A survey of biomedical knowledge graphs and of resources for their construction

Author: Robert Haas

Date: February 6, 2024

Version: 1.0.0

Contents

1 Introduction

Biomedical research generates vast amounts of heterogeneous data, which is stored in many different formats at various locations. Knowledge graphs have emerged as a promising tool for integrating, analyzing and querying this diverse information at scale, thereby paving the way for deriving new insights from a more holistic perspective of biomedical entities and relations. In recent years, there has been a significant increase in the amount of work in this area, making it difficult to keep track of all developments. To help address this issue, this survey presents a comprehensive overview of biomedical knowledge graphs (section 3) and of resources for their construction (section 4). These resources include a) definitions of knowledge graphs and their underlying graph data models (section 4.1), b) file formats to represent and store them (section 4.2), c) databases to retrieve knowledge from (section 4.3), d) ontologies and controlled vocabularies to provide additional domain structure (section 4.4), and e) tools to extract and integrate various kinds of information to generate a knowledge graph (section 4.5).

As supplement to this survey, a website was created that provides all tables of this survey in a more accessible and interactive form: `https://robert-haas.github.io/awesome-biomedical-knowledge-graphs`

Errors and unintended omissions are solely the author's fault. Critical feedback is welcome and encouraged to be sent to robert.haas@protonmail.com.

2   Methodology

To ensure broad coverage, entries were collected from a diverse set of sources, including review articles [6] [33] [84] [101] [115] [201] [242] [370] [448] [451] [514], catalogs [32] [96] [373], results obtained from queries to general-purpose and scholarly search engines, and repositories listed by source code hosting sites.

In current version 1.0.0 of this survey, following exclusion criteria were applied to discovered candidate projects: A content cutoff date has been set as December 31, 2023. This means that projects published afterwards and updates provided later than this date are not covered by this survey. Additionally, projects were excluded if they only had a publication but no accompanying resources such as a website, code repository, or data access in form of downloadable files or queryable APIs. Finally, projects were not included even if their results might qualify by common definitions as a knowledge graph (section 4.1) but they did not refer to them as such, and if they also did not provide them in any file format commonly used for biomedical knowledge graphs (section 4.2).

3 Projects that provide biomedical knowledge graphs

Name	Websites	Publications	Code	Data	Last Update	Organization
ADHD-KG	[424]	[427]	GitHub [425]	.n3 / .nt [425] [426]	2022	School of Computing and Engineering, University of Huddersfield [261]
BIKG	-	[204]	-	-	2022	Company: AstraZeneca [30] [31]
BIOS	[7]	[596]	GitHub [587] [586]	.txt [8]	2022	International Digital Economy Academy [9], Center for Statistical Science, Tsinghua University [111], Individuals from other research institutes
BY-COVID-KG	-	-	GitHub [302]	.pkl / .csv / .graphml / .sif [299]	2022	BeYond-COVID (BY-COVID) project, various European research institutes [43]
Bio2RDF	[129] [127]	[39] [164]	GitHub [130]	.nq / .owl [128] [131], SPARQL endpoint [132]	2014	Individuals from various research institutes
Bio4j	[491]	[428]	GitHub [49] [50]	.dmp (TitanDB) [51] [52]	2014	Company: Era7 Bioinformatics [410]
BioGrakn	[471]	[379]	GitHub [540]	.zip (Grakn) [541]	2018	Company: Vaticle Ltd [542]
BioGraph	[545]	[546]	GitHub [543]	.zip (Neo4J) [544]	2023	Individuals from various research institutes
BioKG	-	[562]	GitHub [274] [273], PyPI [149]	.tsv [275]	2020	Biomedical Discovery Informatics Unit, NUI Galway [202] [150]
Biomedical Data Translator	[10] [20]	[139] [181]	GitHub [134] [133] [135]	REST API [138], Web UI [15]	2023	National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH) [13]

Bioteque	[55]	[184] [183]	GitLab [71]	.tsv [54]	2022	Structural Bioinformatics and Network Biology Group, IRB Barcelona [56]
Biozon	[74]	[76] [77]	Website [75]	.sql (Post-greSQL) [75]	2005	Department of Developmental Biology, Stanford University [66]
COVID-19 Knowledge Graph	[118]	[122]	GitHub [53]	.ttl [119], REST API [120], SPARQL GUI [121]	2022	Department of Computer and Information Sciences, University of Delaware [152]
COVID-19-Net	[525]	-	GitHub [143]	Neo4j Browser [464]	2022	Individuals from various research institutes
COVID-KG	[80]	[563]	-	.csv [81]	2020	BLENDER Lab, University of Illinois [79]
CROssBAR	[380] [505]	[161]	GitHub [4]	.csv / .tsv [3]	2023	Cancer Systems Biology Laboratory (CanSyL), METU [333], Protein Function Development Team, EMBL-EBI [172]
Chem2Bio2RDF	[64]	[116]	-	RDF [65]	2010	School of Informatics and Computing, Indiana University [272]
Clinical Knowledge Graph (CKG)	[485]	[488] [486]	GitHub [334]	.dump (Neo4j) [487]	2021	NNF Center for Protein Research, University of Copenhagen [439]
CovidGraph	[248] [247] [144]	[236]	GitHub [146]	Web UI [377], Neo4j Browser [378]	2022	Covidgraph.org Team, Individuals from various research institutes [145]
CovidPubKG	-	[430]	GitHub [219]	.nt / .ttl [421] [429], SPARQL endpoint [422]	2021	DICE group, Paderborn University [233]
DDI-BLKG	-	[91]	GitHub [89]	.txt [90]	2020	Institute of Informatics and Telecommunications, NCSR Demokritos [276]
DRKG	-	[292]	GitHub [293]	.tsv / .npy [294]	2022	Individuals from various research institutes

DeepKG	[348]	[349]	-	Triplets [348]	2021	Individuals from various research institutes
DisGeNET-RDF	[224] [227]	[431] [449] [432]	Bitbucket [223] [222]	.ttl [228], SPARQL endpoint [230] [229], .tsv / .sql (SQLite) [225], REST API [226]	2020	Integrative Biomedical Informatics (IBI) group, Research Programme on Biomedical Informatics (GRIB) [67]
Drug-CoV	-	[346]	GitHub [344]	.csv [345]	2023	School of Information Technology, Murdoch University [526]
DrugMechDB	[321]	[210] [372]	GitHub [319]	-	2023	Su Lab, Scripps Research Institute [330]
DrugRep-HeSiaGraph	-	[206]	GitHub [205]	.csv [205]	2023	CBRC Lab, Amirkabir University Of Technology [110]
DrugRep-KG	-	[208]	GitHub [207]	.csv / .txt [209]	2023	CBRC Lab, Amirkabir University Of Technology [110]
EmBiology	[169]	[168]	-	Commercial access	2023	Company: Elsevier [167]
EpiGraphDB	[397]	[354]	GitHub [520] [522]	Neo4j build pipeline [523], REST API [521]	2021	MRC Integrative Epidemiology Unit, University of Bristol [95]
GP-KG	-	[203]	File server [103]	.txt [104]	2022	Center for Artificial Intelligence in Drug Discovery, Case Western Reserve University [102]
GenomicKB	[356]	[182]	-	.dump (Neo4j) [355]	2023	Liu Lab, University of Michigan [353]

Google Health Knowledge Graph	[212] [211]	-	-	REST API with JSON-LD responses [214] [215]	2020	Company: Google [213]
GrEDeL	-	[482]	[479]	-	2019	Individuals from various research institutes
HALD	[22]	[585]	GitHub [240]	.json / .csv [584]	2023	Ming Chen's Group of Bioinformatics, Zhejiang University [231]
HemeKG	[264]	[266]	GitHub [262], PyPI [263]	.json (BEL JSON) [265]	2021	Pharmaceutical Biochemistry and Bio-analytics, University of Bonn [83], Department of Bioinformatics, Fraunhofer SCAI [489]
Hetionet	[218]	[258] [257]	Website [217], GitHub [255] [256], PyPI [252] [254]	.json / .tsv / .dump (Neo4j) / .npy (NumPy) [216] Neo4j Browser [253]	2018	Baranzini Lab, UCSF [529], Greene Lab, University of Pennsylvania [329]
HuadingKG	-	[603]	GitHub [602]	RDF (Apache Jena) [602]	2019	Individuals from various research institutes
IASiS Open Data Graph	[438]	[405]	GitHub [504]	MongoDB / Neo4j build pipeline [504]	2020	Institute of Informatics and Telecommunications, NCSR Demokritos [289]
IDSM	[57]	[199]	GitLab [194]	.owl / .ttl / .sql (PostgreSQL) pipeline [195], SPARQL endpoints [196] [197] [198]	2023	Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences [290]

INDRA	[498] [238]	[239] [34]	GitHub [499], PyPI [237]	.tsv [501], .pkl / .json [35], REST API [500]	2023	Sorger Lab, Harvard Medical School [434]
Implicitome	-	[250]	GitHub [70]	.csv / .nq [249]	2017	Individuals from various research institutes
KEGG50k	-	[386]	-	.txt [385]	2018	Data Science Institute, University of Galway [536]
KG-COVID-19	[459]	[460]	GitHub [313]	.nt / .tsv [458], SPARQL endpoint [339]	2023	Individuals from various research institutes
KG-Hub	[341]	[105]	GitHub [539]	.nt [340]	2023	Individuals from various research institutes
KGen	-	[467]	GitHub [465]	.txt / .ttl [466]	2022	Institute of Computing, University of Campinas [288]
KIDS	-	[595]	GitHub [63]	.txt / .xml [62]	2022	Tagkopoulos Lab, University of California [513]
KaBOB	-	[358]	GitHub [357]	RDF pipeline [357]	2015	Computational Bioscience Program, University of Colorado [511]
Knowledge4COVID-19 KG	-	[478]	GitHub [232], Zenodo [298]	SPARQL endpoint [548]	2022	L3S Research Center, University of Hannover [317], Individuals from various research institutes
LMKG	-	[591]	GitHub [589]	.json [590], Neo4j Browser [155]	2023	Department of Electronic Engineering, Tsinghua University [154], Individuals from various research institutes
MDKG	-	[191]	GitHub [189]	.dict / .txt [190]	2020	Central China Normal University [112]

Monarch KG	[389] [390]	[374] [400] [445]	GitHub [391] [388], PyPI [166]	.nt, .jsonl, .tsv, .db (SQLite), .dump (Neo4j) [387]	2023	Department of Biomedical Informatics, University of Colorado [512], Individuals from various research institutes
Mpox Knowledge Graph	[300]	[305]	GitHub [304]	.csv / .graphml / .pkl / .sif [303] [306], Web UI [301]	2022	Fraunhofer Institute for Translational Medicine and Pharmacology (ITMP) [187]
NGLY1 Deficiency KG	-	[450]	GitHub [325] [318]	.csv [323], Neo4j Browser [322]	2019	Su Lab, Scripps Research Institute [331]
NeDRex	[437]	[476]	GitHub [496] [472] [473]	REST API [474], Neo4j API [475], Cytoscape app [147]	2021	Individuals from various research institutes
Neo4COVID-19	[401]	[597]	GitHub [402]	Neo4j Browser [403]	2022	National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH) [14]
OREGANO	-	[85] [88]	GitLab [178]	.tsv [177] [86] [87], SPARQL endpoint [176]	2023	Population Health Research Center, University of Bordeaux [1]
Open Graph Benchmark: ogbl-biokg	[41]	[260]	GitHub [497], PyPI [506]	.csv [40]	2023	Department of Computer Science, Stanford University [527], Individuals from various research institutes
OpenBioLink	[418]	[93]	GitHub [414], PyPI [415]	.tsv [419], .tsv / .nt / .bel [94]	2020	Institute of Artificial Intelligence, Medical University of Vienna [547]
Otter-Knowledge	[200]	[335]	GitHub [369]	.nt [270] [269] [268] [267]	2023	Company: IBM [368]

PGxLOD	[362]	[148] [393]	GitHub [371] [392]	SPARQL end-point [364], FCT Browser [363]	2019	PractiKPharma project, Loria [2], Loria, Joint Research Unit between CNRS, University of Lorraine, INRIA [29]
PharMeBINet	[46]	[316]	GitHub [315]	.db (Neo4j) / .graphml [47] [314]	2022	Bioinformatics / Medical Informatics Department, University Bielefeld [44]
PharmKG	-	[608]	GitHub [382]	.tsv [607]	2020	School of Data and Computer Science, Sun Yat-sen University [528]
PheKnowLator	[509]	[99] [100]	GitHub [98], PyPI [515], Zenodo [159]	.owl / .nt / .gpickle / .pkl / .tsv / .json [97] [158] [157]	2021	Computational Bioscience Program, University of Colorado [535], Individuals from various research institutes
PrimeKG	[620]	[114]	GitHub [381]	.csv / .tab [113]	2022	Zitnik Lab, Harvard University [619]
PubChemRDF	[443]	[192] [308]	-	.ttl [440] [441], REST API [444]	2023	PubChem, NCBI [442]
PyKEEN	[508]	[26] [23] [25]	GitHub [446], PyPI [24]	.csv [507] [447]	2023	Smart Data Analytics Group, University of Bonn [342], Individuals from various research institutes
RDKG-115	-	[617]	GitHub [611]	.csv [612] [613]	2023	Intelligent Medicine Institute, Fudan University [291], Individuals from various research institutes
RNA-KG	-	[109] [108]	GitHub [28]	.nt / .txt [107], SPARQL endpoint [106]	2023	AnacletoLAB, University of Milan [27], Individuals from various research institutes

ROBOKOP	[283] [285]	[396] [78]	GitHub [395]	.dump (Neo4j) / .json / .jsonl [278], Neo4j Browser [279], REST API [284]	2023	Renaissance Computing Institute (RENCI), University of North Carolina [281]
RPath KGs	-	[163]	GitHub [68]	.tsv [69] [162]	2022	Company: Enveda Biosciences [175]
RTX-KG2	-	[583]	GitHub [468]	.json [470] [404]	2023	Development team for the RTX biomedical reasoning tool [469], Individuals from various research institutes
Rare diseases integrative KG	-	[618]	GitHub [17]	Neo4j Browser [16]	2020	Division of Pre-Clinical Innovation, NCATS, NIH [11]
Reactome Graph Database	[454]	[297] [179]	GitHub [455] [456]	.graphdb (Neo4j) [453], REST API [452]	2023	Reactome Team [457], Individuals from various research institutes
Rosalind knowledge graph	[463]	[423]	-	-	2020	Company: ROSALIND, Inc. [462], Company: BenevolentAI [42]
SDKG-11	-	[616]	GitHub [614]	.csv [615]	2022	Institutes of Biomedical Sciences, Fudan University [287], Individuals from various research institutes
SLKG	[598]	[600]	-	Web UI [599]	2021	Individuals from various research institutes
SPOKE	[532] [530] [533] [408]	[394]	GitHub [409]	REST API [534]	2023	SPOKE Team [531], Individuals from various research institutes
SemKG	-	[483]	GitHub [480]	.txt [481]	2021	College of Computer Science and Technology, Dalian University of Technology [510]
SemNet / SemMedDB	[311]	[490] [312]	GitHub [309]	.csv / .sql (MySQL) [375]	2022	Laboratory for Pathology Dynamics, Emory University [310]

StrokeKG	-	[594]	GitHub [593]	.csv / .txt [593], Neo4j browser [592]	2020	College of Computer Science and Technology, National University of Defense Technology [151], Individuals from various research institutes
SynLethKG / KG4SL	[609]	[564] [235]	GitHub [604] [606]	.csv / .npy / .txt [605], .csv / .json / .graphml [610]	2023	Zheng Lab, ShanghaiTech University [494]
TogoGenome	[352]	[307]	GitHub [517]	SPARQL endpoint [351]	2019	Database Center for Life Science, University of Tokyo [350]
TypeDB Bio	[365]	-	GitHub [416]	.csv [417]	2023	Company: Vaticle Ltd [366]
Wikidata Biomedical Knowledge Graph	-	[561]	GitHub [320] [328] [324] [326] [327]	.json / .ttl / .nt / .xml [569], SPARQL endpoint [568]	2023	Su Lab, Scripps Research Institute [332]

Table 1: Projects that provide biomedical knowledge graphs.

4 Resources for creating biomedical knowledge graphs

4.1 Definitions

The phrase "knowledge graph" can be traced back in literature to the 1970s, but its contemporary use was initiated by a Google product announcement in 2012, which soon afterwards inspired similar releases by other companies and in turn led to a renaissance of studying the concept in academia [259]. In light of this historical development, it comes to no surprise that the phrase "knowledge graph" is associated with more than one meaning in current practice. One way to characterize the various incarnations is to understand them as instances of different graph data models that can be defined unambiguously by formal notation from basic set theory. This approach has the advantage that it captures different kinds of knowledge graphs precisely, allows to study the properties that differentiate them, and thereby make informed decisions on what type is best suited for particular use cases.

The following part of the survey provides a collection of several such definitions of the phrase "knowledge graph" in varying degrees of formal rigor. All of them were extracted from biomedical knowledge graph projects listed in the previous section and from references they provide in the descriptions of their data models. Outside of this restricted set of articles, a broader literature review might uncover other useful overviews and discussions, but the purpose of this section is to reflect what definitions are used in practice by projects that construct biomedical knowledge graphs in order to reflect the state of the art in the field of biomedicine. To this end, it is worth to note that 1) not all projects explicitly describe what type of knowledge graph they are constructing but rather require the reader to deduce it from the context and results, and 2) only a few projects provide references to detailed discussions of the graph data model they have chosen for their purposes, leaving it to the reader to identify and acquire presupposed background knowledge.

Of all projects covered by this survey, two stood out by containing references to particularly high-quality discussions of the phrase "knowledge graph". First, OREGANO [88] referred to a comprehensive article that covers various common types of knowledge graphs, languages to query and validate them, semantic frameworks to interpret them, and several other aspects in a highly accessible tutorial-like way [259]. Second, SynLethKG / KG4SL [564] referred to a recent book about graph representation learning which contains clean formal definitions of diverse graph types, moving from basic to more advanced variants, followed by an in depth discussion of various graph embedding approaches [241].

The following content is organized by listing each project that provides a definition of the phrase "knowledge graph" or associated concepts in a separate subsection. Relevant passages were quoted both from the project's own publications and, in the few cases where relevant references were available, also from external articles that contain more detailed elaborations. Bold text highlights were added by the author of this survey to mark relevant terms and phrases, while all other forms of emphasis were reproduced from the original sources. Citations were also kept as they are found in the original texts and therefore adhere to a few different styles.

4.1.1 Bio2RDF

- Bio2RDF publication:

  Each year, NAR [3] publishes a new version of its bioinformatics database list. [...] With such a proliferation of knowledge sources, there is a pressing need for a global multisite search engine and for good **data integration tools**. According to the **data warehouse** approach, such services can be built by collecting information into a **central data repository** [5] and queried with an interface built on top of the repository. However, the warehouse approach does not address the problem of accessing a database outside the warehouse. A system that would be able to query and connect different databases available on Internet would solve that problem. This is one of the goals of the **semantic web approach**: to offer the data warehouse experience without the need of moving first the data into a central repository.
  To address the data integration problem, the semantic web community, led by the W3C, proposed a solution based on a series of standards: the **RDF format** for document [6] and the **OWL language** for ontology specification [7]. RDF and OWL generate a series of entities called 'triple' in the form of a subject, predicate and object. Database systems able

to handle triples are called **triplestore**. [...] We have developed a semantic web application called Bio2RDF to help solve the problem of knowledge integration in bioinformatics. ([39])

- Reference [5] in the Bio2RDF publication:

  There are three main ways in which groups have tried to **integrate biological databases**, which are referred to here as link integration, view integration and data warehouses. [...]

