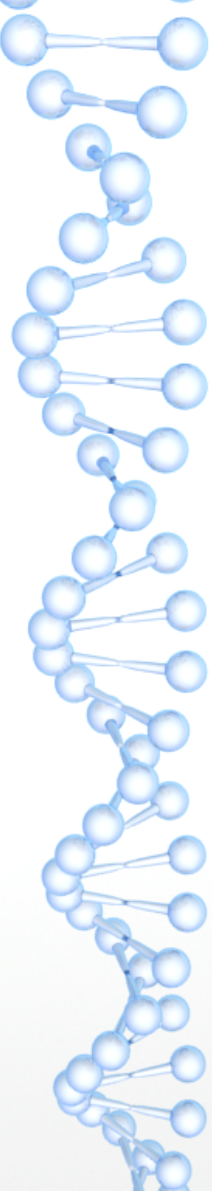


Pandemics Summary

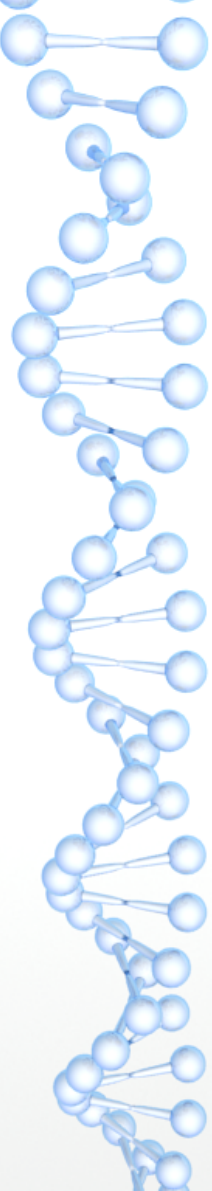
Ken Baclawski
Ontology Summit 2022
30 March 2022



Background

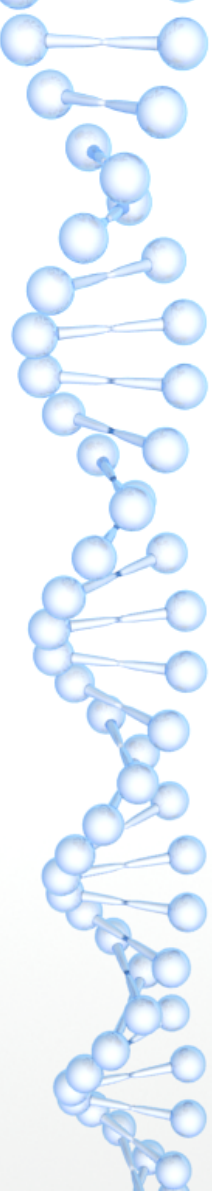
- Pandemics have different characteristics
 - Severity (e.g., cases and fatalities)
 - Period of time
 - Geographic extent
 - Data
- Ongoing global pandemics

Pandemic	Fatalities	Period	Primary Extent
Tuberculosis	1.5M/year	Over 6000 years	Asia
HIV/AIDS	32.7M	40 years	East and Southern Africa
COVID-19	>6.12M	2 years	North America



Pandemic Data is Big Data

- Volume
 - HIV/AIDS 260K papers
 - COVID-19 700+K papers (as of October 2021)
- Variety
 - Lack of interoperability (silos)
 - Hasty publication
 - Ad hoc formats
 - Unspecified semantics
 - Inconsistencies
- Velocity
 - COVID-19 is extremely active
 - Others are slower
- Veracity
 - Mostly accurate
 - Negative correlation with social media sharing
- Value
 - Matthew West will talk about this



Topics

- Goals

- Drug repurposing
- Improved search
- Personalized search
- Contact tracing
- Vaccine distribution

- Challenges

- Lack of interoperability (silos)
- Harmonization
- Limitations of ML methods
- Explainability

- Approaches

- NLP
- Knowledge graphs
- Property graphs
- Embeddings

- Ontologies

- CIDO
- CODO
- COVoc
- IDO-COVID-19
- KG-COVID-19

Disease/Pandemic Design(s)

Upper ontology related terms:
Event, Process, Object, Phenomenon, Action tasks, Plans, Activities, Participation etc.

Core Disease Upper Module

Basic Disease concepts – types, agents, disease situations,
Causal & Participation Disease relations
Risks

Spatial
(GeoSPARQL)

Observation
(SOSA)

Disease Sub-Module 1
Course of disease,
Spread, complications....

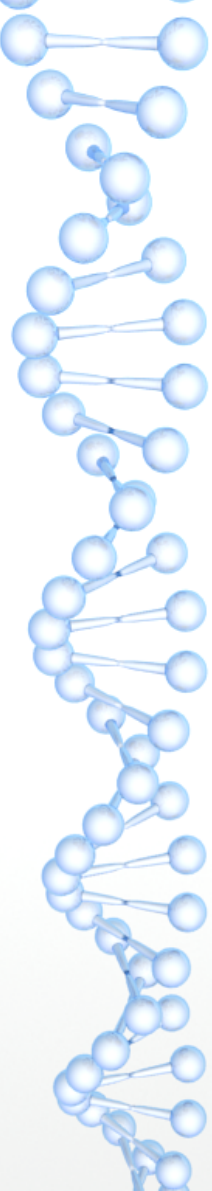
COVID Disease Sub-Module -2....
Interventions.

Temporal
(OWL-Time)

Existing Semantic Resources
Ontologies, Taxonomies, Vocabularies

Applications with their Data

Disease and Pandemic Data Sources
Repositories, LoD, Realtime data, Wikidata etc.



