**CONJECTURAL NOTES**

# Covid-19 knowledge graphs in Health Communication and Information

## Gráficos de conhecimento para Covid-19 em Comunicação e Informação sobre Saúde

## Gráficos de conocimiento Covid-19 en Comunicación e Información de Salud

*Andrew Iliadis[1,a]*
andrew.iliadis@temple.edu | http://orcid.org/0000-0002-8345-6251

[1] Temple University, Lew Klein College of Media and Communication, Department of Media Studies and Production. Philadelphia, PA, United States.

[a] Ph. D in Communication and Philosophy from Purdue University.

## ABSTRACT

This commentary discusses recent developments in 'knowledge graph' technology over the course of the Covid-19 pandemic. Recently experiencing a surge in popularity, knowledge graphs are technologies that assist with data integration through structured metadata modeling. Researchers tag and collate vast amounts of diverse data using knowledge graphs, yet problems related to semantic drift and more salient issues related to the political economy of information and communication technologies persist. Researchers should anticipate that the semantics of these Covid-19 knowledge graphs can change over time. Equally important, researchers should also consider all stakeholders involved, including those stakeholders that might be excluded.

**Keywords**: Covid-19; SARS-CoV-2; knowledge graphs; semantic web; linked data.

## RESUMO

Este comentário discute desenvolvimentos recentes na tecnologia de 'gráficos de conhecimento' durante o curso da pandemia de Covid-19. Gráficos de conhecimento, que vêm tendo um aumento de popularidade, são tecnologias que auxiliam com a integração de dados através de modelamento estruturado de metadados. Pesquisadores rotulam e agregam vastas quantidades de dados diversos usando gráficos de conhecimento, entretanto persistem problemas relacionados a variações semânticas e questões mais salientes relacionadas à economia política de tecnologias de informação e comunicação. Os pesquisadores devem prever que a semântica desses gráficos de conhecimento para Covid-19 pode variar com o tempo. Igualmente importante, os pesquisadores devem também considerar todas as partes interessadas envolvidas, incluindo as que poderiam ser excluídas.

**Palavras-chave:** Covid-19; SARS-CoV-2; gráficos de conhecimento; web semântica; dados ligados.

**RESUMEM**

Este ejemplo analiza los desarrollos recientes en la tecnología de 'gráficos de conocimiento' durante la pandemia de Covid-19. Recientemente experimentando un aumento en popularidad, los gráficos de conocimiento son tecnologías que ayudan a la integración de datos a través del modelado de metadatos estructurados. Los investigadores etiquetan y recopilan grandes cantidades de datos diversos utilizando gráficos de conocimiento, pero persisten los problemas relacionados con la deriva semántica y cuestiones más importantes relacionadas con la economía política de las tecnologías de la información y la comunicación. Los investigadores deben prever que la semántica de estos gráficos de conocimiento de Covid-19 puede cambiar con el tiempo. También es importante que los investigadores consideren a todas las partes interesadas involucradas, incluso las que podrían quedar excluidas.

**Palabras clave:** Covid-19; SARS-CoV-2; gráficos de conocimiento; web semántica; datos vinculados.

On the 40-year anniversary of the publication of the United Nations Educational, Scientific and Cultural Organization's (UNESCO) contentious MacBride Report, which called for a New World Information and Communication Order, it might be helpful to reflect on some of the problems addressed in the report once again. Unequal access to information and communication remains a key global problem, particularly in terms of data sharing and accessibility. In this respect, echoes of MacBride's concerns can be found in something like today's FAIR Data Principles, which argue for scientific data that is Findable, Accessible, Interoperable, and Reusable (WILKINSON *et al.*, 2016). FAIR data has been deeply needed recently against the backdrop of a pandemic, and in the shadow of a Trump administration who vowed to exit the World Health Organization (WHO) when the world needed it most.

Undeterred, over the course of the coronavirus disease (Covid-19) outbreak, scientists and researchers have adapted to the need for greater, globally coordinated efforts regarding health communication and information retrieval related to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). One of the most promising of the proposed SARS-CoV-2 technology projects involves combining international data related to the pandemic into 'knowledge graphs' that offer users easy, open access to information that was previously widely dispersed. Such knowledge graph technologies are an outgrowth of early semantic-web and linked-data initiatives and are quickly becoming one of the most hyped emerging technologies today, at least according to Gartner Inc.'s most recent Hype Cycle for Artificial Intelligence report (GOASDUFF, 2020). It seems that old technologies are becoming new, once again.

