CovidGraph: Integrating COVID-19 Data

CovidGraph: In 2020 SARS-COV-2 began to impact life across the globe on a scale unbeknown to humanity. Over the last two years fast and extensive research in that field generated a vast amount of knowledge about the virus. Research-wise, COVID-19 has been encountered with publications, patents, genome analysis, simulation studies for spread prediction, health studies and the extension of ontology information. One factor for fast and reliable research is commitment to the FAIR guiding principles [1]. CovidGraph offers **f**indable **a**ccessible **i**nteroperable and **r**e-usable COVID-19 data obtained, integrated and connected from open data resources. Data sets from the aforementioned domains are stored in a graph database to offer researchers quick and efficient access to information about COVID-19 (Fig.1). The connections within CovidGraph allow for new types of queries across previously disconnected aspects of the disease.

Domains: CovidGraph comprises information about publications from the COVID-19 Open Research Dataset [2], information about patents [3] and clinical trials [4]. Biomedical entities (e.g. genes, transcripts and proteins) are integrated from a variety of well-established databases [5]. Statistical data is imported from Johns Hopkins University [6]. Simulation models in standard format [7], including a Covid-19 model collection, are integrated from a domain-specific graph database (MaSyMoS, [8]).

CovidGraph – Domains & Connections

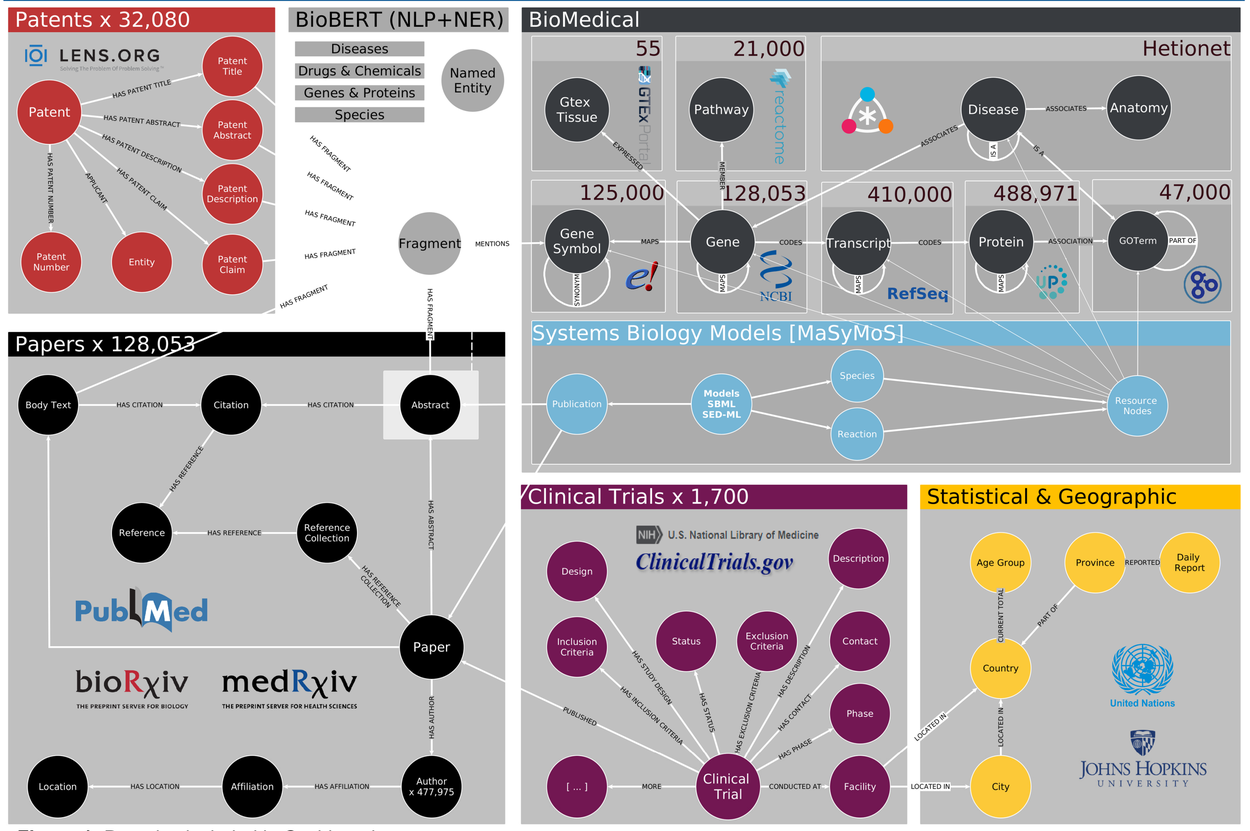

Figure 1: Domains included in Covidgraph

Interfaces

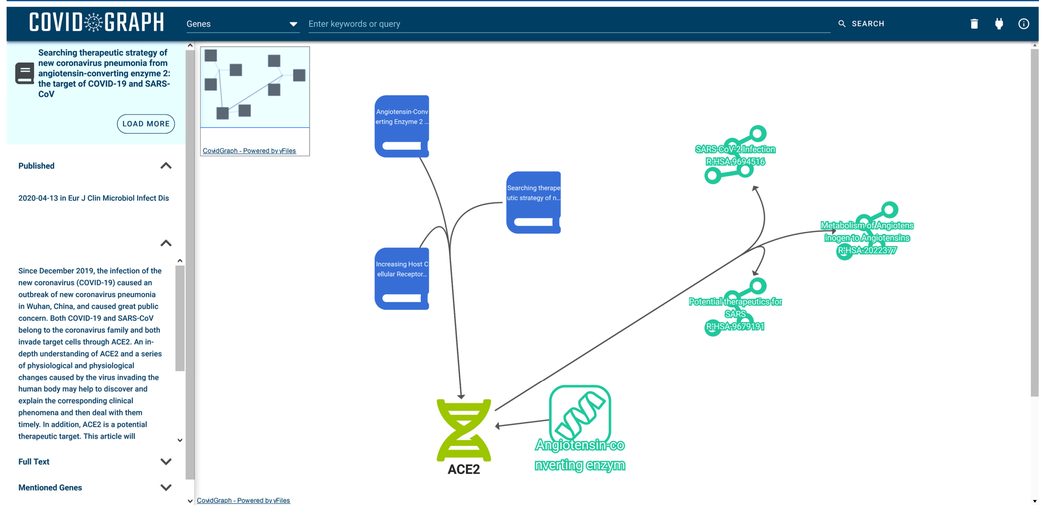

Figure 2: Visual Graph Explorer by yWorks

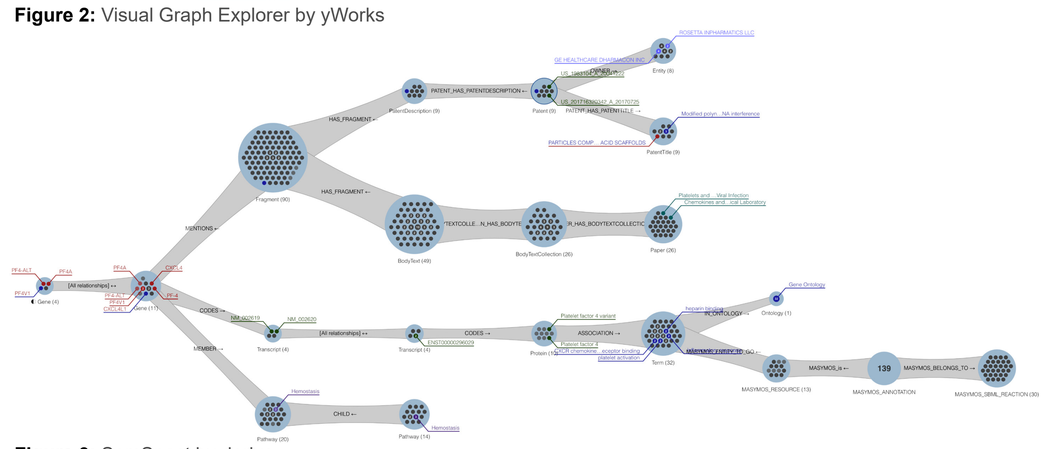

Figure 3: SemSpect by derivo

CovidGraph – How to explore the data?

CovidGraph offers several interfaces for data exploration. The Visual Graph Explorer (Fig. 2) provides predefined views for an intuitive keyword-based graph exploration without prior knowledge of database query languages. SemSpect [9] (Fig. 3) supports drag & drop, expand and filter data items and automatic grouping of similar data items. Hence the graph can easily be traversed and visual representations can be created without detailed knowledge of the data model. Neo4j Bloom (Fig. 4) is an application for graph exploration. It offers semi-natural language queries, rule-based styling and search for phrases.