Session on 9 February 2022

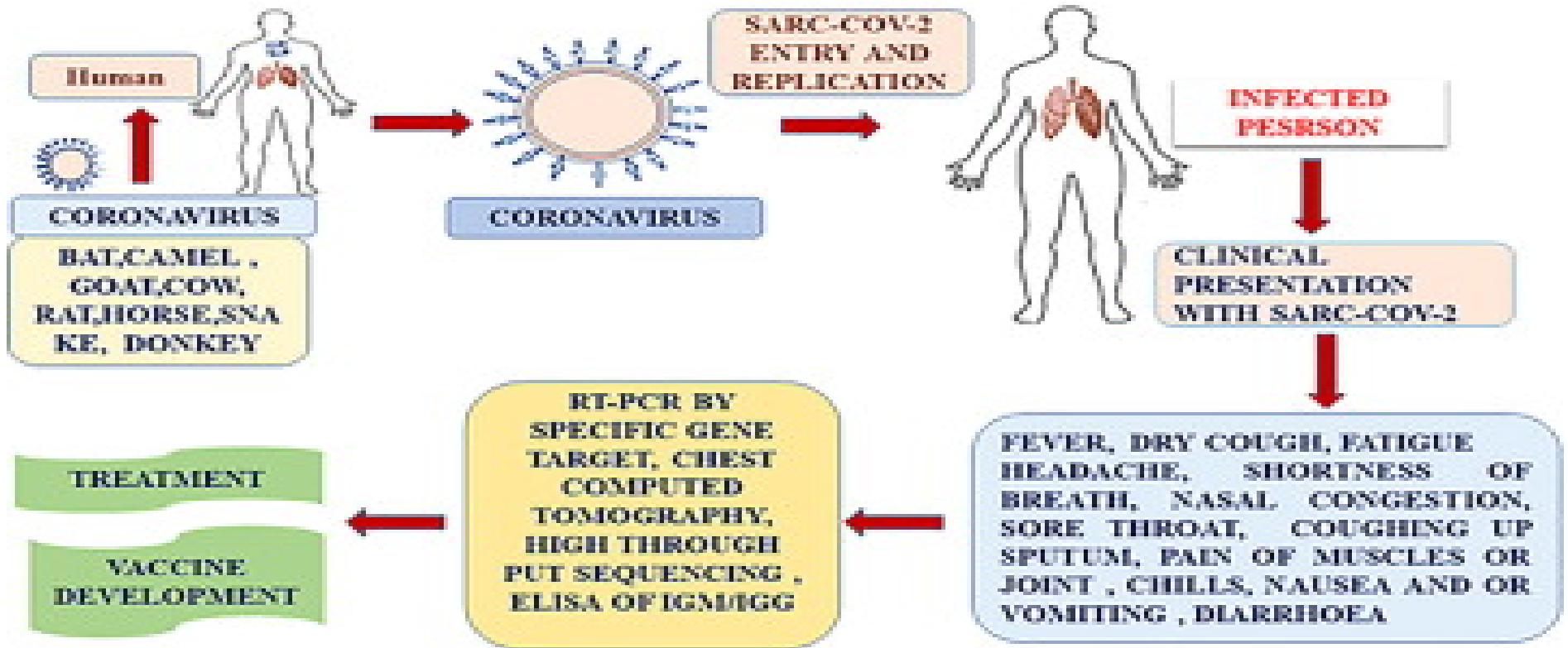
- Chris Mungall and Justin Reese
 - Chris Mungall is Department Head of Biosystems Data Science at Lawrence Berkeley National Laboratory.
 - Justin Reese is a Computer Research Scientist at LBNL who will lead the LBNL team in a new project to develop machine learning tools and best practices to improve the response to COVID-19, leveraging a large range of patient-level COVID-19 data in the National COVID Cohort Collaborative (N3C) Enclave.
- COVID-19 Knowledge Graphs

Pandemics

Session 2 (Feb 9th): Chris Mungall & Justin Reese Covid-19 Knowledge Graphs

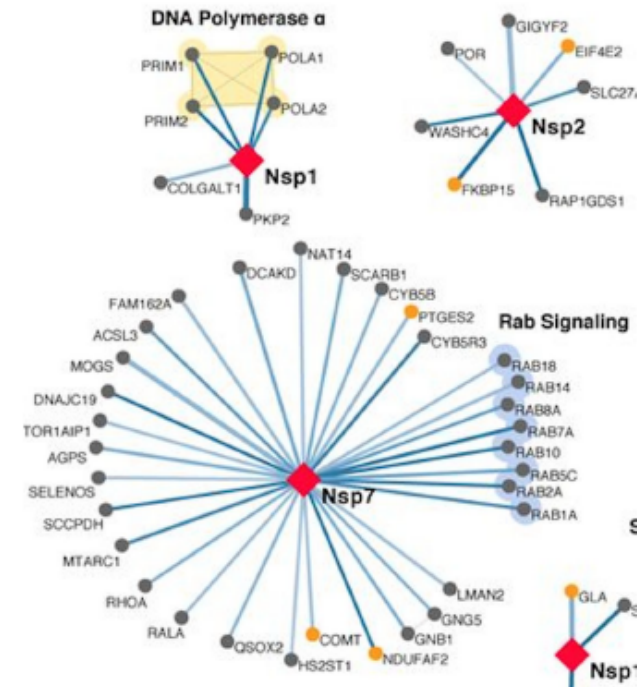
Many semantic resources are available for Disease & Pandemics, but they are scattered, often hastily done and thus somewhat inconsistent.

