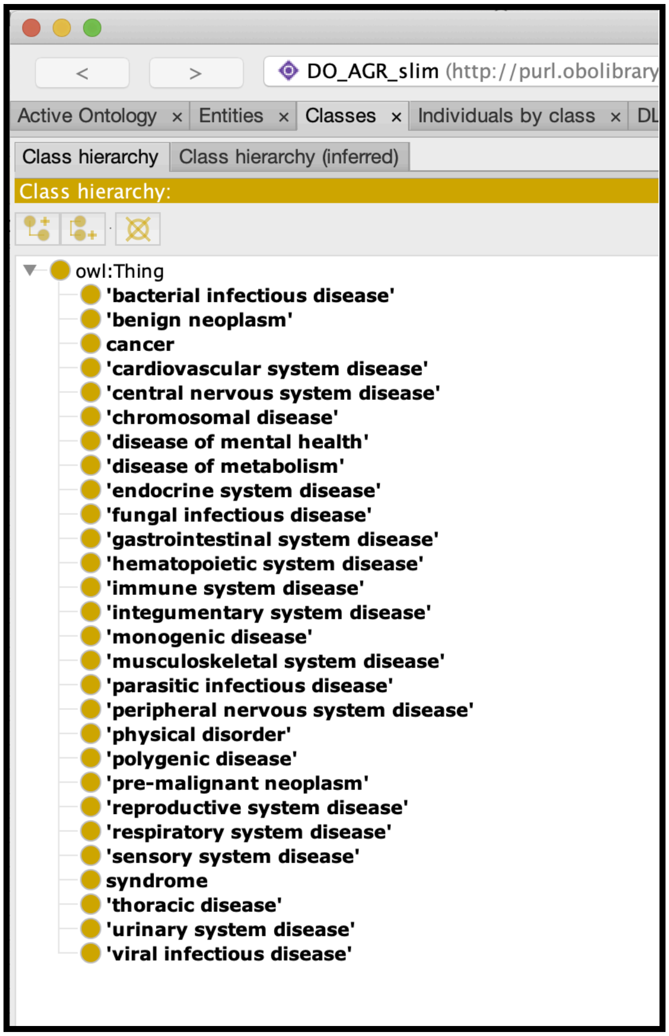
- Homework from last time was to look at UBERON and the current slim and to see if there needs to be changes or even if it would be better to start from scratch. If the current slim is not meeting our needs, we can change it.
- It makes sense to have a process in place to update uberon slim as needed. RFC should include a revision process of a slim. Either as part of each slim RFC or separate RFC focused on slim management.
- Look at MESH anatomy search.Jeremy is going to look at MESH anatomy, will give input if uberon slim update is needed.
- DO has a DO_AGR_slim for high level disease terms, good starting point for disease slim. (https://github.com/DiseaseOntology/HumanDiseaseOntology/blob/main/src/ontology/subsets/DO_AGR_slim.owl). We need to map slim term to granular slim DO ids. Will explore ROBOT to generate disease slim (in contact with Chris Mungall for help).
- Granular terms can map to more than one slim term. Jonathan suggests organizing the list of slim terms such that the system-level ones are together and disease-type ones are together, etc..
- Compressed/uncompressed files will be handled in the next version of C2M2 model.
- C2M2 work in progress - sex (proposed use of SNOMED), race (proposed CDC federal guidelines for categories), ethnicity (proposed CDC federal guidelines for categories), age at recruitment, age at sampling (both age fields will be allow capture of age to two decimal points).
- The ability to check more than one race box will not be in the first release, but is on the list for the next development cycle. These fields are self reported metadata and thus should not be altered from what the subject chose. Currently NULL will represent either that the subject declined to choose an option or if the info is unavailable, but this will be revisited to see about making a distinction between these situations. There are several reasons that it is valuable to capture race/ethnicity metadata. However, slimming race and ethnicity is problematic on many levels - therefore if we restrict to only the federal terms then anytime there is a DCC that is not using the federal set of terms there would be no way to capture that info. Therefore we will include a field for any terms that are

used by a DCC that are not included in the federal list. This will be a free-text field. This field won't be available in the initial role out of race metadata capture in C2M2 but will be incorporated in the following release. Therefore, for the next data submission cycle (after the current one - the current data submission deadline is in October - race/ethnicity info will NOT be part of the model for the October submissions) DCCs that are using the federal terms (we think there are at least 3) will be able to submit their race/ethnicity data. But those DCCs not using those terms won't be able to submit their data until the following data submission cycle.

- There was a question about the sex categories - intersex and transsexual and directionality. (current phenotypic or birth phenotypic sex). We will check SNOMED for more specific child terms. Even if there aren't more specific relevant terms, we will create a solution to capture this precisely and provide a guideline usage doc to tell people what to do.
- HPO is being considered for phenotypes. We need to also think about MPO and whether we need to capture human and other subject info separately.
- Going to add the ability to say "disease not observed" in a sample and subject; "phenotype not observed" in a subject.
- Subjects can be many types of things including human, mouse or cell lines (LINCS). A "subject" is the object of the study.

New Homework:

- Look at the DO AGR slim and determine if you think this could work: https://github.com/DiseaseOntology/HumanDiseaseOntology/tree/main/src/ontology/subsets If it doesn't appear to work, we can make one from scratch - that is not a problem.



- owl:Thing
  - 'bacterial infectious disease'
  - 'benign neoplasm'
  - cancer
  - 'cardiovascular system disease'
  - 'central nervous system disease'
  - 'chromosomal disease'
  - 'disease of mental health'
  - 'disease of metabolism'
  - 'endocrine system disease'
  - 'fungal infectious disease'
  - 'gastrointestinal system disease'
  - 'hematopoietic system disease'
  - 'immune system disease'
  - 'integumentary system disease'
  - 'monogenic disease'
  - 'musculoskeletal system disease'
  - 'parasitic infectious disease'
  - 'peripheral nervous system disease'
  - 'physical disorder'
  - 'polygenic disease'
  - 'pre-malignant neoplasm'
  - 'reproductive system disease'
  - 'respiratory system disease'
  - 'sensory system disease'
  - syndrome
  - 'thoracic disease'
  - 'urinary system disease'
  - 'viral infectious disease'

- Review RFC from Tom on the use of UBERON to capture anatomy for CFDE: https://docs.google.com/document/d/1e14gCuo3K0qEEhitPzS8ybn6HBMTr-gqmggzYJ24Tpk/edit#

Record of the Chat:

11:08:35 From Silverstein, Jonathan to Everyone:

Sorry, just joined, can someone drop the agenda/attendence in chat again?

11:08:55 From Suvarna Nadendla to Everyone:

https://docs.google.com/document/d/1VoHHBeWfol6XNJa3kzOnOFuTaIrcLYbqKYQcOnj1oh4/edit?usp=sharing

11:08:55 From Robert Carter to Everyone:

https://docs.google.com/document/d/1VoHHBeWfol6XNJa3kzOnOFuTaIrcLYbqKYQcOnj1oh4/edit#

11:12:18 From Silverstein, Jonathan to Everyone:

Cool that this AGR was found - it seems to mix a bit of anatomic, and “by system” and “by disease type” (cancer, genetic, etc..) - I wonder if there is one level higher above these to group them in…(would have two-three levels rather than one or two so some downside but this list is hard to parse)

11:34:57 From Silverstein, Jonathan to Everyone:

Agreed these two necessary (as presented totally sensible to me) - Race and Ethnicity, whether as confounders or not per se, depending upon one's viewpoint, associate with "everything" in the U.S. and this is a Federal program also.

11:35:19 From Silverstein, Jonathan to Everyone:

(Short list - DCC has to resolve to Federal seems appropriate to me FWIW)

11:45:15 From Silverstein, Jonathan to Everyone:

Owen's suggestion seems insufficient for folks actively trying to find data for under-represented minorities - I think we should support those goals (I agree with Amanda) with the approach as originally presented here by Arthur/Michelle.

11:45:35 From Raja Mazumder to Everyone:

Agree with Amanda. And also previous comments by Jeremy

11:46:44 From Silverstein, Jonathan to Everyone:

The SF454 form has these also, and yes, point of more of this being asked in Fed, not less, for example PEDP comments required in applications, etc…Amanda right on in my opinion.

11:49:27 From Daniel Clarke to Everyone:

I also think the freetext field makes sense. These terms may in-fact evolve over time

11:50:03 From Raja Mazumder to Everyone:

Agree with free text. Over time maybe some of the terms would make it to non-free text

11:51:23 From Silverstein, Jonathan to Everyone:

By the way: I do think all of these have precise codes in SNOMED, so they could "operate" where folks want to provide them in the same way as 'sex'.

11:52:56 From Daniel Clarke to Everyone:

Note that though free-text means free-text in the C2M2 -- it doesn't mean the original DCC necessarily used free-text -- they came up with a protocol that was scrutinized for their studies.