  Link integration has been by far the most successful, because it lends itself to the haphazard nature of the web. In the familiar and ubiquitous version of this technique, researchers begin their query with one data source, and then follow hypertext links to related information in other data sources. [...]

  View integration leaves the information in its source databases, but builds an environment around the databases that makes them all seem to be part of one large system. [...]

  The last general approach can be broadly described as bringing all the data under one roof in a single database (FIG. 5). The first step in **data warehousing** is to develop a unified data model that can accommodate all the information that is contained in the various source databases. The next step is to develop a series of software programs that will fetch the data from the source databases, transform them to match the unified data model and then load them into the warehouse. The warehouse can then be used as a 'one-stop shop' for answering any of the questions that the source databases can handle, as well as those that require integrated knowledge that the individual sources do not have. [...]

  Ontologies are a sophisticated type of controlled vocabulary that attempt to capture the main concepts in a KNOWLEDGE DOMAIN[13]. [...] Although ontologies do not, by themselves, lead to the integration of biological databases, they are important facilitators. The existence of a shared ontology allows an integrator to merge two databases with some guarantee that a **term** that is used in one database corresponds to the same term used in the other. [...] An important feature of biological ontologies is that the terms are organized in a **hierarchical** manner, so that more specific terms are defined as specializations of more general ones. [...] To support the complex relationships that are common in biology, terms are allowed to have more than one parent in a data structure that is formally called a **directed acyclic graph (DAG)**. [...]

  Globally unique identifiers: Shared ontologies can help bioinformaticians agree on how to describe biological objects, but they do not necessarily help them to agree on how to name them. Recall the issues that the same biological object might have multiple names, and the same name might denote multiple objects. One approach is to designate an **authoritative names commission** to manage the definitive list of such names, as the HUGO Gene Nomenclature Committee is attempting to do with human gene symbols. This rarely works in practice because of the dynamic nature of the field. [...] Although a central naming authority is usually impractical, there are often **de facto authorities** for local subsets of the names. For example, the NCBI can be considered to be authoritative for GenBank identifiers, and WormBase is authoritative for C. elegans gene symbols. ([503])

- Reference [6] in the Bio2RDF publication:

  Resource Description Framework (RDF): RDF is a standard model for **data interchange** on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed.

  RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a "**triple**"). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications.

  This linking structure forms a **directed, labeled graph**, where the edges represent the named link between two resources, represented by the graph nodes. This graph view is

the easiest possible mental model for RDF and is often used in easy-to-understand visual explanations. ([552])

- Reference [7] in the Bio2RDF publication:

  The W3C **Web Ontology Language (OWL)** is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things. OWL is a computational logic-based language such that knowledge expressed in OWL can be exploited by computer programs, e.g., to verify the consistency of that knowledge or to make implicit knowledge explicit. OWL documents, known as ontologies, can be published in the World Wide Web and may refer to or be referred from other OWL ontologies. OWL is part of the **W3C's Semantic Web technology** stack, which includes RDF, RDFS, SPARQL, etc. ([551])

4.1.2 BioKG

- BioKG publication:

  The BioKG **knowledge graph** compiles biological data from different sources in a graph format with focus on data on proteins and chemical drugs. The contents of BioKG knowledge graph can be categorized into three categories: links, properties and metadata. Links represent the connections between the different biological entities, while properties represent the annotations associated to entities. Each biological entity type has its own set of links and properties that describe its activities in biological systems. [...]

  The **links** part of the BioKG data is the core part of BioKG which models the relationships between the biological entities as illustrated in Fig. 1. [...]

  The **properties** part of BioKG contains the associations between the previously discussed biological entities and their different attributes as illustrated in Fig. 5. [...]

  The **metadata** part contains data about biological entities names, types, synonyms, etc. This part of the data is not meant to be used in the training of relational learning models, and it does not contain any attributes or important associations for biological entities. Our objective, however, in this part is to maximize the richness of metadata on each of the included biological entities to facilitate analysing their related insights and to allow for tracking history of changes of ids and synonyms of biological entities' databases entries. ([562])

4.1.3 Bioteque

- Bioteque publication:

  Building the **metagraph**. All gathered data was stored in a **graph database (KG)** in which nodes represent biological or chemical entities and edges represent associations between them.

  Nodes (**entities**). The nodes in the graph can belong to one of 12 types (aka metanodes). For each entity type, we predefined a universe of nodes and chose a reference vocabulary based on standard terminologies. [...]

  Vocabulary mapping. To integrate terminologies, we extracted curated cross-references from the official terminology sources and associated ontologies. As the nomenclatures used to identify diseases and pathways were particularly diverse and rarely cross-referenced, we further increased the mapping of these terms by inferring similarities within concepts as detailed below. [...]

  Edges (associations). Edges in the graph are used to link biological and/or chemical entities. Since two entities may be connected by multiple edge types (i.e., 'compound treats disease' or 'compound causes disease'), we define the associations as **triplets** (metapaths) of entity-relationship-entity (CPD-trt-DIS, CPD-cau-DIS).

  Homogeneous associations are those concerning entities (metanodes) of the same type (e.g.,'gene is co-expressed with gene', GENcex-GEN), while **heterogeneous associations** are related to entities of different types (e.g., 'tissue has cell', TIS-has-CLL). ([184])

4.1.4   CROssBAR

- CROssBAR publication:

  The term **knowledge graph (KG)** defines a specialized data representation structure, in which collections of entities (nodes) are linked to each other (edges) in a semantic context (29). In this study, we chose to represent **heterogeneous biomedical data** using a KG-based structure. In CROssBAR knowledge graphs (CROssBAR-KG), biological components/terms (i.e. drugs, compounds, genes/proteins, bio-processes/pathways, phenotypes and diseases) are represented as nodes, and their known or predicted pairwise relationships are annotated as edges (a protein and its coding gene is treated as one merged term/entry/node). ([161])

- Reference (29) in the CROssBAR publication:

  A **knowledge graph** is a **multi-relational graph** composed of entities as nodes and relations as different types of edges. An instance of edge is a **triplet** of fact *(head entity, relation, tail entity)* (denoted as (h, r, t)). In the past decade, there have been great achievements in building large scale knowledge graphs, however, the general paradigm to support computing is still not clear. Two major difficulties are: (1) A knowledge graph is a symbolic and logical system while applications often involve numerical computing in continuous spaces; (2) It is difficult to aggregate global knowledge over a graph. [...]

  Recently a new approach has been proposed to deal with the problem, which attempts to embed a knowledge graph into a **continuous vector space** while preserving certain properties of the original graph (Socher et al. 2013; Bordes et al. 2013a; Weston et al. 2013; Bordes et al. 2011; 2013b; 2012; Chang, Yih, and Meek 2013). For example, each entity h (or t) is represented as a point $\mathbf{h}$ (or $\mathbf{t}$) in the vector space while each relation r is modeled as an operation in the space which is characterized by an a vector $\mathbf{r}$, such as translation, projection, etc. The representations of entities and relations are obtained by minimizing a global loss function involving all entities and relations. As a result, even the embedding representation of a single entity/relation encodes global information from the whole knowledge graph. Then the embedding representations can be used to serve all kinds of applications. A straightforward one is to complete missing edges in a knowledge graph. For any candidate triplet (h, r, t), we can confirm the correctness simply by checking the compatibility of the representations $\mathbf{h}$ and $\mathbf{t}$ under the operation characterized by $\mathbf{r}$.

  Generally, **knowledge graph embedding** represents an entity as a k-dimensional vector $\mathbf{h}$ (or $\mathbf{t}$) and defines a **scoring function** $f_r(\mathbf{h}, \mathbf{t})$ to measure the plausibility of the triplet (h, r, t) in the embedding space. The score function implies a transformation $\mathbf{r}$ on the pair of entities which characterizes the relation r. ([565])

4.1.5   Drug-CoV

- Drug-CoV publication:

  Knowledge graph. A KG is a type of structured data that represents knowledge as a graph. In this graph, nodes represent entities, and edges represent relationships or connections between them. A KG can be considered a type of **semantic network** that is used to organize and represent knowledge in a machine-readable format [66]. It represents information in the format of triples *(subject, relation, object)*. Notable examples include Wikidata [59] and DrugBank [68]. The notation of the KG in this paper is denoted as $G = (E, G)$, where E is the set of entities (e.g., drugs, diseases and genes) and R is the set of relations (e.g., cause, encode and target) that connect the entities.

  Entity and relation. In a KG, **entities** refer to the real-world objects, concepts, or events that are being represented. **Relations**, on the other hand, describe the connections or interactions between entities in the real world. In a triple $(subject, relation, object)$, the term subject (or object) also can be used interchangeably with the subject entity (or object entity). In this paper, a subject is denoted as $s \in E$, an object is denoted as $o \in E$ and a relation is denoted as

$r \in R$. Multi-relation. In our paper, we adopt the definition of **multi-relations** introduced by [27]. Multi-relations refer to a situation where multiple types of relations (or edges) exist between a pair of entities. ([346])

- Reference [27] in the Drug-CoV publication:

  A **Knowledge Graph (KG)** represents a **graph-structured knowledge base** to encode real-world entities and illustrate the relationship between them. There are two main branches: triple-based KGs and quadruple-based KGs. The former is typically represented as sets of **Resource Description Framework (RDF) triples** (s, r, o), where s is the subject entity, o is the object entity and r is the relation, e.g. (Barack Obama, president of, USA). The latter is often called Temporal KGs (TKGs) represented as quadruples (s, r, o, T), where T denotes the timestamp, e.g. (Barack Obama, president of, USA, 2010). Although typical KGs contain millions of entities and billions of triples, they are far from complete [1]. [...]

  Despite their success in predicting missing links, the triple/path-level learning methods disconnect the diverse aspects of **multi-relations** that are highly semantically related [10]. For an instance extracted from FB15k-237 as shown in Table 1, there are a total of four relations between entity Prince Edward Island and entity Canada, which indicates the positional relationship at different hierarchies. In the **triple-level learning**, models split the example into four different triples and feed them to the score function f : $f(PrinceEdwardIsland, r_a, Canada)$, $f(PrinceEdwardIsland, r_b, Canada)$, $f(PrinceEdwardIsland, r_c, Canada)$ and $f(PrinceEdwardIsland, r_d, Canada)$. In the **path-level learning**, models firstly mine paths between entity $PrinceEdwardIsland$ and entity $Canada$. Then models learn the score function based on the four triples and paths. Either triple-level learning or path-level learning treats this example as four different triples and therefore weakens the semantic connection between the four relations. Obviously, to reflect the relevant knowledge, it is more reasonable to model multi-relations between a pair of entities as an individual vector representation rather than divide them into four different triples [10]. ([347])

- Reference [10] in the previous reference [27]:

  An elementary fact of the **knowledge graph** is represented in the form of a triple with two entities and a relation, i.e., $(head, relation, tail)$ denoted by (h, r, t). For example, $(Obama, bornhere, USA)$ corresponds to the knowledge that Obama was born in the USA. [...] For a given pair of entities, if there is more than one relation simultaneously between them, then each of these relations is called a **hyper-relation**. Conversely, if a relation has never co-occurred with another relation in some entity pairs, it is called a **sole-relation**. ([601])

4.1.6 DrugRep-HeSiaGraph

- DrugRep-HeSiaGraph publication:

  KGs are a proper way to represent information and knowledge in a structural foundation. In order to create a KG, it is essential to address these fundamental aspects: KG construction, node and relationship definition (edge types), and numerical vector embedding, which all play a vital role in its effectiveness. A **knowledge graph** as $G = < V, E, R >$ is defined where V shows the set of nodes, E demonstrates the set of edges, and R is the relationship types between nodes. In KG, each **triplet** is specified as $< h, r, t > \in E$, where $h, t \in V$, and $r \in R$ shows the relationship type r between h, t. ([206])

4.1.7 GenomicKB

- GenomicKB publication:

  Knowledge graphs intuitively represent connected data entities, and have been applied to biological domains (7–12). Compared with traditional tabular-structured data stored at

separate portals, GenomicKB emphasizes the relations between genomic entities at multiple resolutions and from multiple tissues and cell types. Entities from each consortium automatically and explicitly cross-link with one another in the knowledge graph without any operations such as table joining and sorting. In addition, our GenomicKB is rigorously built with well-defined schemata, identities, and ontologies to maintain the data structure, disambiguate genomic concepts, and support future extension. As a result, GenomicKB is not only flexible to adapt updates of nodes, relations, and entire data sources, but also connects with other knowledge graphs in related biomedical domains. [...]

Schemata prescribe high-level structures and semantics that the knowledge graph follows, which reduces data errors and allows reasoning over the data graph (15). [...]

Identity consolidates a set of unique identifiers and disambiguates different genomic identities in the knowledge graph. Since different data sources may follow different conventions to represent the same concept (e.g. ENSG00000223972 and gene DDX11L1), or use the same name to describe different concepts (e.g., gene p53 and protein p53), we use *globally-unique identifiers* and *external identity links* in GenomicKB. For example, for genes, transcripts and exons, we refer to Ensembl (16) IDs for their external identity links. [...]

Ontology is a uniform language to describe scientific terms. Concepts such as cell lines and tissues are represented as ontology URLs and IDs instead of common names to ensure disambiguity and future integration with other knowledge graphs. ([182])

4.1.8   GP-KG

- GP-KG publication:

  The embedding module of KG-Predict is used to transform entities (e.g., drugs, diseases), relations (e.g., Drug–treat–disease), and their features into low-dimensional vector representations whilst maximally preserving properties like graph structure and information. These entity and relation representations are used to predict unseen interactions in the knowledge graph. More specifically, we first define the **knowledge graph** as, $G = (V, E, X, R, S)$ where V and E denote the set of entities and relations, respectively. $T \subseteq V \times E \times V$ denotes the set of **triplets**, X represents features of nodes, R denotes the set of relations, S denotes the initial relation features. KG-Predict takes G as input and learns embeddings of entities and relations by aggregating multi-relational information in the knowledge graph. ([203])

4.1.9   Hetionet

- Hetionet publication:

  We created a general framework and open source software package for representing heterogeneous networks. Like traditional graphs, **heterogeneous networks** consist of nodes connected by edges, except that an additional meta layer defines type. Node type signifies the kind of entity encoded, whereas edge type signifies the kind of relationship encoded. Edge types are comprised of a source node type, target node type, kind (to differentiate between multiple edge types connecting the same node types), and direction (allowing for both directed and undirected edge types). The user defines these types and annotates each node and edge, upon creation, with its corresponding type. The meta layer itself can be represented as a graph consisting of node types connected by edge types. When referring to this graph of types, we use the prefix 'meta'. **Metagraphs** — called **schemas** in previous work [34,35] — consist of metanodes connected by metaedges. In a heterogeneous network, each **path**, a series of edges with common intermediary nodes, corresponds to a **metapath** representing the type of path. A path's metapath is the series of metaedges corresponding to that path's edges. The possible metapaths within a heterogeneous network can be enumerated by traversing the metagraph. ([257])

- Reference [35] in the Hetionet publication:

  An **information network** is defined as a **directed graph** $G = (\mathcal{V}, \mathcal{E})$ with an object type mapping function $\tau : \mathcal{V} \to \mathcal{A}$ and a link type mapping function $\phi : \mathcal{E} \to \mathcal{R}$, where each object $v \in \mathcal{V}$ belongs to one particular object type $\tau(v) \in \mathcal{A}$, each link $e \in \mathcal{E}$ belongs to a particular relation $\phi(e) \in \mathcal{R}$, and if two links belong to the same relation type, the two links share the same starting object type as well as the ending object type.

  Different from the traditional network definition, we explicitly distinguish object types and relationship types in the network. [...] When the types of objects $|\mathcal{A}| > 1$ or the types of relations $|\mathcal{R}| > 1$, the network is called **heterogeneous information network**; otherwise, it is a **homogeneous information network**.

  Given a complex heterogeneous information network, it is necessary to provide its meta level (i.e., schema-level) description for better understanding the object types and link types in the network. Therefore, we propose the concept of **network schema** to describe the **meta structure of a network**. [...]

  The **network schema**, denoted as $T_G = (\mathcal{A}, \mathcal{R})$, is a **meta template for a heterogeneous network** $G = (\mathcal{V}, \mathcal{E})$ with the object type mapping $\tau : \mathcal{V} \to \mathcal{A}$ and the link mapping $\phi : \mathcal{E} \to \mathcal{R}$, which is a directed graph defined over object types $\mathcal{A}$, with edges as relations from $\mathcal{R}$.

  The network schema of a heterogeneous information network has specified type constraints on the sets of objects and relationships between the objects. These constraints make a heterogeneous information network semi-structured, guiding the exploration of the semantics of the network. [...]

  Heterogeneous information networks can be constructed almost in any domain, such as social networks (e.g., Facebook), e-commerce (e.g., Amazon and eBay), online movie databases (e.g., IMDB), and numerous database applications. Heterogeneous information networks can also be constructed from text data, such as news collections, by **entity and relationship extraction** using natural language processing and other advanced techniques.

  Diverse information can be associated with information networks. **Attributes** can be attached to the nodes or links in an information network. For example, location attributes, either categorical or numerical, are often associated with some users and tweets in a Twitter information network. Also, **temporal information** is often associated with nodes and links to reflect the dynamics of an information network. For example, in a bibliographic information network, new papers and authors emerge every year, as well as their associated links. ([257])

4.1.10   KEGG50k

- KEGG50k publication:

  Knowledge graphs are a data representation that model relational information as a graph, where the graph nodes represent knowledge entities and its edges represent relations between them. They model facts as (subject, predicate, object) (SPO) **triples** *e.g. (Aspirin, Drug-Target, COX-1)*, where a subject entity is connected to an object entity through a predicate relation.

  In recent years, knowledge graphs have become a popular means for data representation in the **semantic web** community to create the "web of data", which is a network of interconnected entities that can be easily interpreted by both humans and machines [36], where knowledge graphs are used to model **linked data**. ([386])

4.1.11   KGen

- KGen publication:

  In formal terms, a **Knowledge Graph** $\mathcal{KG} = (\mathcal{V}, \mathcal{E})$ can be represented as a regular graph, containing a set of Vertices $\mathcal{V}$ and Edges $\mathcal{E}$. The vertices express entities or concepts, and the

edges express how such concepts and entities relate to each other.

A **RDF triple** refers to a data entity composed of a subject (s), predicate (p) and an object (o), represented as $t = (s, p, o)$. In KGs, the edges are, then, a set of predicates, such that $\mathcal{E} = \{p_0, p_1, \ldots, p_n\}$. The vertices are, in turn, a set of subjects and objects, such that $\mathcal{V} = \{s_0, s_1, \ldots, s_n, o_0, o_1, \ldots, o_n\}$. In this work, a KG is represented as a set of RDF triples, such that, $\mathcal{KG} = \{t_0, t_1, \ldots, t_n\}$, where $t_0 = (s_0, p_0, o_0)$, $t_1 = (s_1, p_1, o_1), \ldots, t_n = (s_n, p_n, o_n)$.

