Post-Workshop Report of the WCO-2020: **Workshop on COVID-19 Ontologies**

By Asiyah Yu Lin
DATA scholar/Ontologist
NHGRI, NIH

March 2, 2022

Agenda

- Overview of the workshop
- Go through each session, highlight some talks
- Activities afterwards:
  - COVID-19 Ontology harmonization
  - COVID-19 Epidemiological Ontologies harmonization

Workshop on COVID-19 Ontologies

8th Annual Workshop of the Clinical and Translational Science Ontology Group

- Held on Oct. 23, 2020 and Oct. 30, 2020 https://github.com/CIDO-ontology/WCO
- About 60 people from Europe, North America, Asia, and Saudi Arabia joined
- Day 1 Sessions:
  - **Extending existing Ontologies for COVID-19**
  - **Newly developed Ontologies specifically for COVID-19**
  - **COVID-19 ontology resources**
- Day 2 Sessions:
  - **Data standardization**
  - **Drug analysis**
  - **Knowledge and NLP**
  - **Application tools**

Workshop resource

- https://github.com/CIDO-ontology/WCO
- All videos are available at the **Biomedical Ontology World YouTube Channel** https://www.youtube.com/watch?v=5XBYUgMO8WM&list=PLhzFEi0G-n-vjm82aFY1CR8tfuU6MdYtX

Keynotes and invited speaker

- **Dr. Warren Kibbe**, Duke University. Topic: National COVID Cohort Collaborative (N3C) and mapping terminologies in a pandemic
- **Dr. Gregory Pappas,** (currently detailing at ONC) Associate Director National Surveilence, Center for Biologics Evaluation and Research (CBER), FDA. Topic: SHIELD: Extracting Value from Daily Reported COVID-19 Test Data. (SHIELD)
- **Dr. Myles Axton** Editor-in-Chief for Advanced Genetics @ Wiley (now senior advisor at LuminousMindInc.com). Topic: Fair Molecules for COVID-19 Research

Session 1: Extending existing Ontologies for COVID-19

- Homeostasis Imbalance Process (HoIP) Ontology for COVID-19 related infectious process by Yuki Yamagata from RIKEN, Japan
- COVID-19 vaccine clinical trials by Asiyah Lin and Oliver He's lab : extending the Vaccine Ontology(VO) and Clinical Trial Ontology(CTO)
- Medical Action Ontology (MAXO) for COVID-19 treatment terms by Leigh Carmody from Jacksons Lab

Session 2: Newly developed COVID-19 ontologies

- *COVoc: a COVID-19 ontology to support literature triage* by the European Bioinformatics Institute (EBI)
- *A semantic data model for the WHO COVID-19 Case Report Form* by Go Fair group
- *CODO: An Ontology to Capture Data on the Covid-19 Pandemic* from India
- *Coordinating Coronavirus Research with IDO-COVID-19* by John Beverley and Barry Smith
- *CIDO: The Community-based Coronavirus Infectious Disease Ontology* led by Oliver He

Session 4: Data standardization (featuring diagnostic testing and epidemiological data)

- *The DataHarmonizer: A fast, secure application for the harmonization of SARS-CoV-2 contextual data* from Canada
- *Towards a Minimum Information Standard About a Viral Control (MIAViC)* by collaborative group: Asiyah Lin, AdveritasDx and Marc L. Salit
- **Keynote**: SHIELD: Interoperability, Laboratory Data, And Real World Evidence (RWE): Where We Are? By Dr. Pappas