September 8, 2021

Objective	see agenda	Time	Every other Wednesday 11am Eastern
Leader(s):	Michelle Giglio Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;Suvarna Nadendla (HMP) ; ; ; Michelle Giglio; ; ; ; ; ;Eryk Kropiwnicki (IDG) ;John Erol Evangelista (LINCS, IDG) ; Daniel Clarke (LINCS); ; ; Sherry Xie (LINCS); ; ; ; ; Jeremy Yang (IDG); ; ; ; ; ; ; ; ; ;Srinivasan Ramachandran (MW) ; Mano Maurya (MW); ; ; ; Mark Musen (HuBMAP); ; ; ; ;Vincent Metzger (IDG) ; ; ; ;Steve Mathias (IDG) ; ; ; ; ; ; ; ; ; Haluk Resat; ; ; ; ; ; ; ; Owen White; ; ; ; ; Philippe Rocca-Serra; ; ; ; ; Jonathan Silverstein (HuBMAP); ; ; ; ; ; ; ; Chris Kinsinger; ; ; ; ; ; ; ;Dan Lyman (GlyGen) ; ; ; ; ; ; ; ;Bob Carter ; ; ; ; ; ; ; ; ; Raja Mazumder (GlyGen); ; ; ; ; ; ; ; ; ;Tom Gillespie (SPARC) ;		

Agenda Item	Action Items	Owner
Updates - RFCs - NTRs https://docs.google.com/spreadsheets/d/1zoHaSpJ4W5q_tVtiWMdLdMr9REHmRZPWb1-WmAKIquk/edit?usp=sharing - guidelines - uberon slim comments?		
- https://docs.nih-cfde.org/en/latest/c2m2/draft-C2M2_specification/#c2m2-technical-specification - Recent C2M2 implementation of ○ Disease table ○ links from Disease to subject and biosample	-	
- change to QC scripts such that terms imported into ontologies (from other ontologies) are valid		
- Upcoming C2M2 plans ○ PubChem ○ genes ○ 6 fields of clinical metadata ○ analysis_type		

Notes:

Discussion of uberon draft slim

Not many comments from group. Some questions about the random terms like ‘neck’ that were included. Michelle explained that those didn’t have any corresponding system-level group to go into. Thus they were kind of left hanging. Group was encouraged to review and bring any suggestions for alternate arrangements to the next meeting.

There were suggestions to have use cases to help in determining what is the best slim. Michelle encouraged the group to suggest use cases at next meeting to help guide discussion.

Recognition by group that slims may (most likely will) change over time in response to use cases or other things. The RFCs will document the changes.

New fields/tables in C2M2

Michelle reviewed for the group the new tables in C2M2 for associating disease with either subject or biosample - biosample_disease and subject_disease. Multiple diseases can be attached to each.

Refinements requested by the group:

ability to indicate that a sample did NOT have a disease, was a control, or otherwise “healthy”

ability to describe relationships such as “has_disease” and “is_model_of_disease”

Another related suggestion was the ability to capture pathology.

Update to QC process

Terms from other ontologies that are imported to be actual terms in the first ontology (like CHMO terms imported into OBI as true terms) will now pass QC.

Reviewed C2M2 expansion plans being worked on for this quarter.

August 25, 2021 - Meeting was canceled

August 11, 2021

Objective	see agenda	Time	Every other Wednesday 11am Eastern

Leader(s):	Michelle Giglio Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;Owie ; Suvarna Nadendla (HMP); ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;Steve Mathias (IDG) ; ; ; ; ; ; ; ; ;Jeremy Yang ; ; ; ; ; ; ; ; Sherry Xie (LINCS); ; ; ; ;John Erol Evangelista(LINCS, IDG) ; Daniel Clarke (LINCS); ;Dan Lyman (GlyGen) ; ; ;Eryk Kropiwnicki (IDG) ; ; ; ; ; ; ;Mano Maurya (UCSD, MW) ; ; ; ; Philippe Rocca-Serra; ; ; ; ; ; ; ; ; ; ; Arthur Brady; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;Michelle Giglio ; ; ; ; ;		

Agenda Item	Action Items	Owner
● **Updates** ○ **NTRs** ○ **RFCs**		
● **Uberon Slim discussion**	●	

Notes:

Updates on RFC status:

- Under review by all of CFDE:
  - Review time ends this Friday:
    - EDAM for data_type and file_format (Michelle Giglio)
    - Disease Ontology (Philippe Rocca-Serra)
  - Just sent to Amanda for sharing with all of CFDE:
    - PubChem for chemicals (Jeremy Yang)
    - OBI for assay types (Michelle Giglio)
- Under development:
  - Uberon for anatomy (Bernard de Bono)
- Next up:
  - Slim RFCs

Updates on new term status

-LINCS terms official release in OBI is imminent

-HuBMAP and Metabolomics terms have been assigned OBI ids and can be used; Alex and Ivan are reviewing HuBMAP terms, Mano will review Metabolomics terms

-Still to do are terms for Kids First

-We have a New Term Status tracking page: https://docs.google.com/spreadsheets/d/1zoHaSpJ4W5q_tVtiWMdLdMr9REHmRZPWb1-WmAKIquk/edit?usp=sharing

This provides a central location for tracking status. It also is where Arthur pulls info from to mark terms as provisional in the portal data.

Slim work

-Suvvi and Michelle (mostly Suvvi) made a first version of a slim for Uberon. This was needed quickly so that engineers could begin building and testing the ability of the Portal to deal with slim data.

-We started off using “subsets” that uberon developed and have available on github. These are at a level of body systems like “nervous system” and “respiratory system” and seemed a good level to start with. About 2000 uberon terms didn’t map to any of these subsets. Of >14,000 terms in uberon, 213 are currently used in DCC CFDE data. We checked for terms in the 213 that didn’t map to any subsets, there were about 100. Suvvi and I manually mapped these to either the system level categories or to other terms that we added to the slim term list. We ended up with 38 terms in this slim. Anything else in uberon (outside of the 213 terms used by DCCs) that didn’t map to a subset was assigned to “anatomical entity” for now.

-This slim can be a starting point for discussion and development of a final slim. There are many ways we could decide to slim things, these system-level terms are just one way. We may also need/want some slim terms that are more specific than these system-level ones. Again, this slim is a starting point for discussion.

-Uberon slim info can be found here:
https://docs.google.com/spreadsheets/d/16x-6dTsmxWZH2nnyabSZF4-H3PlEiUPHJClGwuhC1_k/edit?usp=sharing

There are 3 tabs in this sheet. The first one has all 14,000 uberon term mappings to a slim term. The second one has just the list of 38 slim terms. The third one has the 213 uberon terms currently used in DCC data and their mappings to the slim terms. Obviously that 213 will continue to change over time and we’ll need to make sure that any new terms that enter the set of DCC-used terms are appropriately assigned a slim term either from one of the subsets (ideally) or manually.

-New slims can be substituted into use in the portal at any time. So, once we decide on a slim it can immediately go into production.

-Homework for everyone: look at this slim in the context of the larger uberon ontology, think about whether these system levels look like a good aggregation point, see if other categories might make more sense, see if additional slim categories could be used to cluster any of the existing 38, think about how your data will look in this slim. We will discuss more on future calls.

Things still in progress or on the to-do list:

-model how to relate disease-anatomy-phenotype and think about genes and chemicals in this mix too. (Michelle is supposed to draft a strawperson model to start discussion)

-collect term usage guidelines

July 28th meeting was canceled

July 14, 2021

Objective	see agenda	Time	Every other Wednesday 11am Eastern
Leader(s):	Michelle Giglio Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	; Suvarna Nadendla (HMP); ;Owen White ;Srinivasan Ramachandran (MW) ; Mano Maurya (MW, UCSD);Eryk Kropiwnicki (IDG) ; ; ; ; ; ; ; John Erol Evangelista (LINCS, IDG); Cristian Bologa; ; ; ; ; ; ; ; ; ; ; ; ; ;Steve Mathias (IDG) ;Amanda Charbonneau ; ; ; ; ; ; Bernard de Bono ; ; Haluk Resat; ; ;Jeremy Yang (IDG) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Raja Mazumder; ; Susanna-Assunta Sansone (CFDE-CC); ; ; ; ; ; Arthur Brady; ; ; ; ; Daniel Clarke (LINCS); ; ;Daniel Lyman (GlyGen)		

Agenda Item	Action Items	Owner
● **Updates** ○ **github issue tracker: https://github.com/nih-cfde/ontologyWG/issues** ○ **NTRs**		
I finally posted the RFCs to the whole CFDE, sorry it took so long	**- Use of EDAM for "file_format and "data_type" fields in the file table of the C2M2** **- Use of DOID for "disease" field in the C2M2 model**	amanda
● **Discussion of RFC comments** ○ **PubChem** ○ **OBI**	● **ALL: Please put in final comments/suggestions for these two RFCs within the next week**	
● **Discussion of Homework results regarding capture of phenotype info**	● **Michelle, Bernard, Tom, Arthur, Suvvi, look into example cases and the basic fields and relations we need - try to make a draft of something to share**	

Notes:

(There was no meeting held on June 30th)

Updates:

- New CFDE-ontology Github to keep track of work to do and done
- **https://github.com/nih-cfde/ontologyWG/issues**

New term requests

- OBI
  - LINCS batch is officially in , although not released
  - almost ready to pre-register HUBMAP terms to get IDs
  - Change will be made to C2M2 this week that allows terms imported into OBI (like CHMO) to be used directly as values in the table
  - If a CHMO terms is not imported, we will ask OBI about it. In time, we may change things to be able to use non-imported terms, but this isn't the case yet.
- Establishing trust with the ontology communities we need, so we will not start off with pull requests but rather work through the ontology's NTR systems
- Anatomy terms requests can be processed right way through InterLex - CFDE will be added as a "community"
- Other term requests should be made into issues in the dedicated CFDE-ontology github - **https://github.com/nih-cfde/ontologyWG/issues**
- Chemical terms: PubChem a vocabulary is also integrated in ChEBI ontology, PubChem also include MeSH terms; PubChem has a process for requesting new substances/compounds to be included.