Unlike algorithmic and statistical techniques that might focus on analyzing strings of data (looking for patterns of 1s and 0s, for example), knowledge graphs focus on identifying common-sense concepts related to items such as people, places, actions, and things. These concepts are then used to tag different types of distributed information before bringing the tagged information together in a graph data model. The graphed data adds a layer of semantic expressivity and context to the original information, thus making that original information findable, interoperable, translatable and sharable. There are recent and upcoming texts that explain the utility of these knowledge graph technologies in different contexts (PAN *et al.*, 2017; KEJRIWAL, 2019; FENSEL *et al.*, 2020; QI *et al.*, 2021; KEJRIWAL; KNOBLOCK; SZEKELY, 2021). Primarily, they are of keen interest to open government and health researchers.

More formally, knowledge graphs identify subjects, predicates, and objects in information through metadata markup. Things like scientific articles, data that those articles contain, authors, publishers, universities, databases, research teams, funding agencies, their betweenness and relationships, and more, are all tagged using some serialization or another and then visualized in a graph explorer. While the field of health communication and information has used semantic technologies like knowledge graphs for some time, these technologies were further popularized by Google in 2012 (Google published a famous blog article stating that they would shift to focus on 'things' and not 'strings') (SINGHAL, 2012) and are now even more commonly used by other platform companies such as Amazon and Microsoft, but also governments who participate in open data initiatives, journalists who work on data-centered news stories, legal researchers who commonly use databases, and cultural analytics projects which include things like digital cultural heritage and digital humanities. Librarians and archivists are all familiar with these tools in their archiving and preservation work.

Early in the pandemic, several Covid-19 knowledge graph projects began and have led to publications (CHEN *et al.*, 2020; DOMINGO-FERNÁNDEZ *et al.*, 2020; REESE *et al.*, 2020; WANG *et al.*, 2020; WISE *et al.*, 2020). Included among these are CovidGraph, "a non-profit collaboration of researchers, software developers, data scientists and medical professionals" (c2020) who have "built a research and communication platform that encompasses over 40,000 publications, case statistics, genes and functions, molecular data and much more" (c2020). CovidGraph integrates data from several sources, including the

United States National Center for Biotechnology Information's Gene Database, the Gene Ontology project, the United Nations World Population Prospects, and the Covid-19 Data Repository managed by the Center for Systems Science and Engineering at Johns Hopkins University, among others. The team involved in building CovidGraph includes representatives from organizations like the German Center for Diabetes Research, yWorks, LinkedIn, Neo4j, Structr, and Linkurious, among others. The project is soliciting more partners and datasets for integration.

Another project, the Covid-19 Knowledge Graph, is described as "a computable, multi-modal, cause-and-effect knowledge model of Covid-19 pathophysiology" (2020). Led by the Department of Bioinformatics at the Fraunhofer Institute for Algorithms and Scientific Computing in Sankt Augustin, Germany, the project aims to create a model of biological mechanisms related to Covid-19 to improve therapies. As Anadiotis (2020) notes, others at the Open Research Knowledge Graph project, including lead Dr. Sören Auer, Director of the TIB (the Leibniz Information Centre for Science and Technology and University Library), are also looking at FAIR data for "facilitating scientific research in general, and coronavirus research in particular... to describe research papers in a structured manner, making them easier to find and compare" (2020). There are several other Covid-19 knowledge graph projects under development, along with many papers and conferences.

While such knowledge graph initiatives are surely needed during these times, researchers might also be attentive to issues relating to semantic drift and nuances of meaning and more salient issues related to the global political economy of information and communication technologies. Researchers must remain alert and anticipate that the semantics of these Covid-19 knowledge graphs will change over time, and should ensure that archiving and provenance be adequately documented. Equally important, researchers should consider all stakeholders involved. For example, knowledge graph initiatives should consider including a variety of languages in the development stage so that knowledge can be shared in languages other than English. Relatedly, certain terms included in the knowledge graphs might be culturally and/or ethically sensitive, or terms may change over time. What gets included in the knowledge graph is also a political decision. What terms are used, which databases are integrated, and, most importantly, which individuals are members of the research teams that build these knowledge graph technologies and have they reached out to scholars and researchers in the Global South for their perspectives or to participate? As of right now, contributions from the Global South are not very visible in these knowledge graph initiatives (as far as I can tell). Outreach needs to be increased so that knowledge may be shared from all parts of the world and included in the design stages. More data from the Global South will enrich these knowledge graphs and help in the fight towards improving vaccines.

