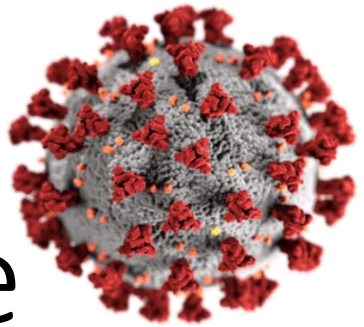


CODO: An Ontology to Capture Data on the Covid-19 Pandemic

[1]**Biswanath Dutta** and [2]Michael DeBellis

[1]Associate Professor, DRTC, Indian Statistical Institute, Bangalore, INDIA

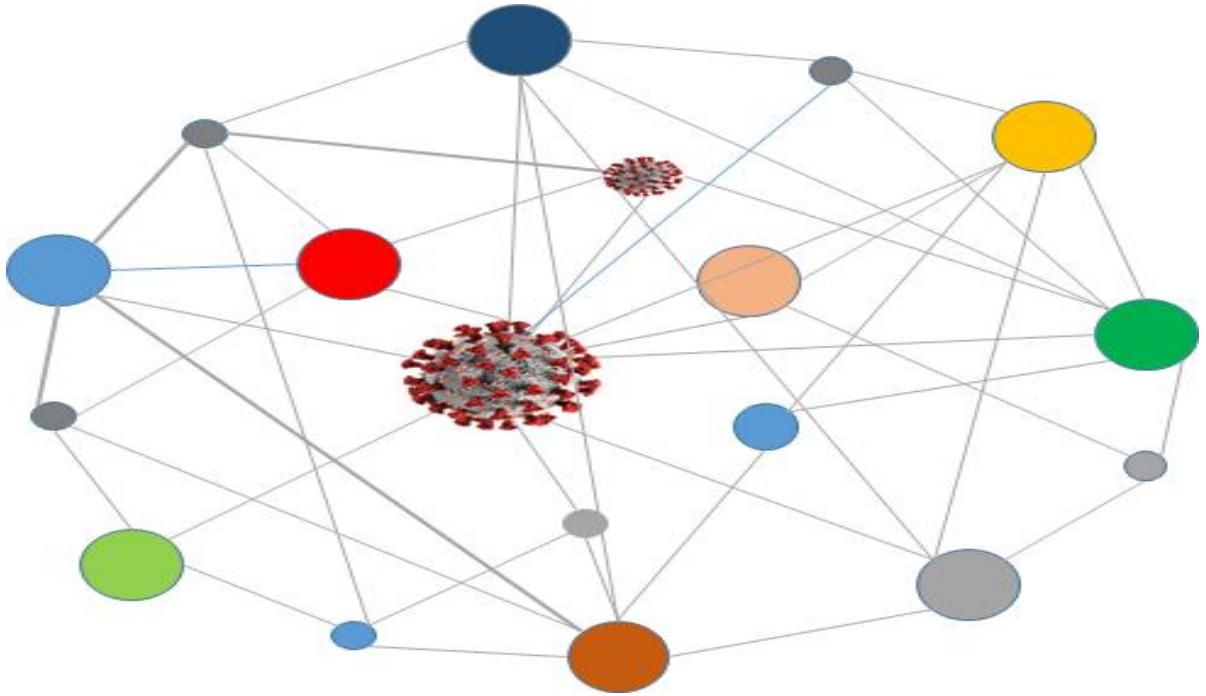
[1]Email: dutta2005@gmail.com, bisu@isibang.ac.in

[2]https://www.michaeldebellis.com/

[2]Email: mdebellissf@gmail.com

Outline

- Background
- CODO Ontology
- CODO Knowledge Graph

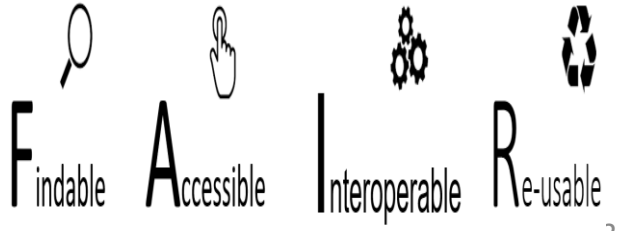


Introduction

- COVID-19 pandemic is a worldwide crisis jeopardizing the health of everyone on the planet.

- One of the tools to combat the pandemic is the **collection and analysis of data** using FAIR principles.

- Organizing data with technology based on FAIR principles can provide **open, federated data sources that will provide healthcare workers with the critical information** required to track and eventually control the growth of the pandemic.



Introduction

A	B	C	D	E	F	G	H
Case	Date	Age	Sex	City	State	Cluster	Reason
1	2020-03-09T00:00:00	41	Male	Bangalore	Karnataka	From USA	Texas US
2	2020-03-10T00:00:00	0	Female	Bangalore	Karnataka	From USA	Spouse
3	2020-03-10T00:00:00	13	Female	Bangalore	Karnataka	From USA	Daughter
4		0		Bangalore	Karnataka	From Unit	London
5	2020-03-13T00:00:00	26	Male	Bangalore	Karnataka	From the	Greece
6	2020-03-12T00:00:00	76	Male	Kalburgi	Karnataka	From Mid	Saudi Ara
7		0			Karnataka	Unknown	No detail
8	2020-03-17T00:00:00	32	Male	Bangalore	Karnataka	From Unit	Co passer
9	2020-03-17T00:00:00	63	Male	Kalburgi	Karnataka	From Unit	Co passer
10	2020-03-17T00:00:00	20	Female	Bangalore	Karnataka	From Unit	UK

Situation by Country, Territory & Area

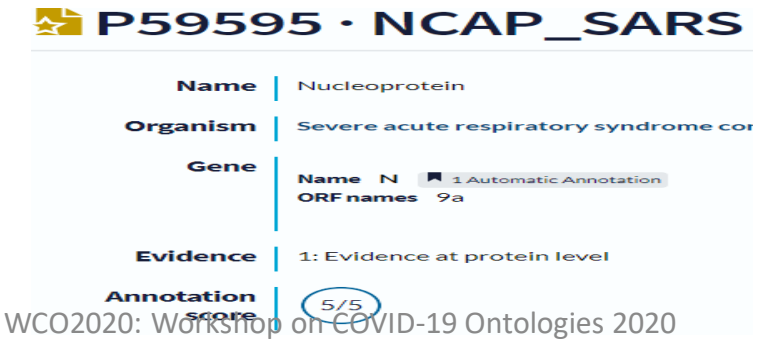
Name	Cases - cumulative total	Cases - newly reported in last 24 hours	Deaths - cumulative total
Global	**23,518,343**	**206,382**	**810,492**
United States o...	5,649,928	37,765	175,813
Brazil	3,605,783	23,421	114,744
India	3,167,323	60,975	58,390
Russian Feder...	966,189	4,696	16,568

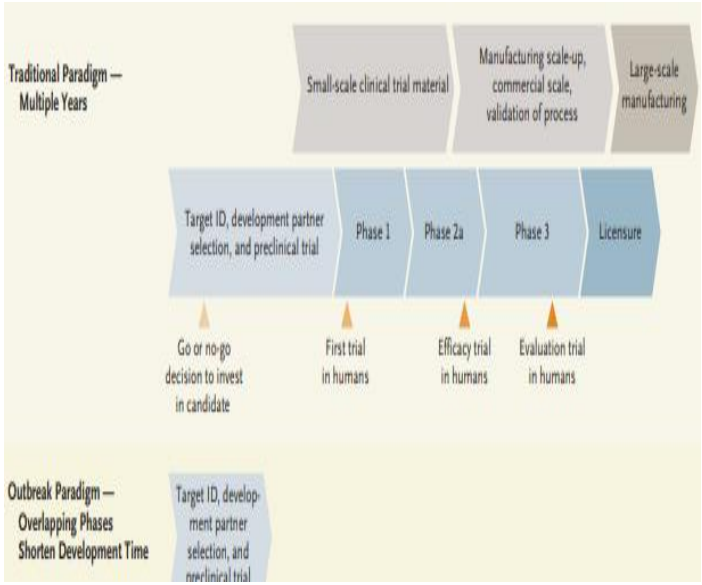
In-depth: Excess mortality

Excess mortality refers to the number of deaths *from all causes* above and beyond what we would have expected to see under 'normal' conditions.[1] In this case, we're interested in how deaths during the COVID-19 pandemic compare to the average number of deaths over the same period in previous years.

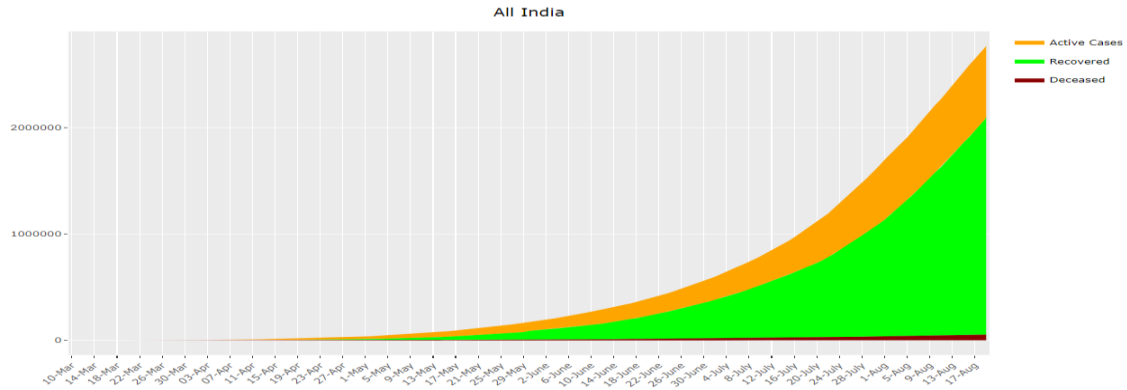
Looking at excess mortality is helpful for understanding the total impact of the pandemic on deaths – both direct and indirect. It helps us understand the direct impact by capturing deaths caused by COVID-19 that were not correctly diagnosed and reported, for example because no test for the virus was conducted. It helps us understand the indirect