RFCs - final revisions

- OBI-RFC - Michelle
- ChemicalOntology-PubChem-RFC - Jeremy
  - Other relevant resources:
    - Main RefMet page
    - Download RefMet
    - Structure search
- **ACTION: All -** Please put in final comments/suggestions for both of these within the next week

Phenotype

- Homework for discussion
  - Alliance of Genomics Resources is focused on linking orthologs to human disease phenotypes as well as capturing the relationships (via RO) of genotypes, environmental exposure, or chemicals to the disease/phenotype
    - Previous concerns about this work: representing phenotypes for different organisms was challenging but seems to have been solved, but only because they are focusing on a narrow range of things around orthologous genes
    - For the CFDE portal, we can use a similar modelling structure as what AGR has done, but only use it at the high level
    - **ACTION**: Michelle, Bernard, Tom, Arthur, Suvvi, look into example cases and the basic fields and relations we need

---

June 30, 2021

No meeting

June 16, 2021

Objective	see agenda	Time	Every other Wednesday 11am Eastern
Leader(s):	Michelle Giglio Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	; Suvarna Nadendla (HMP); ; ; Eryk Kropiwnicki (IDG); ; ; ; ; Sherry Xie (LINCS); ; ; ; ; ; Arthur Brady; ; Philippe Rocca-Serra; Susanna-A Sansone ;Marisa Lim (CFDE-CC) ; ; ; ; ; ;Steve Mathias (IDG) ; ; Michelle Giglio ; ; ; ; ; ; ; ; ; ; ; ;Mano Maurya (MW, UCSD) ; ; ; ; ; ;Dan Lyman (GlyGen) ; ; ; ; ; ;Chris Kinsinger (NIH) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;Srinivasan Ramachandran (MW) ; ; ; ; ; ; ; ; ; ; ; ; ;Jessica Binder (IDG) ; ; ; ; ; ; ; ; ; ; ; ; ; ;Jeremy Yang (IDG) ; ; Bernard de Bono ; ; ; ; ; ; ; John Erol Evangelista (LINCS, IDG); ; ; ; ; ; Ellen Quardokus (Borner/IU); ; ; ; ; ; ; ; ;Raja Mazumder (GlyGen)		

Agenda Item	Action Items	Owner
● RFC updates		
● choosing an ontology for phenotypes	● Provide use case(s) detailing use of phenotypical informations	DCCs
● next steps on phenotype-anatomy-disease modeling		
● drafting slims and guidelines		

Notes:

RFC updates:

- Two new RFCs in Ontology WG folder
  - On PubChem ontology and OBI.

Ontology for phenotypes - user survey:

- How many DCC go granular at cell phenotype level? Is this richness relevant to the scope of the CFDE Portal? What implications are on the C2M2 model?
- Srini Ramachandran: should it include clinical phenotype as described using SNOMED in the context of molecular characterization by metabolomics?
- Do we need to include model organism phenotype ontologies (ZPO,XPO,DPO)?
- AB: request to have DCC provide a use case detailing how phenotypic information is stored and used in the DCC. This is to be used to shape future development.

- JY: maybe we should also consider *not* including the notion of phenotype at all because it is a big area of development. OW expands on the fact this is not a solved pb and is a community problem.
- AB: we don't need an all or nothing approach either, we can index what is provided by the dcc.

JY subgroup:

- working on a manuscript about aspects of interoperation and enhancing FAIRness on their resources: Glygen + CHEBI + Pubchem -> to be made available via biorxiv

Important note: There will **NOT** be a call on June 30th. Our next call will be July 14th.

During the next 4 weeks, please work on the following homework:

A. Review and make comments on the two new draft RFCs:

- Use of PubChem: https://docs.google.com/document/d/1JV_xMWEV5bl3wWw3s1feomZKr2wlARFEaHw1tLYI7DY/edit?usp=sharing
    - Relevant to review of this RFC is the attached manuscript from Raja's group.
- Use of OBI: https://docs.google.com/document/d/1fTA2O71QkQD_yPmvGr0uz7Vr8FMTzvY3FyW66Y3qQYo/edit?usp=sharing

B. Ponder the below questions regarding capture of phenotypes:

1. What level of granularity is absolutely needed? Would it need to be at the cell level? Always keeping in mind the goals of the CFDE.
2. Could HPO (Human Phenotype Ontology) or MP (Mammalian Phenotype Ontology) work for 80% of DCC needs? (even if there might be a loss of some specificity - again keeping in mind the goals of CFDE)
3. Consider the option of NOT capturing phenotype at all.
4. Explore what the Alliance of Genome Resources is doing (Michelle will inquire but others should of course feel free to explore this too.)
5. How does/should PATO fit in?
6. Please post a few examples of your use of phenotype terms. If you are using an ontology other than MP, HPO, PATO please give a one-line description of that ontology.

Post your thoughts/answers to these questions in the google doc here: https://docs.google.com/document/d/1Lu7BNJM-QX6BOALymepAnLag45pXMbCkbRfh_eOJ4mE/edit?usp=sharing

June 2, 2021

Objective	see agenda	Time	Every other Wednesday 11am Eastern
Leader(s):	Michelle Giglio Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	; Suvarna Nadendla (HMP); ; ; ; ; Owen!; ; ; ; ;Steve Mathias (IDG) ; ; ; ; ; ;Arthur Brady ;Eryk Kropiwnicki (IDG) ; ; Chris Kinsinger (NIH); ; ; ; ; ;Jose Sanchez(CFDE-CC) ; ; ; ;George Papanicolaou ; ; ; ; ; ; ; ; ; ; ; ; Sherry Xie (LINCS); ; ; ; ; Tom Gillespie (SPARC); ; ; ; ; ; ; ; ; ; ; ; ; ; ;Philippe Rocca-Serra (CFDE-CC) ; ; Susanna Sansone (CFDE-CC); ; ; ; ; ; ;Dan Lyman (GlyGen) ; ; ; ; ; Bernard de Bono (SPARC); ;Asiyah Lin (NHGRI) ; ; ; ; ; ; ; ; Haluk Resat; ; ; ; ;John Erol Evangelisa (LINCS, IDG) ; ; ; ; Jeremy Yang (IDG); ; ;Mano Maurya (MW) ; ; ; ;Marisa Lim (CFDE-CC) ; ; ; ; ;Jonathan Silverstein ; ; ; ; ; ; ; ; Cristian Bologa; Raja Mazumder (GlyGen)		

Agenda Item	Action Items	Owner
● **updates** ○ **RFCs** ○ **new terms**	need RFCs for process for new term requests	
● **guidelines on term use for consistency** ○ **who will work on this?** ○ **subgroup?**		
● **follow up from anatomy discussion** ○ **use of Uberon** ○ **incorporation of InterLex**	RFC for use of Uberon: produce an initial draft	**Bernard**
● **generation of slims** ○ **plan for initial work on portal** ○ **which ontologies first?** ○ **who will help?**		
● **ontology for chemicals** ○ **there is a proposal to use PubChem** ○ **are there counter proposals?** ○ **how do we proceed to a decision?**	Jeremy Y volunteered to draft RFC.	

Notes and actions:

- Guidelines for consistency
  - will do an automated check after the June submissions
  - should give the DCCs a heads up this is coming
  - internal DCC consistency needs to happen first
  - will publish guidelines for new DCCs to use when they join

- Biomarkers - how to capture this
    - to get involved contact **Raja**
        - (C2M2 workspace) Slack: **#biomarker-datamodel-ontology**
- RFC for use of Uberon
    - **Bernard** to produce an initial draft
- InterLex:
    - used for anatomy but it is also domain agnostic
    - Already accepts new terms across different projects, and we just need tags to track requests from CFDE
- Are there other term needs that have not been yet communicated to this group?
    - Jonathan flags that this is an ongoing process
        - HUBMAP OBI needs have been communicated to Michelle
    - Anatomy terms will go via the Anatomy WG for vetting first
    - please let Michelle know if you have new term needs
- Slims
    - to drive queries and displays in the portal only - they do <u>not</u> affect submissions - DCCs should submit the most granular terms that make sense for their data
    - The portal group is discussing implementations of the slims
        - In this ontology WG we will start to create the slims, via RFCs to create particular one;
        - for portal development CFDE-CC is starting with ‘test slims’ to drive the implementation in the portal
    - Priority: Anatomy, Assays; also explore what other slims are already available.
- Chemicals:
    - PubChem (a vocabulary) is integrating ChEBI (an ontology) - it appears that the two will be able to work together quite well
    - **Jeremy** to write the RFC; bounce off Raja, then bring to this group
- Next steps:
    - connect anatomy, phenotypes and disease and how these are represented in the C2M2
    - Bernard and the AWG have already done a lot of work in this area
    - we’ll need to determine what is the appropriate level to model this in C2M2 so as to capture what is needed for CFDE goals, but not try to capture the entire complexity that may be beyond the scope of CFDE
- Announcement for a relevant workshop in July: WSBO2021: *Workshop* on Synergizing Biomedical *Ontologies*

May 19, 2021

Objective	see agenda	Time	Every other Wednesday 11am Eastern
Leader(s):	Michelle Giglio Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	; Suvarna Nadendla (HMP); ; ; Michelle Giglio; ;Sherry Xie (LINCS) ; ;Jose Sanchez (CFDE-CC) ; ; ; ; ; ; ; ; Bernard de Bono ; ;Steve Mathias (IDG) ;Philippe Rocca-Serra (CFDE-CC) ; ; ; ; ; ; Chris Kinsinger (NIH); ; ; ; ; ; ; ; ; Jeremy Yang (IDG); ; ; ; ;Jeremy Walter (CFDE-CC) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;Marisa Lim (CFDE-CC) ;Eryk Kropiwnicki (IDG) ; ; ; ; ; ; ; Dan Lyman (GlyGen); Susanna-Assunta Sansone (CFDE-CC); Mano Maurya (UCSD, MW); Srinivasan Ramachandran (MW, UCSD); ; ; ;		

Agenda Item	Action Items	Owner
● updates ○ DOID RFC progress		
● Anatomy WG report by Bernard de Bono and Tom Gillespie	●	

Notes:

Bernard de Bono: creation of annotation pipeline for anatomy.