An **ontology** describes a real-world domain in terms of concepts, attributes, relationships and axioms [54]. Formally, an ontology $\mathcal{O}$ is represented as a set of classes $\mathcal{C}_\mathcal{O}$ interrelated by directed relations $\mathcal{R}$, and a set of attributes $\mathcal{A}_\mathcal{O}$, i.e., $\mathcal{O} = (\mathcal{C}_\mathcal{O}, \mathcal{R}_\mathcal{O}, \mathcal{A}_\mathcal{O})$.

In this sense, we may consider an **ontology-linked knowledge graph** $\mathcal{KG}' = (\mathcal{V}', \mathcal{E}') = \{t'_0, t'_1, \ldots, t'_n\}$, having some of its constituents as instances of classes, relations, and attributes of a given ontology $\mathcal{O}' = (\mathcal{C}_{\mathcal{O}'}, \mathcal{R}_{\mathcal{O}'}, \mathcal{A}_{\mathcal{O}'})$. A given predicate $p' \in E'$ may be an instance of a relation $r' \in \mathcal{R}_{\mathcal{O}'}$. A given subject $s' \in V'$ and an object $o' \in V'$ may be instances of, either a class $c' \in \mathcal{C}_{\mathcal{O}'}$, or an attribute $a' \in \mathcal{A}_{\mathcal{O}'}$.

We introduce our KGen (a shorthand for Knowledge Graph Generation) method and tool implementation to generate ontology-linked KGs. ([467])

- Reference [54] in the KGen publication:

  An **ontology** is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an Ontology is a systematic account of Existence. For AI systems, what "exists" is that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the **universe of discourse**. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. Thus, in the context of AI, we can describe the ontology of a program by defining a set of representational terms. In such an ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names mean, and formal axioms that constrain the interpretation and well-formed use of these terms. Formally, an ontology is the statement of a **logical theory**.[1] [...]

  Formal ontologies are designed. When we choose how to represent something in an ontology, we are making design decisions. To guide and evaluate our designs, we need objective criteria that are founded on the purpose of the resulting artifact, rather than based on a priori notions of naturalness or Truth. Here we propose a preliminary set of **design criteria** for ontologies whose purpose is knowledge sharing and interoperation among programs based on a shared conceptualization. ([234])

4.1.12 NGLY1 Deficiency KG

- NGLY1 Deficiency KG publication:

  Knowledge graphs are computer-readable semantic representations of relational information, where concepts are encoded as nodes, and the relationships between those concepts are represented as edges. Knowledge graphs make it easy to integrate information from many sources, to explore **heterogeneous information** within a single data model and to infer new relationships via efficient queries. Knowledge graphs have been used to organize background knowledge for data interpretation and hypothesis generation in a wide variety of contexts (6–9,4,10–12). ([450])

4.1.13 OREGANO

- OREGANO publication:

  The data used by learning algorithms must be represented in such a way that machines can learn a general model. One of our assumptions is that the more information there is to

characterise a drug, the better the pattern. It is therefore necessary to integrate as much drug data as possible. One of the concerns that emerges in this context is that the data are very heterogeneous and it is difficult to characterise drugs in the same way. In order to learn effectively on such **heterogeneous data**, it is necessary to find a way to represent the drug data and to leverage this representation. The type of representation that can be used at this stage is a **knowledge graph**, which has been defined by Hogan et al. as: "*a graph of data intended to accumulate and convey knowledge of real world, whose nodes represent nodes of interest and whose edges represent links between nodes*"[9]. A knowledge graph therefore accumulates knowledge of the real world in which the nodes represent notions of interest and whose edges represent links between them. A knowledge graph is thus a set of nodes (or entities) N and labelled links (or relations or predicates) L represented as **triples** of the form: (N_x, L_1, N_y). The edges describe the binary links between two nodes and they are generally oriented and meaningful. In this case, the nodes are differentiated: the subject is the source node of the relationship, the object is the target node resulting in a **triplet** expressed as (subject, predicate, object).

From the perspective of defining a graph for drug repositioning, it is then possible to represent each drug, target, disease or other related entities by a node in the graph and then link these nodes together.

Biomedical data are well suited to be stored in knowledge graphs because they are scattered over many knowledge sources without being linked to each other. In this context, the "***Semantic Web*** *initiative*" offers an idealised vision of the Web, with the idea that resources on the Web should be connected by **semantic links** (as opposed to hyperlinks) and that the meaning of these resources should be exploitable by machines[10]. Following this paradigm, various initiatives aiming to interconnect existing knowledge sources have emerge[11,12] [sic]. ([88])

- Reference [9] in the OREGANO publication:

  Though the phrase "**knowledge graph**" has been used in the literature since at least 1972 [118], the modern incarnation of the phrase stems from the 2012 announcement of the Google Knowledge Graph [122], followed by further announcements of knowledge graphs by Airbnb, Amazon, eBay, Facebook, IBM, LinkedIn, Microsoft, Uber, and more besides [57, 95]. The growing industrial uptake of the concept proved difficult for academia to ignore, with more and more scientific literature being published on knowledge graphs in recent years [32, 77, 100, 105, 106, 140, 144]. [...]

  The definition of a "**knowledge graph**" remains contentious [13, 15, 32], where a number of (sometimes conflicting) definitions have emerged, varying from specific technical proposals to more inclusive general proposals.[1] Herein, we define a knowledge graph as a *graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent potentially different relations between these entities.* The graph of data (a.k.a. *data graph*) conforms to a graph-based data model, which may be a *directed edge-labelled graph*, a *heterogeneous graph*, a *property graph*, and so on (we discuss these models in Section 2). [...]

  A **directed edge-labelled graph**, or del graph for short (also known as a *multi-relational graph* [9, 17, 93]) is defined as a set of nodes — such as Santiago, Arica, 2018-03-22 12:00 — and a set of directed labelled edges between those nodes, such as Santa Lucía—city→Santiago. In knowledge graphs, nodes represent entities (the city Santiago; the hill Santa Lucía; noon on March 22nd, 2018; etc.) and edges represent binary relations between those entities (e.g., Santa Lucía is in the city Santiago). [...] Modelling data in this way offers more flexibility for integrating new sources of data, compared to the standard relational model, where a schema must be defined upfront and followed at each step. While other structured data models such as trees (XML, JSON, etc.) would offer similar flexibility, graphs do not require organising the data hierarchically (should venue be a parent, child, or sibling of type, for example?). They also allow cycles to be represented and queried (e.g., in Figure 1, note the directed cycle in the routes between Santiago, Arica, and Viña del Mar).

A standard data model based on del graphs is the **Resource Description Framework (RDF)** [24]. RDF defines three types of nodes: *Internationalised Resource Identifiers (IRIs)*, used for globally identifying entities and relations on the Web; *literals*, used to represent strings and other datatype values (integers, dates, etc.); and *blank nodes*, used to denote the existence of an entity. [...]

A **heterogeneous graph** [61, 142, 154] (or heterogeneous information network [128, 129]) is a graph where each node and edge is assigned one type. Heterogeneous graphs are thus akin to del graphs—with edge labels corresponding to edge types—but where the type of node forms part of the graph model itself, rather than being expressed as a special relation, as seen in Figure 2. An edge is called ***homogeneous*** if it is between two nodes of the same type (e.g., borders); otherwise it is called ***heterogeneous*** (e.g., capital). Heterogeneous graphs allow for partitioning nodes according to their type, for example, for the purposes of machine learning tasks [61, 142, 154]. However, unlike del graphs, they typically assume a one-to-one relation between nodes and types (notice the node Santiago with zero types and EID15 with multiple types in the del graph of Figure 1). [...]

A **property graph** allows a set of property–value pairs and a label to be associated with nodes and edges, offering additional flexibility when modelling data [4, 84]. [...] Though not yet standardised, property graphs are used in popular graph databases, such as **Neo4j** [4, 84]. While the more intricate model offers greater flexibility in terms of how to encode data as a property graph (e.g., using property graphs, we can continue modelling flights as edges in Figure 3(b)) potentially leading to a more intuitive representation, these additional details likewise require more intricate query languages, formal semantics, and inductive techniques versus simpler graph models such as del graphs or heterogeneous graphs. ([259])

4.1.14 Otter-Knowledge

- Otter-Knowledge publication:

  A **Multimodal Knowledge Graph (MKG)** is a **directed labeled graph** where labels for nodes and edges have well-defined meanings, and each node has a modality, a particular mode that qualifies its type (text, image, protein, molecule, etc.). We consider two node subsets: *entity nodes* (or entities), which corresponds to concepts in the knowledge graph (for example protein, or molecule), and *attribute nodes* (or attributes), which represent qualifying attributes of an entity (for example the mass of a molecule, or the description of a protein). We refer to an edge that connects an entity to an attribute as *data property*, and an edge that connects two entities as *object property*. Each node in the graph has a unique identifier, and a unique modality (specified as a string). [...]

  The framework that builds the MKG ensures that each **triple** is unique, and it automatically merges entities having the same **unique identifier**, but whose data is extracted from different data sources. It is also possible to use the special relation `sameAs`[1] to indicate that two entities having different unique identifiers are to be considered as the same entity. The `sameAs` relation is useful when creating a MKG from multiple partially overlapping data sources; when the graph is built. Additionally, it is possible to build an MKG incrementally, by merging two or more graphs built using different **schemas**; The merge operation automatically combines entities with matching unique identifiers or distinct attributes (e.g., proteins with the same sequence). ([335])

4.1.15 PharmKG

- PharmKG publication:

  For a decade, basic networks such as **undirected and unirelational graphs** were used to model intricate interactions in biomedical systems [4–8]. Despite the impressive performances of these models, these networks failed to capture the semantics within different types

of relationships between biomedical entities. For example, drug–protein interactions modeled with basic networks cannot distinguish between different kinds of interactions such as inhibition, activation, binding, etc. Because of this, many recent works have since switched to using **multi-relational networks**, i.e. **knowledge graphs (KGs)**, where **KG embedding (KGE)** [9–12] approaches were utilized to map graphs into a low-dimensional space while maximally preserving its topological properties. As such, downstream tasks such as relation prediction, clustering and visualization can be done by typical non-network-based models [13–15]. [...]

KGs are **multi-relational, directed graphs** in which nodes represent entities and edges represent their relations. [...]

The first major biomedical KG work was published by Belleau et al. [22], where **semantic web technologies** were applied to convert publicly available bioinformatics databases into **RDF formats**. From the processed RDF file, the biomedical **triplets** (entity, relation and entity) could be subsequently obtained to construct a biomedical KG. Unfortunately, this kind of KG contains a significant number of **metadata relations** that can interfere with the performance of link prediction algorithms, and special care was needed to exclude trivially inferable statements from the test set [23]. Since then, efforts have been focused on constructing **task-oriented KGs** and applying them to downstream biomedical applications, such as drug repositioning [15, 24, 31], with only a few KGs focused on the construction of annotated, clarified biomedical knowledge networks. ([608])

4.1.16 PheKnowLator

- PheKnowLator publication:

  Multiple definitions of KGs have been proposed in the literature, all sharing the assumption that KGs are more than simple large-scale graphs.[13–15] Existing definitions are best summarized by Ehrlinger's and Wöb's (2016) definition: "A **knowledge graph** acquires and integrates information into an ontology and applies a reasoner to derive new knowledge".[13] We provide an alternative definition and consider a KG a **graph-based data structure** representing a variety of **heterogeneous entities** and **multiple types of relationships** between them and serving as an abstract framework that is able to infer new knowledge (as well as reveal and resolve discrepancies or contradictions) to address a variety of applications and use cases. [...]

  1. Knowledge Model. Following **Semantic Web standards**,[110] PKT-KG defines a **KG** as $K = \langle T, A \rangle$, where T is the TBox and A is the ABox. The TBox represents the **taxonomy** of a particular domain.[111,112] It describes classes, properties/relationships, and assertions that are assumed to generally hold within a domain (e.g., a gene is a heritable unit of DNA located in the nucleus of cells [Figure 7a]). The ABox describes attributes and roles of instances of classes (i.e., individuals) and assertions about their membership in classes within the TBox (e.g., A2M is a type of gene that may cause Alzheimer's Disease [Figure 7b]).[111,112] [...]

  2. Relation Strategy. PKT-KG provides two relation strategies. The first strategy is standard or **directed relations**, through a single directed edge (e.g., "gene causes phenotype"). The second strategy is inverse or **bidirectional relations**, through inference if the relation is from an ontology like the RO (e.g., "chemical participates in pathway" and "pathway has participant chemical") or through inferring implicitly symmetric relations for edge types that represent biological interactions (e.g., gene-gene interactions).

  3. Semantic Abstraction. KGs built using expressive languages like **OWL** are structurally complex and composed of **triples** or edges that are logically necessary but not biologically meaningful (e.g., anonymous subclasses used to express TBox assertions with all-some quantification). PKT-KG currently uses the **OWL-NETS**[41] semantic abstraction algorithm to convert or transform **complex KGs** into **hybrid KGs**. OWL-NETS v2.0[113] includes additional functionality that harmonizes a semantically abstracted KG to be consistent with a class- or instance-based knowledge model. ([100])

4.1.17 PubChemRDF

- PubChemRDF publication:

    Semantic Web technologies and standards include the trio of the **Resource Description Framework (RDF)**, **Web Ontology Language (OWL)**, and **SPARQL query language** [16]. RDF is a standard model that uses machine-understandable metadata to describe the type and relation of any Web resource, which can be anything that has an identity, such as a document, a person, a datum, or an operation. RDF uses an abstract model to decompose information into small pieces with well-defined semantics (meaning), so as to express knowledge in a general, yet simple and flexible way. Each small piece of information is represented as an **RDF statement**, also called a "**triple**" of subject-predicate-object, and the RDF model can be expressed as a collection of triples. The semantics and syntax in a given RDF model are defined in **controlled vocabularies** or **ontologies**, and **OWL** is widely used to create domain-specific ontologies with increased expressivity. It is worth noting that ontologies are not only vocabularies that define a set of common and shared terms in a hierarchical structure to describe domain knowledge, they are also computable by enabling first-order logical reasoning, i.e. the statements asserted to the parent classes can be inherited by the child classes. The logic-based inference can be used to derive new RDF statements that are not explicitly asserted, and logic rules can be used to identify conflict statements on behalf of consistency checking. Hence, ontologies designed for automated inference must be carefully formulated according to the semantics of the language and as such are distinct from informal knowledge organization systems such as taxonomy and thesaurus. **SPARQL** serves as an RDF query language and data access protocol for the Semantic Web with the ability to locate and retrieve specific information across widespread databases as well as generate query reports that can be directly analyzed by network visualization and data mining applications. SPARQL may be used to query relational databases [17, 18] as well as RDF databases (triple stores) [19, 20], and may increase in popularity in the near future with the rapidly increasing scalability of RDF databases. ([192])

4.1.18 PyKEEN

- PyKEEN publication 1:

    In the last two decades, representing factual information as **knowledge graphs (KGs)** has gained significant attention. KGs have been successfully applied to tasks such as link prediction, clustering, and question answering. In the context of this paper, a KG is a **directed, multi-relational graph** that represents entities as nodes, and their relations as edges, and can be used as an abstraction of the real world. Factual information contained in KGs is represented as **triples** of the form (h, r, t), where h and t denote the head and tail entities, and r denotes their respective relation. Prominent examples of KGs are DBpedia [18], Wikidata [25], Freebase [5], and Knowledge Vault [10]. Traditionally, KGs have been processed in their essential form as symbolic systems, but recently, **knowledge graph embedding models (KGEs)** have become popular that encode the nodes and edges of KGs into low-dimensional continuous vector spaces while best preserving the structural properties of the KGs. ([26])

- PyKEEN publication 2:

    Knowledge graphs (KGs) encode knowledge as a set of **triples** $\mathcal{K} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ where $\mathcal{E}$ denotes the set of entities and $\mathcal{R}$ the set of relations. **Knowledge graph embedding models (KGEMs)** learn representations for entities and relations of KGs in vector spaces while preserving the graph structure. The learned embeddings can support machine learning tasks such as entity clustering, link prediction, entity disambiguation, as well as downstream tasks such as question answering and item recommendation (Nickel et al., 2015; Wang et al., 2017; Ruffinelli et al., 2020; Kazemi et al., 2020). ([25])

4.1.19 RNA-KG

- RNA-KG publication:

  In this phase, we identified the KG representation and the kind of storage system to adopt. **RDF triples** have turned out to be suitable because of their common, flexible, and uniform data model. These properties result in an ontologically-graounded knowledge graph for conducting different kinds of analysis and reasoning.

  Since a standardized formal definition for the concept of a KG is still lacking, we considered the one adopted by Callahan and colleagues[19] where a **KG** is a pair $\langle T, A \rangle$, where T is the TBox and A the ABox. The TBox represents the **taxonomy** of a particular domain including classes, properties/relationships, and assertions that are assumed to generally hold within a domain (e.g., a miRNA is a small regulatory ncRNA located in an exosome as depicted in Figure 4). The ABox describes **attributes** and roles of class instances (i.e., individuals) and assertions about their membership in classes within the TBox (e.g., hsa-miR-125b-5p is a type of miRNA that may cause leukemia). Non-ontological entities (i.e., entities from a data source that are not compliant to a given set of ontologies such as RNA molecules) can be integrated with ontologies using either a TBox (i.e., class-based) or ABox (i.e., instance-based) knowledge model. For the class-based approach, each database entity is represented as `subClassOf` an existing ontology class, while for the instance-based approach it is represented as `instanceOf` an existing ontology class. ([109])

- Reference [19] in the RNA-KG publication is the PheKnowLator publication covered previously.

4.1.20 ROBOKOP

- ROBOKOP publication 1:

  A **knowledge graph (KG)** uses an appropriate ontology to express domain knowledge as a graph of relationships (edges) between entities (nodes), with related nodes connected by edges. **KG databases** such as **Neo4j** allow KGs to be queried using a query language such as Cypher that is designed to find matching relationship paths or sub-graphs within the KG. ([396])

- ROBOKOP publication 2:

  ROBOKOP KG is based on a set of semantic types, as defined in the **BioLink data model**.[6] (A simplified version of the ROBOKOP **KG database schema** is provided as a graphical abstract, with nodes representing entities in the BioLink data model, and edges representing predicates or relationships between connected entities.) This model defines the high-level concepts between which relationships can be made, as well as a series of properties belonging to these concepts. The BioLink data model is hierarchical, with more specific concepts deriving from more general ones; for instance "cellular component" is derived from "anatomical entity". The model also contains union terms that provide capabilities to, for example, group diseases and phenotypes. Entities are identified with conceptual terms from biomedical **ontologies** such as Gene Ontology (GO).[7] ROBOKOP KG requires all nodes to be an instance of one or more BioLink data model types. ([78])

4.1.21 SemKG

- SemKG publication:

  Knowledge graph is a **multi-relational graph** composed of entities as nodes and relations as different types of edges. In this work, we constructed a biomedical knowledge graph, called SemKG, with the predications which are extracted from PubMed abstracts by SemRep. In the SemKG, let $E = \{e_1, e_2, \ldots, e_N\}$ denote the set of n entities, $R = \{r_1, r_2, \ldots, r_M\}$ denote the set of relations between entities and $T = \{t_1, t_2, \ldots, t_K\}$ denote semantic type of entities.