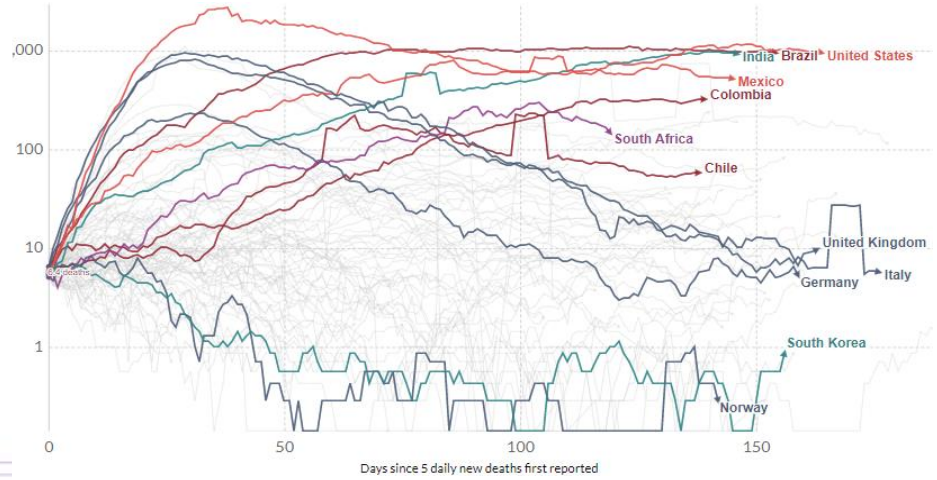
Data

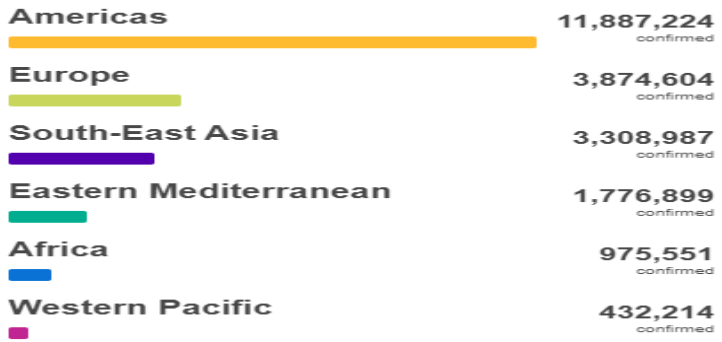


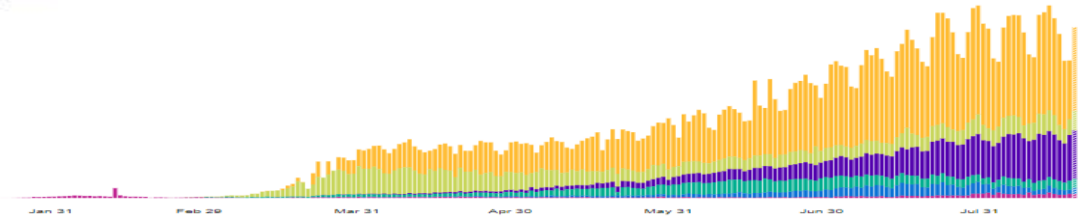


Introduction

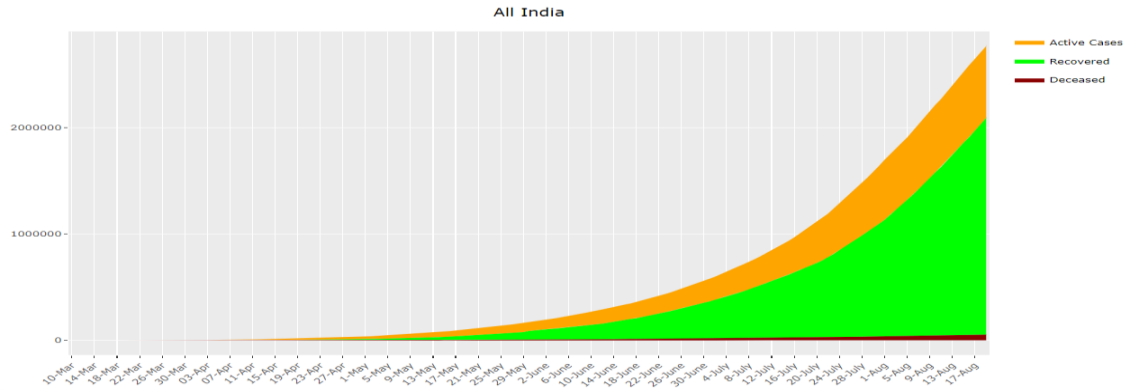




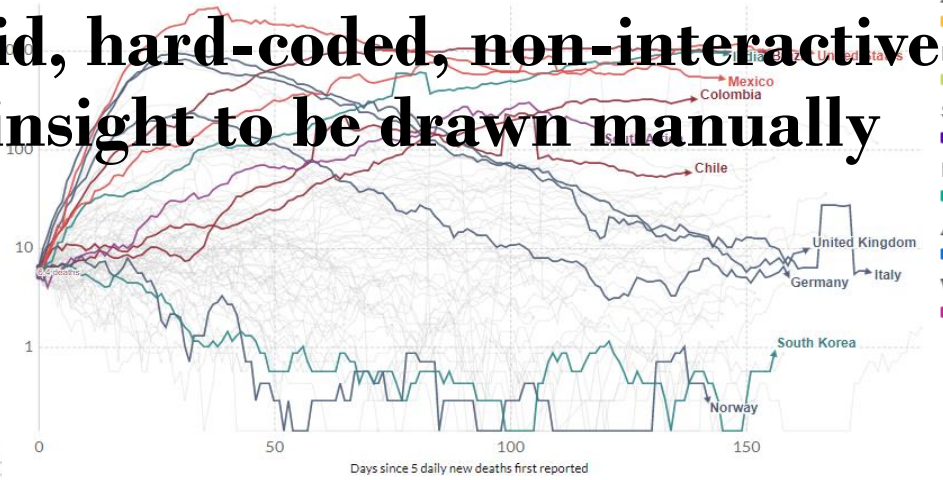


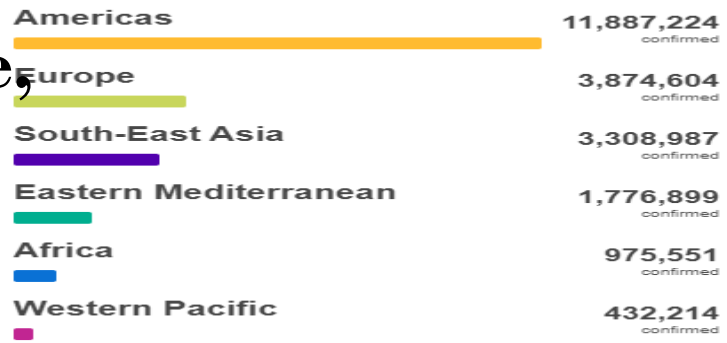


Where are the problems?

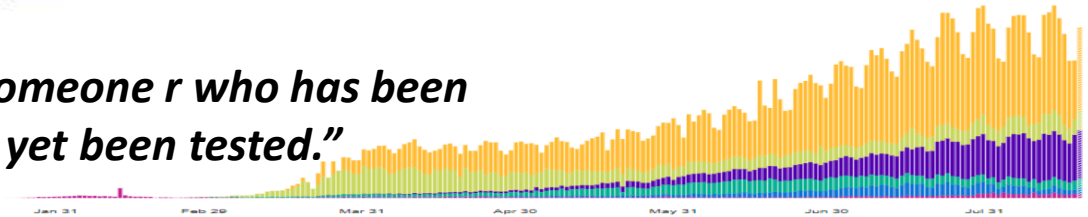


rigid, hard-coded, non-interactive, insight to be drawn manually





"find all people p who are related to someone r who has been diagnosed with COVID-19 and has not yet been tested."



CODO Ontology

A first step at utilizing knowledge graph technology to help combat the pandemic

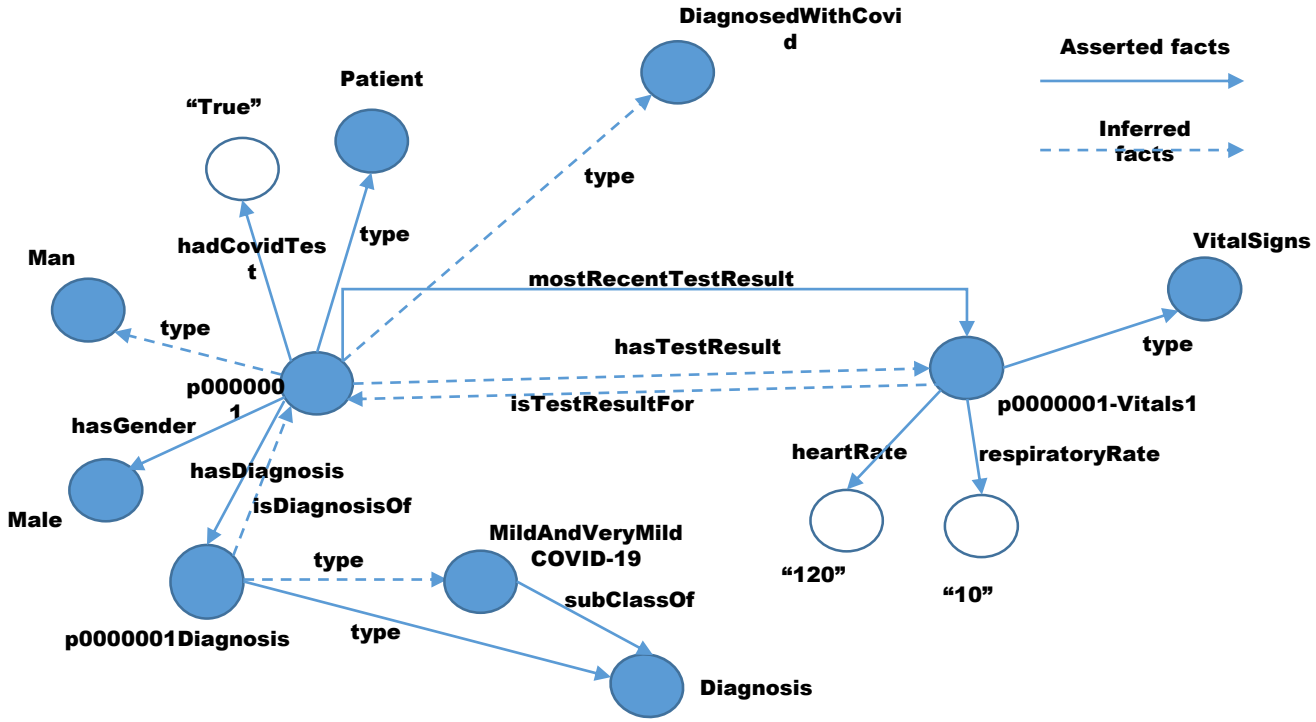
A KG is a manifestation of an intelligent **Web of Data** informed by an **ontology**.