We can find best sources, explain why these are good & provide ideas on improvements



- Chris Mungall and Justin Reese's work on Covid-19 Knowledge Graphs

KG-COVID-19 Knowledge Graph for COVID-19 Response



Gordon et al, 2020: A SARS-CoV-2 protein interaction map reveals targets for drug repurposing

Goal:

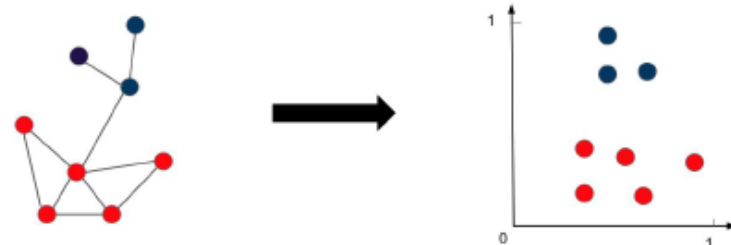
- Use Machine Learning (ML) to make useful COVID related predictions
 - e.g. drug repositioning

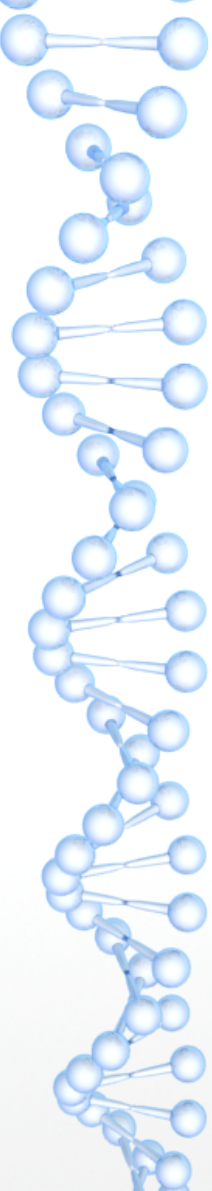
Challenge:

- Data is siloed
- Traditional ML methods don't take into account interconnectedness

Approach:

- Create a COVID-19 *Knowledge Graph (KG)*
- Develop graph *embeddings*





Topics

- Goals

- Drug repurposing
- Improved search
- Personalized search
- Contact tracing
- Vaccine distribution

- Challenges

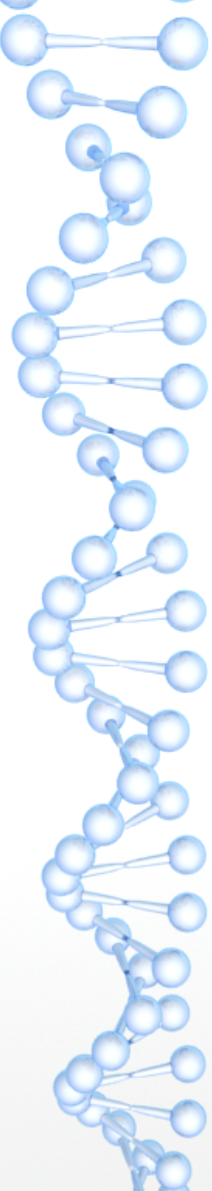
- Lack of interoperability (silos)
- Harmonization
- Limitations of ML methods
- Explainability

- Approaches

- NLP
- Knowledge graphs
- Property graphs
- Embeddings

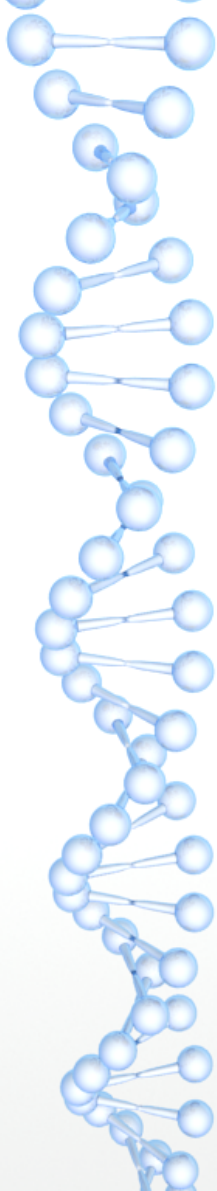
- Ontologies

- CIDO
- CODO
- COVoc
- IDO-COVID-19
- KG-COVID-19
- Other COVID-19 KGs



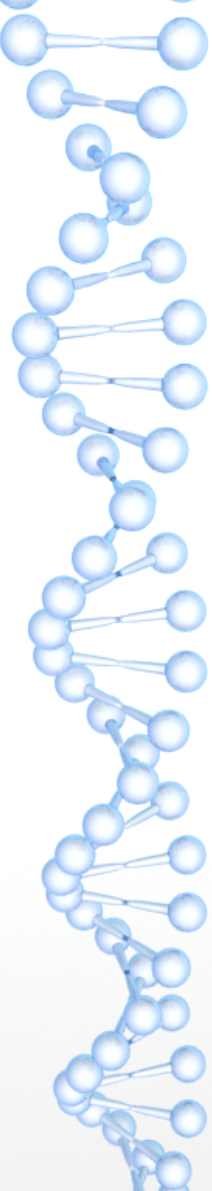
Session on 23 February 2022

- Mark Fox
 - ISO/IEC WG11 Data Standards for Public Health Emergencies and Related Ontologies



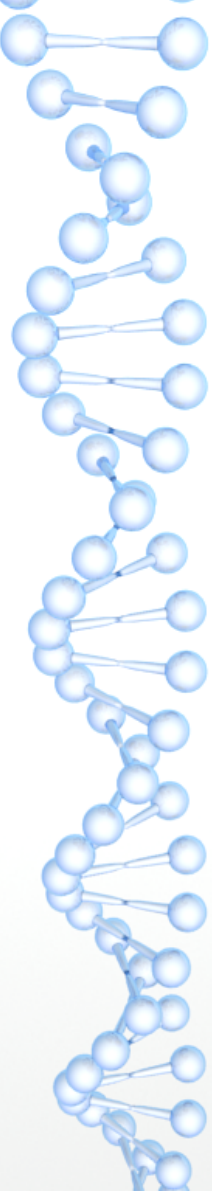
Why a PHE Data Standard?