The elements of R and T are all from the UMLS semantic network. The edge between entities e_i and e_j is weighted by the number of predications that have been extracted. Besides, the **attribute** of edge includes the abstracts' PubMed ID (pmid) from where the predications are extracted. ([483])

4.1.22 SynLethKG / KG4SL

- SynLethKG / KG4SL publication:

  Knowledge graphs (KGs) are a type of **multi-relational graph**, where nodes and edges have different types. A KG is denoted by $G = (V, E)$, where edges in set E are defined as **triplets** $e = (h, \tau, t)$ indicating a particular relationship $\tau \in T$ between two nodes (Hamilton, 2020). [...]

  The KG SynLethKG is denoted by $G = (V_e, V_r)$, which contains a set of entities V_e and a set of relationships V_r. Each edge in the KG is defined as a triplet $T = (h, r, t)$, which shows a relationship of type r between head entity h and tail entity t, where $h, t \in V_e$ and $r \in V_r$. ([564])

- Reference (Hamilton, 2020) in the SynLethKG / KG4SL publication:

  Formally, a **graph** $G = (\mathcal{V}, \mathcal{E})$ is defined by a set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$ between these nodes. We denote an edge going from node $u \in \mathcal{V}$ to node $v \in \mathcal{V}$ as $(u, v) \in \mathcal{E}$. In many cases we will be concerned only with **simple graphs**, where there is at most one edge between each pair of nodes, no edges between a node and itself, and where the edges are all undirected, i.e., $(u, v) \in \mathcal{E} \leftrightarrow (v, u) \in \mathcal{E}$.

  A convenient way to represent graphs is through an *adjacency matrix* $\mathbf{A} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$. To represent a graph with an adjacency matrix, we order the nodes in the graph so that every node indexes a particular row and column in the adjacency matrix. We can then represent the presence of edges as entries in this matrix: $\mathbf{A}[u, v] = 1$ if $(u, v) \in \mathcal{E}$ and $\mathbf{A}[u, v] = 0$ otherwise. If the graph contains only **undirected** edges then $\mathbf{A}$ will be a symmetric matrix, but if the graph is **directed** (i.e., edge direction matters) then $\mathbf{A}$ will not necessarily be symmetric. Some graphs can also have **weighted** edges, where the entries in the adjacency matrix are arbitrary real-values rather than $\{0, 1\}$. For instance, a weighted edge in a protein-protein interaction graph might indicated the strength of the association between two proteins.

  Multi-relational graphs: Beyond the distinction beween undirected, directed, and weighted edges, we will also consider graphs that have different *types* of edges. For instance, in graphs representing drug-drug interactions, we might want different edges to correspond to different side effects that can occur when you take a pair of drugs at the same time. In these cases we can extend the edge notation to include an edge or relation type τ, e.g., $(u, \tau, v) \in \mathcal{E}$, and we can define one adjacency matrix $\mathbf{A}_\tau$ per edge type. We call such graphs *multi-relational*, and the entire graph can be summarized by an adjacency tensor $\mathcal{A}^{|\mathcal{V}| \times |\mathcal{R}| \times |\mathcal{V}|}$, where $\mathcal{R}$ is the set of relations. Two important subsets of multi-relational graphs are often known as *heterogeneous* and *multiplex* graphs.

  Heterogeneous graphs: In heterogeneous graphs, nodes are also imbued with *types*, meaning that we can partition the set of nodes into disjoint sets $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2 \cup \ldots \cup \mathcal{V}_k$ where $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset, \forall i \neq j$. Edges in heterogeneous graphs generally satisfy constraints according to the node types, most commonly the constraint that certain edges only connect nodes of certain types, i.e., $(u, \tau_i, v) \in \mathcal{E} \rightarrow u \in \mathcal{V}_j, v \in \mathcal{V}_k$. For example, in a heterogeneous biomedical graph, there might be one type of node representing proteins, one type of representing drugs, and one type representing diseases. Edges representing "treatments" would only occur between drug nodes and disease nodes. Similarly, edges representing "polypharmacy side-effects" would only occur between two drug nodes. **Multipartite** graphs are a well-known special case of heterogeneous graphs, where edges can only connect nodes that have different types, i.e., $(u, \tau_i, v) \in \mathcal{E} \rightarrow u \in \mathcal{V}_j, v \in \mathcal{V}_k \wedge j \neq k$.

Multiplex graphs: In multiplex graphs we assume that the graph can be decomposed in a set of k *layers*. Every node is assumed to belong to every layer, and each layer corresponds to a unique relation, representing the *intra-layer* edge type for that layer. We also assume that *inter-layer* edges types can exist, which connect the same node across layers. Multiplex graphs are best understood via examples. For instance, in a multiplex transportation network, each node might represent a city and each layer might represent a different mode of transportation (e.g., air travel or train travel). Intra-layer edges would then represent cities that are connected by different modes of transportation, while inter-layer edges represent the possibility of switching modes of transportation within a particular city.

Feature information: Last, in many cases we also have ***attribute* or *feature* information** associated with a graph (e.g., a profile picture associated with a user in a social network). Most often these are node-level attributes that we represent using a real-valued matrix $\mathbf{X} \in \mathbb{R}^{|\mathcal{V}| \times m}$, where we assume that the ordering of the nodes is consistent with the ordering in the adjacency matrix. In heterogeneous graphs we generally assume that each different type of node has its own distinct type of attributes. In rare cases we will also consider graphs that have real-valued edge features in addition to discrete edge types, and in some cases we even associate real-valued features with entire graphs.

[...]

In knowledge graph completion, we are given a **multi-relational graph** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the edges are defined as tuples $e = (u, \tau, v)$ indicating the presence of a particular relation $\tau \in \mathcal{T}$ holding between two nodes. Such multirelational graphs are often referred to as ***knowledge graphs***, since we can interpret the tuple (u, τ, v) as specifying that a particular "fact" holds between the two nodes u and v. For example, in a biomedical knowledge graph we might have an edge type $\tau = \text{TREATS}$ and the edge (u, TREATS, v) could indicate that the drug associated with node u treats the disease associated with node v. Generally, the goal in knowledge graph completion is to predict missing edges in the graph, i.e., relation prediction, but there are also examples of node classification tasks using multi-relational graphs [Schlichtkrull et al., 2017]. ([241])

4.2 File formats

Knowledge graphs are stored and shared in various file formats, some of which are associated with a particular graph data model. Table 2 covers formats and models encountered in the domain of biomedicine, which is represented by the projects collected in section 3.

Format name	Extensions	Websites	Specification	Graph model	Organization
BEL	.bel	[126] [37]	[38]	BEL network	BEL.bio organization [36]
CSV	.csv	[570]	[493]	Various models	IETF Network Working Group [492]
GraphDB	.graphdb	[577]	[412]	RDF graph	Company: Ontotext [413]
GraphML	.graphml	[571]	[221]	Various models	GraphML Project Group [220]
JSON-LD	.jsonld, .json	[572]	[549]	RDF graph	W3C [560]
KGX	.json, .jsonl, .tsv, .ttl	[398]	[60]	Property graph	Individuals from various research institutes
N-Quads	.nq	[574]	[554]	RDF graph	W3C [560]
N-Triples	.nt	[573]	[555]	RDF graph	W3C [560]
Neo4j	.dump	[575]	[406]	Property graph	Company: Neo4j, Inc. [407]
Notation3	.n3	[576]	[550]	RDF graph	W3C [560]
RDF/JSON	.json	[185]	[553]	RDF graph	W3C [560]
RDF/XML	.rdf	[578]	[558]	RDF graph	W3C [560]
RDFa	.html, .svg, .xml	[579]	[559]	RDF graph	W3C [560]
SIF	.sif	[433]	[137]	Various models	Cytoscape Consoritum [136]
TriG	.trig	[580]	[556]	RDF graph	W3C [560]
TriX	.trix	[581]	[251]	RDF graph	W3C [560]
Turtle	.ttl	[582]	[557]	RDF graph	W3C [560]

Table 2: File formats for representing knowledge graphs.

4.3 Databases

There are thousands of databases that collect data related to biomedicine in varying degrees of quality and recency. Listing them is beyond the scope of this survey, however, table 3 contains a list of projects that provide comprehensive collections of biomedical databases in form of web interfaces with search functionality. In addition, some of these projects provide annotations for each entry in their collection. For example, NAR db status [45] indicates whether a database is online, when it was last updated and whether a full download of its contents is provided by the host.

Database name	Website	Publications	Number of databases	Organization
Bio.tools	[48]	[295]	2377	ELIXIR [165]
DaTo	[384]	[343]	35400 (count includes tools)	Ming Chen's Lab, Zhejiang University [383]
Database Commons	[123]	[367]	6389	China National Center for Bioinformation (CNCB), Chinese Academy of Sciences [58]
FAIRsharing	[180]	[484]	2075	The FAIRsharing team, University of Oxford [420]
MetaBase	[140]	[82]	1802 (last online in 2014)	Individuals from various research institutes
NAR db status	[45]	[188]	2246	Bioinformatics/Medical Informatics department, Bielefeld University [524]
NAR online Molecular Biology Database Collection	[435]	[461]	1764	Nucleic Acids Research (NAR) Journal, University of Oxford [436]
Online Bioinformatics Resources Collection (OBRC)	[246]	[117]	2408 (count includes tools)	Health Sciences Library System, University of Pittsburgh [245]
Online Resource Finder For Life Sciences (OReFiL)	[518]	[588]	N/A	Database Center for Life Science, University of Tokyo [277]
The Bioinformatics Link Directory	[160]	[186] [92]	621 (last online in 2017)	The Canadian Bioinformatics Workshops (CBW) [5]

Table 3: Projects that provide listings of biomedical databases.

4.4 Ontologies and controlled vocabularies

There are hundreds of ontologies in the domain of biomedicine. As with databases, it is beyond the scope of this survey to list them here directly. Instead, table 4 contains projects that provide collections of biomedical ontologies in searchable web interfaces.

Ontology name	Website	Publications	Number of Ontologies	Organization
BioPortal	[519]	[567]	1077	Center for Biomedical Informatics Research (BMIR), Stanford University [376]
Ontobee	[243]	[411]	262	He Group, University of Michigan Medical School [244]
Ontology Lookup Service (OLS)	[174]	[142] [141]	251	Samples, Phenotypes and Ontologies Team (SPOT), EMBL-EBI [173]
Open Biomedical Ontologies (OBO) Foundry	[566]	[296]	258	Individuals from various research institutes

Table 4: Projects that provide listings of biomedical ontologies.

4.5 Tools

The creation and usage of knowledge graphs involves a series of steps that are sometimes combined in a dedicated tool. Table 5 provides an overview of such tools that were either built or mentioned by the projects collected in section 3.

Creator name	Websites	Publications	Code	Last Update	Organization
AgreementMakerLight	[336]	[495]	GitHub [21]	2023	LASIGE Computer Science and Engineering Research Centre, University of Lisbon [337]
BioCypher	[156]	[361]	GitHub [360], PyPI [359]	2023	Saez-Rodriguez Group, Heidelberg University [477]
BioDBLinker	-	[562]	GitHub [273], PyPI [149]	2020	Biomedical Discovery Informatics Unit, NUI Galway [202] [150]
DemKG	-	[516]	GitHub [153]	2023	Individuals from various research institutes
IASiS Open Data Graph	[438]	[405]	GitHub [504]	2020	Institute of Informatics and Telecommunications, NCSR Demokritos [289]
INDRA	[498] [238]	[239] [34]	GitHub [499], PyPI [237]	2023	Sorger Lab, Harvard Medical School [434]
K-BiOnt	-	[502]	GitHub [73]	2022	Biomedical Text Mining Team (BioTM), LASIGE, University of Lisbon [72] [338]
KG-Hub	[341]	[105]	GitHub [539]	2023	Individuals from various research institutes
KGTK	[125]	[271]	GitHub [537], PyPI [124]	2023	Information Sciences Institute, University of Southern California [286]
KGX	[399] [59]	-	GitHub [61], PyPI [538]	2023	Individuals from various research institutes
KGen	-	[467]	GitHub [465]	2022	Institute of Computing, University of Campinas [288]

ORION	-	-	GitHub [280]	2023	Renaissance Computing Institute (RENCI), University of North Carolina [282]
PheKnowLator	[509]	[99] [100]	GitHub [98], PyPI [515], Zenodo [159]	2021	Computational Bioscience Program, University of Colorado [535], Individuals from various research institutes
ROBOKOP Knowledge Graph Builder (KGB)	[283]	[78]	GitHub [395]	2023	Renaissance Computing Institute (RENCI), University of North Carolina [281]
SEmantic Modeling machIne (SeMi)	[170]	[193]	GitHub [171]	2020	Company: Elsevier
Stitcher	-	[19]	GitHub [18]	2023	Division of Pre-Clinical Innovation, NCATS, NIH [12]

Table 5: Projects that provide tools for creating biomedical knowledge graphs.

References

[1] Bordeaux population health – Université de Bordeaux. *AHeaD — Bordeaux population health*. URL: https://www.bordeaux-population-health.center/the-teams/ahead.

[2] French National Research Agency (ANR). *PractiKPharma Project Homepage*. URL: https://practikpharma.mystrikingly.com.

[3] Cancer Systems Biology Laboratory (CanSyL). *CROssBAR integrated knowledge graphs (KG) · cansyl/CROssBAR*. URL: https://github.com/cansyl/CROssBAR/tree/master/CROssBAR_INTEGRATED_KGs.

[4] Cancer Systems Biology Laboratory (CanSyL). *CROssBAR: Comprehensive Resource of Biomedical Relations with Deep Learning Applications and Knowledge Graph Representations*. URL: https://github.com/cansyl/CROssBAR.

[5] The Canadian Bioinformatics Workshops (CBW). *About Bioinformatics.ca*. URL: https://bioinformatics.ca/about/about-bioinformatics.

[6] Bilal Abu-Salih. "Domain-specific knowledge graphs: A survey". In: *Journal of Network and Computer Applications* 185 (July 1, 2021), p. 103076. DOI: 10.1016/j.jnca.2021.103076.

[7] International Digital Economy Academy. *Biomedical Informatics Ontology System — BIOS@IDEA*. URL: https://bios.idea.edu.cn/lang-en.

[8] International Digital Economy Academy. *Biomedical Informatics Ontology System — BIOS@IDEA*. URL: https://bios.idea.edu.cn/download.

[9] International Digital Economy Academy. *International Digital Economy Academy (IDEA)*. URL: https://www.idea.edu.cn.

[10] National Center for Advancing Translational Sciences. *Biomedical Data Translator — National Center for Advancing Translational Sciences*. URL: https://ncats.nih.gov/research/research-activities/translator.

[11] National Center for Advancing Translational Sciences. *Divisions and Offices — National Center for Advancing Translational Sciences*. URL: https://ncats.nih.gov/about/divisions-and-offices.

[12] National Center for Advancing Translational Sciences. *Divisions and Offices — National Center for Advancing Translational Sciences*. URL: https://ncats.nih.gov/about/divisions-and-offices.

[13] National Center for Advancing Translational Sciences. *National Center for Advancing Translational Sciences*. URL: https://ncats.nih.gov.

[14] National Center for Advancing Translational Sciences. *National Center for Advancing Translational Sciences*. URL: https://ncats.nih.gov.

[15] National Center for Advancing Translational Sciences. *NCATS Biomedical Data Translator*. URL: https://ui.transltr.io.

[16] National Center for Advancing Translational Sciences. *Neo4j Browser*. URL: https://disease.ncats.io.

[17] National Center for Advancing Translational Sciences. *Stitcher*. URL: https://github.com/ncats/stitcher.

[18] National Center for Advancing Translational Sciences. *stitcher*. URL: https://github.com/ncats/stitcher.

[19] National Center for Advancing Translational Sciences. *stitcher/paper at master · ncats/stitcher*. URL: https://github.com/ncats/stitcher/tree/master/paper.

[20] National Center for Advancing Translational Sciences. *Translator Developer Documentation*. URL: https://ncatstranslator.github.io/TranslatorTechnicalDocumentation.

[21] AgreementMakerLight. *AML-Project*. URL: https://github.com/AgreementMakerLight/AML-Project.

[22] Zexu Wu et al. *HALD*. URL: https://bis.zju.edu.cn/hald.

[23] Mehdi Ali et al. "BioKEEN: a library for learning and evaluating biological knowledge graph embeddings". In: *Bioinformatics* 35.18 (Sept. 15, 2019), pp. 3538–3540. DOI: `10.1093/bioinformatics/btz117`.

[24] Mehdi Ali et al. *pykeen · PyPI*. URL: `https://pypi.org/project/pykeen`.

[25] Mehdi Ali et al. "PyKEEN 1.0: a Python library for training and evaluating knowledge graph embeddings". In: *The Journal of Machine Learning Research* 22.1 (Jan. 1, 2021), 82:3723–82:3728. URL: `http://jmlr.org/papers/v22/20-825.html`.

[26] Mehdi Ali et al. "The KEEN Universe". In: *The Semantic Web – ISWC 2019*. Ed. by Chiara Ghidini et al. Lecture Notes in Computer Science. Springer International Publishing, Oct. 17, 2019, pp. 3–18. DOI: `10.1007/978-3-030-30796-7_1`.

[27] AnacletoLAB. *AnacletoLAB - Computational Biology and Bioinformatics*. URL: `https://anacletolab.di.unimi.it`.

[28] AnacletoLab. *RNA-KG*. URL: `https://github.com/AnacletoLAB/RNA-KG`.

[29] LORIA Laboratoire lorrain de recherche en informatique et ses applications. *Loria*. URL: `https://www.loria.fr/en`.

[30] AstraZeneca. *AI and knowledge graphs for drug discovery*. URL: `https://www.astrazeneca.com/what-science-can-do/topics/data-science-ai/connected-medicines-innovations-data-science-ai.html`.

[31] AstraZeneca. *AstraZeneca*. URL: `https://github.com/AstraZeneca`.

[32] AstraZeneca. *Awesome Drug Discovery Knowledge Graphs*. URL: `https://github.com/AstraZeneca/awesome-drug-discovery-knowledge-graphs`.

[33] Samira Babalou, Sheeba Samuel, and Birgitta König-Ries. *Reproducible Domain-Specific Knowledge Graphs in the Life Sciences: a Systematic Literature Review*. Sept. 15, 2023. DOI: `10.48550/arXiv.2309.08754`.

[34] John A Bachman, Benjamin M Gyori, and Peter K Sorger. "Automated assembly of molecular mechanisms at scale from text mining and curated databases". In: *Molecular Systems Biology* 19.5 (May 9, 2023), e11325. DOI: `10.15252/msb.202211325`.

[35] John A. Bachman, Benjamin M. Gyori, and Peter K. Sorger. *INDRA assembly Benchmark Corpus*. Jan. 22, 2023. DOI: `10.5281/zenodo.7559353`.

[36] BEL.bio. *About BEL.bio*. URL: `https://bel.bio/about`.

[37] BEL.bio. *BEL: Biological Expression Language*. URL: `https://bel.bio`.

[38] BEL.bio. *bel_specifications*. URL: `https://github.com/belbio/bel_specifications`.

[39] François Belleau et al. "Bio2RDF: Towards a mashup to build bioinformatics knowledge systems". In: *Journal of Biomedical Informatics*. Semantic Mashup of Biomedical Data 41.5 (Oct. 1, 2008), pp. 706–716. DOI: `10.1016/j.jbi.2008.03.004`.