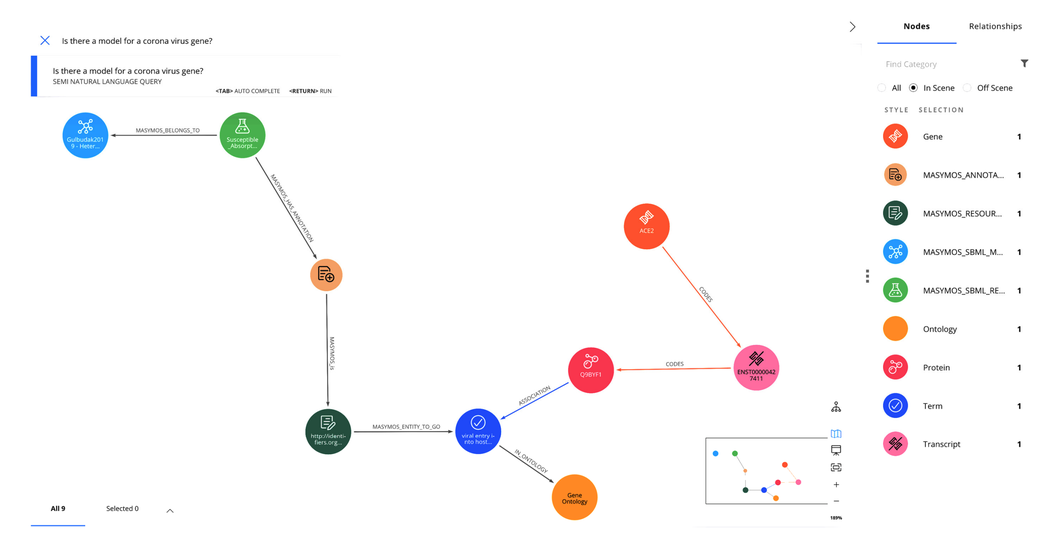

Figure 4: Bloom by Neo4J

References

[1] Wilkinson, M, Dumontier, M, Aalbersberg, I, Appleton, G, Axton, M, Baak, A, Blomberg, N, Boiten, JW, Silva Santos, L, Bourne, P, others. "The FAIR Guiding Principles for scientific data management and stewardship". Scientific data 2016; 3(1):1–9.
[2] Wang, L, Lo, K, Chandrasekhar, Y, Reas, R, Yang, J, Eide, D, Funk, K, Kinney, R, Liu, Z, Merrill, W, others. "CORD-19: the Covid-19 Open Research Dataset". arXiv 2020.
[3] The Lens. About The Lens. https://about.lens.org/covid-19.
[4] Zarin, D, Tse, T, Williams, R, Califf, R, Ide, N. "The ClinicalTrials.gov results database - update and key issues". New England Journal of Medicine 2011; 364(9):852–860.
[5] Henkel, R. "Biomedical Repositories for Simulation Studies". In: Reference Module in Biomedical Sciences. Elesvir, 2020. Available from: https://doi.org/10.1016/b978-0-12-801238-3.11684-8
[6] Dong, E, Du, H, Gardner, L. "An interactive web-based dashboard to track COVID-19 in real time". The Lancet Infectious Diseases 2020; 20(5):533–534.
[7] Malik-Sheriff, R, Glont, M, Nguyen, T, Tiwari, K, Roberts, M, Xavier, A, Vu, M, Men, J, Maire, M, Kananathan, S, others. "BioModels - 15 years of sharing computational models in life science". Nucleic Acids Research 2020; 48(D1):D407–D415.
[8] Henkel, R, Wolkenhauer, O, Waltemath, D. "Combining computational models, semantic annotations and simulation experiments in a graph database". Database 2015; 2015:1–6.
[9] Liebig, T, Vialard, V, Opitz, M Connecting the Dots in Million-Nodes Knowledge Graphs with SemSpect. In International Semantic Web Conference (Posters, Demos & Industry Tracks) 2017

Ron Henkel · Lea Gütebier · Dagmar Waltemath

Medizininformatik · Universitätsmedizin Greifswald
https://www.medizin.uni-greifswald.de/medizininformatik/