Session 5: Drug analysis

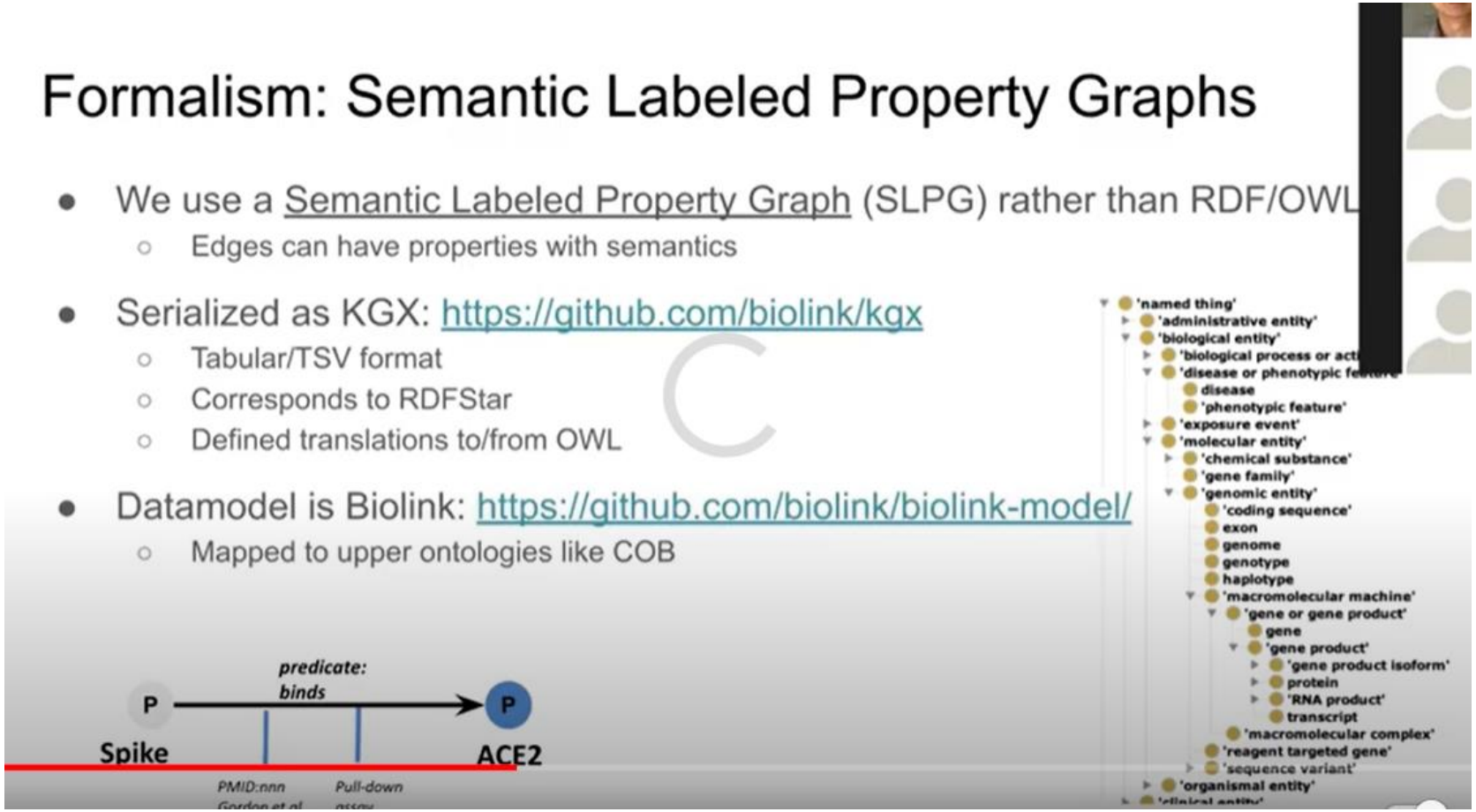
- *AI Powered drug repositioning for treating COVID-19 (*FDA NCTR drug repurposing using AI)
- *A knowledge graph for COVID-19* (Chris Mungal's group, Justine Reese)
- *Candidate COVID-19 Drugs Prediction* (KAUST ontology and AI group led by Robert Hoehndorf in Saudi)

AI Powered drug repositioning for treating COVID-19: network analysis, omics and experimental validation