CODO Ontology Goals

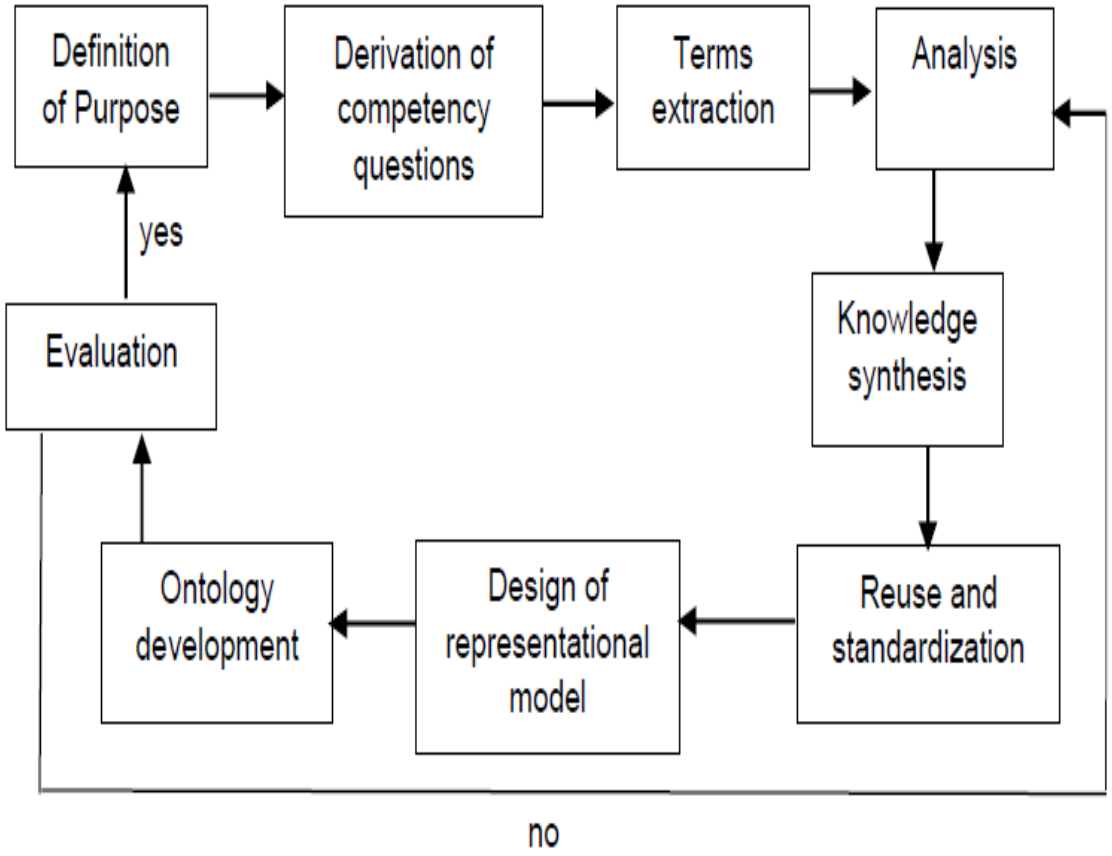
- To serve as an explicit ontology for use by data and service providers to publish COVID-19 data as FAIR data
- To develop and offer distributed, heterogeneous, semantic services and applications
  - E.g., decision support system, advanced analytics, such as COVID-19 contact tracing, factors of disease transmission, etc.
- To provide a standards-based reusable vocabulary for the use of various organizations (e.g., government agencies, hospitals) to annotate and describe COVID-19 information

CODO Ontology Use cases

- Knowledge Graph creation
- Annotation of COVID-19 literature
- Application design (e.g., COVID-19 risk detection system)



CODO Ontology Design Approach



Dutta, B. and DeBellis, M.(2020). CODO: an ontology for collection and analysis of COVID-19 data. In Proc. of 12th Int. Conf. on Knowledge Engineering and Ontology Development (KEOD), 2-4 November 2020 (accepted)

S2: **Competency questions**

i. Find all People p who are related to someone r who has been diagnosed with COVID-19 and who has not yet been tested.

ii. Give the primary reasons i for the maximum number of COVID-19 patients p.

iii. Give the most prevalent symptoms s of Severe COVID-19 d.

iv. Find all patients where it is known who they contracted the virus from.

v. Gives all the patients who have contracted the virus from 2 (or more) possible patients

vi. Find chains of patients.

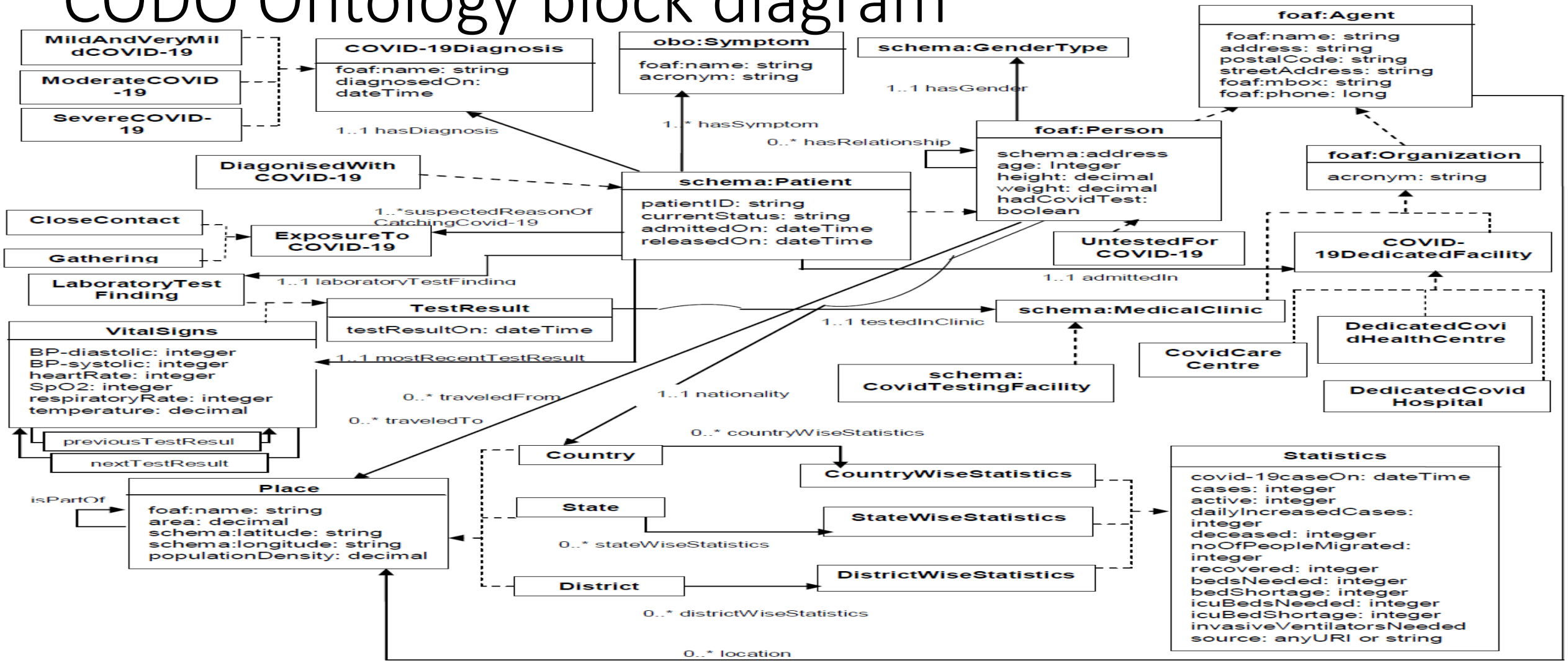
vii. Give all patients where we know the reason they caught the virus.

viii. Count the number of patients in age groups.

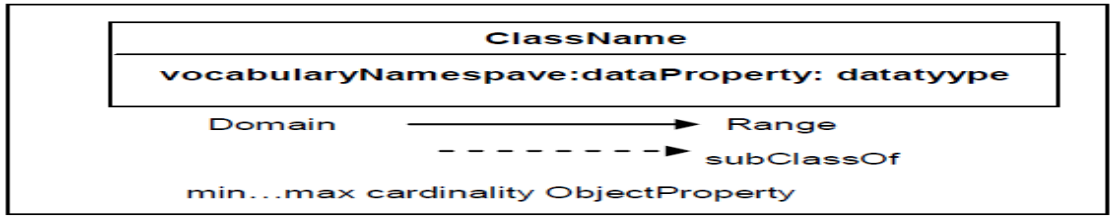
ix. Find all people diagnosed with Covid who are in family relations.

x. Give the travel history of a patient p?