[40] Open Graph Benchmark. *Index of /ogb/data/linkproppred*. URL: `https://snap.stanford.edu/ogb/data/linkproppred`.

[41] Open Graph Benchmark. *Link Property Prediction*. URL: `https://snap-stanford.github.io/ogb-web/docs/linkprop/#ogbl-biokg`.

[42] BenevolentAI. *BenevolentAI — AI Drug Discovery — AI Pharma*. URL: `https://www.benevolent.com`.

[43] BeYond-COVID (BY-COVID). *BY-COVID*. URL: `https://by-covid.org`.

[44] Universität Bielefeld. *Bioinformatics / Medical Informatics Department*. URL: `https://www.techfak.uni-bielefeld.de/ags/bi`.

[45] Bioinformatics/Medical Informatics department at Bielefeld University. *NAR db status*. URL: `https://nardbstatus.kalis-amts.de`.

[46] Department Bioinformatics Bielefeld University and medical informatics. *PharMeBINet*. URL: https://pharmebi.net.

[47] Department Bioinformatics Bielefeld University and medical informatics. *PharMeBINet: Downloads*. URL: https://pharmebi.net/#/download.

[48] *bio.tools · Bioinformatics Tools and Services Discovery Portal*. URL: https://bio.tools.

[49] Bio4j. *Bio4j bioinformatics graph data platform*. URL: https://github.com/bio4j/bio4j.

[50] Bio4j. *Bio4j Titan*. URL: https://github.com/bio4j/bio4j-titan.

[51] Bio4j. *bio4j-titan/docs/Bio4jAWSReleases.md at master · bio4j/bio4j-titan*. URL: https://github.com/bio4j/bio4j-titan/blob/master/docs/Bio4jAWSReleases.md.

[52] Bio4j. *bio4j-titan/docs/ImportingTitanBio4j.md at master · bio4j/bio4j-titan*. URL: https://github.com/bio4j/bio4j-titan/blob/master/docs/ImportingTitanBio4j.md.

[53] Center for Bioinformatics and Computational Biology at UD. *covid19kg_rdf*. URL: https://github.com/udel-cbcb/covid19kg_rdf.

[54] Structural Bioinformatics and Network Biology Group. *BIOTEQUE: Downloads*. URL: https://bioteque.irbbarcelona.org/downloads.

[55] Structural Bioinformatics and Network Biology Group. *BIOTEQUE: Metapath Explorer*. URL: https://bioteque.irbbarcelona.org.

[56] Structural Bioinformatics and Network Biology Group at IRB Barcelona. *Structural Bioinformatics and Network Biology Group*. URL: https://sbnb.irbbarcelona.org.

[57] Bioinformatics @ IOCB Prague. *IDSM - Integrated Database of Small Molecules*. URL: https://idsm.elixir-czech.cz.

[58] China National Center for Bioinformation. *China National Center for Bioinformation (CNCB)*. URL: https://www.cncb.ac.cn/?lang=en.

[59] biolink. *KGX Format*. URL: https://github.com/biolink/kgx/blob/master/specification/kgx-format.md.

[60] biolink. *kgx/specification/kgx-format.md*. URL: https://github.com/biolink/kgx/blob/master/specification/kgx-format.md.

[61] biolink. *Knowledge Graph Exchange*. URL: https://github.com/biolink/kgx.

[62] Integrative Biology and Predictive Analytics Group. *KIDS/kg_constructor/data at master · IBPA/KIDS · GitHub*. URL: https://github.com/IBPA/KIDS/tree/master/kg_constructor/data.

[63] Integrative Biology and Predictive Analytics Group. *Knowledge Integration and Decision Support (KIDS)*. URL: https://github.com/IBPA/KIDS.

[64] Semantic Systems Chemical Biology. *chem2bio2rdf*. URL: http://cheminfov.informatics.indiana.edu:8080/c2b2r.

[65] Semantic Systems Chemical Biology. *Wayback machine: chem2bio2rdf datasets*. URL: https://web.archive.org/web/20150121065022/http://cheminfov.informatics.indiana.edu:8080/datasets.html.

[66] Stanford Developmental Biology. *Stanford Developmental Biology*. URL: https://devbio.stanford.edu.

[67] Research Unit on Biomedical Informatics (GRIB). *Integrative Biomedical Informatics - Research — Grib*. URL: http://grib.imim.es/research/integrative-biomedical-informatics.

[68] Enveda Biosciences. *RPath*. URL: https://github.com/enveda/RPath.

[69] Enveda Biosciences. *RPath/data/kg at master · enveda/RPath · GitHub*. URL: https://github.com/enveda/RPath/tree/master/data/kg.

[70] BiosemanticsDotOrg. *GeneDiseasePaper*. URL: https://github.com/BiosemanticsDotOrg/GeneDiseasePaper.

[71] Bioteque. *bioteque / Bioteque*. URL: https://gitlabsbnb.irbbarcelona.org/bioteque/bioteque.

[72]   BioTM. *BioTM, LASIGE, Fac. Ciências, Universidade de Lisboa*. URL: https://github.com/lasigeBioTM.

[73]   BioTM. *K-BiOnt: Biomedical Relation Extraction with Knowledge Graph-based Recommendations*. URL: https://github.com/lasigeBioTM/K-BiOnt.

[74]   Biozon.org. *Biozon*. URL: https://web.archive.org/web/20180307024822/http://biozon.org.

[75]   Biozon.org. *Biozon : Download files and software*. URL: https://web.archive.org/web/20180831035044/http://biozon.org/ftp/downloads.html.

[76]   Aaron Birkland and Golan Yona. "BIOZON: a hub of heterogeneous biological data". In: *Nucleic Acids Research* 34 (suppl_1 Jan. 1, 2006), pp. D235–D242. DOI: 10.1093/nar/gkj153.

[77]   Aaron Birkland and Golan Yona. "BIOZON: a system for unification, management and analysis of heterogeneous biological data". In: *BMC Bioinformatics* 7.1 (Feb. 15, 2006), p. 70. DOI: 10.1186/1471-2105-7-70.

[78]   Chris Bizon et al. "ROBOKOP KG and KGB: Integrated Knowledge Graphs from Federated Sources". In: *Journal of Chemical Information and Modeling* 59.12 (Dec. 23, 2019), pp. 4968–4973. DOI: 10.1021/acs.jcim.9b00683.

[79]   BLENDER Lab, CS @ Illinois. *BLENDER Lab*. URL: https://blender.cs.illinois.edu.

[80]   BLENDER Lab, CS @ Illinois. *BLENDER Lab — Software*. URL: https://blender.cs.illinois.edu/covid19.

[81]   BLENDER Lab, CS @ Illinois. *Wayback Machine: KG.zip*. URL: https://web.archive.org/web/20210124013751/http://159.89.180.81/demo/covid/KG.zip.

[82]   Dan M. Bolser et al. "MetaBase—the wiki-database of biological databases". In: *Nucleic Acids Research* 40 (D1 Jan. 1, 2012), pp. D1250–D1254. DOI: 10.1093/nar/gkr1099.

[83]   Universität Bonn. *Department of Pharmaceutical Biochemistry and Bioanalytics*. URL: https://www.pharma.uni-bonn.de/pharmazeutische-biochemie-und-bioanalytik/en.

[84]   Stephen Bonner et al. "A review of biomedical datasets relating to drug discovery: a knowledge graph perspective". In: *Briefings in Bioinformatics* 23.6 (Nov. 1, 2022), bbac404. DOI: 10.1093/bib/bbac404.

[85]   Marina Boudin. "Computational Approaches for Drug Repositioning: Towards a Holistic Perspective based on Knowledge Graphs". In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. CIKM '20. Association for Computing Machinery, Oct. 19, 2020, pp. 3225–3228. DOI: 10.1145/3340531.3418510.

[86]   Marina Boudin et al. *The OREGANO knowledge graph for computational drug repurposing*. Nov. 10, 2023. DOI: 10.5281/zenodo.10103842.

[87]   Marina Boudin et al. *The OREGANO knowledge graph for computational drug repurposing*. July 12, 2023. DOI: 10.6084/m9.figshare.23553114.v3.

[88]   Marina Boudin et al. "The OREGANO knowledge graph for computational drug repurposing". In: *Scientific Data* 10.1 (Dec. 6, 2023), p. 871. DOI: 10.1038/s41597-023-02757-0.

[89]   Konstantinos Bougiatiotis. *Drug-Drug Interaction Prediction on a Biomedical Literature Knowledge Graph (DDI-BLKG)*. URL: https://github.com/kbogas/DDI_BLKG.

[90]   Konstantinos Bougiatiotis. *ownCloud - A safe home for all your data*. URL: https://owncloud.skel.iit.demokritos.gr:443/index.php/s/WFpHQ6aegYK1J7M.

[91]   Konstantinos Bougiatiotis et al. "Drug-Drug Interaction Prediction on a Biomedical Literature Knowledge Graph". In: *Artificial Intelligence in Medicine*. Ed. by Martin Michalowski and Robert Moskovitch. Lecture Notes in Computer Science. Springer International Publishing, Sept. 26, 2020, pp. 122–132. DOI: 10.1007/978-3-030-59137-3_12.

[92]   Michelle D. Brazas et al. "A decade of web server updates at the bioinformatics links directory: 2003–2012". In: *Nucleic Acids Research* 40 (W1 July 1, 2012), W3–W12. DOI: 10.1093/nar/gks632.

[93]   Anna Breit et al. "OpenBioLink: a benchmarking framework for large-scale biomedical link prediction". In: *Bioinformatics* 36.13 (July 13, 2020), pp. 4097–4098. DOI: `10.1093/bioinformatics/btaa274`.

[94]   Anna Breit et al. *OpenBioLink2020*. Apr. 27, 2020. DOI: `10.5281/zenodo.5361324`.

[95]   University of Bristol. *MRC Integrative Epidemiology Unit — University of Bristol*. URL: `https://www.bristol.ac.uk/integrative-epidemiology`.

[96]   Tiffany Callahan. *Open Source Biomedical KG - GitHub Scraper*. Oct. 30, 2023. DOI: `10.5281/zenodo.10052114`.

[97]   Tiffany J. Callahan. *Archived Builds*. URL: `https://github.com/callahantiff/PheKnowLator/wiki/Archived-Builds`.

[98]   Tiffany J. Callahan. *PheKnowLator*. URL: `https://github.com/callahantiff/PheKnowLator`.

[99]   Tiffany J. Callahan et al. *A Framework for Automated Construction of Heterogeneous Large-Scale Biomedical Knowledge Graphs*. May 2, 2020. DOI: `10.1101/2020.04.30.071407`.

[100] Tiffany J. Callahan et al. *An Open-Source Knowledge Graph Ecosystem for the Life Sciences*. July 11, 2023. DOI: `10.48550/arXiv.2307.05727`.

[101] Tiffany J. Callahan et al. "Knowledge-Based Biomedical Data Science". In: *Annual Review of Biomedical Data Science* 3.1 (2020), pp. 23–41. DOI: `10.1146/annurev-biodatasci-010820-091627`.

[102] Case Western Reserve University. *Center for Artificial Intelligence in Drug Discovery — Case Western Reserve University*. URL: `https://case.edu/medicine/aicenter`.

[103] Case Western Reserve University. *Index of /public/data/GPKG-Predict*. URL: `http://nlp.case.edu/public/data/GPKG-Predict`.

[104] Case Western Reserve University. *Index of /public/data/GPKG-Predict/data*. URL: `http://nlp.case.edu/public/data/GPKG-Predict/data`.

[105] J. Harry Caufield et al. *KG-Hub – Building and Exchanging Biological Knowledge Graphs*. Jan. 31, 2023. DOI: `10.48550/arXiv.2302.10800`.

[106] Emanuele Cavalleri. *Blazegraph Workbench*. URL: `http://fievel.anacleto.di.unimi.it:9999/blazegraph/#splash`.

[107] Emanuele Cavalleri. *RNA-KG*. Dec. 22, 2023. DOI: `10.5281/zenodo.10078876`. URL: `https://zenodo.org/records/10078876`.

[108] Emanuele Cavalleri et al. "A Meta-Graph for the Construction of an RNA-Centered Knowledge Graph". In: *Bioinformatics and Biomedical Engineering*. Ed. by Ignacio Rojas et al. Lecture Notes in Computer Science. Springer Nature Switzerland, 2023, pp. 165–180. DOI: `10.1007/978-3-031-34953-9_13`.

[109] Emanuele Cavalleri et al. *RNA-KG: An ontology-based knowledge graph for representing interactions involving RNA molecules*. Nov. 30, 2023. DOI: `10.48550/arXiv.2312.00183`.

[110] CBRC Lab Amirkabir University. *CBRC*. URL: `http://bioinformatics.aut.ac.ir`.

[111] Center for Statistical Science, Tsinghua University. *Center for Statistical Science*. URL: `http://www.stat.tsinghua.edu.cn/en`.

[112] Central China Normal University. *Central China Normal University*. URL: `https://www.ccnu.edu.cn`.

[113] Payal Chandak. *PrimeKG*. May 2, 2022. DOI: `10.7910/DVN/IXA7BM`.

[114] Payal Chandak, Kexin Huang, and Marinka Zitnik. "Building a knowledge graph to enable precision medicine". In: *Scientific Data* 10.1 (Feb. 2, 2023), p. 67. DOI: `10.1038/s41597-023-01960-3`.

[115] Avishek Chatterjee et al. "Knowledge Graphs for COVID-19: An Exploratory Review of the Current Landscape". In: *Journal of Personalized Medicine* 11.4 (Apr. 2021), p. 300. DOI: `10.3390/jpm11040300`.

[116] Bin Chen et al. "Chem2Bio2RDF: a semantic framework for linking and data mining chemogenomic and systems chemical biology data". In: *BMC Bioinformatics* 11.1 (May 17, 2010), pp. 1–13. DOI: 10.1186/1471-2105-11-255.

[117] Yi-Bu Chen et al. "The Online Bioinformatics Resources Collection at the University of Pittsburgh Health Sciences Library System—a one-stop gateway to online bioinformatics databases and software tools". In: *Nucleic Acids Research* 35 (suppl_1 Jan. 1, 2007), pp. D780–D785. DOI: 10.1093/nar/gkl781.

[118] Chuming Chen et al. *COVID-19 Knowledge Graph*. URL: https://research.bioinformatics.udel.edu/covid19kg.

[119] Chuming Chen et al. *COVID-19 Knowledge Graph - RDF Dumps*. URL: https://research.bioinformatics.udel.edu/covid19kg/rdfdumps.

[120] Chuming Chen et al. *COVID-19 Knowledge Graph - RESTful API*. URL: https://research.bioinformatics.udel.edu/covid19kg/api.

[121] Chuming Chen et al. *COVID-19 Knowledge Graph - YASGUI SPARQL GUI*. URL: https://research.bioinformatics.udel.edu/covid19kg/yasgui.

[122] Chuming Chen et al. "COVID-19 Knowledge Graph from semantic integration of biomedical literature and databases". In: *Bioinformatics* 37.23 (Dec. 7, 2021), pp. 4597–4598. DOI: 10.1093/bioinformatics/btab694.

[123] Chinese Academy of Sciences China National Center for Bioinformation / Beijing Institute of Genomics. *Database Commons*. URL: https://ngdc.cncb.ac.cn/databasecommons.

[124] ISI CKGs. *kgtk · PyPI*. URL: https://pypi.org/project/kgtk.

[125] ISI CKGs. *KGTK documentation*. URL: https://kgtk.readthedocs.io.

[126] BEL community. *Biological Expression Language*. URL: https://biological-expression-language.github.io.

[127] Bio2RDF Consortium. *Bio2RDF Home*. URL: https://github.com/bio2rdf/bio2rdf-scripts/wiki.

[128] Bio2RDF Consortium. *Bio2RDF Release 3*. URL: https://download.bio2rdf.org/files/release/3/release.html.

[129] Bio2RDF Consortium. *Bio2RDF v2.7a*. URL: https://bio2rdf.org.

[130] Bio2RDF Consortium. *Bio2RDF-scripts*. URL: https://github.com/bio2rdf/bio2rdf-scripts.

[131] Bio2RDF Consortium. *File Browser*. URL: https://download.bio2rdf.org.

[132] Bio2RDF Consortium. *OpenLink Virtuoso SPARQL Query Editor*. URL: https://bio2rdf.org/sparql.

[133] Biomedical Data Translator Consortium. *Knowledge Graph Exchange (KGE) Archive and Registry Working Group*. URL: https://github.com/NCATSTranslator/Knowledge_Graph_Exchange_Registry.

[134] Biomedical Data Translator Consortium. *ReasonerAPI*. URL: https://github.com/NCATSTranslator/ReasonerAPI.

[135] Biomedical Data Translator Consortium. *TranslatorTechnicalDocumentation*. URL: https://github.com/NCATSTranslator/TranslatorTechnicalDocumentation.

[136] Cytoscape Consortium. *Cytoscape Consortium*. URL: https://github.com/cytoscape.

[137] Cytoscape Consortium. *Cytoscape User Manual*. URL: https://cytoscape.org/manual/Cytoscape2_5Manual.html#SIF%20Format.

[138] SmartAPI consortium. *SmartAPI — Translator*. URL: https://smart-api.info/portal/translator.

[139] The Biomedical Data Translator Consortium. "The Biomedical Data Translator Program: Conception, Culture, and Community". In: *Clinical and Translational Science* 12.2 (Nov. 9, 2018), pp. 91–94. DOI: 10.1111/cts.12592.

[140]   MetaBase contributors. *MetaBase*. URL: https://web.archive.org/web/20140118114727/http://metadatabase.org/wiki/Main_Page.

[141]   Richard Côté et al. "The Ontology Lookup Service: bigger and better". In: *Nucleic Acids Research* 38 (suppl_2 July 1, 2010), W155–W160. DOI: 10.1093/nar/gkq331.

[142]   Richard G. Côté et al. "The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries". In: *BMC Bioinformatics* 7.1 (Dec. 2006), pp. 1–7. DOI: 10.1186/1471-2105-7-97.

[143]   COVID-19-Net. *Covid-19-Community*. URL: https://github.com/covid-19-net/covid-19-community.

[144]   Covidgraph.org Team. *CovidGraph Documentation*. URL: https://github.com/covidgraph/documentation/wiki.

[145]   Covidgraph.org Team. *Covidgraph.org Team*. URL: https://github.com/covidgraph.

[146]   Covidgraph.org Team. *documentation*. URL: https://github.com/covidgraph/documentation.

[147]   *Cytoscape App Store - NeDRex*. URL: https://apps.cytoscape.org/apps/nedrex.

[148]   Kevin Dalleau et al. "Learning from biomedical linked data to suggest valid pharmacogenes". In: *Journal of Biomedical Semantics* 8.1 (Apr. 20, 2017), pp. 1–12. DOI: 10.1186/s13326-017-0125-1.

[149]   INSIGHT Centre for Data Analytics. *biodblinker · PyPI*. URL: https://pypi.org/project/biodblinker.

[150]   Biomedical Discovery Informatics unit at Data Science Institute of NUI Galway. *DSI - Biomedical Discovery Informatics*. URL: https://github.com/dsi-bdi.

[151]   National University of Defense Technology. *Colleges & Institutes - National University of Defense Technology*. URL: https://english.nudt.edu.cn/nav/Academics/CollegesandInstitutes.