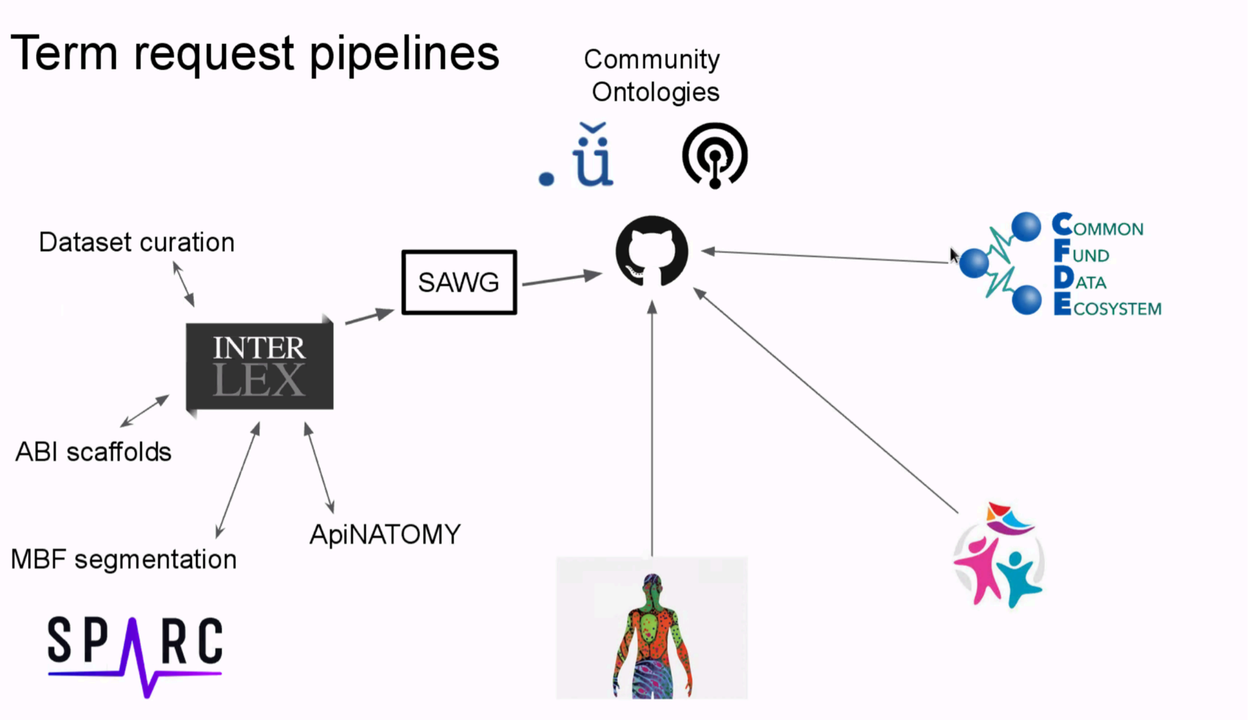
- Document available here: cAWG Anatomy KM Report https://docs.google.com/document/d/1aVvYj6Vi_hbkcC7dWLHE1HGoMn-FRh0-Rzqpu5AIsQs/edit#
  - 2 main messages:

  -standardize as much as position

  -be aware of the relation between entities
- Anatomy terms surveyed over 7 dcc.
- Hubmap uses ML based annotation
- Kids' first : create a pediatric extension to cancer terminology and birth defects
- MW: advanced representation based on snomed-ct
- MotrPAc: initially thought to have limited needs (muscle, fat, blood) but more complex (more on this later) -> main use case about 'exercise physiology

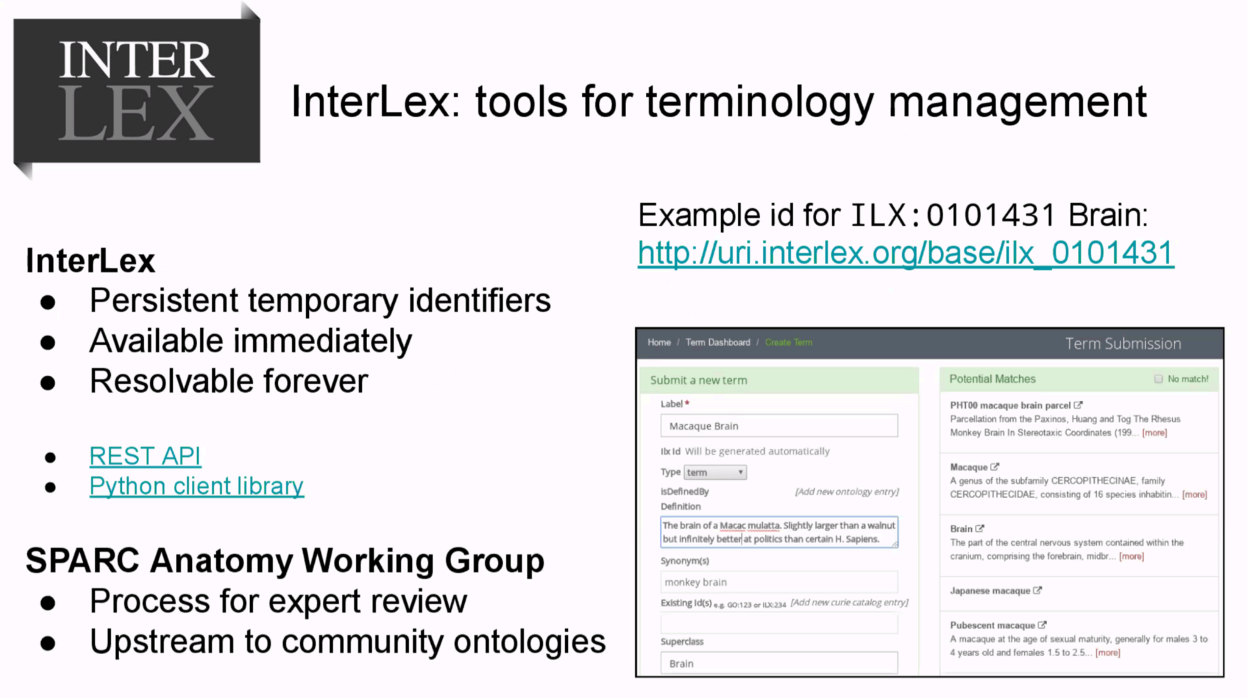
- Most need: crosstalks between uberon and HPO
- Pb down to implicit partonomy of organism part which does not sit well with the phenotypic is_a explicit hierarchy. This undermines calculation cross dcc
- This became apparent when dealing with MotrPac data. SOP recorded many physiological parameters. VO2peak is a complex phenotype, in fact a derivative measure from other phenotypic observations. -> 10 phenotypes need to be recorded to calculate VO2peak.
- Term requests pipeline needs to be allow all DCC to know about temp identifiers
- Need to relate phenotypic measurement to anatomical locations.
- Questions: if we are to use the power of reasoning, how does this sit with the search engine capability of the current service.

Bernard de Bono slides:
cAWG Report discussion slides
Tom Gillespie slides:
SPARC term request pipelines and InterLex


Term request pipelines

- Positive feedback loop with community ontologies.
- However the turn around time for including new term request remains an issue and warrants creation of temporary ids.

- In our pipeline, any new terms comes to the ‘interlex’ element which makes it available for use while the term is also sent to the SA working group for processing (anatomists), for a subset of those, terms will end up into community ontologies.
- Questions: How to coordinate these terms requests and share this pipeline with other dcc so we are using similar infrastructure.



InterLex: tools for terminology management

Example id for ILX:0101431 Brain:
http://uri.interlex.org/base/ilx_0101431

InterLex
- Persistent temporary identifiers
- Available immediately
- Resolvable forever

- REST API
- Python client library

SPARC Anatomy Working Group
- Process for expert review
- Upstream to community ontologies

Home / Term Dashboard / Create Term

Term Submission

Submit a new term

Label *

Macaque Brain

Ilx Id Will be generated automatically

Type term

IsDefinedBy [Add new ontology entry]

Definition

The brain of a Macac mulatta. Slightly larger than a walnut but infinitely better at politics than certain H. Sapiens.

Synonym(s)

monkey brain

Existing Id(s) e.g. GO:123 or ILX:234 [Add new curie catalog entry]

Superclass

Brain

Potential Matches ☐ No match!

PHT00 macaque brain parcel
Parcellation from the Paxinos, Huang and Tog The Rhesus Monkey Brain In Stereotaxic Coordinates (199... [more]

Macaque
A genus of the subfamily CERCOPITHECINAE, family CERCOPITHECIDAE, consisting of 16 species inhabitin... [more]

Brain
The part of the central nervous system contained within the cranium, comprising the forebrain, midbr... [more]

Japanese macaque

Pubescent macaque
A macaque at the age of sexual maturity, generally for males 3 to 4 years old and females 1.5 to 2.5... [more]

- Interlex: A suite of tools for terminology management.

- REST api used by 2 groups at least in the neuroimaging domain.

http://uri.interlex.org/base/ilx_0101431

- MG: this is indeed the type of tools we’d like to consider/use for the case we are facing.