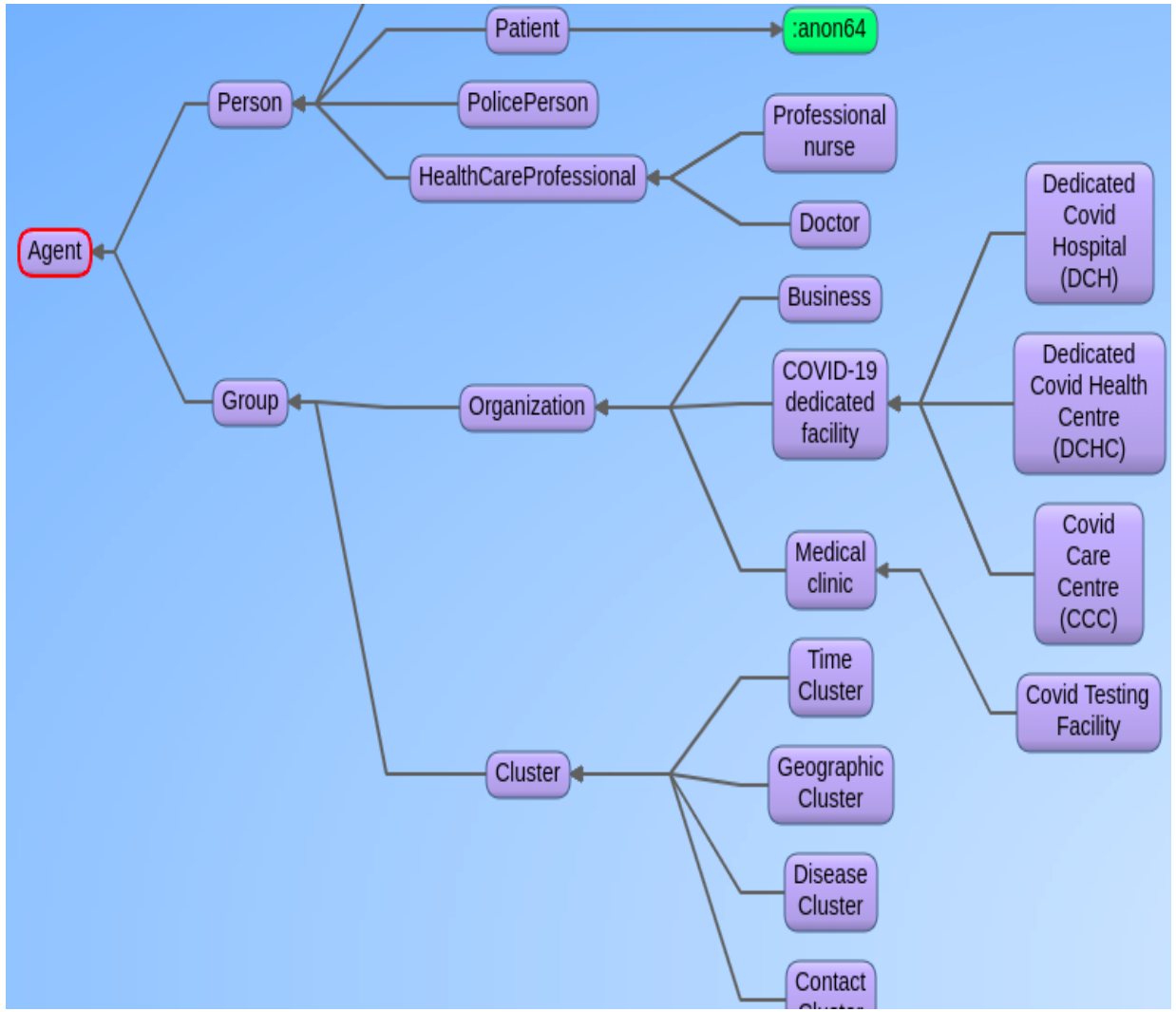
CODO Ontology block diagram

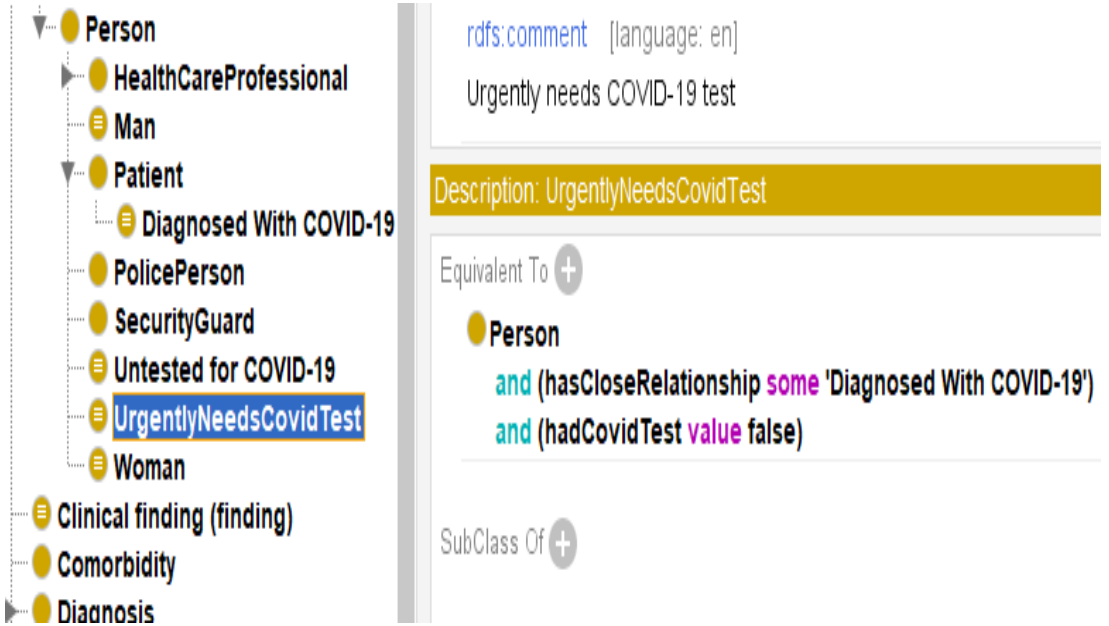


Showing **CODO v 1.0** (https://github.com/biswanathdutta/CODO)

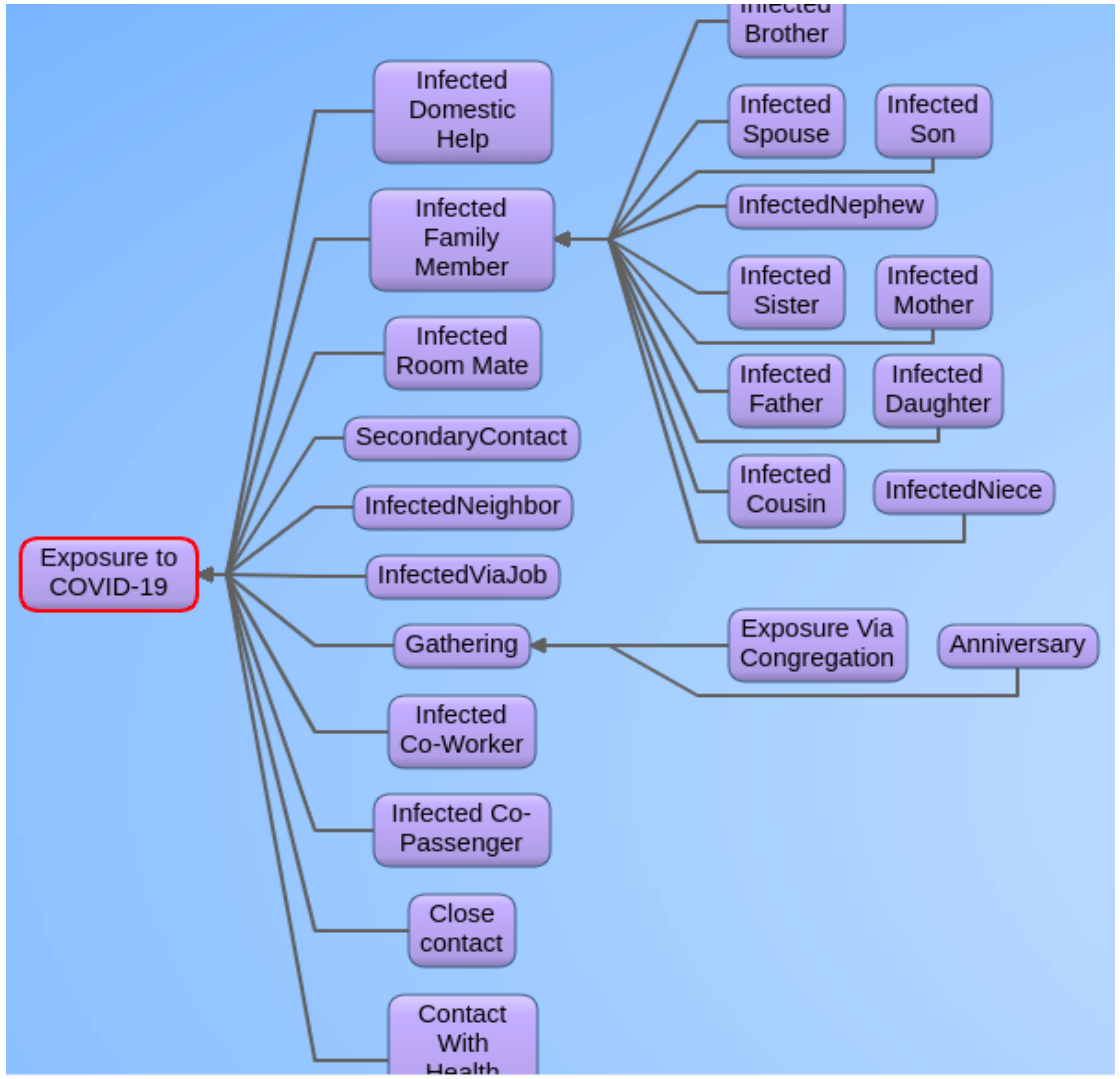


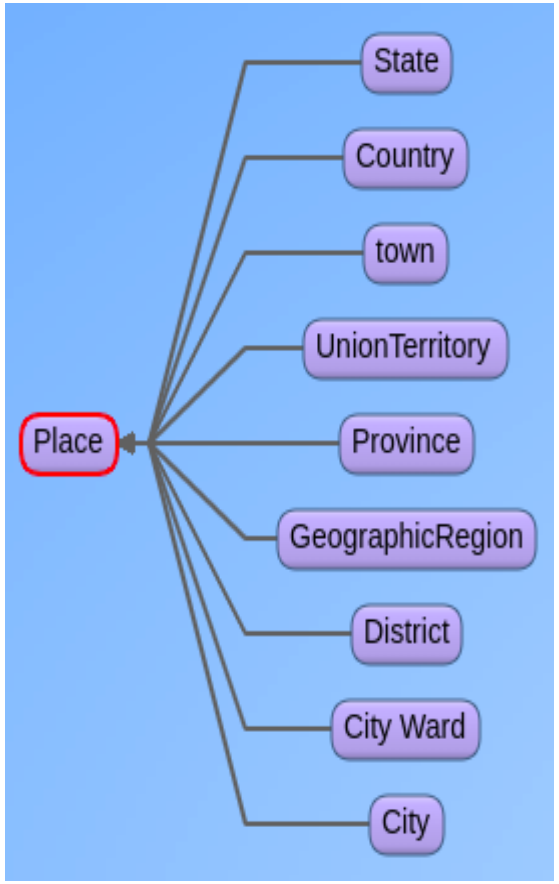
CODO v1.3 [some glimpses (1 of 3)]