- We just shut down the world for the last two years because of a PHE!
- In a world of limited resources and time, we need to better understand, coordinate and efficiently/effectively respond to the challenges that cities, states and nations face.
- Students of global cities, like myself, have a deep understanding of the extent to which legacy systems have created data silos across the departments/divisions.
- A first step to breaking down the silos is the creation of data standards.
- Stating that “Data Standards for X” is the most important is not helpful.



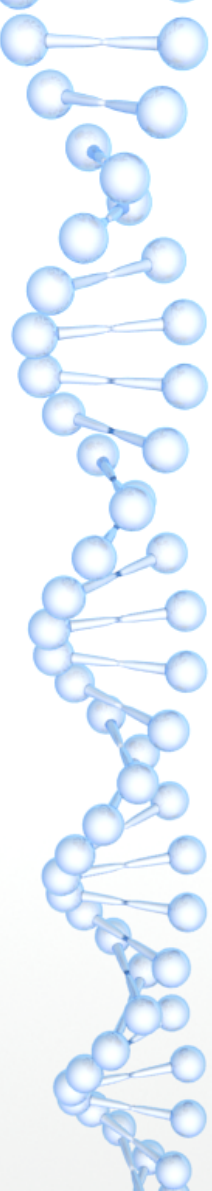
5087-4 Use Cases

1. Contact Tracing
2. Disease Tracing
3. Drinking Water Pollution Management
4. Emergency Medical Resource Management
5. Food Poisoning Management
6. Infection Prevention for Medical Personnel
7. Mental Health Management
8. Stakeholder Collaboration/Volunteer Management
9. Testing
10. Transportation Management
11. Vaccine Distribution Management
12. Vaccine Efficacy Management
13. Vaccination Management



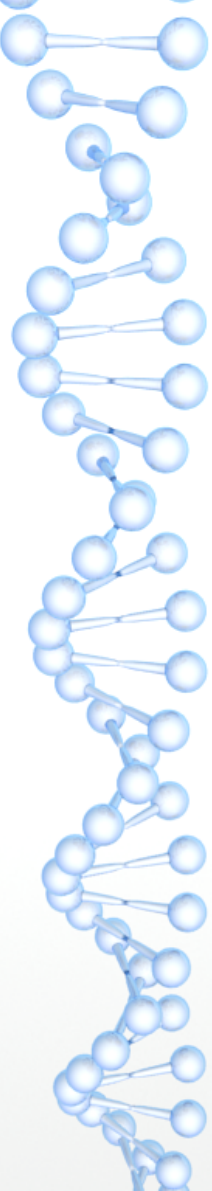
5087-4 Concept Patterns

- | | |
|-------------------------------|-------------------------|
| 1. City | 7. Medical Personnel |
| 2. Disease | 8. Micro-Organism |
| 3. Emergency Medical Resource | 9. Organization |
| 4. Government | 10. Person |
| 5. Logistics and Distribution | 11. Public Health Event |
| 6. Medical Activity | 12. Vaccination |



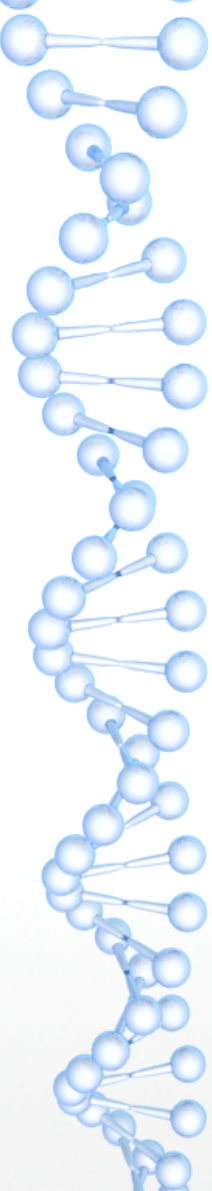
Topics

- Goals
 - Drug repurposing
 - Improved search
 - Personalized search
 - Contact tracing
 - Vaccine distribution
- Challenges
 - Lack of interoperability (silos)
 - Harmonization
 - Limitations of ML methods
 - Explainability
- Approaches
 - NLP
 - Knowledge graphs
 - Property graphs
 - Embeddings
- Ontologies
 - CIDO
 - CODO
 - COVoc
 - IDO-COVID-19
 - KG-COVID-19



Session on 2 March 2022

- Asiyah Yu Lin
- Post-workshop report of WCO-2020: Workshop on COVID-19 Ontologies



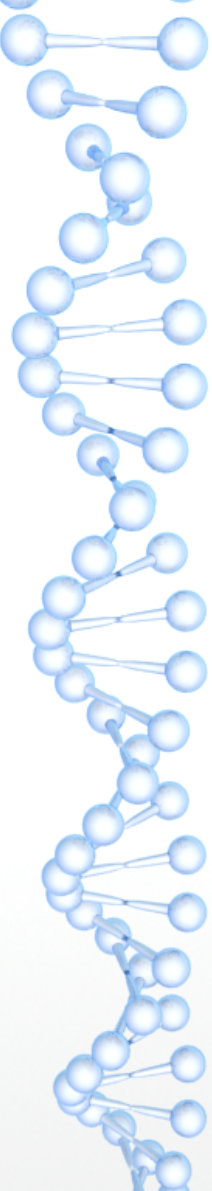
Motivation: as a community, produce as little ontologies as possible for end users.