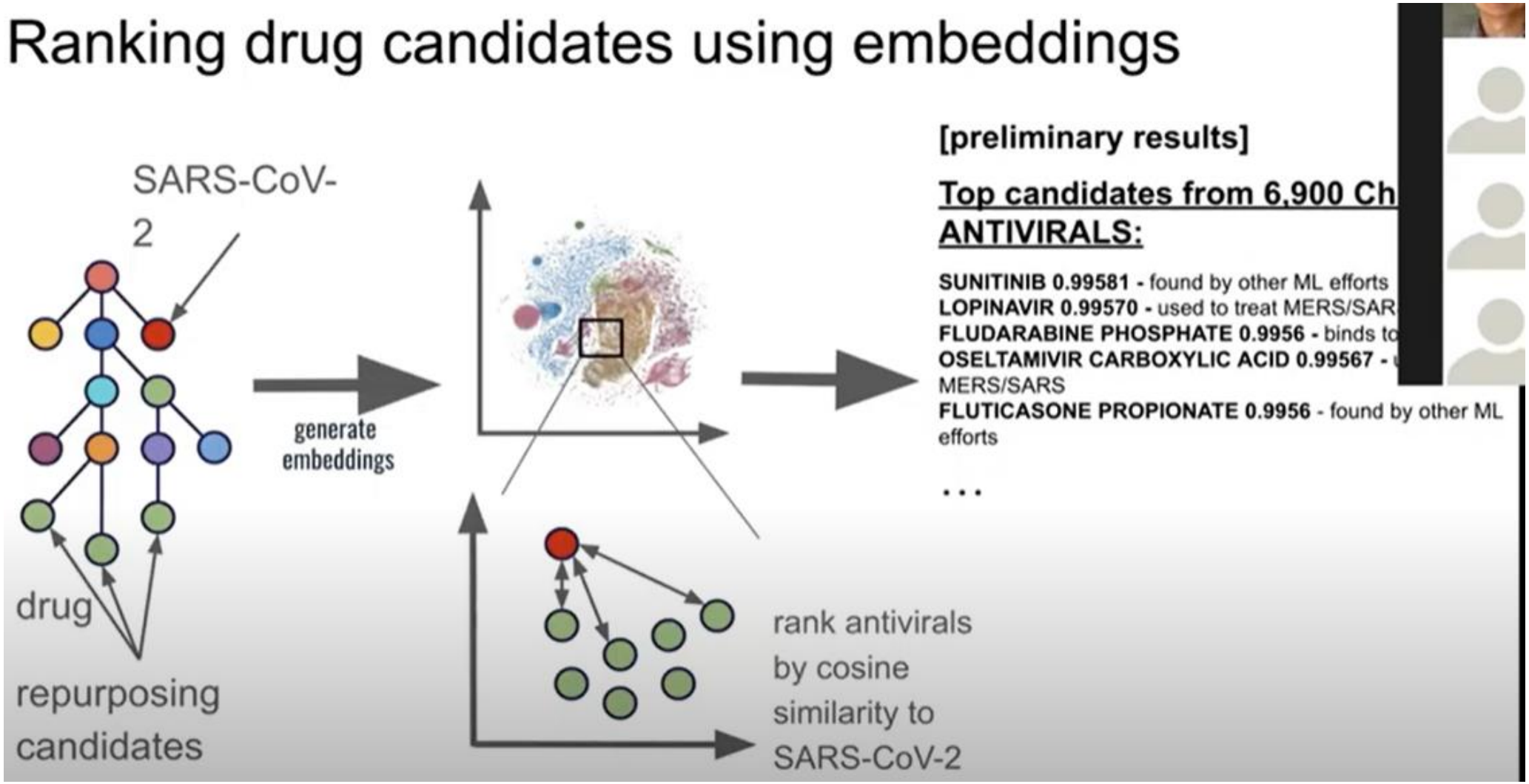
Take-home Messages

- Publicly available ontologies and biological data with AI allows to systematic survey the marketed drugs for their potential to treat COVID19
  - Network pharmacology to prioritize the marketed drugs.
  - Safety of repositioning candidates was prioritized by molecule profiling
- We proposed 24 repositioning candidates for mild and severe patients, respectively.
  - Some of the candidates are already on the clinical trials.
  - These candidates have been tested on two in vitro assays.
- Other repositioning approaches will be further explored in combination with the current findings to provide a better result.