May 5, 2021

Objective	see agenda below	Time	Every other Wednesday 11am Eastern
Leader(s):	Michelle Giglio Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	; ; ; Owen; ; ; Suvarna Nadendla (HMP); ; ; ; Michelle Giglio ; ; ; ; ; Jeremy Yang; ; ; ; ; ; ; ; ; ; Raja Mazumder (GlyGen); ; ; ; ;Philippe Rocca-Serra (CFDE-CC) ; Susanna-Assunta Sansone (CFDE-CC); ; ; ; George Papanicolaou; ; ; ; ; ; ;Mark Musen (HuBMAP) ; ; ; ; ; ; ;Jessica Binder (IDG-UNM) ; ; ; ; ; ; ; ; ; ; ; ;Dan Lyman (GlyGen) ; ; ; ; ; ; ; Marisa Lim (CFDE-CC); ; ; ; ; ; ; ; ; ; ; ; ; Sherry Xie (LINCS); ; ; ; ; ; ;Arthur Brady (CC) ; ; ; ; ; ; ; ; Steve Mathias (IDG) ; ; ; Daniel J. B. Clarke (LINCS); ; ;John Erol Evangelista (LINCS, IDG) ;Eryk Kropiwnicki (IDG) ; ;Mano Maurya (MW) ; Srinivasan Ramachandran (MW); ; ; ; ; ; Haluk Resat; ; ; ;		

Agenda Item	Action Items	Owner
• updates ○ DOID RFC review ○ slim progress	one week to make comments	PRS
• Increasing consistency in metadata submissions through term use guidelines	•	
• Choosing an ontology for chemicals ○ presentation from Jeremy Yang on PubChem ○ discussion, invite others to present at future meetings		
• Coming May 19th ○ report from Bernard on Anatomy group work		

Notes:

- Slim progresses
    - AB started working on generating slims by semi-automated processes. Ongoing work,WIP to be discussed in future calls.
    - JY: slims are to be used in front-end.
    - MG: yes to ease search to avoid overwhelming users with long lists of terms.
- Metadata annotation
    - Addressing the issue of annotation inconsistencies within and across DCC (e.g. EDAM term selection for `data type` or `file format`)

    - ***Agreed***: for compressed archives holding a single file, the data format used should be that of the uncompressed file. Add a field to C2M2 that is compressed file format.
        - Comments:
            - It would break downstream use (e.g. in workflow and programmatic use)
            - More than one mime-types should be supplied, one for the compressed form and that of the data file . limitations caused by mime-type themselves which aren't rich enough to represent the diversity of file format.
            - AB: we should add a 'compression type' attribute, to allow specifications of that information.
- **Presentation by Jeremy Yang: {link}**
    - Discussion about which resources to use for chemicals.
        - Depending on the use cases, the best resource to use will vary
        - Pubchem way has huge uptake, mapping to many other resources, RDF representation but there are also limitations. CHEBI for instance provides an ontological way to navigate data which can be beneficial. CHEBI has flexibility to represent glycan (owing to the ability to accept uncertainty, absence of linkage (a requirement for submission to Pubchem). But Chebi has a less established maintenance infrastructure.
            - Probably to work with both resources and have a hybrid approach
        - AB: creation of a molecular entity object in c2m2 model. Ask the group for list of attributes required for this entity.

April 21, 2021

Objective	see agenda below	Time	Every other Wednesday 11am Eastern
Leader(s):	Michelle Giglio Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;Michelle Giglio ;PRS ;Amanda Charbonneau ;Susanna Sansone ; ; ; ; Mano Maurya; ;Arthur Brady ; Eryk Kropiwnicki ; ; ; ;Steve Mathias (IDG) ; ; ; ; ; ; ; ; Owen White ; ; ;Bernard de Bono ; Jeremy Walter (CFDE-CC); ; ; ; ; ;Chris Kinsinger ; ; ;Jessica Binder(IDG) ; ; ; ; ; ; Sherry Xie (LINCS); ; ; John Erol Evangelista (LINCS, IDG); ; Daniel Clarke (LINCS); ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;Jeremy Yang ; ; ;Marisa Lim (CFDE-CC) ; ; ; ; Dan Lyman (GlyGen); Raja Mazumder (GlyGen) ; ; ; ;George Papanicolaou ; ; ; ; ;		

Agenda Item	Action Items	Owner

● updates ○ EDAM RFC ○ new term requests		
● revisit decision on ontology for disease		
● draft RFC for disease (pending outcome of above item) https://docs.google.com/document/d/1OS_69jvdexMvH9KSptGDAKAXCzyDtca2M0VjsGXImEE/edit?usp=sharing	●	
● Exploring metadata in the CFDE Portal ○ improving visualization ○ increasing consistency		
● next areas to work on ○ chemicals ○ data use rules ○ anatomy		

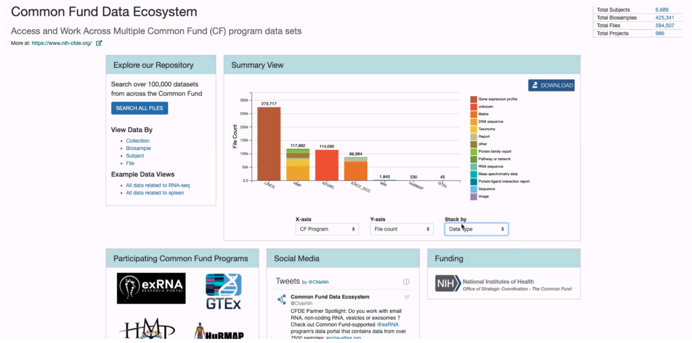
Notes:

- **Ontology for disease**
  Discussion ongoing on which resource to use (DO or MONDO).
  - JS raises the point that the question of which resource to use would be relevant when contrasting UMLS vs DO but less obvious when comparing DO and MONDO?
  - MM: what is the longevity of these artefacts? Given the complexity of MONDO, how will it be maintained in the future? So that should be a criteria for selecting a resource.
  - MG: DO is currently funded by NHGRI but it is fair to acknowledge this is an issue for these resources.
  - JS: "fruit salad" vs "pineapple" pb.
  - SNOMED-cT is the reference but is complex and DO has been specifically developed to help in the problems faced in cfde. Since DO xref Mondo, the choice should be to go for DO with the understanding of the constraints defined by CFDE.
  - AB: we are ready to learn and iterate (to echo the request by Jeremy Yang)
  - RM: document edge cases and document a path for accommodating those (e.g relying on HPO for specific phenotypic description
    - **Motion to use DO - *agreed***:
      OW: strongly in support of reaching a decision .
      JY: we are comfortable with the decision :).
      DC: there were a couple of questions on slack hence revisiting the decision in order to allow people to voice any concerns.

- **Draft RFC for disease**:

    - BdB to add comments/questions to the DO -RFC

- **The CFDE Portal**
    - AC presents dataportal UI summary view and some ontology dependent searches (Anatomy, Data Type)



Common Fund Data Ecosystem

Access and Work Across Multiple Common Fund (CF) program data sets

More at: https://www.nih-cfde.org/

Total Subjects	6,689
Total Biosamples	425,341
Total Files	594,507
Total Projects	986

Explore our Repository

Search over 100,000 datasets from across the Common Fund

SEARCH ALL FILES

View Data By

- Collection
- Biosample
- Subject
- File

Example Data Views

- All data related to RNA-seq
- All data related to spleen

Summary View

X-axis
CF Program
Y-axis
File count
Stack by
Data type

Participating Common Fund Programs

Social Media

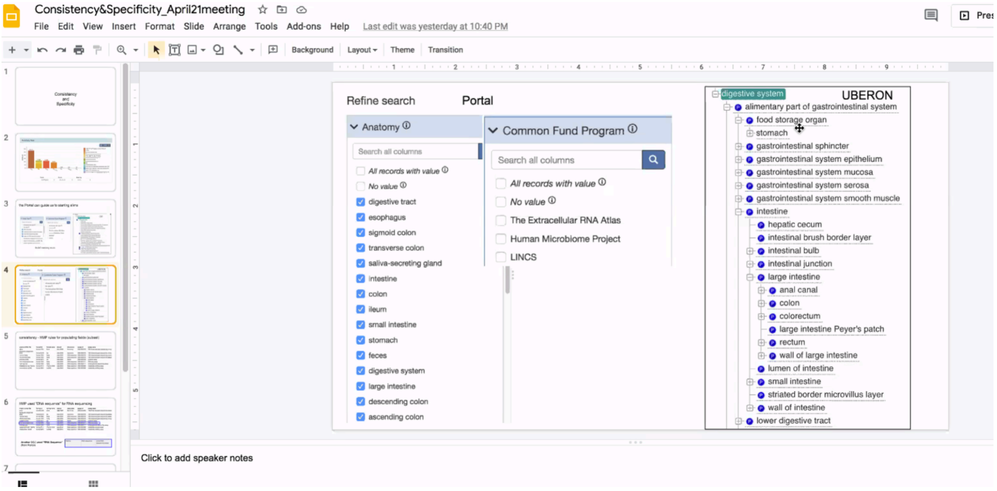
Tweets by @CfdeNih

Common Fund Data Ecosystem @CfdeNih
CFDE Partner Spotlight: Do you work with small RNA, non-coding RNA, vesicles or exosomes ? Check out Common Fund-supported @exRNA program's data portal that contains data from over 7500 samples: exrna-atlas.org