[152]   University of Delaware. *Computer & Information Sciences at the University of Delaware*. URL: https://www.cis.udel.edu.

[153]   demkg-framework. *kg-transform*. URL: https://github.com/demkg-framework/kg-transform.

[154]   Department of electronic engineering, Tsinghua University. *Department Of Electronic Engineering Tsinghua University*. URL: https://www.ee.tsinghua.edu.cn/en.

[155]   Department of electronic engineering, Tsinghua University. *Neo4j Browser*. URL: http://43.140.200.9:7474/browser.

[156]   BioCypher developers. *Democratising Knowledge Graphs*. URL: https://biocypher.org.

[157]   PheKnowLator Ecosystem Developers. *pheknowlator – Bucket-Details – Cloud Storage – Google Cloud Console*. URL: https://console.cloud.google.com/storage/browser/pheknowlator.

[158]   PheKnowLator Ecosystem Developers. *PheKnowLator Human Disease Knowledge Graph Benchmarks Archive*. Nov. 1, 2023. DOI: 10.5281/zenodo.10065431.

[159]   PheKnowLator Ecosystem Developers. *Search PheKnowLator*. URL: https://zenodo.org/communities/pheknowlator-ecosystem.

[160]   Canadian Bioinformatics.ca Links Directory. *Bioinformatics Links Directory — Bioinformatics.ca Links Directory*. URL: https://web.archive.org/web/20170608065028/https://bioinformatics.ca/links_directory.

[161]   Tunca Doğan et al. "CROssBAR: comprehensive resource of biomedical relations with knowledge graph representations". In: *Nucleic Acids Research* 49.16 (Sept. 20, 2021), e96. DOI: 10.1093/nar/gkab543.

[162]   Daniel Domingo-Fernández et al. *Causal reasoning over knowledge graphs leveraging drug-perturbed and disease-specific transcriptomic signatures for drug discovery*. Jan. 17, 2022. DOI: 10.5281/zenodo.5865857.

[163]   Daniel Domingo-Fernández et al. "Causal reasoning over knowledge graphs leveraging drug-perturbed and disease-specific transcriptomic signatures for drug discovery". In: *PLOS Computational Biology* 18.2 (Feb. 25, 2022), e1009909. DOI: 10.1371/journal.pcbi.1009909.

[164] Michel Dumontier et al. "Bio2RDF release 3: a larger connected network of linked data for the life sciences". In: *Proceedings of the 2014 International Conference on Posters & Demonstrations Track - Volume 1272*. ISWC-PD'14. CEUR-WS.org, Oct. 21, 2014, pp. 401–404. URL: https://dl.acm.org/doi/10.5555/2878453.2878554.

[165] ELIXIR. *ELIXIR*. URL: https://elixir-europe.org.

[166] Glass Elsarboukh, Tim E Putman, and Kevin Schaper. *monarch-py · PyPI*. URL: https://pypi.org/project/monarch-py.

[167] Elsevier. *Elsevier — An Information Analytics Business*. URL: https://www.elsevier.com.

[168] Elsevier. *EmBiology*. URL: https://assets.ctfassets.net/o78em1y1w4i4/1o4TSxumH9ZFQFc2cd27Qf/16dd4d6afb7621db4d5655478f00d3e5/EmBiology_Fact_Sheet.pdf.

[169] Elsevier. *EmBiology — Biological data structured for insights — Elsevier*. URL: https://www.elsevier.com/products/embiology.

[170] Elsevier. *SeMi - SEmantic Modeling machIne*. URL: https://github.com/giuseppefutia/semi/blob/master/README.md.

[171] Elsevier. *SeMi - SEmantic Modeling machIne*. URL: https://github.com/ElsevierSoftwareX/SOFTX_2019_258.

[172] EMBL. *Protein Function Development – Services for protein sequences and their function*. URL: https://www.ebi.ac.uk/about/teams/protein-function-development.

[173] EMBL. *Samples, phenotypes and ontologies – Delivering databases and interoperability tools for mammalian models of disease, GWAS, PGS and Polygenic Scores*. URL: https://www.ebi.ac.uk/about/teams/samples-phenotypes-ontologies.

[174] EMBL-EBI. *Ontology Lookup Service (OLS)*. URL: https://www.ebi.ac.uk/ols4.

[175] Inc. Enveda Therapeutics. *Enveda Biosciences — Plant-Powered Drug Discovery*. URL: https://www.envedabio.com.

[176] ERIAS. *Blazegraph Workbench*. URL: http://91.121.148.199:8889/bigdata/#query.

[177] ERIAS. *Data_OREGANO/Graphs · oregano · GitLab*. URL: https://gitub.u-bordeaux.fr/erias/oregano/-/tree/master/Data_OREGANO/Graphs.

[178] ERIAS. *oregano · GitLab*. URL: https://gitub.u-bordeaux.fr/erias/oregano.

[179] Antonio Fabregat et al. "Reactome graph database: Efficient access to complex pathway data". In: *PLOS Computational Biology* 14.1 (Jan. 29, 2018), e1005968. DOI: 10.1371/journal.pcbi.1005968.

[180] FAIRsharing. *FAIRsharing*. URL: https://fairsharing.org.

[181] Karamarie Fecho et al. "Progress toward a universal biomedical data translator". In: *Clinical and Translational Science* 15.8 (May 25, 2022), pp. 1838–1847. DOI: 10.1111/cts.13301.

[182] Fan Feng et al. "GenomicKB: a knowledge graph for the human genome". In: *Nucleic Acids Research* 51 (D1 Jan. 6, 2023), pp. D950–D956. DOI: 10.1093/nar/gkac957.

[183] Adrià Fernández-Torras et al. "BQsupports: systematic assessment of the support and novelty of new biomedical associations". In: *Bioinformatics* 39.9 (Sept. 1, 2023), btad581. DOI: 10.1093/bioinformatics/btad581.

[184] Adrià Fernández-Torras et al. "Integrating and formatting biomedical data as pre-calculated knowledge graph embeddings in the Bioteque". In: *Nature Communications* 13.1 (Sept. 9, 2022), p. 5304. DOI: 10.1038/s41467-022-33026-0.

[185] The Apache Software Foundation. *RDF JSON - docs.api*. URL: https://jena.apache.org/documentation/io/rdf-json.html.

[186] Joanne A. Fox et al. "The Bioinformatics Links Directory: a Compilation of Molecular Biology Web Servers". In: *Nucleic Acids Research* 33 (suppl_2 July 1, 2005), W3–W24. DOI: 10.1093/nar/gki594.

[187] Fraunhofer Institute for Translational Medicine and Pharmacology ITMP. *Fraunhofer ITMP*. URL: https://www.itmp.fraunhofer.de/en.html.

[188] Marcel Friedrichs and Cassandra Königs. "A web-based platform for the annotation and analysis of NAR-published databases". In: *PLOS ONE* 18.10 (Oct. 23, 2023), e0293134. DOI: 10.1371/journal.pone.0293134.

[189] Chengcheng Fu. *MDKG*. URL: https://github.com/ccszbd/MDKG.

[190] Chengcheng Fu. *MDKG/dataset.rar at master · ccszbd/MDKG · GitHub*. URL: https://github.com/ccszbd/MDKG/blob/master/dataset.rar.

[191] Chengcheng Fu et al. "An Integrated Knowledge Graph for Microbe-Disease Associations". In: *Health Information Science*. Ed. by Zhisheng Huang et al. Lecture Notes in Computer Science. Springer International Publishing, Oct. 20, 2020, pp. 79–90. DOI: 10.1007/978-3-030-61951-0_8.

[192] Gang Fu et al. "PubChemRDF: towards the semantic annotation of PubChem compound and substance databases". In: *Journal of Cheminformatics* 7.1 (July 14, 2015), p. 34. DOI: 10.1186/s13321-015-0084-4.

[193] Giuseppe Futia, Antonio Vetrò, and Juan Carlos De Martin. "SeMi: A SEmantic Modeling machIne to build Knowledge Graphs with graph neural networks". In: *SoftwareX* 12 (July 1, 2020), p. 100516. DOI: 10.1016/j.softx.2020.100516.

[194] Jakub Galgonek. *chemdb · GitLab*. URL: https://ip-147-251-124-124.flt.cloud.muni.cz/chemdb.

[195] Jakub Galgonek. *chemdb / loaders · GitLab*. URL: https://ip-147-251-124-124.flt.cloud.muni.cz/chemdb/loaders.

[196] Jakub Galgonek. *IDSM / SPARQL GUI*. URL: https://idsm.elixir-czech.cz/sparql.

[197] Jakub Galgonek. *IDSM / SPARQL GUI*. URL: https://idsm.elixir-czech.cz/chemweb.

[198] Jakub Galgonek. *YASGUI*. URL: https://idsm.elixir-czech.cz/sparql/endpoint/idsm.

[199] Jakub Galgonek and Jiří Vondrášek. "IDSM ChemWebRDF: SPARQLing small-molecule datasets". In: *Journal of Cheminformatics* 13.1 (May 12, 2021), p. 38. DOI: 10.1186/s13321-021-00515-1.

[200] Marcos Martinez Galindo. *Otter-Knowledge*. URL: https://medium.com/ibm-data-ai/otter-knowledge-2400afe3fe17.

[201] Ylenia Galluzzo. "A Review: Biological Insights on Knowledge Graphs". In: *New Trends in Database and Information Systems*. Ed. by Silvia Chiusano et al. Communications in Computer and Information Science. Springer International Publishing, Sept. 8, 2022, pp. 388–399. DOI: 10.1007/978-3-031-15743-1_36.

[202] University of Galway. *Data Science Institute - University of Galway*. URL: https://www.universityofgalway.ie/dsi.

[203] Zhenxiang Gao, Pingjian Ding, and Rong Xu. "KG-Predict: A knowledge graph computational framework for drug repurposing". In: *Journal of Biomedical Informatics* 132 (Aug. 1, 2022), p. 104133. DOI: 10.1016/j.jbi.2022.104133.

[204] David Geleta et al. *Biological Insights Knowledge Graph: an integrated knowledge graph to support drug development*. Nov. 1, 2021. DOI: 10.1101/2021.10.28.466262.

[205] Zahra Ghorbanali et al. *DrugRep-HeSiaGraph*. URL: https://github.com/CBRC-lab/DrugRep-HeSiaGraph.

[206] Zahra Ghorbanali et al. "DrugRep-HeSiaGraph: when heterogenous siamese neural network meets knowledge graphs for drug repurposing". In: *BMC Bioinformatics* 24.1 (Oct. 3, 2023), p. 374. DOI: 10.1186/s12859-023-05479-7.

[207] Zahra Ghorbanali et al. *DrugRep-KG*. URL: https://github.com/CBRC-lab/DrugRep-KG.

[208] Zahra Ghorbanali et al. "DrugRep-KG: Toward Learning a Unified Latent Space for Drug Repurposing Using Knowledge Graphs". In: *Journal of Chemical Information and Modeling* 63.8 (Apr. 24, 2023), pp. 2532–2545. DOI: 10.1021/acs.jcim.2c01291.

[209] Zahra Ghorbanali et al. *DrugRep-KG/DRP-KGE-Data and Code at main · CBRC-lab/DrugRep-KG · GitHub*. URL: https://github.com/CBRC-lab/DrugRep-KG/tree/main/DRP-KGE-Data%20and%20Code.

[210] Adriana Carolina Gonzalez-Cavazos et al. "DrugMechDB: A Curated Database of Drug Mechanisms". In: *Scientific Data* 10.1 (Sept. 16, 2023), p. 632. DOI: 10.1038/s41597-023-02534-z.

[211] Google. *A reintroduction to our Knowledge Graph and knowledge panels*. URL: https://blog.google/products/search/about-knowledge-graph-and-knowledge-panels.

[212] Google. *A remedy for your health-related questions: health info in the Knowledge Graph*. URL: https://blog.google/products/search/health-info-knowledge-graph.

[213] Google. *Google - About Google, Our Culture & Company News*. URL: https://about.google.

[214] Google. *Google Knowledge Graph Search API*. URL: https://developers.google.com/knowledge-graph.

[215] Google. *Google Knowledge Graph Search API — Enterprise Knowledge Graph*. URL: https://cloud.google.com/enterprise-knowledge-graph/docs/search-api.

[216] et al Greene Lab Baranzini Lab. *Het.io - Explore*. URL: https://het.io/explore/#download-dataset.

[217] et al Greene Lab Baranzini Lab. *Het.io - Software*. URL: https://het.io/software.

[218] et al Greene Lab Baranzini Lab. *Hetionet - An integrative network of biomedical knowledge*. URL: https://het.io.

[219] DICE group. *CovidPubKG: A FAIR Knowledge Graph of COVID-19 Publications*. URL: https://github.com/dice-group/COVID19DS.

[220] GraphML Working Group. *About GraphML*. URL: http://graphml.graphdrawing.org/about.html.

[221] GraphML Working Group. *The GraphML File Format*. URL: http://graphml.graphdrawing.org.

[222] IBI group. *disgenet-app — Bitbucket*. URL: https://bitbucket.org/ibi_group/disgenet-app.

[223] IBI group. *disgenet2r — Bitbucket*. URL: https://bitbucket.org/ibi_group/disgenet2r.

[224] Integrative Biomedical Informatics Group. *DisGeNET - a database of gene-disease associations*. URL: https://www.disgenet.org.

[225] Integrative Biomedical Informatics Group. *DisGeNET - a database of gene-disease associations*. URL: https://www.disgenet.org/downloads.

[226] Integrative Biomedical Informatics Group. *DisGeNET Programmatic Access*. URL: https://www.disgenet.org/api.

[227] Integrative Biomedical Informatics Group. *DisGeNET RDF*. URL: https://www.disgenet.org/rdf.

[228] Integrative Biomedical Informatics Group. *Index of /download*. URL: http://rdf.disgenet.org/download.

[229] Integrative Biomedical Informatics Group. *SPARQL Explorer*. URL: http://rdf.disgenet.org/lodestar/sparql.

[230] Integrative Biomedical Informatics Group. *Virtuoso SPARQL Query Editor*. URL: http://rdf.disgenet.org/sparql.

[231] Ming Chen's Bioinformatics Group. *Ming Chen's Group of Bioinformatics!* URL: https://bis.zju.edu.cn/binfo.

[232] Scientific Data Management Group. *Knowledge4COVID-19*. URL: https://github.com/SDM-TIB/Knowledge4COVID-19.

[233] The Data Science (DICE) group. *The Data Science (DICE) group*. URL: https://dice-research.org.

[234] Thomas R. Gruber. "Toward principles for the design of ontologies used for knowledge sharing?" In: *International Journal of Human-Computer Studies* 43.5 (Nov. 1, 1995), pp. 907–928. DOI: 10.1006/ijhc.1995.1081.

[235] Jing Guo, Hui Liu, and Jie Zheng. "SynLethDB: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets". In: *Nucleic Acids Research* 44 (D1 Jan. 4, 2016), pp. D1011–D1017. DOI: 10.1093/nar/gkv1108.

[236] Lea Gütebier et al. "CovidGraph: a graph to fight COVID-19". In: *Bioinformatics* 38.20 (Oct. 15, 2022), pp. 4843–4845. DOI: 10.1093/bioinformatics/btac592.

[237] Benjamin Gyori. *indra · PyPI*. URL: https://pypi.org/project/indra.

[238] Benjamin M Gyori and John A Bachman. *INDRA documentation — INDRA 1.22.0 documentation*. URL: https://indra.readthedocs.io.

[239] Benjamin M Gyori et al. "From word models to executable models of signaling networks using automated assembly". In: *Molecular Systems Biology* 13.11 (Nov. 1, 2017), p. 954. DOI: 10.15252/msb.20177651.

[240] hald. *hald*. URL: https://github.com/zexuwu/hald.

[241] William L. Hamilton. *Graph Representation Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Springer International Publishing, 2020. DOI: 10.1007/978-3-031-01588-5.

[242] Katrin Hänsel et al. "From Data to Wisdom: Biomedical Knowledge Graphs for Real-World Data Insights". In: *Journal of Medical Systems* 47.1 (May 17, 2023), p. 65. DOI: 10.1007/s10916-023-01951-2.

[243] University of Michigan Medical School He Group. *Ontobee*. URL: https://ontobee.org.

[244] University of Michigan Medical School He Group. *Welcome to He Group Webpage*. URL: https://www.hegroup.org.

[245] University of Pittsburgh Health Sciences Library System. *HSLS*. URL: https://www.hsls.pitt.edu.

[246] University of Pittsburgh Health Sciences Library System. *OBRC: Online Bioinformatics Resources Collection — HSLS*. URL: https://www.hsls.pitt.edu/obrc.

[247] HealthECCO. *CovidGraph - a COVID-19 Knowledge Graph - HealthECCO*. URL: https://web.archive.org/web/20230327061808/https://healthecco.org/covidgraph.

[248] HealthECCO. *HealthECCO*. URL: https://healthecco.org.

[249] Kristina M. Hettne et al. *Dryad — Data – The implicitome: a resource for rationalizing gene-disease associations*. URL: https://datadryad.org/stash/dataset/doi:10.5061/dryad.gn219.

[250] Kristina M. Hettne et al. "The Implicitome: A Resource for Rationalizing Gene-Disease Associations". In: *PLoS ONE* 11.2 (Feb. 26, 2016), e0149621. DOI: 10.1371/journal.pone.0149621.

[251] Hewlett-Packard Development Company, L.P. *TriX : RDF Triples in XML*. URL: https://www.hpl.hp.com/techreports/2004/HPL-2004-56.html.

[252] Daniel Himmelstein. *hetio · PyPI*. URL: https://pypi.org/project/hetio.

[253] Daniel Himmelstein. *Hetionet in Neo4j*. URL: https://neo4j.het.io/browser.

[254] Daniel Himmelstein. *hetnetpy · PyPI*. URL: https://pypi.org/project/hetnetpy.

[255] Daniel Himmelstein et al. *hetio/hetionet: Data repository containing Hetionet downloads*. URL: https://github.com/hetio/hetionet.

[256] Daniel Himmelstein et al. *hetio/hetnetpy repository: Hetnets in Python*. URL: https://github.com/hetio/hetnetpy.

[257] Daniel S. Himmelstein and Sergio E. Baranzini. "Heterogeneous Network Edge Prediction: A Data Integration Approach to Prioritize Disease-Associated Genes". In: *PLOS Computational Biology* 11.7 (July 9, 2015), e1004259. DOI: 10.1371/journal.pcbi.1004259.

[258] Daniel Scott Himmelstein et al. "Systematic integration of biomedical knowledge prioritizes drugs for repurposing". In: *eLife* 6 (Sept. 22, 2017), e26726. DOI: 10.7554/eLife.26726.

[259] Aidan Hogan et al. "Knowledge Graphs". In: *ACM Computing Surveys* 54.4 (July 2, 2021), 71:1–71:37. DOI: 10.1145/3447772.

[260] Weihua Hu et al. *Open Graph Benchmark: Datasets for Machine Learning on Graphs*. Feb. 24, 2021. DOI: 10.48550/arXiv.2005.00687.

[261] University of Huddersfield. *School of Computing and Engineering*. URL: https://pure.hud.ac.uk/en/organisations/school-of-computing-and-engineering.

[262] Farah Humayun and Daniel Domingo-Fernández. *HemeKG*. URL: https://github.com/hemekg/hemekg.