A knowledge graph for COVID-19: Knowledge Graph approach



Ranking drug candidates using embeddings



Take Homes

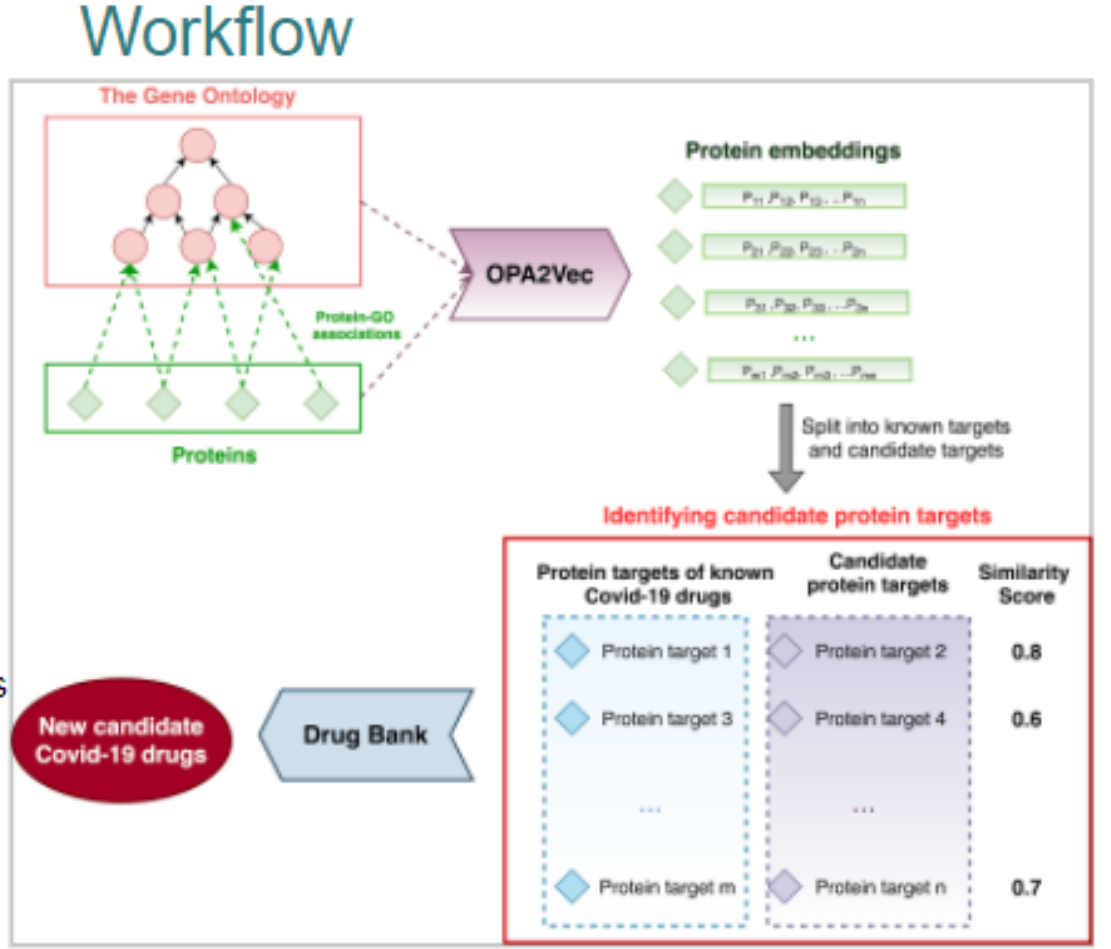
- Knowledge Graphs (KGs) provide ideal substrate for Machine Learning and embedding
- KGs can be used to rank candidates for drug repositioning
  - SLPG (Semantic Labeled Property Graphs) provide an appropriate formalism for operations on KGs
  - More work required on combining different modes of inference (OWL, ML)
- We need unified efforts to build the KG of life
  - We leverage OBO for all ontologies, but need to go beyond ontologies for named entities and linkages
  - Need a parallel effort for KGs
  - Polyproteins pose challenges for building molecular parts list in viruses
- Links
  - https://twitter.com/chrismungall/status/1291463463206715392
  - https://dougsouccdi.wordpress.com/2019/03/14/biological-knowledge-graph-modeling-design-patterns/

Candidate COVID-19 Drugs Prediction: Ontology based ML method

- Using CIDO and OPA2Vec (Ontology-based embeddings)
- Identification of new candidate drugs for Covid-19 based on protein target similarity.

9

- **Step 1:** Generating vector embeddings of proteins using ML-based OPA2Vec.
- **Step 2:** Calculating similarity between known protein targets of Covid-19 drugs and all other proteins to identify *candidate protein targets*.
- **Step 3**: Identifying candidate drugs based on association with the predicted protein targets using the Drug Bank.