Funding

National Institutes of Health
Office of Strategic Coordination - The Common Fund

    - The various lists of ontology terms used for each of the placeholder are available from: https://drive.google.com/drive/u/0/folders/1YzegpRP8GauSORUQBLVOb0Vas3KAqI7A
    - Leading into the options/needs to produce ontology 'slims'
    - MG presents the following:



Consistency&Specificity_April21meeting

Last edit was yesterday at 10:40 PM

Refine search — Portal

Anatomy

Search all columns

- □ *All records with value*
- □ *No value*
- ☑ digestive tract
- ☑ esophagus
- ☑ sigmoid colon
- ☑ transverse colon
- ☑ saliva-secreting gland
- ☑ intestine
- ☑ colon
- ☑ ileum
- ☑ small intestine
- ☑ stomach
- ☑ feces
- ☑ digestive system
- ☑ large intestine
- ☑ descending colon
- ☑ ascending colon

Common Fund Program

Search all columns

- □ *All records with value*
- □ *No value*
- □ The Extracellular RNA Atlas
- □ Human Microbiome Project
- □ LINCS

UBERON

- digestive system
    - alimentary part of gastrointestinal system
        - food storage organ
            - stomach
        - gastrointestinal sphincter
        - gastrointestinal system epithelium
        - gastrointestinal system mucosa
        - gastrointestinal system serosa
        - gastrointestinal system smooth muscle
        - intestine
            - hepatic cecum
            - intestinal brush border layer
            - intestinal bulb
            - intestinal junction
            - large intestine
                - anal canal
                - colon
                - colorectum
                - large intestine Peyer's patch
                - rectum
                - wall of large intestine
            - lumen of intestine
            - small intestine
            - striated border microvillus layer
            - wall of intestine
        - lower digestive tract

Click to add speaker notes

    - AB: a suggestion: unclear about the need to make a slim. We could have an algorithmic approach to this so the search hierarchy is updated automatically when new data is added.
    - MG: based on prior exp building slims, i am unsure it would result in biologically meaningful representations.
    - **TODO:** dig up references/bibliography about slims.
      RM: Cancer slim ->https://pubmed.ncbi.nlm.nih.gov/25841438/

JY: +1 to AB's suggestions. We have done something like that when working on Proteins. It seems to make sense.

OW: was AB concern mainly about curational time.

March 24, 2021

Objective	see agenda below	Time	Every other Wednesday 11am Eastern
Leader(s):	Michelle Giglio and Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	; ; ; ; ; ; ; ; ; ; ; ; Owen White; ; ; ; Michelle Giglio (CFDE-CC/HMP); Suvarna Nadendla (HMP); ; ; Srinivasan Ramachandran (MW); ; Susanna-Assunta Sansone (CFDE-CC); ; ; ; ; ; ; ; ;Steve Mathias (IDG/UNM) ; ; ; ; ; ;Jeremy Walter (CFDE-CC) ; ; ; ; ; ; ; ; ; ; ; ; ;Chris Kinsinger (NIH) ; ; ; ; ; Jessica Binder; ; ; ; ; ; Marisa Lim (CFDE-CC) ; ; ; ; ; ; ; ;Bernard de Bono ; ; ; ; John Erol Evangelista(LINCS, IDG); Daniel Clarke (LINCS); ;Mano Maurya (UCSD) ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ;Arthur Brady ; ; Haluk Resat; Eryk Kropiwnicki (IDG); ; ; ; ;Sherry Xie (LINCS) ; ; ; ; ; ; ; ; ; ;George Papanicolaou (NIH OD) ; ; ; ; ; ; ; ; ; ;		

Agenda Item	Action Items	Owner
● finalize EDAM RFC for circulation to rest of CFDE	Completed, pending edit on the process (see notes)	
● updates from DCCs on how attempts at use of DO, OBI, and EDAM are going	● Add a information about how contents of tar (and other archive) files will be captured as the C2M2 work progresses ● Look into why EDAM does not have these data file types	Arthur Michelle
● discuss OBI/BAO mappings	● Check OBI cross-references, and explore route to link BAO and OBI more closely	Michelle
● ontology slims - which ones are needed?, who will work on this?		

● what next? - possibly anatomy in collaboration with anatomy group? relationship between anatomy disease and phenotype	● Present the work of the Anatomy WG to this Ontology WG, on (or after) April 21st	Bernardo
● new term needs ○ Metabolomics - liquid chromatography mass spec assay terms in OBI, general mass spec issues in OBI with regard to planned process vs assay ○ LINCS OBI terms - pending ids in use, finalization in progress (ball in Michelle's court)		

Notes:

- EDAM RFC completed
  - Revise the text on EDAM process and submission of new terms
- Experience on attempts at use of DO, OBI, and EDAM are going
  - OBI: address lack of MS terms for metabolomics
  - DO: has anyone done the mapping with this ontology?
    - Jeremy: existing datasets are also not well defined, so the mapping will be an iterative work
    - More work will be done when the clinical WG is up and running
    - Disease and phenotype: DO, DOID, MONDO and HPO
      - Tom G: we have used MONDO as there is integration with OBO Foundry ontologies
      - Jeremy: separate the discussion about the various disease ontologies, and which one to use, from the rest of clinical ontologies
      - Allison (via slack): review of disease ontologies; also on MONDO "a couple quick ones come to mind that i'm aware of folks looking at recently, one is lung cancers and the structures you find in NCIt: https://www.ebi.ac.uk/ols/ontologies/ncit/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FNCIT_C4878 (and MONDO brings a lot of this in as well: https://www.ebi.ac.uk/ols/ontologies/mondo/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FMONDO_0008903) which includes a lot of knowledge of both the anatomical information, but also about staging systems and other clinically relevant knowledge, and then others are some of the congenital heart defects like tetralogy of fallot: https://www.ebi.ac.uk/ols/ontologies/mondo/terms?iri=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FMONDO_0008542 which

provides deeper information about the anatomical relationships, but also that it's a mendelian disease and pretty extensive linking to other code systems”.

        - Bernard: after April 21st the DCCs can report more on this, when Anatomy WG meets; we will progressively need to link anatomy to phenotypes and phenotypes to diseases
    - Data file types: we need capability to describe the file type and what it is under a compressed file
        - Current C2M2 model only allows to define the size but not the type, such capability will be added for people to add a manifesto to explain what is in the compressed file
        - This info is key for computational pipelines
        - **Action**: ***Arthur*** add a manifesto (to explain what is in the compressed file) as the C2M2 work progresses
        - **Action**: ***Michelle*** will look into why EDAM does not have ziped data file types
- OBI/BAO mappings
    - Automated tools do not surface a lot of overlaps
    - We will continue review this point as we get more clarity on mapping
    - For the moment we will continue to use OBI, unless DCCs flags issues, but this does not stop DCCs to use BAO
    - Jonathan S: suggests we put terms from BAO into OBI as external references
        - **Action**: ***Michelle*** check OBI does with cross-references, and explore route to create more linkages
- Ontology slims - which ones are needed?, who will work on this?
    - Tom G: we maintain slims, if you have clear use case it is worth, otherwise maintenance is costly
    - Jonathan S: worth doing only if we have specific use cases
    - Michelle: general strategy would be to select of high level terms and slim up the rest to those terms; question is what do you do with terms down in the tree that do not fit in the slimmed version; primary slims’ use case is display and UI
    - Owen: help search and presenting the data more practically; collapsing and displaying info in the UI
    - Jeremy: high level grouping and classifications
    - **Action?**: ***?*** put together slims for the UI
- Relationship between anatomy disease and phenotype
    - **Action: *Bernard***: will present the work of the Anatomy WG to this Ontology WG, on or after the April 21st;
    - Tom G: we also need a term request workflow, and educational around this
    - Bernard: Human anatomy ontology is for adult, and this is an issue for Kids First

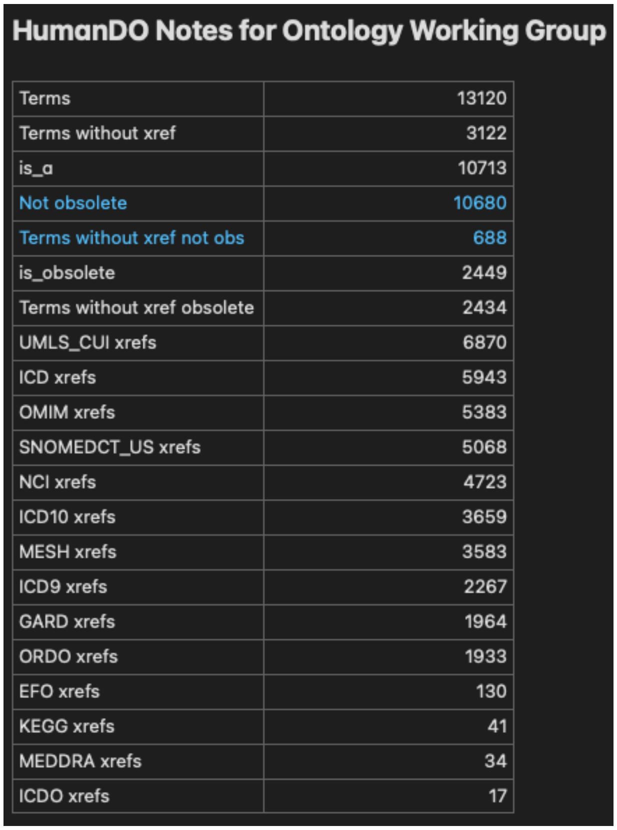
March 10, 2021

Objective	see agenda items below	Time	Every other Wednesday 11am Eastern
Leader(s):	Michelle Giglio and Philippe Rocca-Serra	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	;Owen White;Eryk Kropiwnicki (IDG-CFDE);;;;;;;;;;;;Sherry Xie (LINCS);;;;;;;;;;;;;Suvarna Nadendla (HMP);Philippe Rocca-Serra;;Michelle Giglio ; ; ;; ;; ;Arthur Brady; ;; ;; ;; Srinivasan Ramachandran;;;;Mano Maurya;;;;;;;;;;;Jonathan nSilverstein;;;;;;;;;Bernard de Bono;;;;;;;Jeffrey Grethe (SPARC);;;;;;;;;;;;;;;; ; Marisa Lim (CFDE-CC) ;; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Jessica Binder; ; ; ; ; ;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;Daniel Clarke (LINCS);John Erol Evangelista (LINCS, CFDE);;;;;;;Chris Kinsinger; Susanna-Assunta Sansone (CFDE-CC); ;;Jeremy Walter (CFDE-CC); Raja Mazumder (GlyGen)