[263] Farah Humayun and Daniel Domingo-Fernández. *hemekg · PyPI*. URL: https://pypi.org/project/hemekg.

[264] Farah Humayun and Daniel Domingo-Fernández. *HemeKG v0.0.1-dev*. URL: https://hemekg.github.io/hemekg.

[265] Farah Humayun and Daniel Domingo-Fernández. *hemekg/hemekg at master · hemekg/hemekg · GitHub*. URL: https://github.com/hemekg/hemekg/tree/master/hemekg.

[266] Farah Humayun et al. "A Computational Approach for Mapping Heme Biology in the Context of Hemolytic Disorders". In: *Frontiers in Bioengineering and Biotechnology* 8 (Mar. 6, 2020). DOI: 10.3389/fbioe.2020.00074.

[267] IBM. *ibm/otter_dude · Datasets at Hugging Face*. URL: https://huggingface.co/datasets/ibm/otter_dude.

[268] IBM. *ibm/otter_primekg · Datasets at Hugging Face*. URL: https://huggingface.co/datasets/ibm/otter_primekg.

[269] IBM. *ibm/otter_stitch · Datasets at Hugging Face*. URL: https://huggingface.co/datasets/ibm/otter_stitch.

[270] IBM. *ibm/otter_uniprot_bindingdb_chembl · Datasets at Hugging Face*. URL: https://huggingface.co/datasets/ibm/otter_uniprot_bindingdb_chembl.

[271] Filip Ilievski et al. "KGTK: A Toolkit for Large Knowledge Graph Manipulation and Analysis". In: *The Semantic Web – ISWC 2020*. Ed. by Jeff Z. Pan et al. Lecture Notes in Computer Science. Springer International Publishing, 2020, pp. 278–293. DOI: 10.1007/978-3-030-62466-8_18.

[272] The Trustees of Indiana University. *IU School of Informatics, Computing, and Engineering*. URL: https://luddy.indiana.edu/index.html.

[273] DSI - Biomedical Discovery Informatics. *biodblinker*. URL: https://github.com/dsi-bdi/biodblinker.

[274] DSI - Biomedical Discovery Informatics. *biokg*. URL: https://github.com/dsi-bdi/biokg.

[275] DSI - Biomedical Discovery Informatics. *Release biokg-v1.0.0 · dsi-bdi/biokg*. URL: https://github.com/dsi-bdi/biokg/releases/tag/v1.0.0.

[276] Institute of Informatics and Telecommunications. *National Centre for Scientific Research "Demokritos"*. URL: https://www.iit.demokritos.gr.

[277] The University Of Tokyo Future Center Initiative. *Database Center for Life Science – The University Of Tokyo Future Center Initiative*. URL: https://www.satellite.u-tokyo.ac.jp/en/resident/about_dbcls.

[278] Renaissance Computing Institute. *Index of /var/plater/bl-3.5.4/RobokopKG/929354295ba7d43c/*. URL: https://stars.renci.org/var/plater/bl-3.5.4/RobokopKG/929354295ba7d43c.

[279] Renaissance Computing Institute. *Neo4j Browser*. URL: http://robokopkg.renci.org/browser.

[280] Renaissance Computing Institute. *ORION*. URL: https://github.com/RobokopU24/ORION.

[281] Renaissance Computing Institute. *RENCI - Renaissance Computing Institute at UNC*. URL: https://renci.org.

[282] Renaissance Computing Institute. *RENCI - Renaissance Computing Institute at UNC*. URL: https://renci.org.

[283] Renaissance Computing Institute. *ROBOKOP*. URL: https://robokop.renci.org.

[284] Renaissance Computing Institute. *ROBOKOP KG*. URL: https://robokop.renci.org/api-docs/docs/automat/robokop-kg.

[285] Renaissance Computing Institute. *ROBOKOP U24 · GitHub*. URL: https://github.com/RobokopU24.

[286] USC Information Sciences Institute. *Information Sciences Institute*. URL: https://www.isi.edu.

[287] Fudan University Institute of Biomedical Sciences. *Institute of Biomedical Sciences, Fudan University*. URL: https://ibs.fudan.edu.cn/ibsen.

[288] Institute of Computing - State University of Campinas. *Institute of Computing*. URL: https://ic.unicamp.br/en.

[289] Institute of Informatics and Telecommunications — National Centre for Scientific Research "Demokritos". *National Centre for Scientific Research "Demokritos"*. URL: https://www.iit.demokritos.gr.

[290] Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences. *IOCB Prague — Institute of Organic Chemistry and Biochemistry of the CAS*. URL: https://www.uochb.cz/en.

[291] Fudan University Intelligent Medicine Institute. *Intelligent Medicine Institute, Fudan University*. URL: https://imien.fudan.edu.cn.

[292] Vassilis N. Ioannidis et al. *DRKG/DRKG Drug Repurposing Knowledge Graph.pdf at master · gnn4dr/DRKG*. URL: https://github.com/gnn4dr/DRKG/blob/master/DRKG%20Drug%20Repurposing%20Knowledge%20Graph.pdf.

[293] Vassilis N. Ioannidis et al. *Drug Repurposing Knowledge Graph (DRKG)*. URL: https://github.com/gnn4dr/DRKG.

[294] Vassilis N. Ioannidis et al. *GitHub - gnn4dr/DRKG: A knowledge graph and a set of tools for drug repurposing*. URL: https://github.com/gnn4dr/DRKG#download-drkg.

[295] Jon Ison et al. "Tools and data services registry: a community effort to document bioinformatics resources". In: *Nucleic Acids Research* 44 (D1 Jan. 4, 2016), pp. D38–D47. DOI: 10.1093/nar/gkv1116.

[296] Rebecca Jackson et al. "OBO Foundry in 2021: operationalizing open data principles to evaluate ontologies". In: *Database* 2021 (Sept. 29, 2021), baab069. DOI: 10.1093/database/baab069.

[297] G. Joshi-Tope et al. "Reactome: a knowledgebase of biological pathways". In: *Nucleic Acids Research* 33 (suppl_1 Jan. 1, 2005), pp. D428–D432. DOI: 10.1093/nar/gki072.

[298] Samaneh Jozashoori et al. *Knowledge4COVID-19*. Apr. 19, 2021. DOI: 10.5281/zenodo.4701817. URL: https://zenodo.org/records/4701817.

[299] Reagon Karki. *BY-COVID Knowledge Graph: A comprehensive network integrating various data resources of BY-COVID project — BioModels*. URL: https://www.ebi.ac.uk/biomodels/MODEL2209020001#Files.

[300] Reagon Karki. *Monkeypox Knowledge Graph: A comprehensive representation embedding chemical entities and associated biology of Monkeypox — BioModels*. URL: https://www.ebi.ac.uk/biomodels/MODEL2208040001.

[301] Reagon Karki. *Mpox Knowledge Graph - A comprehensive representation embedding chemical entities and associated biology of Mpox*. URL: https://doi.org/10.18119/N9SG7D.

[302] Reagon Karki et al. *COVID-19 Knowledge Graph: A comprehensive representation of chemical and biological entities associated with COVID-19*. URL: https://github.com/Fraunhofer-ITMP/BY-COVID-KG.

[303] Reagon Karki et al. *Monkeypox Knowledge Graph: A comprehensive representation embedding chemical entities and associated biology of Monkeypox — BioModels*. URL: https://www.ebi.ac.uk/biomodels/MODEL2208040001#Files.

[304] Reagon Karki et al. *Monkeypox Knowledge Graph: A comprehensive representation of chemical and biological entities associated with Monkeypox*. URL: https://github.com/Fraunhofer-ITMP/mpox-kg.

[305] Reagon Karki et al. "Mpox Knowledge Graph: a comprehensive representation embedding chemical entities and associated biology of Mpox". In: *Bioinformatics Advances* 3.1 (Jan. 1, 2023), vbad045. DOI: 10.1093/bioadv/vbad045.

[306] Reagon Karki et al. "Mpox Knowledge Graph: A comprehensive representation of chemical and biological entities associated with Mpox". In: (May 17, 2023). URL: https://zenodo.org/records/7944120.

[307] Toshiaki Katayama et al. "TogoGenome/TogoStanza: modularized Semantic Web genome database". In: *Database* 2019 (Jan. 1, 2019), bay132. DOI: 10.1093/database/bay132.

[308] Sunghwan Kim et al. "PubChem 2023 update". In: *Nucleic Acids Research* 51 (D1 Jan. 6, 2023), pp. D1373–D1380. DOI: 10.1093/nar/gkac956.

[309] Anna Kirkpatrick. *SemNet*. URL: https://github.com/pathology-dynamics/semnet-2.

[310] Anna Kirkpatrick, David Kartchner, and Stephen Allegri. *Pathology Dynamics*. URL: https://sites.gatech.edu/cassie-mitchell-lab.

[311] Anna Kirkpatrick, David Kartchner, and Stephen Allegri. *SemNet 2.0*. URL: https://sites.gatech.edu/cassie-mitchell-lab/semnet-2-0.

[312] Anna Kirkpatrick et al. "Optimizations for Computing Relatedness in Biomedical Heterogeneous Information Networks: SemNet 2.0". In: *Big Data and Cognitive Computing* 6.1 (Mar. 2022), p. 27. DOI: 10.3390/bdcc6010027.

[313] Knowledge-Graph-Hub. *kg-covid-19*. URL: https://github.com/Knowledge-Graph-Hub/kg-covid-19.

[314] Cassandra Königs. *PharMeBINet*. May 14, 2022. DOI: 10.5281/zenodo.6578218.

[315] Cassandra Königs. *PharMeBINet: The heterogeneous pharmacological medical biochemical network*. URL: https://github.com/ckoenigs/PharMeBINet.

[316] Cassandra Königs, Marcel Friedrichs, and Theresa Dietrich. "The heterogeneous pharmacological medical biochemical network PharMeBINet". In: *Scientific Data* 9.1 (July 11, 2022), p. 393. DOI: 10.1038/s41597-022-01510-3.

[317] L3S Research Center. *L3S Research Center - All Projects*. URL: https://www.l3s.de/research-at-l3s/all-projects.

[318] Su Lab. *BioKnowledge reviewer library*. URL: https://github.com/SuLab/bioknowledge-reviewer.

[319] Su Lab. *DrugMechDB*. URL: https://github.com/SuLab/DrugMechDB.

[320] Su Lab. *genewikiworld*. URL: https://github.com/SuLab/genewikiworld.

[321] Su Lab. *Indication Mechanism of Action Database — DrugMechDB*. URL: https://sulab.github.io/DrugMechDB.

[322] Su Lab. *Neo4j Browser*. URL: https://web.archive.org/web/20211204195413/http://ngly1graph.org/browser.

[323] Su Lab. *ngly1-graph/neo4j-graphs at master · SuLab/ngly1-graph*. URL: https://github.com/SuLab/ngly1-graph/tree/master/neo4j-graphs.

[324] Su Lab. *scheduled-bots*. URL: https://github.com/SuLab/scheduled-bots.

[325] Su Lab. *The NGLY1 Deficiency Knowledge Graph*. URL: https://github.com/SuLab/ngly1-graph.

[326] Su Lab. *WD-rephetio-analysis*. URL: https://github.com/SuLab/WD-rephetio-analysis.

[327] Su Lab. *Wikidata-phenomizer*. URL: https://github.com/SuLab/Wikidata-phenomizer.

[328] Su Lab. *WikidataIntegrator*. URL: https://github.com/SuLab/WikidataIntegrator.

[329] The Greene Lab. *Greene Lab*. URL: https://greenelab.com.

[330] The Su Lab. *The Su Lab — Quantitative methods in biomedical discovery*. URL: https://sulab.org.

[331] The Su Lab. *The Su Lab — Quantitative methods in biomedical discovery*. URL: https://sulab.org.

[332] The Su Lab. *The Su Lab — Quantitative methods in biomedical discovery*. URL: https://sulab.org.

[333] Kanser Sistem Biyolojisi Laboratuvar. *CanSyL — Cancer Systems Biology Laboratory*. URL: https://cansyl2.metu.edu.tr/en.

[334] Mann Labs. *CKG*. URL: https://github.com/MannLabs/CKG.

[335] Hoang Thanh Lam et al. *Otter-Knowledge: benchmarks of multimodal knowledge graph representation learning from different sources for drug discovery*. Oct. 19, 2023. DOI: 10.48550/arXiv.2306.12802.

[336] LASIGE. *AgreementMakerLight — LASIGE*. URL: https://www.lasige.pt/tools/agreementmakerlight.

[337] LASIGE. *LASIGE*. URL: https://www.lasige.pt.

[338] LASIGE. *LASIGE*. URL: https://www.lasige.pt.

[339] Lawrence Berkeley National Laboratory, Berkeley Bioinformatics Open-source Projects (BBOP). *Blazegraph Workbench*. URL: http://kg-hub-rdf.berkeleybop.io/blazegraph/#query.

[340] Lawrence Berkeley National Laboratory, Berkeley Bioinformatics Open-source Projects (BBOP). *KG-Hub Index*. URL: https://kghub.io.

[341] Lawrence Berkeley National Laboratory, Berkeley Bioinformatics Open-source Projects (BBOP). *Knowledge Graph Hub*. URL: https://kghub.org.

[342] Jens Lehmann. *Smart Data Analytics*. URL: https://sda.tech.

[343] Qilin Li et al. "DaTo: an atlas of biological databases and tools". In: *Journal of Integrative Bioinformatics* 13.4 (June 1, 2016), pp. 30–38. DOI: 10.1515/jib-2016-297.

[344] Sirui Li. *kg*. URL: https://github.com/SRL94/kg.

[345] Sirui Li. *kg/Drug-CoV at main · SRL94/kg*. URL: https://github.com/SRL94/kg/tree/main/Drug-CoV.

[346] Sirui Li et al. "Drug-CoV: a drug-origin knowledge graph discovering drug repurposing targeting COVID-19". In: *Knowledge and Information Systems* 65.12 (Dec. 1, 2023), pp. 5289–5308. DOI: 10.1007/s10115-023-01923-5.

[347] Sirui Li et al. "Modelling Multi-relations for Convolutional-based Knowledge Graph Embedding". In: *Procedia Computer Science*. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 26th International Conference KES2022 207 (Jan. 1, 2022), pp. 624–633. DOI: 10.1016/j.procs.2022.09.117.

[348] Zongren Li et al. *COVID-19 KG*. URL: https://web.archive.org/web/20230930133838/http://covidkg.ai.

[349] Zongren Li et al. "DeepKG: an end-to-end deep learning-based workflow for biomedical knowledge graph extraction, optimization and applications". In: *Bioinformatics* 38.5 (Mar. 1, 2022), pp. 1477–1479. DOI: 10.1093/bioinformatics/btab767.

[350] Database Center for Life Science. *DBCLS — Database Center for Life Science*. URL: https://dbcls.rois.ac.jp/index-en.html.

[351] Database Center for Life Science (DBCLS). *sparql-proxy*. URL: http://togogenome.org/sparql.

[352] Database Center for Life Science (DBCLS). *TogoGenome*. URL: http://togogenome.org.

[353] Jie Liu. *Liu Lab*. URL: https://jieliu6.github.io.

[354] Yi Liu et al. "EpiGraphDB: a database and data mining platform for health data science". In: *Bioinformatics* 37.9 (June 9, 2021), pp. 1304–1311. DOI: 10.1093/bioinformatics/btaa961.

[355] Department of Computational Medicine Liu Lab and Bioinformatics. *Database Download :: GKB Documentation*. URL: https://liu-bioinfo-lab.github.io/GenomicKB_documentation/documentation/db_download/index.html.

[356] Department of Computational Medicine Liu Lab and Bioinformatics. *GenomicKB*. URL: https://gkb.dcmb.med.umich.edu.

[357] Kevin Livingston. *KaBOB*. URL: https://github.com/drlivingston/kabob.

[358] Kevin M. Livingston et al. "KaBOB: ontology-based semantic integration of biomedical databases". In: *BMC Bioinformatics* 16.1 (Apr. 23, 2015), p. 126. DOI: 10.1186/s12859-015-0559-3.

[359] Sebastian Lobentanzer. *biocypher · PyPI*. URL: https://pypi.org/project/biocypher.

[360] Sebastian Lobentanzer, BioCypher Consortium, and Julio Saez-Rodriguez. *Democratizing knowledge representation with BioCypher*. URL: https://github.com/biocypher/biocypher.

[361] Sebastian Lobentanzer et al. "Democratizing knowledge representation with BioCypher". In: *Nature Biotechnology* 41.8 (Aug. 2023), pp. 1056–1059. DOI: 10.1038/s41587-023-01848-y.

[362] Loria. *PGxLOD*. URL: https://pgxlod.loria.fr.

[363] Loria. *Precision Search & Find*. URL: https://pgxlod.loria.fr/fct.

[364] Loria. *Virtuoso SPARQL Query Editor*. URL: https://pgxlod.loria.fr/sparql.

[365] Vaticle Ltd. *TypeDB Blog: Accelerating drug discovery with applied knowledge engineering and TypeDB*. URL: https://typedb.com/blog/accelerating-drug-discovery-with-applied-knowledge-engineering-and-typedb.

[366] Vaticle Ltd. *TypeDB: The polymorphic database powered by types*. URL: https://typedb.com.

[367] Lina Ma et al. "Database Commons: A Catalog of Worldwide Biological Databases". In: *Genomics, Proteomics & Bioinformatics* (Dec. 23, 2022). DOI: 10.1016/j.gpb.2022.12.004.

[368] International Business Machines. *International Business Machines · GitHub*. URL: https://github.com/IBM.

[369] International Business Machines. *Otter Knowledge*. URL: https://github.com/IBM/otter-knowledge.

[370] Finlay MacLean. "Knowledge graphs and their applications in drug discovery". In: *Expert Opinion on Drug Discovery* 16.9 (Sept. 2, 2021), pp. 1057–1069. DOI: 10.1080/17460441.2021.1910673.

[371] Yassine Marzougui. *RDF-data-mining*. URL: https://github.com/yassmarzou/pgx-lod-mining.

[372] Michael Mayers et al. "Design and application of a knowledge network for automatic prioritization of drug mechanisms". In: *Bioinformatics* 38.10 (May 13, 2022), pp. 2880–2891. DOI: 10.1093/bioinformatics/btac205.

[373] John P. McCrae. *LOD Cloud*. URL: https://lod-cloud.net/datasets.

[374] Julie A McMurry et al. "Navigating the Phenotype Frontier: The Monarch Initiative". In: *Genetics* 203.4 (Aug. 1, 2016), pp. 1491–1495. DOI: 10.1534/genetics.116.188870.

[375] National Library of Medicine. *Access to SemRep/SemMedDB/SKR Resources*. URL: https://lhncbc.nlm.nih.gov/ii/tools/SemRep_SemMedDB_SKR.html.

[376] Stanford Medicine. *Center for Biomedical Informatics Research (BMIR)*. URL: https://bmir.stanford.edu.

[377] The covidgraph.org project members. *CovidGraph Explorer*. URL: https://live.yworks.com/covidgraph.

[378] The covidgraph.org project members. *Neo4j Browser*. URL: https://db.covidgraph.org/browser.

[379] Antonio Messina et al. "BioGrakn: A Knowledge Graph-Based Semantic Database for Biomedical Sciences". In: *Complex, Intelligent, and Software Intensive Systems*. Advances in Intelligent Systems and Computing. Springer International Publishing, July 5, 2017, pp. 299–309. DOI: 10.1007/978-3-319-61566-0_28.