Candidate drugs

Drug	CHEBI ID	Confidence score	Clinical trials
camostat	CHEBI_135632	0.89	11
gemcitabine	CHEBI_175901	0.79	0
tamoxifen	CHEBI_41774	0.77	2
imatinib	CHEBI_45783	0.77	5
dasatinib	CHEBI_49375	0.74	0
chlorpromazine	CHEBI_3647	0.74	3
indinavir	CHEBI_44032	0.71	0
azithromycin	CHEBI_2955	**0.69**	114
sirolimus	CHEBI_9168	0.68	6

Session 6: Knowledge and NLP

- *Dark Matter: A Health Canada Solution for Information Retrieval* (Health Canada)
- *Code System Development Protocol for Computable Expression of Scientific Knowledge* (Brian Alper)
- *Standardizing COVID-19 clinical concepts for N3C analytics (Hongfang Liu from Mayo*)

Other sessions: COVID-19 resource and ontology tools

Resource:

Enrique Wulff. *Covid-19 ontology resources for libraries*

Tools:

- Victor Lee. *Value Set-based Approach to Authoring Ontologies*
- ***OntoloBridge*** *the FAIR-bridge between users of scientific vocabulary and ontology creators (*John Turner from Stephan Schurer's lab, Miami University)
- *Extending Machine Learning Capabilities of* ***BioHarmony Annotator*** (Hande Küçük from Collaborative Drug Discovery)