Analyze ontologies:
differences/commonalities

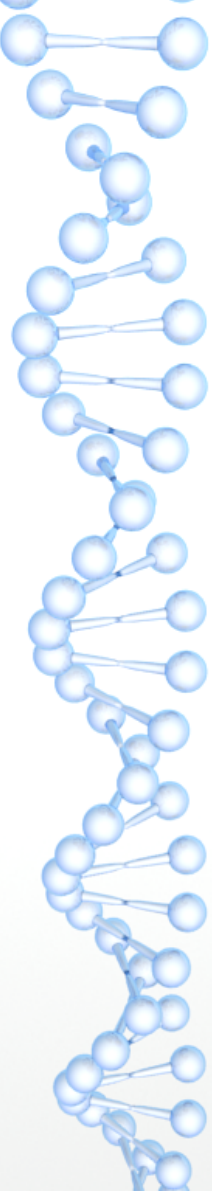
- Align with BFO
- Reuse OBO ontologies
- Community interactions and teamwork

Goal: merge harmonizable ontologies



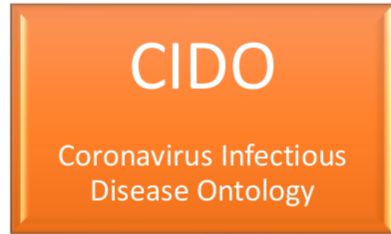
Three tiers of harmonization

- 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???



The community

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



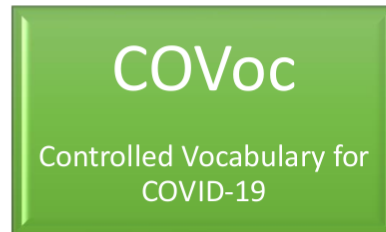
:: Yuki Yamagata
:: Tatsuya Kushida
:: Hiroshi Masuya



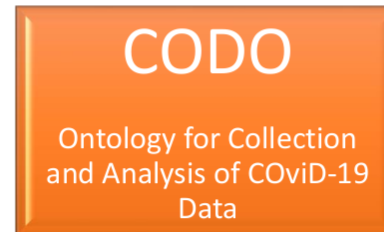
:: John Beverley
:: IDO team



:: John Beverley
:: IDO team



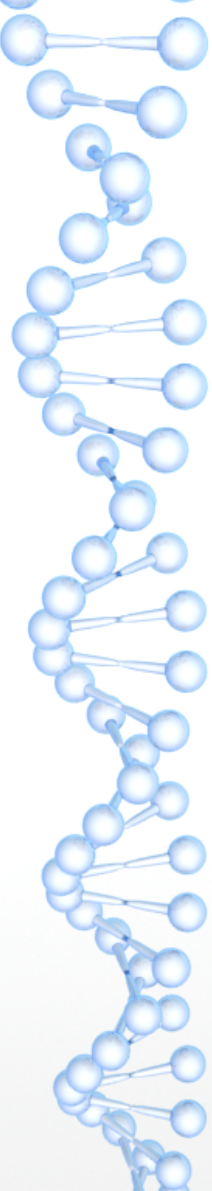
:: Zoë May Pendlington
:: Paola Roncaglia



:: Biswanath Dutta
:: Michael DeBellis

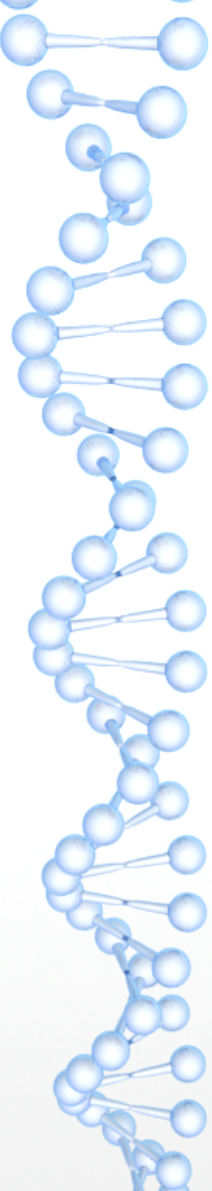


:: Leigh C. Carmody



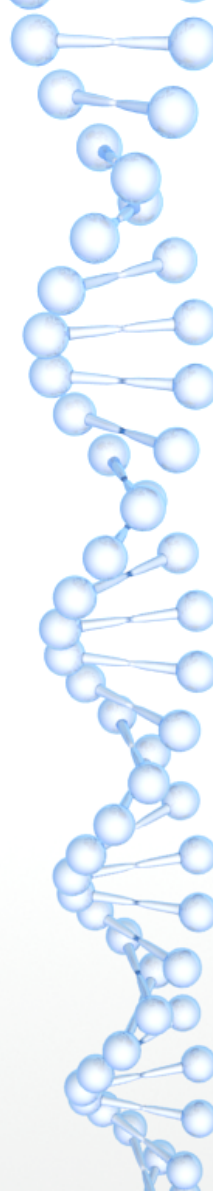
Take Homes

- 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 operation 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://douroucoul.wordpress.com/2019/03/14/biological-knowledge-graph-modeling-design-patterns/>



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.