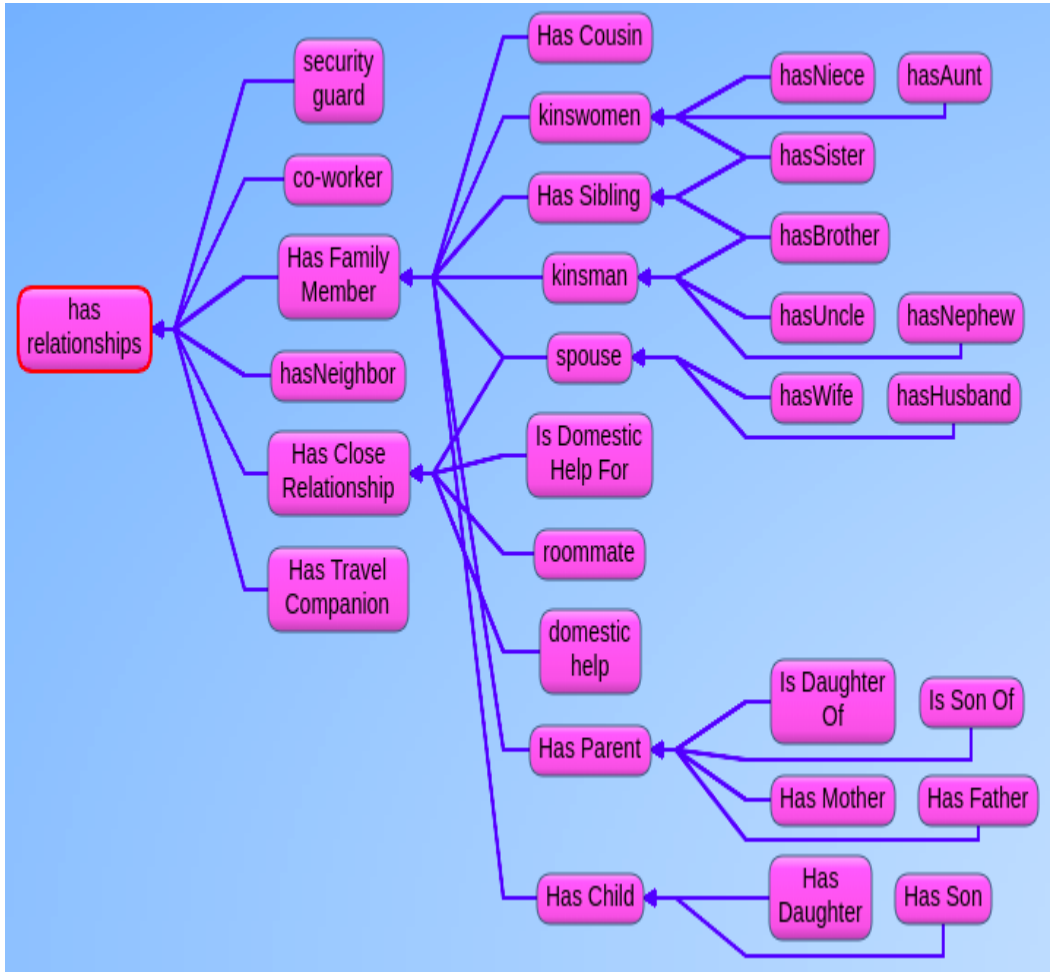


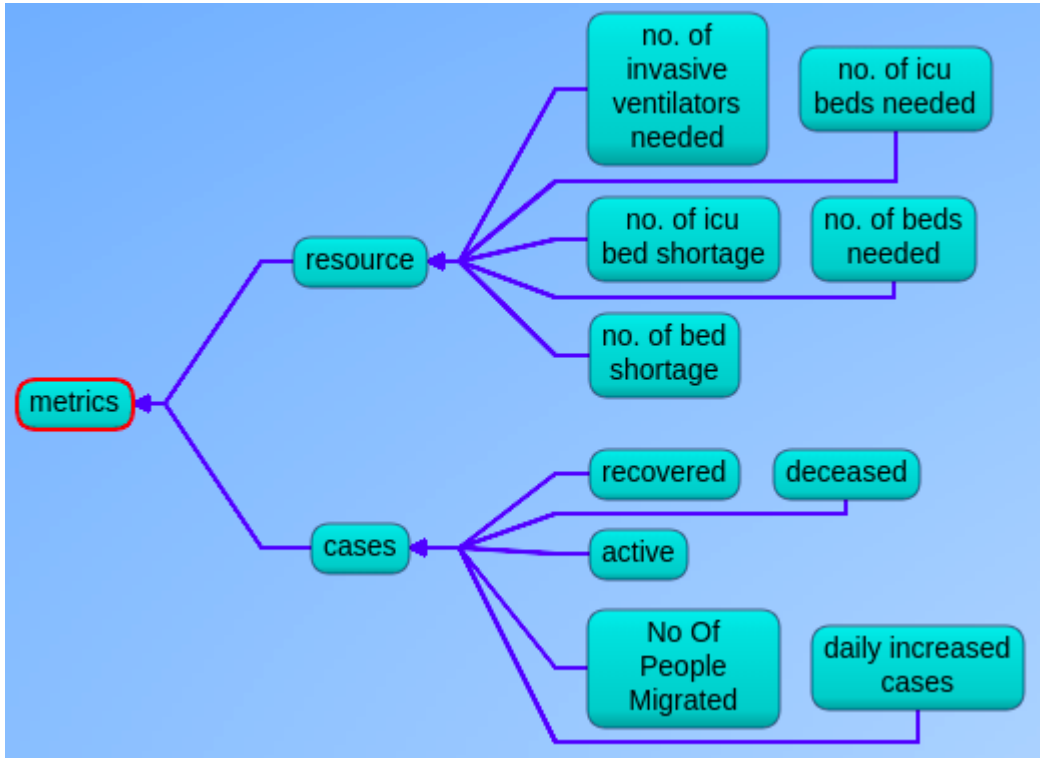
CODO v1.3 [some glimpses (2 of 3)]

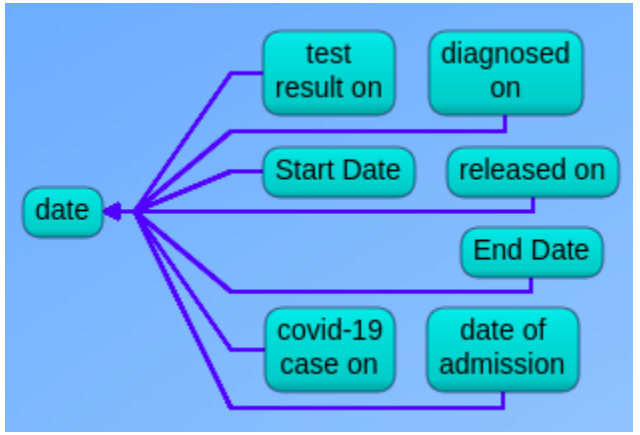




CODO v1.3 [some glimpses (3 of 3)]







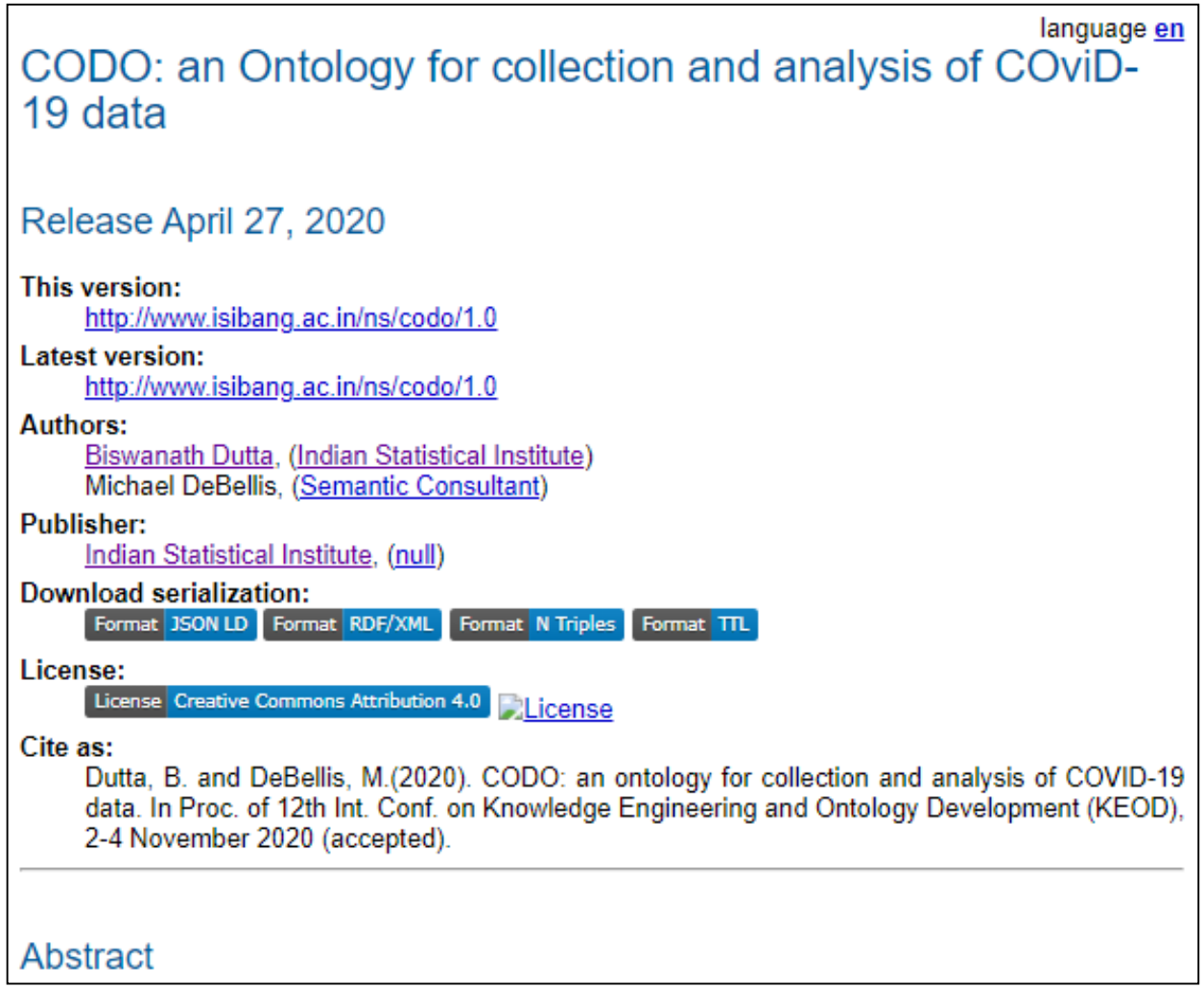
CODO Ontology

CODO v1.3 consists of

# of classes: 90

# of object property: 73

# of data property: 50



language en

CODO: an Ontology for collection and analysis of COviD-19 data

Release April 27, 2020

This version:
http://www.isibang.ac.in/ns/codo/1.0

Latest version:
http://www.isibang.ac.in/ns/codo/1.0

Authors:
Biswanath Dutta, (Indian Statistical Institute)
Michael DeBellis, (Semantic Consultant)

Publisher:
Indian Statistical Institute, (null)

Download serialization:
Format JSON LD Format RDF/XML Format N Triples Format TTL

License:
License Creative Commons Attribution 4.0 License

Cite as:
Dutta, B. and DeBellis, M.(2020). CODO: an ontology for collection and analysis of COVID-19 data. In Proc. of 12th Int. Conf. on Knowledge Engineering and Ontology Development (KEOD), 2-4 November 2020 (accepted).