Agenda Item	Action Items	Owner
● propose use of slack chat instead of zoom chat		
● review and finalize OWG Charter	●	
● discuss use of DOID in light of information collected as part of homework from last meeting	●	
● discuss need for OBI-BAO mappings and possible branch of BAO that could be used intact	●	
● review draft of RFC for use of EDAM	● https://docs.google.com/document/d/1HQPtk6381Yncxp8Apzxyv8YwPNYIcvWEAH0gKraqKHE/edit	
● new term requests	●	

Notes:

- No objections to using slack channel rather the zoom chat for centralizing discussions.
- Creation of a Clinical Data working group
  - need to liaise, and Bernard de Bono indicates he could play that role (Work with Shankar, Srini at MW)
- Jonathan Silverstein produced the results of mapping of DOID to other resources

- 

  HumanDO Notes for Ontology Working Group

Terms	13120
Terms without xref	3122
is_a	10713
Not obsolete	10680
Terms without xref not obs	688
is_obsolete	2449
Terms without xref obsolete	2434
UMLS_CUI xrefs	6870
ICD xrefs	5943
OMIM xrefs	5383
SNOMEDCT_US xrefs	5068
NCI xrefs	4723
ICD10 xrefs	3659
MESH xrefs	3583
ICD9 xrefs	2267
GARD xrefs	1964
ORDO xrefs	1933
EFO xrefs	130
KEGG xrefs	41
MEDDRA xrefs	34
ICDO xrefs	17

  - DO is heavily cross-referenced that it essentially answers the question of which resource to consider for c2m2
  - Very few DOID entities have no xref. (less that 7%)
- Bernard de Bono questions: what kind of relation are used to describe the nature of the equivalence between terms .
- Suvvi presenting DOID vs Mondo mapping (output of EMBL-EBI OXO mapping service).
  - Inspecting the non-overlapping elements from each of the resources
  - Question about any insights about closest neighbour or presence of dangling class
    - Suggestion of considering ontology “slims” (as practiced for a long time with Gene Ontology (GO)
  - JS: knowledge graph approach taken allows effective navigation. DOID owing to the extensive xref justifies the selection of the resource.
  - Todo: need to check which version of DO has been used in the “mapping”.
    - https://mondo.monarchinitiative.org/pages/faq/ It appears a new release comes out every month. (from Mano Maurya)
  - DC: would MONDO satisfy the requirements from the clinical side of things based on xref present in MONDO. From the LINCS point of view, DOID is sufficient.
  - OW: DOID contains the right level the xref and can deliver the service needed. The PI for DOID is very responsive and the added benefit is the proximity (at U Maryland)
- MM: DOID is included in UMLS so “we get all the mappings for free”.
- OW: reminder that all RFC are visibles to DCCs and all participants.

    - AI: ask the wider group to vote on the choice and ask for further comments.
- DC: what about more granular clinical stuff. MG: -> delegate to specialized resources when needed.
- JS: DOID is not “exactly” part of the distribution of UMLS but it is heavily used to xref so many UMLS terms have ‘landing spot” in DOID.
    - This is an “interoperation” point that DOID brings (in contrast to other resources)(DOID covers the ‘diagnosis’ aspect)
    - Not all clinical metadata element need to be coming from DOID especially for descriptors outside the ‘diagnosis’ description.
- From slack:
    - Jonathan Silverstein 4:47 PM

  note, one other opportunity is i can run the process i ran for Uberon and CL against UMLS to discover additional xRefs DO may want to incorporate (i.e. ones that are obvious in strict - conservative - automatic methods) - that is in process i had mentioned before that MAY add about 2600 UMLS xRefs of UMLS CUIs to Uberon and CL (one those groups review the automated suggestions)

    - Raja Mazumder (GlyGen) 4:51 PM

  Here is a DO Cancer Slim work we did which might provide some ideas about Slims within the DO context https://academic.oup.com/database/article/doi/10.1093/database/bav032/2433164

    - allison 4:54 PM

  I'll just note we found DO pretty lacking for cancer types (esp on the pediatric side), we typically use NCIt (the OBO edition makes it a lot more usable), would be great to think of slims there: https://www.ebi.ac.uk/ols/ontologies/ncit

    -

February 24, 2021

Objective	focus on these metadata types: data format, data type, assay type, disease, phenotype	Time	Feb. 24, 2021 11am Eastern
Leader(s):	Michelle Giglio	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	Michelle Giglio (CFDE-CC, HMP);;Steve Mathias (IDG);Suvarna Nadendla (HMP);;;;;Srinivasan Ramachandran (MW, UCSD);; ;Mano Maurya (MW, UCSD); ;;;;;;;;Chris Kinsinger (NIH);;;;Marisa Lim (CFDE-CC);;;;Jeremy Walter (CFDE-CC;;;;;;;Mark Musen (HuBMAP, Stanford);;;;;;;;Daniel Clarke (LINCS);;;Sherry Xie (LINCS);;;;John Erol Evangelista (LINCS, IDG);;;;Amanda Charbonneau (CFDE-CC);;;;;Haluk Resat;;;;;Jeremy Yang ;;;;;;Abhijna Parigi(CFDE-CC) George Papanicolaou (NIH)		

Agenda Item	Action Items	Owner
● review homework results at https://docs.google.com/document/d/1uX42D54i_z75MfgK4PxzZ9xc4WlVoz-i_Hx3-87jD2A/edit?usp=sharing	●	
● discuss data format and data type	● EDAM	
● discuss assay type	● OBI	
● discuss disease	● DCCs will take their disease terms and map them to DOID for coverage, find what MONDO has that DOID does not have	

Notes:

- Data type
    - HMP - EDAM terms are used. EDAM came as the best scoring match at NCBO. Next is ITO. NCIT - the terms are located all over the place and not defined the way HMP wanted.
    - LINC - similar results as HMP
    - Metabolomics - EDAM, CHMO, AFO
    - SPARC - NCIT (imaging, electrophysiology). They will look more into EDAM to see if there is terminology mismatch.
    - HuBMAP - have not searched the terms but are happy to follow the choices being made by the group. (from chat “On assay types, HuBMAP has a number of the newer single-cell ones that haven’t made it to OBI yet, but not clear they are anywhere yet. We previously discussed providing those to be submitted to CFDE then to OBI - there are some considerations we may have for submitting them to OBI directly instead of via CFDE but effect is same.”)
    - IDG - will go with the group's choice.
      Probably EDAM is the best option. Phillippe - EDAM is used to annotate CWL workflows.
      How to handle the situation with delayed EDAM release cycles?
      Use local ids if EDAM does not respond.
- Assays
    - HMP - OBI is used, LABO is another ontology in NCBO search (includes OBI)
    - LINCS - BAO, new terms for LINCS are given provisional OBI ids
    - Metabolomics - OBO and OBI, ONSTR
    - SPARC - no OBI, NCIT top hit. Electrophysiology ontology developed for their terms. They are fine with the idea of getting their terms into OBI.
    - Jeremy (IDG) will look at OBI, but sounds ok to him.
      OBI looks like the best choice.

- Disease
  - HMP - DOID will be using
  - IDG - MONDO (incorporates DOID) and DOID
  - LINCS - MONDO and HPO are used. In searches NCIT, MONDO, DOID came up (from chat “LINCS also has DOID identifiers, and we did a MONDO mapping”)
  - Metabolomics - MONDO/DOID might not work for them. They have clinical profiles (clinical expressions are captured). Will have to discuss their situation separately.
  - HuBMAP (from chat ”We need to be committed IMHO to ensure we enable the concept cross-referencing of DOID to UMLS CUIs - there are many ways to do this, but the CLINICAL world does not use DOID, it uses SNOMED, ICD, CPT so its very important we link via UMLS CUIs which resolve this. Happy to provide linking of these via CUIs but we must be committed to address those that don’t already link.”) and (“Note that we are currently in a licensing discussion with SNOMED for world-wide use of international and us version both for HuBMAP and potentially for Common Fund (we should discuss before it is finalized!!”))
    Mark Musen - For higher granularity DOID is better, for clinical data - SNOMED is better.
    There will be situations with one to many mappings which will be “mapping hell”.
    Jonathan - is going to look into DOID and see if there are mappings to SNOMED/UMLS.
    (from chat Mark - BioPortal automatically maps terms across ontologies using a lexical analyzer. You can look at these mappings directly in BioPortal and see how DOID and MONDO map to SNOMED, ICD, and other clinical terminologies.)

Homework:

- DCCs will take their disease terms and map them to DOID for coverage,
- find what MONDO has that DOID does not have.
- Jonathan/Mark is going to do some analysis to see what are explicitly mapped to UMLS from DOID.