Topics

- Goals

- Drug repurposing
- Improved search
- Personalized search
- Contact tracing
- Vaccine distribution

- Challenges

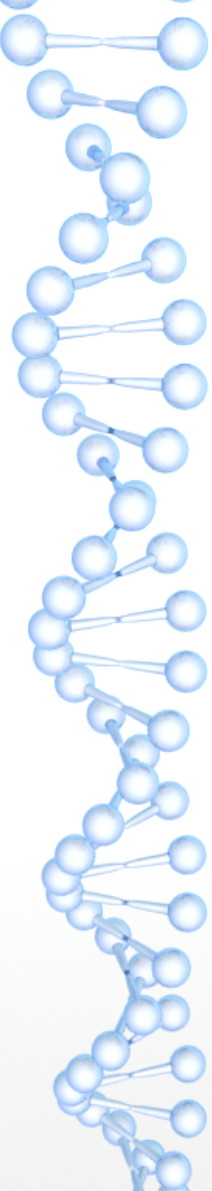
- Lack of interoperability (silos)
- Harmonization
- Limitations of ML methods
- Explainability

- Approaches

- NLP
- Knowledge graphs
- Property graphs
- Embeddings

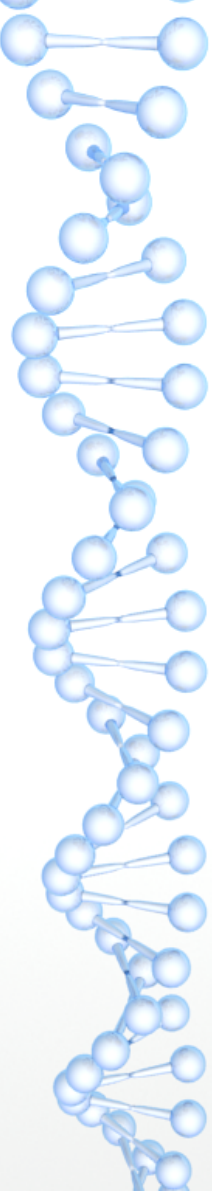
- Ontologies

- CIDO
- CODO
- COVoc
- IDO-COVID-19
- KG-COVID-19
- Other COVID-19 KGs



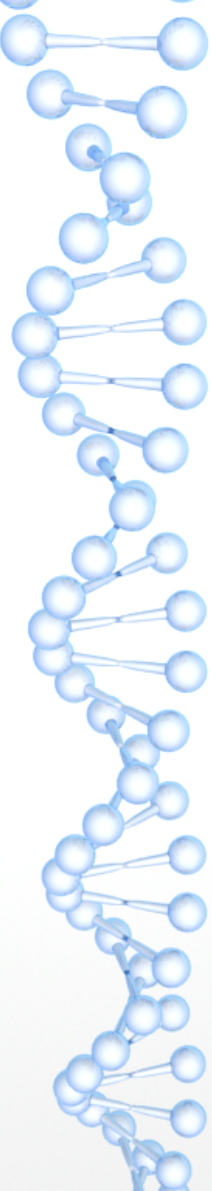
Session on 9 March 2022

- Michael DeBellis and Biswanath Dutta
- COviD-19 Ontology for cases and patient information (CODO)
 - › B. Dutta, M. DeBellis, CODO: An Ontology for Collection and Analysis of Covid-19 Data, available on [arXiv.org](https://arxiv.org)
- They have also written on how Semantic Web technology was used to combat the COVID-19 pandemic
 - › This should be published in March or April.



WE IDENTIFIED EIGHT TYPES OF SYSTEMS

1. Clinical Trials: Tools to search and analyze papers and data on Covid-19 clinical trials
2. Data Analysis: Papers that utilized semantic technology to integrate data from various data sources and to provide additional understanding of the data.
3. Drug Repurposing: The process of taking a medication initially used to treat one disease and utilize it to treat a different disease. Example: Vyvanse is an ADHD medication but many Multiple Sclerosis patients take it to combat symptoms of MS such as chronic fatigue and “brain fog”.
4. Harmonization. One of the biggest problems in medical research are the disparate taxonomies and vocabularies utilized to describe symptoms, diseases, treatments, lab results, etc. Whereas some industries suffer from a lack of standards, healthcare has the opposite problem. Semantic technology can be used to provide integrated models for medical domain models such as SNOMED CT, LOINC, ICD, UMLS, HL7 FHIR, etc.