After the workshop

Ontology harmonization

Collaborations

A community effort for COVID-19 Ontology Harmonization

Presented by Asiyah Yu Lin

Data and Technology Advancement (DATA) Scholar,

NIH/NHGRI

September 17, 2021

ICBO 2021

The 12th International Conference on Biomedical Ontologies

The community

:: Yongqun “Oliver” He
:: William Duncan
:: Asiyah Yu Lin, and et.al

CIDO

Coronavirus Infectious Disease Ontology

:: Yuki Yamagata
:: Tatsuya Kushida
:: Hiroshi Masuya

HOIP

Homeostasis imbalance process ontology

:: John Beverley
:: IDO team

VIDO

Virus Infectious Disease Ontology

:: John Beverley
:: IDO team

COVID-19-IDO

VIDO extension

COVoc

Controlled Vocabulary for COVID-19

:: Zoë May Pendlington
:: Paola Roncaglia

CODO

Ontology for Collection and Analysis of COviD-19 Data

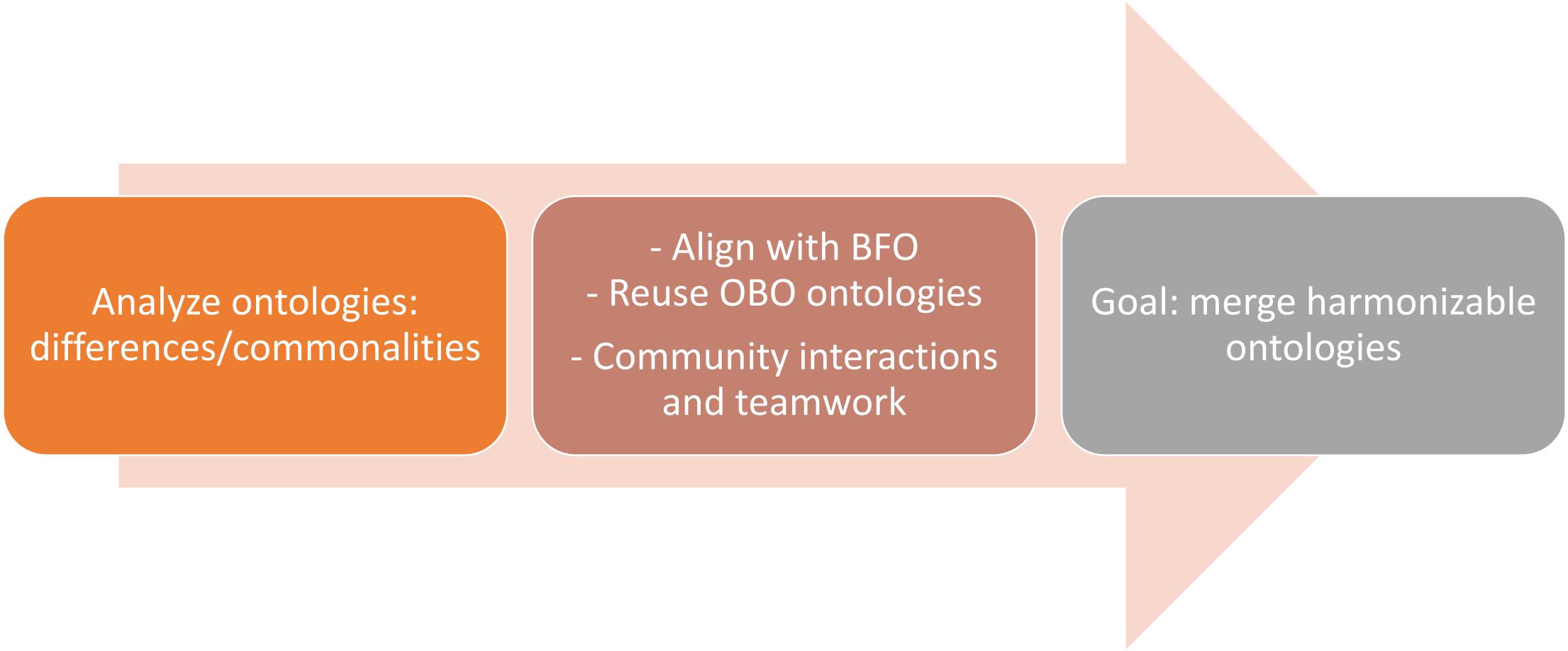
:: Biswanath Dutta
:: Michael DeBellis

MAxO

Medical Action Ontology

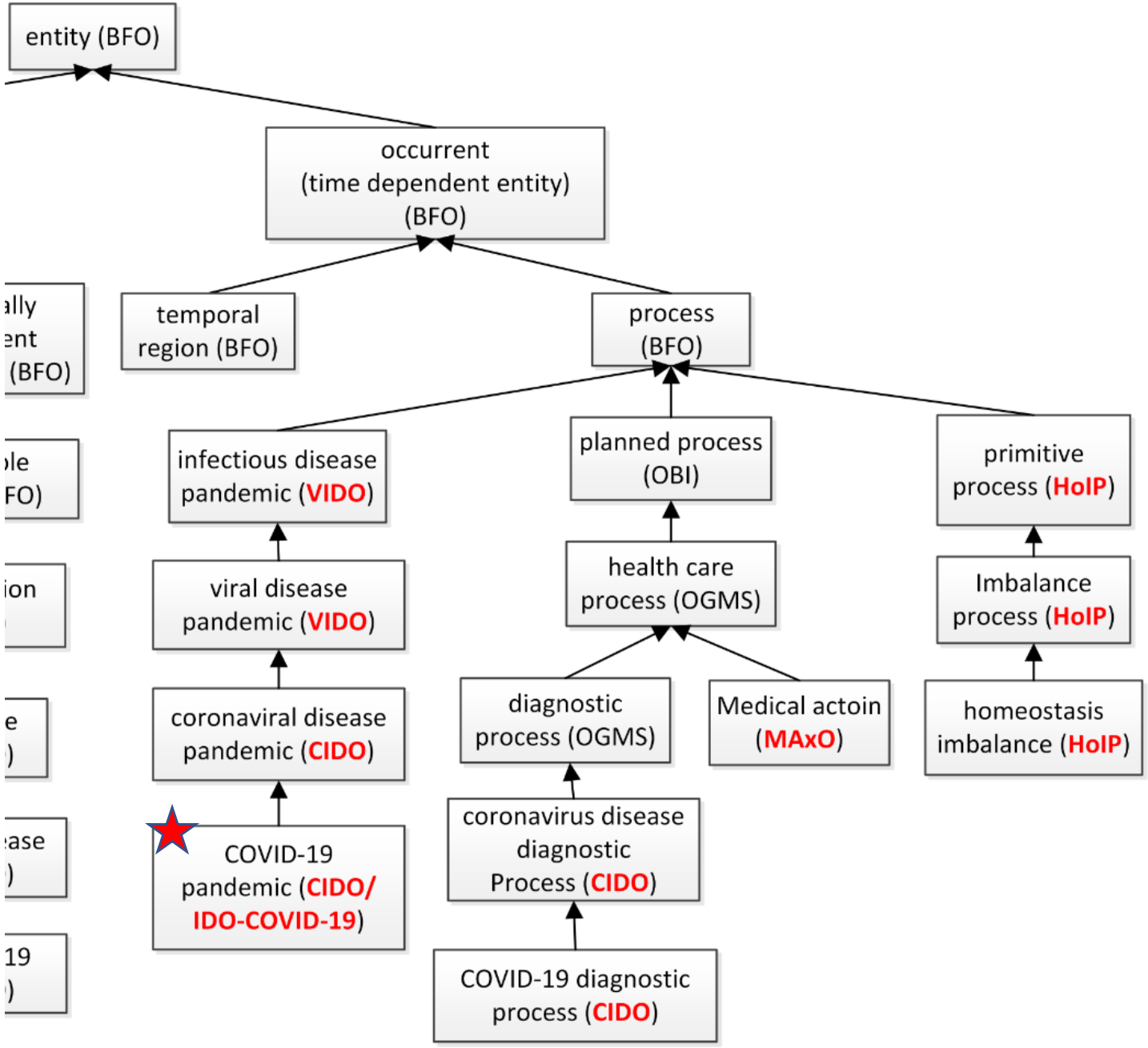
:: Leigh C. Carmody

Motivation: as a community, produce as little ontologies as possible for end users.



Achievements

- CIDO and IDO-COVID-19 merged.
- VIDO and CIDO/IDO-COVID-19 harmonized.
- COVoc submits and reuses CIDO terms.
- HoIP mapping to CIDO.
- CODO aligning to BFO.
- Built a collaborative community aims to harmonization for COVID-19 ontologies



Three tiers of harmonization

- Tier 1: merge harmonizable ontologies
  - individual ontology needs to give up (COVoc and IDO-COVID19)
- Tier 2: build BFO conformed/aligned ontologies
  - CODO and HoIP
- Tier 3: interoperable ontologies (non-BFO and BFO based)
  - Basically, mapping terms
  - Lowest level of harmonization
  - More popular choice
  - Knowledge graph’s rising may help???

Another harmonization activity

- Epidemiological ontologies harmonization (https://github.com/CIDO-ontology/COVID-19-ontology-harmonization/tree/main/EPI-harmonization)
- Mini-hackathon scheduled in March 15th, 2022
- Ontologies:
  - COVID-19 Epidemiology and Monitoring Ontology (CEMO) - Canada https://github.com/NuriaQueralt/covid19-epidemiology-ontology