Abstract

Available from https://w3id.org/codo
https://github.com/biswanathdutta/CODO

CODO Knowledge Graph: a use case of CODO ontology

Primarily with the following two goals:

1. Transforming COVID-19 data as **FAIR Semantic** data
2. CODO ontology evaluation

- **# of axioms**: 1054856
- **# of individuals**: 75759 (patients)
- **# of classes**: 90
- **# of object properties**: 73
- **# of data properties**: 50

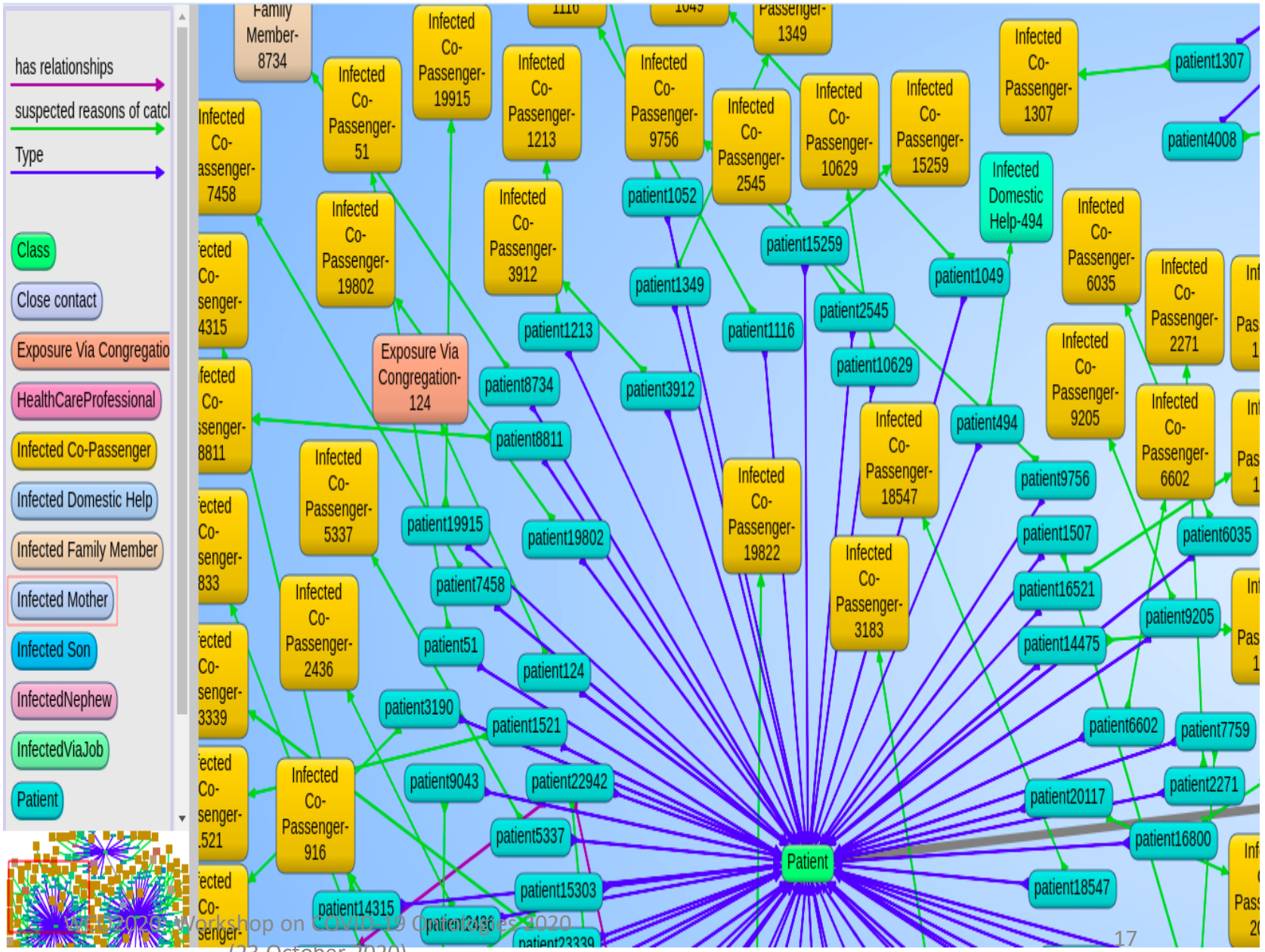
A	B	C	D	E	F	G	H	I	J	K	L	M
Case	Date	Age	Sex	City	State	Cluster	Reason	Nationality	RD	P	C	RE
15	19-Mar	35	Male	Madikeri in K	Karnataka	From Middle East	Dubai	India	C	0	0	
16	21-Mar	53	Female	Bangalore-Ur	Karnataka	From United Kingdom	Domestic help	India	C	11	0	
17	21-Mar	39	Male	Bangalore-Ur	Karnataka	From the rest of Europe	Amsterdam in Netherlan	India	C	0	1	
18	21-Mar	21	Male	Bangalore-Ur	Karnataka	From the rest of Europe	Edinburgh in Scotland	India	C	0	0	
19	21-Mar	31	Male	Chikballarpu	Karnataka	From Middle East	Mecca in Saudi Arabia	India	C	0	1	
20	21-Mar	35	Male	Mysore	Karnataka	From Middle East	Dubai	India	C	0	0	
21	22-Mar	35	Male	Dharwad	Karnataka	From Middle East	Dubai	India	C	0	0	
22	22-Mar	64	Female	Gowribidanu	Karnataka	From Middle East	Mother	India	C	19	0	
23	22-Mar	36	Female	Bangalore-Ur	Karnataka	From the rest of Europe	Switzerland and France	India	C	0	0	
24	22-Mar	27	Male	Bangalore-Ur	Karnataka	From the rest of Europe	Germany	India	C	0	0	
25	22-Mar	51	Male	Bangalore-Ur	Karnataka	From United Kingdom	London	India	C	0	1	F68
26	23-Mar	22	Male	Mangalore bu	Karnataka	From Middle East	Dubai	India	C	0	0	
27	23-Mar	46	Male	Bangalore-Ur	Karnataka	From Middle East	Dubai	India	C	0	0	
28	23-Mar	38	Male	Bangalore-Ur	Karnataka	From Middle East	Dubai	India	C	0	0	
							ondon	India	C	0	0	
							pouse	India	C	17	0	

B	C	D	E	F	G	H	I	J	K	L
District	State P No	Age In	Sex	Descriptic	Symptoms	Co-Morbiditie	DateOfAdmiss	DateOfDecea	MediaBulletin	Notes
Bengaluru	6553	53	F	ILI	Fever & Cold	Thyroid diseas	10/6/2020	20-06-2020	21-06-2020	
Bengaluru	8872	62	M	ILI	Fever	Diabetes Mell	16-06-2020	16-06-2020	21-06-2020	
Bengaluru	8880	55	M	SARI	Pneumonia	Respiratory di	18-06-2020	20-06-2020	21-06-2020	
Bidar	9149	70	M	Contact of	Loss of Appeti	Weakness, Hy	18-06-2020	18-06-2020	21-06-2020	
Bidar	9150	46	M	Contact ur	Weakness	Convulsions,	17-06-2020	18-06-2020	21-06-2020	
Ballari	7732	60	Femal	ILI	Fever & Cougl	uncontrolled I	17-06-2020	22-06-2020	22-06-2020	
Ramanaga	9237	90	M	Contact ur	Asymptomati	Hypertension		20-06-2020	22-06-2020	Died at his residence
Bengaluru	9276	45	Male	SARI	Fever & Cougl	TB	12/6/2020	15-06-2020	22-06-2020	