There have been several studies showing that knowledge graphs like the one provided by Google contain errors (VANG, 2013), while other papers have discussed the political consequences of web semantics in terms of problematic issues related to identity and verification (GINSBERG, 2008), and more generally in political operationalization (ILIADIS, 2019). Knowledge graphs have consequences when they result in restrictive or erroneous identification and verification, and they can be used in contexts that serve to control and manipulate—perhaps not in the initial stages, but issues of this type might appear later in the knowledge graph's lifecycle. Data might also lose its provenance once it is taken up in a knowledge graph and lose the social context in which it was developed. As Ford and Graham (2016) have noted in the context of Google's graphing of geographic locations in its knowledge graph, "there is an ethical argument to be made for reconnecting facts to the social contexts from which they are derived... a lack of provenance continues to strip the digital layers of place of important context" (p. 967). Proper knowledge graphs should include detailed historical records of provenance, data lineage, and contextualization related to the data so that future users may observe the archived and recorded changes.

How can we think about funding and initiatives to support truly global knowledge graphs that are not tethered, on the one hand, to single geographic locations (United States, Europe, etc.) in terms of management or, on the other hand, to large corporations and private companies with clear economic and financial incentives? What would such an initiative look like in terms of an equitable governance structure and democratic list of participants? What rules could be used to ensure a just selection of datasets, languages, and concepts? Some countries can begin seriously thinking about these questions, but government reaction to the Covid-19 crisis has been uneven, some have been disastrous, and certain individuals are not up to the task.

While the MacBride Report may have been a little overzealous in its positive characterization of the liberatory potential of computers and the internet (RODRÍGUEZ; ILIADIS, 2019), unprepared for the grip that global internet companies would have on all aspects of life, unequal access to information and communication remains a global concern. Knowledge graphs are a step in the right direction towards making data findable, accessible, interoperable, and reusable all around the globe, even with their attendant problems. Yet, there are powers that would prefer to retreat from such global initiatives. Just as the United States withdrew from UNESCO almost 40 years ago in 1984 only to rejoin in 2003, before once again leaving in 2018, all for political purposes, the United States has recently once again withdrawn from a global body, this time the WHO (effective on July 6, 2021), due to political nativism and a refusal to engage in globally coordinated cooperation and collaboration efforts. It is unfortunate that, during this pandemic and rise in the need for health communication and information access, some are turning their backs to the world once again. Hopefully, with the arrival of the Biden administration and the signing of new executive orders to rejoin the WHO, the course will be corrected.

## REFERENCES

ANADIOTIS, George. Graph analytics and knowledge graphs facilitate scientific research for COVID-19. **ZDNet**, [*s. i.*]: ZDNet, 2020. Available from: https://www.zdnet.com/article/graph-analytics-and-knowledge-graphs-facilitate-scientific-research-for-covid-19/. Retrieved: October 9, 2020.

COVIDGRAPH. [*S. l.*]: The CovidGraph Project, c2020. Available from: https://covidgraph.org/. Retrieved: October 9, 2020.

COVID-19 knowledge graph: a knowledge graph around COVID-19 in Biological Expression Language (BEL). [*S. l.*]: Github: c2020. Available from: https://covidgraph.org/. Retrieved: October 9, 2020.

CHEN, Chongyan; EBEID, Islam Akef; BU, Yi; DING, Ying. Coronavirus knowledge graph: a case study. In: KDD2020: ACM KNOWLEDGE DISCOVERY IN DATABASES, [26]., August 23-27, 2020, San Diego, California. **Proceedings […]**, New York, NY: ACM, 2020. Available from: http://arxiv.org/abs/2007.10287. Retrieved: October 9, 2020.

DOMINGO-FERNÁNDEZ, Daniel. COVID-19 Knowledge Graph: a computable, multi-modal, cause-and-effect knowledge model of COVID-19 pathophysiology. **Bioinformatics**, [*S. l.*] btaa834, Sept 2020. DOI: https://doi.org/10.1093/bioinformatics/btaa834. Available from: https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btaa834/5911625. Retrieved: October 9, 2020.

FENSEL, Dieter *et al.* **Knowledge graphs:** methodology, tools and selected use cases. [*S. l.*]: Springer International Publishing: 2020. *E-book*. DOI: https://doi.org/10.1007/978-3-030-37439-6.

FORD, Heather; GRAHAM, Mark. Provenance, power and place: linked data and opaque digital geographies. **Environment and Planning D: Society and Space**, [*s. l.*], v. 34, n. 6, p. 957-970, 2016. DOI: https://doi.org/10.1177%2F0263775816668857. Available from: https://journals.sagepub.com/doi/full/10.1177/0263775816668857. Retrieved: October 9, 2020.