  - Genomic Epidemiology Ontology (GenEpiO) - Europe https://github.com/GenEpiO/genepio
- Will base on OBI's data related modeling.

Diagnostic Testing collaborations

- APHL (Association of Public Health Laboratories)
- AdveritasDx (https://adveritasdx.com)and fellow collaborators (volunteer group)
- Myself – CIDO – diagnostic testing – FDA EUA testings
- SHIELD project LIVD table
  - https://aspe.hhs.gov/shield-standardization-lab-data-enhance-patient-centered-outcomes-research-value-based-care
  - https://www.cdc.gov/csels/dls/sars-cov-2-livd-codes.html
- Goals:
  - Automated info extraction from FDA EUA website
  - Data harmonization

Other outcomes, collaborations

- Book chapter published by Enrique Wulff
  Features and Limitations of Ontologies for Coronavirus Data Management in Libraries
- Health Canada and FDA interaction on ontology and NLP
- COVID-19 vaccine knowledge base (Cov19VaxKB) by Oliver He's lab
  http://www.violinet.org/cov19vaxkb/
  https://pubmed.ncbi.nlm.nih.gov/34981039/
- Another workshop on Ontology Tools: **WSBO2021: Workshop on Synergizing Biomedical Ontologies** with University of Miami and CDD
  https://www.youtube.com/watch?v=Qa31-0oEO4w&list=PLhzFEi0G-n-v6IlpOXdNGN9yvRsWF0hpF
  https://github.com/OntoloBridge/WSBO
- FDA-NIH ontology round table group: explore ontologies to be included in the FDA data standard catalog*

Final thoughts

- NIH Fundings towards ontology development is limited.
  - I am conducting a NIH ontology investment portfolio analysis.
  - It is time to get commercial partners involved. – SciBite, Pistoia Alliance, etc.
- Ontologists should continue building ontologies to connect data in a meaningful way – it will be certainly used.
- Build FAIR ontologies.
- Harmonize ontologies as much as possible – less ontologies are better.
- Advocate ontology as a data asset in an organization.
- Never an ontology be ready to use off the shelf, the users of ontologies always need to customize public ontologies for their own use cases.

Thank you!

My contact info:

linikujp@gmail.com, asiyah.lin@nih.gov