Feb. 10, 2021

Objective	finalize Charter, set initial tasks for the group	Time	Feb. 10, 2021 11am Eastern
Leader(s):	Michelle (and co-leader?)	Where	https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	Jose Sanchez; Owen White (Maryland); Suvarna Nadendla (HMP); ; Philippe Rocca-Serra (CFDE-CC); Susanna-Assunta Sansone (CFDE-CC) ; ; Sherry Xie (LINCS); Erk Kropiwnincki (IDG) ; ; ; Michelle Giglio; ; ; ; ; Jeffrey Grethe (SPARC);		

	; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; ; Amanda Charbonneau (CFDE-CC); ; ; ;Mark Musen (HuBMAP) ; ; Chris Kinsinger (NIH); ; ; ; ; ; ; Daniel Clarke (LINCS); John Erol Evangelista (LINCS, IDG) ; ; ; ; ; ; ; Marisa Lim (CFDE-CC) ; ; ; ; ; ; ; ; ; Jeremy Walter (CFDE-CC); ; ; ; ; ; Haluk Resat; ; ; ; ; Arthur Brady (CFDE-CC); ; ; ; ; ; ; Mano Maurya (UCSD, MW); ;Srinivasan Ramachandran (MW) ; ;Jeremy Yang (IDG) ;Sherry Jenkins (LINCS) ; Tom Gillespie (SPARC);

Agenda Item	Action Items	Owner
• volunteers to co-lead	• Please contact Michelle to volunteer to help lead the group	
• Review and finalize Charter https://docs.google.com/document/d/1Pa1imd3wUIsmd03qCkf0-zzPt4kvJe2AAN9veN5XdhU/edit?usp=sharing	•	
• Review priorities from HW, set initial tasks	• **DCCs, please run your disease, phenotype, data format, data type and assay terms through BioPortal's Recommender service and share the results (as screenshots) with Michelle**	
• group communication	• **Slack + e-mail**	
• future meetings	• **Feb 24, 2021 (11.00 EST)**	

Notes:

- Reach out to cAWg and User Interface Group with Charter.
- Top 2 areas of metadata to focus
  - Phenotype/Clinical metadata and disease (Owen and Michelle) and existing fields in C2M2
  - Anatomy
  - Methods
  - Genes
  - Data types
  - Specimen Type and Location
  - existing C2M2 fields: anatomy, assay type, file type, data type, taxonomy
  - Analysis type
  - Study designs
  - Cellular phenotypes in relation to disease

- First focus is to finalize the existing C2M2 metadata ontologies.
- FHIR describes Clinical Metadata, covers many different aspects. We should be confined to limited aspects.
    - CFDE will not replicate everything that DCCs have. We should focus on metadata that is relevant to CFDE.
- Starting off with Phenotype/disease.Are diagnostic codes used by DCCs? CFDE will look into adding diagnostic codes as a search feature. If this is useful to the broader community then can be a priority. Based on the responses, MONDO and DO for diseases; PATO, NPO, HPO and MPO for phenotypes will be considered for choosing.
- Mark Musen will show us the Ontology Recommenders Service in BioPortal. A list of disease terms are given as input into the service and it will output a list of ontologies having most of the terms based on some metrics.

- Mark demos Ontology Recommenders Service - https://bioportal.bioontology.org/recommender
  Here's a link to the paper on the Recommender service:

https://jbiomedsem.biomedcentral.com/articles/10.1186/s13326-017-0128-y

- HW for DCCs: DCCs use the Recommenders service for their disease terms and phenotypes and see what results they get. Michelle will let DCCs share their results so that they can be discussed in the next meeting.

- Is there a way to export the bioportal recommender output table? It can be assessed through an API and results can be downloaded. But if doing on the website, then it has to be a screenshot.
- DCCs should also use BioPortal's Recommender service for data types. data formats, anatomy and assay types.

As per the responses from DCCs - Assay: BAO, OBI, while HuBMAP is looking at OBI for their list of assay terms.

Feb. 3, 2021 - First OWG meeting

Objective	Get organized	Time	11am EST
Leader(s):	Michelle Giglio, plus TBD	Where	Zoom link: https://zoom.us/j/93940503153?pwd=cWdQQ3VxaWFxVHVxaVp5OVk2N0svUT09
Participants: SIGN IN: Name & Affiliation	Mark Musen (HuBMAP); Suvarna Nadendla (HMP); Michelle Giglio (CFDE-CC) ; Philippe Rocca-Serra (CFDE-CC); Sherry Xie (LINCS); Susanna-Assunta Sansone (CFDE-CC); ; ; ; Jonathan Silverstein(HuBMAP); ;Steve Mathias (IDG) ; ;Sherry Jenkins (LINCS) ; ; ;Jeremy Walter(CFDE-CC) ; ; ; Raja Mazumder		

	(GlyGen); ; ;Arthur Brady (CFDE-CC) ; Cristian Bologa (IDG); ; Amanda Charbonneau (CFDE-CC); Marisa Lim(CFDE-CC) ; ; ; ; ;Jeffrey Grethe (SPARC) ; ; ; Haluk Resat (NIH); ; ; ; ; Daniel J. B. Clarke (LINCS) ; John Erol Evangelista (LINCS & IDG); ; Srinivasan Ramachandran (MW); Rafael Gonçalves (HuBMAP); ;Bob Carter (CFDE-CC) ; Jeremy Yang (IDG);Jose Sanchez (CFDE-CC) ;Tom Gillespie (SPARC); Mano Maurya (UCSD, MW)

Agenda Item	Action Items	Owner
● People introductions (brief) - see what DCCs are represented and who is attending from each	Please note your affiliation with your name above	
● Overview and status of work so far in CFDE-CC	C2M2 is evolving model- even the current levels (0 &1). There is room to add supported ontologies to the existing levels as well as future levels	
● Discuss scope, goals, and priorities	We need to choose which ontologies to support in the C2M2 - poll DCCs for current ontology use - Identify commonalities, overlaps, clashes (maybe have everyone do local translations) - Add new terms to existing ontologies - Expand model to use new ontologies	
Choosing ontologies	Want living ontologies that are still under development. Have experience with adding terms to ontologies to make them fit CFDE better. Takes work, but is useful work.	
Setting initial goals		
● establish charter document		
● establish frequency of meetings	Next meeting 2/10 8am PST/11am EST, same zoom link as today's call	
● Track work in github with issues?		

Notes:

- In discussion of options for when there is a clash (DCCs using different ontologies/CVs for the same metadata type) - MG suggested: 1. decide on one for all to use; 2. keep both (or very limited number) and map between them. It was pointed out that this mapping can be done at the CFDE portal level or it could be done locally at the DCC site - that is the DCC would convert their data to the CFDE standard before submission.
- What does ontology mean here?
    - CV, ontologies, anything standardized is included
    - Want to choose existing ontologies
    - Other metadata schemas/models (not CFDE C2M2) are excluded
        - schemas, model structures, relationships between modeled resources, etc. are right now being digested, designed, prototyped and tested by a second WG dedicated to overall coordination of the C2M2 itself
        - (this C2M2 group will produce an RFC in Q2 2021 fully describing C2M2 Level 1, after which point the group governing Level 1 schema design will expand to include a wider cross section of the CFDE)
    - This group (ontology):
        - will focus on selection, policy, usage, [any other relevant ramifications] covering all C2M2-embedded usage of any **standardized metadata annotation methods**, typically in the form of controlled vocabularies and ontologies
        - oversees considerable overlap with the modeling group in terms of system design, and will work closely with them to coordinate consensus on relevant technical interfaces
- What should we not do next?
    - Strains- no current accepted ontologies
    -
        - Maybe want to treat like variants?
- Including the NIH Common Data Elements to the resource list would be good https://cde.nlm.nih.gov/home
    - Agree that CDEs are important. NCI CDEs are in CEDAR, which. Makes it easy to incorporate CDEs into metadata.
- HuBMAP has about a dozen assays (perhaps two dozen) not in OBI - mostly single-cell - are you recommending we leverage you/CFDE to request those in OBI rather than do ourselves (which would be appreciated) - if so, do we go to you, Michelle?
    - Yes. Would like to do some pre-work to figure out how they relate to what already exists. Then would need to make definitions. Michelle can take it to OBI WG, or you can.
- IDG has specific things they'd like to see in level 1. How do we get changes into there? An RFC?
    - Probably want to add it to level 2 not 1 so we can freeze level 1 and give people a single target to hit.
    - IDG doesn't have biosamples or subjects so can't do level 1, stuck at level 0.
    - Adding new entities will be done by RFC yes. The only constraint is that we are trying to keep level 1 as entities that are broadly useful, level 2 is where we start

    doing more focused entities. So if there are multiple DCCs that can use the entity, it can likely go into level 1.
        - Level 0: shows things exist
        - Level 1: lowest common denominator, lets you see what is available across CF in very basic terms
        - Level 2: Actually model all the useful things we need
- Who is going to do the mapping of terms from two different ontologies for similar concepts? Is that discussed?
    - BioPortal has Ontology Recommenders Service which takes a list of terms and outputs a list of ontologies related to those terms.
    - BioPortal also already has a lot of mappings done, and other resources
    - If a DCC is doing mapping, the CFDE-CC stands ready to help with mapping tasks.

Homework:

1. come up with two things they think are of high priority that we should start with
2. review the draft Ontology WG Charter document - edit, comment
3. let Michelle know if you prefer communication via slack, email, or both
4. (added via email after the meeting) - let Michelle know if you are interested in co-leading the group