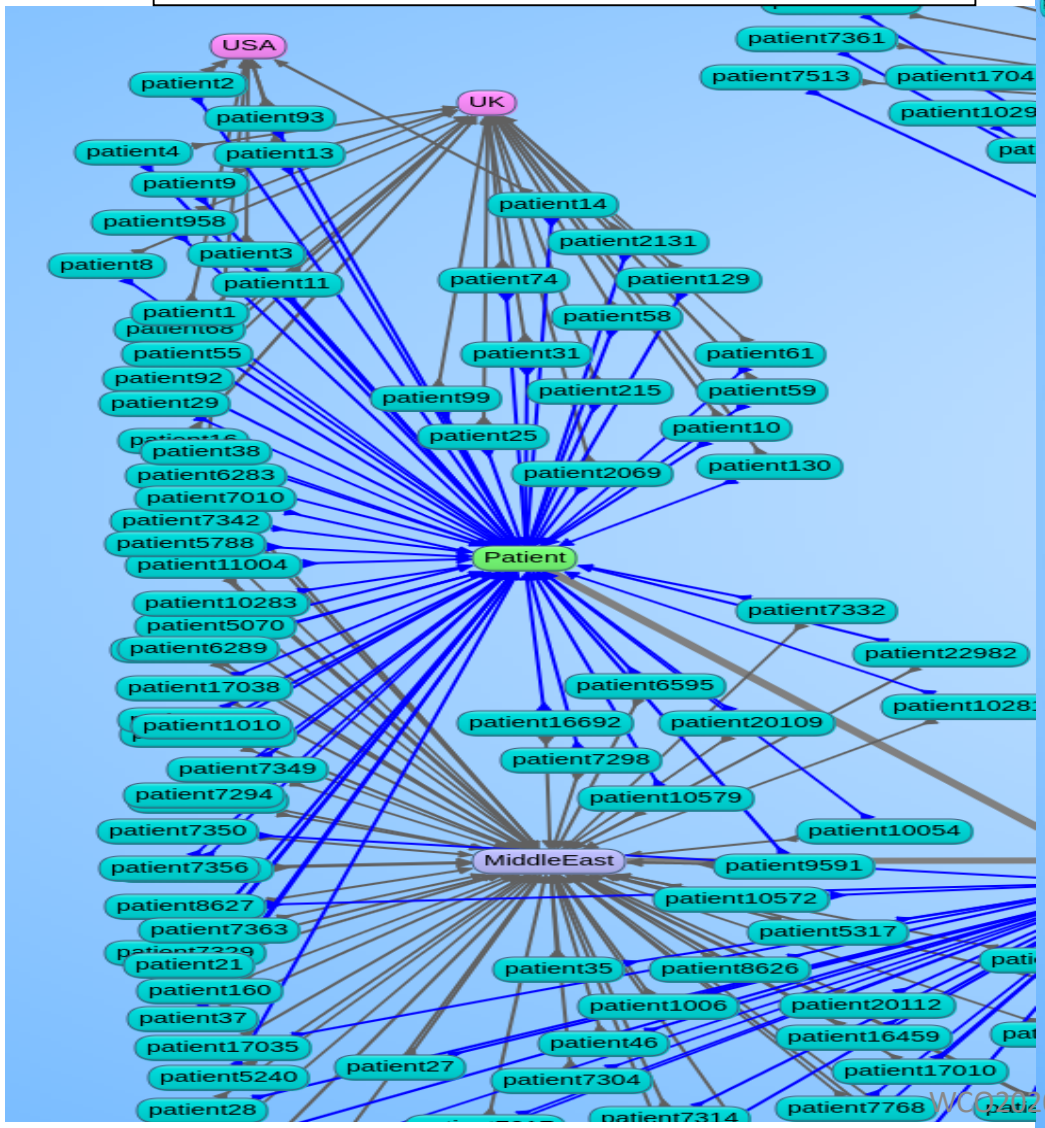
CODO Knowledge Graph

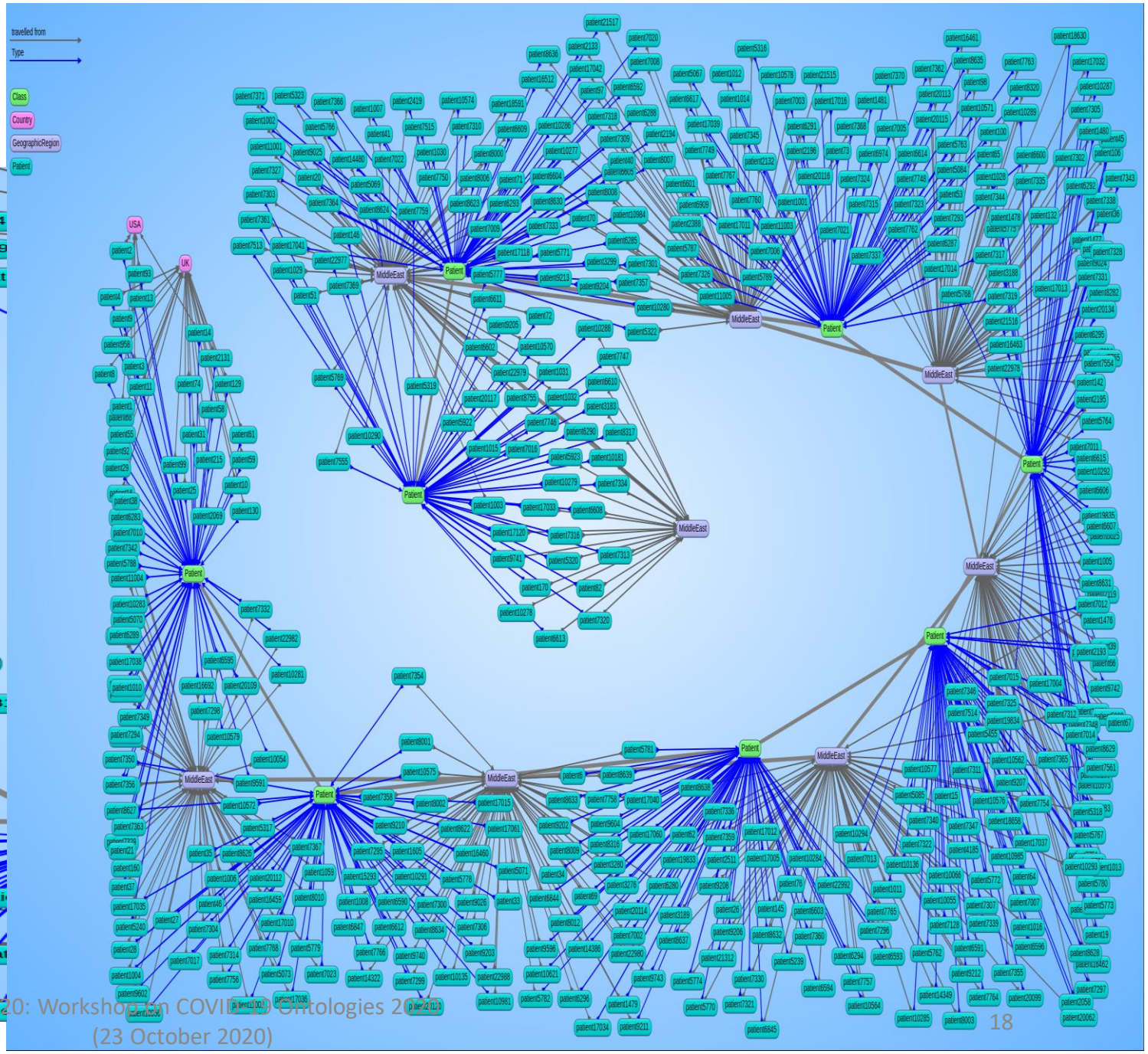
#Show the patients with the possible reasons of catching COVID-19. Also, display the relationships between the patients, if any.

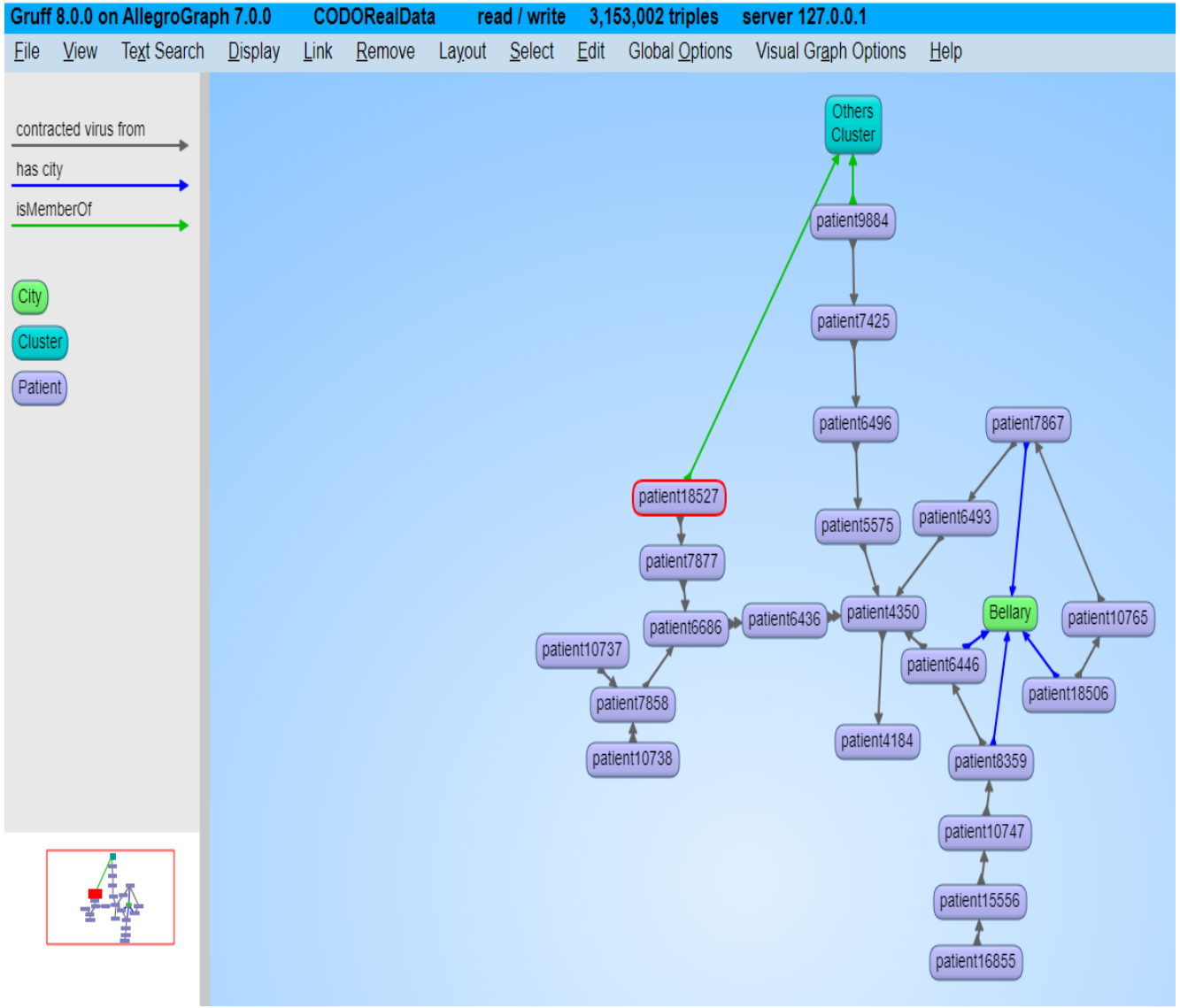
```
SELECT ?p ?r ?l

WHERE {

?p rdf:type schema:Patient.

?p
codo:suspectedReasonOfCatchingCovid
-19 ?r.

OPTIONAL{?p codo:hasRelationship
?l.}

} LIMIT 150
```