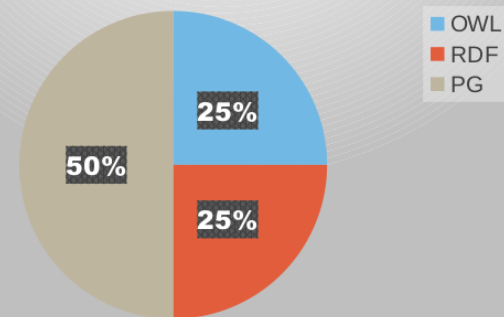
WE IDENTIFIED EIGHT TYPES OF SYSTEMS

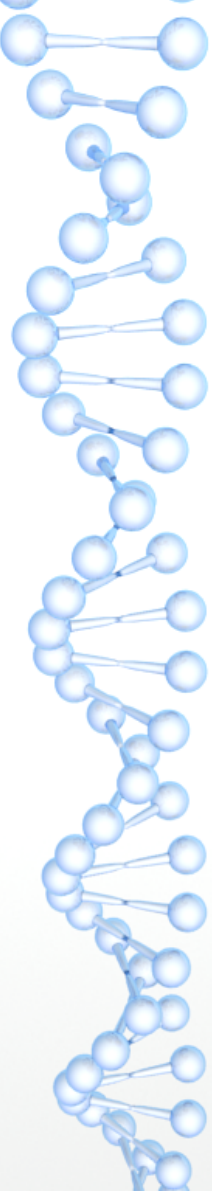
5. Semantic Search: One of the first and most common knowledge graph applications in Industry (e.g., see Pool Party case studies). Semantic search allows researchers to search for papers related to the pandemic using ontologies rather than keywords.
6. Statistics: Statistical models take data and represent it in a graph model. The data is provided by governments and public health organizations on the number of people who have contracted the virus in a specific area over given time intervals. In addition, they record data such as the number of patients who recover, die, have been vaccinated, etc.
7. Surveillance: Surveillance systems aid in monitoring patients. These papers focus on an Internet of Things (IoT) architecture. They take input from mobile devices as well as stationary medical devices. This includes the domain of risk detection: systems that identify people in specific communities (e.g., older people in residential care facilities) who are most at risk from the pandemic.
8. Vocabularies: Ontologies designed to model specific domains such as the Covid-19 virus (e.g., CIDO) in order to be reused by other researchers.

SUMMARY OF FINDINGS

- Half of the systems utilized non-standard knowledge graph models (Property Graphs) and 75% did not use OWL
- 72% of the systems where Machine Learning (ML) was relevant (excluding Statistics, Vocabularies, and Harmonization) used ML.

Knowledge Representation Technologies





Topics

- Goals

- Drug repurposing
- Improved search
- Personalized search
- Contact tracing
- Vaccine distribution

- Challenges

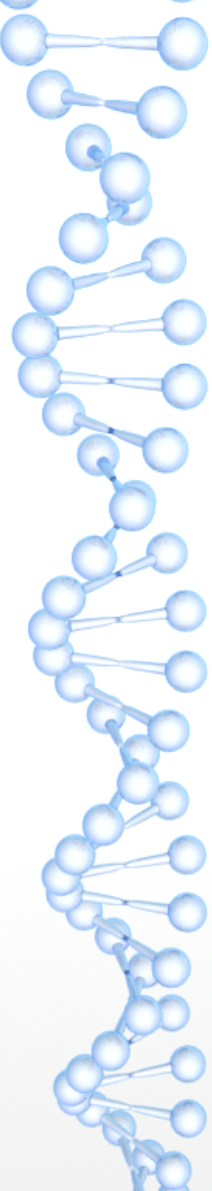
- Lack of interoperability (silos)
- Harmonization
- Limitations of ML methods
- Explainability

- Approaches

- NLP
- Knowledge graphs
- Property graphs
- Embeddings

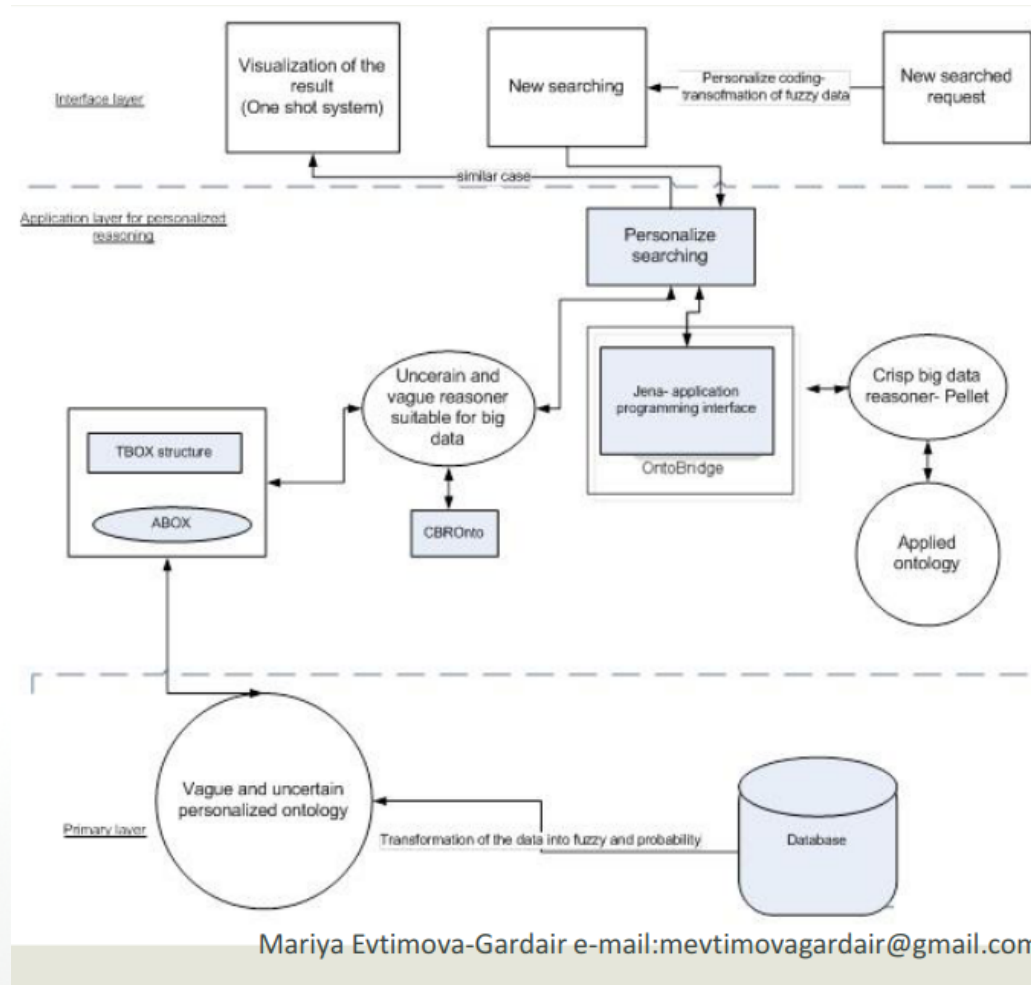
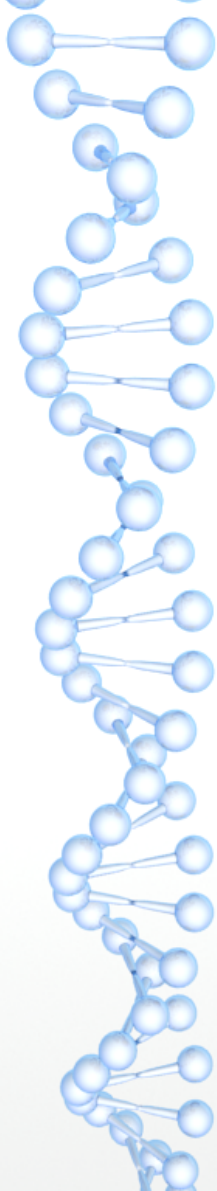
- Ontologies