GINSBERG, Allen. Ontological indeterminacy and the semantic web. **International Journal on Semantic Web and Information Systems**, [*s. l.*], v. 4, n. 2, p. 19-48, 2008. DOI: http://www.doi.org/10.4018/jswis.2008040102. Available from: https://www.igi-global.com/gateway/article/2848. Retrieved: October 9, 2020.

GOASDUFF, Laurence. 2 Megatrends Dominate the Gartner Hype Cycle for Artificial Intelligence. **Smarter With Gartner**, [*s. l.*]: Gartner Inc, 2020. Available from: https://www.gartner.com/smarterwithgartner/2-megatrends-dominate-the-gartner-hype-cycle-for-artificial-intelligence-2020/. Retrieved: October 9, 2020.

ILIADIS, Andrew. The Tower of Babel problem: making data make sense with Basic Formal Ontology. **Online Information Review**, [*s. l.*], v. 43, n. 6, p. 1021-1045, 2019 DOI: www.doi.org/10.1108/OIR-07-2018-0210. Available from: https://www.emerald.com/insight/content/doi/10.1108/OIR-07-2018-0210/full/html. Retrieved: October 9, 2020.

KEJRIWAL, Mayank. **Domain-specific knowledge graph construction**. [*S. l.*]: Springer International Publishing: 2019. *E-book*. DOI: https://doi.org/10.1007/978-3-030-12375-8.

KEJRIWAL, Mayank; KNOBLOCK, Craig A.; SZEKELY, Pedro. **Knowledge graphs:** fundamentals, techniques, and applications. Cambridge, MA: MIT Press; 2021.

PAN, Jeff Z.; VETERE, Guido; GOMEZ-PEREZ, Jose Manuel; WU, Honghan (ed.). **Exploiting linked data and knowledge graphs in large organisations**. [*S. l.*]: Springer International Publishing, 2017. *E-book.* DOI: https://doi.org/10.1007/978-3-319-45654-6.

QI, Gillien *et al.* **Knowledge graph**. [*S. l.*]: Springer International Publishing, 2017. *E-book.*

REESE, Justin *et al.* KG-COVID-19: a framework to produce customized Knowledge Graphs for COVID-19 Response. **BioRxiv**, [*S. l.*], 2020. *Preprint.* DOI: https://doi.org/10.1101/2020.08.17.254839. Available from: https://www.biorxiv.org/content/10.1101/2020.08.17.254839v1.article-info. Retrived: October 9, 2020.

RODRÍGUEZ, Clemencia; ILIADIS, Andrew. The MacBride Report legacy and media democracy today. **Media Development**, [*s. l.*], v. 3, p. 17-24. Available from: https://waccglobal.org/the-macbride-report-legacy-and-media-democracy-today/. Retrieved: October 9, 2020.

SINGHAL, Amit. Introducing the knowledge graph: things, not strings. *In*: GOOGLE. **The Keyword**. [*S. l.*], May 16, 2012. Available from: https://blog.google/products/search/introducing-knowledge-graph-things-not/. Retrieved: October 9, 2020.

VANG, Katrine Juel. Ethics of Google's Knowledge Graph: some considerations. **Journal of Information, Communication and Ethics in Society**, [*S. l.*], v. 11, n. 4, p. 245-260, 2013. DOI: https://doi.org/10.1108/JICES-08-2013-0028. Available from: https://www.emerald.com/insight/content/doi/10.1108/JICES-08-2013-0028/full/html. Retrieved: October 9, 2020.

WILKINSON, Michael D. *et al*. The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, [London], v. 3, p. 160018, 2016. DOI: https://doi-org.libproxy.temple.edu/10.1038/sdata.2016.18. Available at: https://pubmed.ncbi.nlm.nih.gov/26978244/. Retrieved: October 9, 2020.

WANG, Quinyun *et al*. COVID-19 literature knowledge graph construction and drug repurposing report generation [Submitted on 1 Jul 2020 (v1), last revised 20 Jul 2020 (this version, v3)]. **ArXiv**, [*s. i.*], 2020. *Preprint.* Available from: http://arxiv.org/abs/2007.00576. Retrieved: October 9, 2020.

WISE, Colby *et al*. COVID-19 knowledge graph: accelerating information retrieval and discovery for scientific literature. *In*: KNOWLEDGEABLE NATURAL LANGUAGE PROCESSING WORKSHOP, 1. December 7, 2020, Suzhou, China. **Proceedings** […]. Stroudsburg: ACL, 2020. Available from: http://arxiv.org/abs/2007.12731. Retrieved: October 9, 2020.