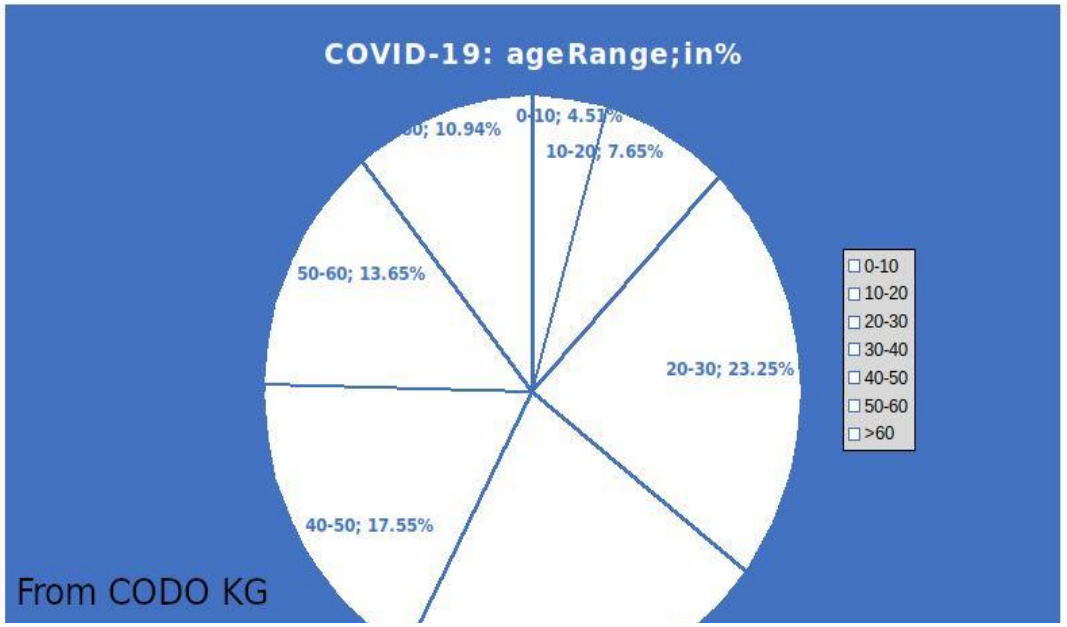


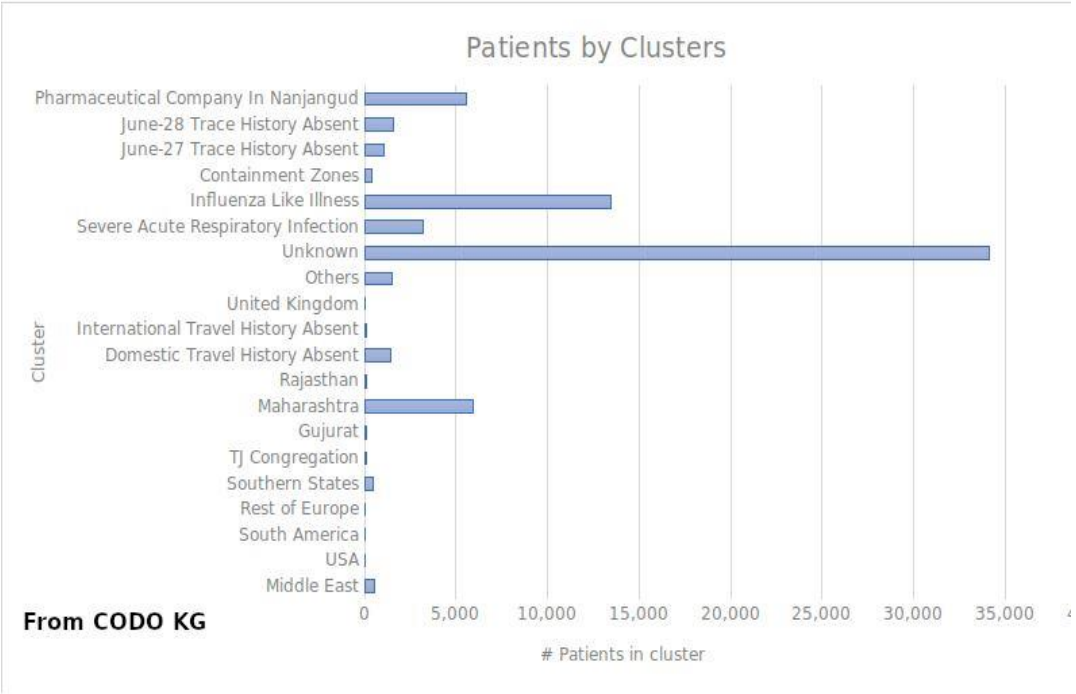
```
SELECT ?p ?tf ?tt
WHERE {
        ?p rdf:type schema:Patient.
        ?p codo:travelledFrom ?tf.
  OPTIONAL{?p codo:travelledTo ?tt.}
} LIMIT 2000
```






Gruff 8.0.0 on AllegroGraph 7.0.0 CODORealData read / write 3,153,002 triples server 127.0.0.1
File View Text Search Display Link Remove Layout Select Edit Global Options Visual Graph Options Help
contracted virus from
has city
isMemberOf
City
Cluster
Patient
Others Cluster
patient9884
patient7425
patient6496
patient7867
patient18527
patient5575
patient6493
patient7877
patient6686
patient6436
patient4350
Bellary
patient10765
patient10737
patient6446
patient18506
patient7858
patient4184
patient8359
patient10738
patient10747
patient15556
patient16855


COVID-19: ageRange;in%
0-10; 4.51%
10-20; 7.65%
20-30; 23.25%
40-50; 17.55%
50-60; 13.65%
0-10
10-20
20-30
30-40
40-50
50-60
>60
From CODO KG


Patients by Clusters
Pharmaceutical Company In Nanjangud
June-28 Trace History Absent
June-27 Trace History Absent
Containment Zones
Influenza Like Illness
Severe Acute Respiratory Infection
Unknown
Others
United Kingdom
International Travel History Absent
Domestic Travel History Absent
Rajasthan
Maharashtra
Gujurat
TJ Congregation
Southern States
Rest of Europe
South America
USA
Middle East
Cluster
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
Patients in cluster
From CODO KG

Challenges

- Data availability (e.g., clinical data, radiological imaging data)
- Standard format for data capture and communication
- Most of the datasets are not directly consumable, not suitable to the graph
  - This made the data transformation complicated, time consuming
- Inconsistent data (e.g., sometimes P1342, sometimes 1342. Sometimes P132-P134 and sometimes "P132 and P133 and P134")
- Typo errors
  - Data misplacement
  - Spelling errors (mostly the place names and this complicates the linkage with the external resources)

Conclusion

- The CODO ontology is only the first step in providing a knowledge graph model for COVID-19 based on FAIR data principles.
- The CODO ontology has already found its use in a real world project called **Ping**: *a COVID-19 risk detection system for older people in residential aged care using ontology and machine learning technology* (https://bioportal.bioontology.org/projects/Ping)

- Future plan
  - Enhance the current CODO ontology by integrating many more COVID-19 datasets available on the Web, both from India and world-wide.
  - Publish CODO using a triplestore database published as a SPARQL endpoint.

Acknowledgement

- Thanks to Franz Inc. (http://www.allegrograph.com) for their generous help with AllegroGraph and Gruff.
- This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health. For more information on Protégé: https://protege.stanford.edu/

References

1. https://medium.com/virtuoso-blog/linked-data-ontologies-and-knowledge-graphs-a3d0ad6d6f66
2. Singhal, A. (2012). Introducing the Knowledge Graph: things, not strings. https://www.blog.google/products/search/introducing-knowledge-graph-things-not/
3. Idehen, Kingsley U. (2020). Linked Data, Ontologies, and Knowledge Graphs. https://www.linkedin.com/pulse/linked-data-ontologies-knowledge-graphs-kingsley-uyi-idehen/
4. Blumauer, A. and Kiryakov, A. (2020). Knowledge Graphs: 5 Use Cases and 10 Steps to Get There. (https://www.ontotext.com/knowledgehub/webinars/knowledge-graphs-5-use-cases-and-10-steps-to-get-there/)
5. W3C, 2014. RDF 1.1: Concepts and Abstract Syntax. W3C Recommendation. https://www.w3.org/TR/rdf11-concepts/
6. W3C, 2014a. RDF Schema 1.1. W3C Recommendation. https://www.w3.org/TR/rdf-schema/
7. W3C, 2012. Web Ontology Language Document Overview (Second Edition). W3C Recommendation. https://www.w3.org/TR/owl2-overview/
8. Gruber, T.R. (1993), “A translation approach to portable ontologies”, Knowledge Acquisition, Vol. 5 No. 2, pp. 199-220.
9. DuCharme, Bob, 2011. Learning SPARQL. O’Reilly.

References

10. W3C, 2004. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C Member Submission. https://www.w3.org/Submission/SWRL/
11. Aasman, J. (2020). Stanford CS 520 Knowledge Graphs. https://web.stanford.edu/class/cs520/abstracts/Aasman.pdf
12. Dutta, B., Chatterjee, U. and Madalli, D. P. (2015). YAMO: Yet Another Methodology for Large-scale Faceted Ontology Construction. In Emerald Journal of Knowledge Management. Vol. 19, no. 1, pp. 6 – 24.
13. Graph: Enterprise Knowledge Graph. 2020. (Available from https://www.youtube.com/watch?time_continue=5&v=MJuRnuA0hrM&feature=emb_logo)
14. Bender, Edward A.; Williamson, S. Gill (2010). Lists, Decisions and Graphs. With an Introduction to Probability.
15. Natasha Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, And Jamie Taylor (2019). Five diverse technology companies show how it's done. ACM.
16. Dieter Fensel, [...], Alexander Wahler (2020). Knowledge Graphs: methodology, tools and selected use cases. Springer.
17. Andreas Blumauer and Helmut Nagy (2020). The Knowledge Graph Cookbook: recipes that work. Edition Mono/Monochrom, Vienna, Austria.
18. Dutta, B. (2020). Knowledge Graph and the current pandemic COVID-19. Available from: https://www.isibang.ac.in/~bisu/paper/KEOD%20CODO%207August2020-Preprint.pdf

Thank you!!!

- **Access to CODO Ontology and CODO Knowledge Graph**
  - GitHub (https://github.com/biswanathdutta/CODO)
  - Browse CODO Ontology (https://w3id.org/codo)
  - https://bioportal.bioontology.org/ontologies/CODO

Dr. Biswanath Dutta

Email: dutta2005@gmail.com

bisu@isibang.ac.in

Twitter: @biswanath_dutta