- CIDO
- CODO
- COVoc
- IDO-COVID-19
- KG-COVID-19
- Other COVID-19 KGs

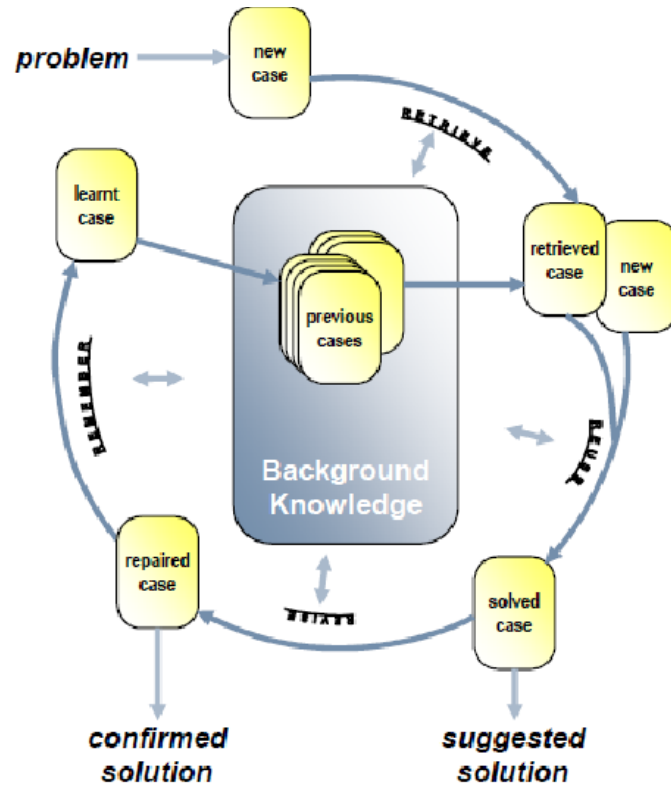


Session on 16 March 2022

- Mariya Evtimova-Gardair
- Semantic agents system for personalized searching for COVID information



Architecture for using
crisp (logical)
reasoning to answer
vague queries on Big
Data sources

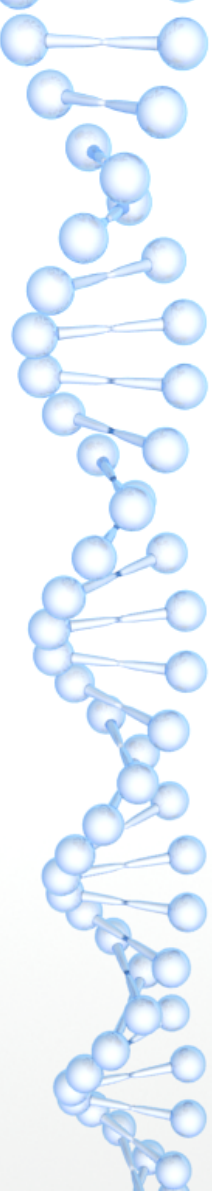


CBR is:

- ✓ Successful AI subfield
- ✓ Mature & established

Building CBR applications requires:

- Formalization
- Methodologies
- Implementation assistance



Topics

- Goals

- Drug repurposing
- Improved search
- Personalized search
- Contact tracing
- Vaccine distribution

- Challenges

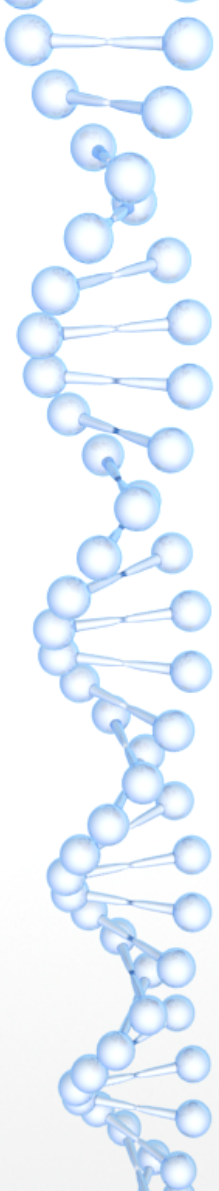
- Lack of interoperability (silos)
- Harmonization
- Limitations of ML methods
- Explainability

- Approaches

- NLP
- Knowledge graphs
- Property graphs
- Embeddings

- Ontologies

- CIDO
- CODO
- COVoc
- IDO-COVID-19
- KG-COVID-19
- Other COVID-19 KGs



Open Synthesis Discussion