[380] European Bioinformatics Institute (EMBL-EBI) Middle East Technical University (METU). *CROssBAR Project*. URL: https://cansyl.metu.edu.tr/crossbar.

[381] mims-harvard. *PrimeKG*. URL: https://github.com/mims-harvard/PrimeKG.

[382] MindRank-Biotech. *PharmKG*. URL: https://github.com/MindRank-Biotech/PharmKG.

[383] Zhejiang University Ming Chen's Lab. *BIS: Bioinformatics Server @ Zhejiang University*. URL: https://bis.zju.edu.cn.

[384] Zhejiang University Ming Chen's Lab. *DaTo*. URL: https://bis.zju.edu.cn/dato.

[385] Sameh K. Mohamed. *KEGG50k-drugtargets*. URL: https://figshare.com/s/bbfc7b82d17e0b8b6a43.

[386] Sameh K. Mohamed, Aayah Nounu, and Vít Nováček. "Drug target discovery using knowledge graph embeddings". In: *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*. SAC '19. Association for Computing Machinery, Apr. 8, 2019, pp. 11–18. DOI: 10.1145/3297280.3297282.

[387] Monarch Initiative. *Directory index for https://data.monarchinitiative.org/*. URL: https://data.monarchinitiative.org.

[388] Monarch Initiative. *Monarch App*. URL: https://github.com/monarch-initiative/monarch-app.

[389] Monarch Initiative. *Monarch Initiative*. URL: https://monarchinitiative.org.

[390] Monarch Initiative. *monarch-py*. URL: https://monarch-app.monarchinitiative.org.

[391] Monarch Initiative. *monarch_ingest*. URL: https://github.com/monarch-initiative/monarch-ingest.

[392] Pierre Monnin. *pgxlod-reconciliation-rules*. URL: https://github.com/practikpharma/pgxlod-reconciliation-rules.

[393] Pierre Monnin et al. "PGxO and PGxLOD: a reconciliation of pharmacogenomic knowledge of various provenances, enabling further comparison". In: *BMC Bioinformatics* 20.4 (Apr. 18, 2019), pp. 1–16. DOI: 10.1186/s12859-019-2693-9.

[394] John H Morris et al. "The scalable precision medicine open knowledge engine (SPOKE): a massive knowledge graph of biomedical information". In: *Bioinformatics* 39.2 (Feb. 1, 2023), btad080. DOI: 10.1093/bioinformatics/btad080.

[395] Kenneth Morton et al. *robokop*. URL: https://github.com/NCATS-Gamma/robokop.

[396] Kenneth Morton et al. "ROBOKOP: an abstraction layer and user interface for knowledge graphs to support question answering". In: *Bioinformatics* 35.24 (Dec. 15, 2019), pp. 5382–5384. DOI: 10.1093/bioinformatics/btz604.

[397] MRC Integrative Epidemiology Unit. *EpiGraphDB*. URL: http://epigraphdb.org.

[398] Chris Mungall et al. *KGX Format*. URL: https://kgx.readthedocs.io/en/latest/kgx_format.html.

[399] Chris Mungall et al. *Welcome to the KGX documentation — kgx 1.5.1 documentation*. URL: https://kgx.readthedocs.io.

[400] Christopher J. Mungall et al. "The Monarch Initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species". In: *Nucleic Acids Research* 45 (D1 Jan. 4, 2017), pp. D712–D722. DOI: 10.1093/nar/gkw1128.

[401] National Center for Advancing Translational Sciences (NCATS). *Neo4COVID-19*. URL: https://web.archive.org/web/20210621192816/https://neo4covid19.ncats.io.

[402] National Center for Advancing Translational Sciences (NCATS). *neo4covid19*. URL: https://github.com/ncats/neo4covid19.

[403] National Center for Advancing Translational Sciences (NCATS). *Neo4j Data Browser*. URL: https://aspire.covid19.ncats.io:7473.

[404] ncats. *translator-lfs-artifacts/files at main · ncats/translator-lfs-artifacts*. URL: https://github.com/ncats/translator-lfs-artifacts/tree/main/files.

[405] Anastasios Nentidis et al. "iASiS Open Data Graph: Automated Semantic Integration of Disease-Specific Knowledge". In: *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*. July 28, 2020, pp. 220–225. DOI: 10.1109/CBMS49503.2020.00049.

[406] Inc. Neo4j. *Create a DBMS dump file - Neo4j Desktop*. URL: https://neo4j.com/docs/desktop-manual/current/operations/create-dump.

[407] Inc. Neo4j. *Neo4j Graph Database & Analytics — The Leader in Graph Databases*. URL: https://neo4j.com.

[408] Cyberinfrastructure for Network Science Center. *Spoke*. URL: https://cns-iu.github.io/spoke-vis.

[409] Cyberinfrastructure for Network Science Center at Indiana University. *Envisioning SPOKE: Visualizing 3M Nodes and 30M Edges.* URL: https://github.com/cns-iu/spoke-vis.

[410] *Oh no sequences! Research Group - Era7 Bioinformatics.* URL: https://era7bioinformatics.com/en/page_id_392_title_oh-no-sequences_-research-group.html.

[411] Edison Ong et al. "Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration". In: *Nucleic Acids Research* 45 (D1 Jan. 4, 2017), pp. D347–D352. DOI: 10.1093/nar/gkw918.

[412] Ontotext. *Ontotext GraphDB.* URL: https://www.ontotext.com/products/graphdb.

[413] Ontotext. *Ontotext Home.* URL: https://www.ontotext.com.

[414] OpenBioLink. *OpenBioLink.* URL: https://github.com/OpenBioLink/OpenBioLink.

[415] OpenBioLink. *openbiolink · PyPI.* URL: https://pypi.org/project/openbiolink.

[416] TypeDB OSI. *TypeDB Bio: Biomedical Knowledge Graph.* URL: https://github.com/typedb-osi/typedb-bio.

[417] TypeDB OSI. *typedb-bio/dataset at master · typedb-osi/typedb-bio · GitHub.* URL: https://github.com/typedb-osi/typedb-bio/tree/master/dataset.

[418] Simon Ott, Adriano Barbosa da Silva, and Matthias Samwald. *OpenBioLink2021 Challenge.* URL: https://openbiolink.github.io.

[419] Simon Ott, Adriano Barbosa da Silva, and Matthias Samwald. *OpenBioLink2021 Challenge.* URL: https://openbiolink.github.io/dataset.

[420] The University of Oxford. *FAIRsharing.* URL: https://researchdata.ox.ac.uk/fairsharing.

[421] The DICE group at Paderborn University. */datasets/CovidPubGraph/archive/.* URL: https://files.dice-research.org/datasets/CovidPubGraph/archive.

[422] The DICE group at Paderborn University. *Virtuoso SPARQL Query Editor.* URL: https://covid-19ds.data.dice-research.org/sparql.

[423] Saee Paliwal et al. "Preclinical validation of therapeutic targets predicted by tensor factorization on heterogeneous graphs". In: *Scientific Reports* 10.1 (Oct. 26, 2020), p. 18250. DOI: 10.1038/s41598-020-74922-z.

[424] Emmanuel Papadakis. *ADHD-KG.* URL: https://emmanuelpapadakis.github.io/ADHD-KG.

[425] Emmanuel Papadakis. *ADHD-KG.* URL: https://github.com/emmanuelpapadakis/ADHD-KG.

[426] Emmanuel Papadakis. *ADHD-KG v1.0.* Dec. 10, 2022. DOI: 10.5281/zenodo.7422526. URL: https://zenodo.org/records/7422526.

[427] Emmanuel Papadakis et al. "ADHD-KG: a knowledge graph of attention deficit hyperactivity disorder". In: *Health Information Science and Systems* 11.1 (Nov. 11, 2023), p. 52. DOI: 10.1007/s13755-023-00253-8.

[428] Pablo Pareja-Tobes et al. *Bio4j: a high-performance cloud-enabled graph-based data platform.* Mar. 20, 2015. DOI: 10.1101/016758.

[429] Svetlana Pestryakova et al. *CovidPubGraph: A FAIR Knowledge Graph of COVID-19 Publications.* Mar. 31, 2021. DOI: 10.5281/zenodo.4650261.

[430] Svetlana Pestryakova et al. "CovidPubGraph: A FAIR Knowledge Graph of COVID-19 Publications". In: *Scientific Data* 9.1 (July 8, 2022), p. 389. DOI: 10.1038/s41597-022-01298-2.

[431] Janet Piñero et al. "DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes". In: *Database* 2015 (Jan. 1, 2015), bav028. DOI: 10.1093/database/bav028.

[432] Janet Piñero et al. "The DisGeNET knowledge platform for disease genomics: 2019 update". In: *Nucleic Acids Research* 48 (D1 Jan. 8, 2020), pp. D845–D855. DOI: 10.1093/nar/gkz1021.

[433] BioUML platform. *SIF (file format) - BioUML platform.* URL: https://wiki.biouml.org/index.php/SIF_(file_format).

[434] President and Fellows of Harvard College. *Sorger Lab.* URL: https://sorger.med.harvard.edu.

[435] Oxford University Press. *Database Summary Paper Categories.* URL: https://www.oxfordjournals.org/nar/database/c.

[436] Oxford University Press. *Nucleic Acids Research — Oxford Academic.* URL: https://academic.oup.com/nar.

[437] Prof. Dr. Jan Baumbach, Chair of Computational Systems Biology. *NeDRex.* URL: https://nedrex.net.

[438] Project IASiS. *The iASiS project Knowledge Graph — IASIS.* URL: https://project-iasis.eu/node/178.

[439] Novo Nordisk Foundation Center for Protein Research. *Novo Nordisk Foundation Center for Protein Research.* URL: https://www.cpr.ku.dk.

[440] PubChem. *FTP Download.* URL: https://pubchem.ncbi.nlm.nih.gov/docs/rdf-ftp.

[441] PubChem. *Loading PubChemRDF.* URL: https://pubchem.ncbi.nlm.nih.gov/docs/rdf-load.

[442] PubChem. *PubChem.* URL: https://pubchem.ncbi.nlm.nih.gov.

[443] PubChem. *PubChemRDF.* URL: https://pubchem.ncbi.nlm.nih.gov/docs/rdf.

[444] PubChem. *RESTful Interface.* URL: https://pubchem.ncbi.nlm.nih.gov/docs/rdf-rest.

[445] Tim E Putman et al. "The Monarch Initiative in 2024: an analytic platform integrating phenotypes, genes and diseases across species". In: *Nucleic Acids Research* (Nov. 24, 2023), gkad1082. DOI: 10.1093/nar/gkad1082.

[446] PyKEEN. *PyKEEN.* URL: https://github.com/pykeen/pykeen.

[447] PyKEEN. *pykeen/src/pykeen/datasets at master · pykeen/pykeen · GitHub.* URL: https://github.com/pykeen/pykeen/tree/master/src/pykeen/datasets.

[448] Tongqi Qian, Shijia Zhu, and Yujin Hoshida. "Use of big data in drug development for precision medicine: an update". In: *Expert Review of Precision Medicine and Drug Development* 4.3 (May 4, 2019), pp. 189–200. DOI: 10.1080/23808993.2019.1617632.

[449] Núria Queralt-Rosinach et al. "DisGeNET-RDF: harnessing the innovative power of the Semantic Web to explore the genetic basis of diseases". In: *Bioinformatics* 32.14 (July 15, 2016), pp. 2236–2238. DOI: 10.1093/bioinformatics/btw214.

[450] Núria Queralt-Rosinach et al. "Structured reviews for data and knowledge-driven research". In: *Database* 2020 (Jan. 1, 2020), baaa015. DOI: 10.1093/database/baaa015.

[451] Enayat Rajabi and Somayeh Kafaie. "Knowledge Graphs and Explainable AI in Healthcare". In: *Information* 13.10 (Sept. 28, 2022), p. 459. DOI: 10.3390/info13100459.

[452] Reactome. *Content Service - Reactome Pathway Database.* URL: https://reactome.org/dev/content-service.

[453] Reactome. *Download - Reactome Pathway Database.* URL: https://reactome.org/download-data.

[454] Reactome. *Graph Database - Reactome Pathway Database.* URL: https://reactome.org/dev/graph-database.

[455] Reactome. *Reactome Graph Database.* URL: https://github.com/reactome/graph-core.

[456] Reactome. *Reactome Graph Database Batch Importer.* URL: https://github.com/reactome/graph-importer.

[457] Reactome. *Team - Reactome Pathway Database.* URL: https://reactome.org/about/team.

[458] Justin Reese. *KG-COVID-19.* URL: https://github.com/Knowledge-Graph-Hub/kg-covid-19/wiki.

[459] Justin Reese. *KG-COVID-19 documentation — kg-covid-19 1.0.0 documentation.* URL: https://kghub.org/kg-covid-19.

[460] Justin T. Reese et al. "KG-COVID-19: A Framework to Produce Customized Knowledge Graphs for COVID-19 Response". In: *Patterns* 2.1 (Jan. 2021), p. 100155. DOI: 10.1016/j.patter.2020.100155.

[461] Daniel J Rigden and Xosé M Fernández. "The 2023 Nucleic Acids Research Database Issue and the online molecular biology database collection". In: *Nucleic Acids Research* 51 (D1 Jan. 6, 2023), pp. D1–D8. DOI: 10.1093/nar/gkac1186.

[462] Inc. ROSALIND. *Discovery Platform & Data Hub for Scientists — ROSALIND.* URL: https://www.rosalind.bio.

[463] Inc. ROSALIND. *Knowledge Graph & Search.* URL: https://www.rosalind.bio/knowledge-graph.

[464] Peter Rose and Kaushik Ram Ganapathy. *reader@neo4j://132.249.238.185:7687 - Neo4j Browser.* URL: http://132.249.238.185:7474/browser/?connectURL=neo4j://reader:demo@132.249.238.185:7687.

[465] Anderson Rossanez. *KGen.* URL: https://github.com/rossanez/KGen.

[466] Anderson Rossanez. *KGen/examples/biomedical at master · rossanez/KGen · GitHub.* URL: https://github.com/rossanez/KGen/tree/master/examples/biomedical.

[467] Anderson Rossanez et al. "KGen: a knowledge graph generator from biomedical scientific literature". In: *BMC Medical Informatics and Decision Making* 20.4 (Dec. 14, 2020), p. 314. DOI: 10.1186/s12911-020-01341-5.

[468] Development team for the RTX biomedical reasoning tool. *KG2: the second-generation RTX knowledge graph.* URL: https://github.com/RTXteam/RTX-KG2.

[469] RTXteam. *Development team for the RTX biomedical reasoning tool · GitHub.* URL: https://github.com/RTXteam.

[470] RTXteam. *RTX KG2.* URL: http://rtx-kg2-public.s3-website-us-west-2.amazonaws.com.

[471] Tomas Sabat. *Tomas Sabat.* URL: https://medium.com/@tasabat.

[472] Sepideh Sadegh. *NeDRexApp.* URL: https://github.com/repotrial/NeDRexApp.

[473] Sepideh Sadegh. *nedrexdb.* URL: https://github.com/repotrial/nedrexdb.

[474] Sepideh Sadegh and James Skelton. *NeDRexDB.* URL: https://api.nedrex.net.

[475] Sepideh Sadegh and James Skelton. *NeDRexDB - Neo4j.* URL: http://neo4j.nedrex.net.

[476] Sepideh Sadegh et al. "Network medicine for disease module identification and drug repurposing with the NeDRex platform". In: *Nature Communications* 12.1 (Nov. 25, 2021), p. 6848. DOI: 10.1038/s41467-021-27138-2.

[477] Saezlab. *Saez-Rodriguez Group.* URL: https://saezlab.org.

[478] Ahmad Sakor et al. "Knowledge4COVID-19: A semantic-based approach for constructing a COVID-19 related knowledge graph from various sources and analyzing treatments' toxicities". In: *Journal of Web Semantics* 75 (Jan. 1, 2023), p. 100760. DOI: 10.1016/j.websem.2022.100760.

[479] Shengtian Sang. *GitHub - ShengtianSang/GrEDeL: GrEDeL.* URL: https://web.archive.org/web/20200914193006/https://github.com/ShengtianSang/GrEDeL.

[480] Shengtian Sang. *SemaTyP: a knowledge graph based literature mining method for drug discovery.* URL: https://github.com/ShengtianSang/SemaTyP.

[481] Shengtian Sang. *SemaTyP/data at main · ShengtianSang/SemaTyP · GitHub.* URL: https://github.com/ShengtianSang/SemaTyP/tree/main/data.

[482] Shengtian Sang et al. "GrEDeL: A Knowledge Graph Embedding Based Method for Drug Discovery From Biomedical Literatures". In: *IEEE Access* 7 (Dec. 12, 2018), pp. 8404–8415. DOI: 10.1109/ACCESS.2018.2886311.

[483] Shengtian Sang et al. "SemaTyP: a knowledge graph based literature mining method for drug discovery". In: *BMC Bioinformatics* 19.1 (May 30, 2018), p. 193. DOI: 10.1186/s12859-018-2167-5.

[484] Susanna-Assunta Sansone et al. "FAIRsharing as a community approach to standards, repositories and policies". In: *Nature Biotechnology* 37.4 (Apr. 2019), pp. 358–367. DOI: 10.1038/s41587-019-0080-8.

[485] Alberto Santos, Ana R. Colaço, and Annelaura B. Nielsen. *Clinical Knowledge Graph — ClinicalKnowledgeGraph 1.0 documentation*. URL: https://ckg.readthedocs.io.

[486] Alberto Santos et al. "A knowledge graph to interpret clinical proteomics data". In: *Nature Biotechnology* 40.5 (Jan. 31, 2022), pp. 692–702. DOI: 10.1038/s41587-021-01145-6.

[487] Alberto Santos et al. "Clinical Knowledge Graph". In: (). DOI: 10.17632/mrcf7f4tc2.

[488] Alberto Santos et al. *Clinical Knowledge Graph Integrates Proteomics Data into Clinical Decision-Making*. May 10, 2020. DOI: 10.1101/2020.05.09.084897.

[489] Fraunhofer SCAI. *Bioinformatics - Fraunhofer SCAI*. URL: https://www.scai.fraunhofer.de/en/business-research-areas/bioinformatics.html.

[490] Andrew R. Sedler and Cassie S. Mitchell. "SemNet: Using Local Features to Navigate the Biomedical Concept Graph". In: *Frontiers in Bioengineering and Biotechnology* 7 (July 3, 2019). DOI: 10.3389/fbioe.2019.00156.

[491] oh no sequences! *Bio4j*. URL: https://web.archive.org/web/20160215040814/https://bio4j.com.

[492] Shafranovich. *Working Groups*. URL: https://www.ietf.org/how/wgs.

[493] Yakov Shafranovich. *Common Format and MIME Type for Comma-Separated Values (CSV) Files*. Request for Comments RFC 4180. Internet Engineering Task Force, Oct. 2005. DOI: 10.17487/RFC4180.

[494] ShanghaiTech University. *Jie Zheng Lab at ShanghaiTech*. URL: https://shanghaitechzhengjielab.github.io/Zhenglab.github.io.

[495] Marta Contreiras Silva, Daniel Faria, and Catia Pesquita. "Matching Multiple Ontologies to Build a Knowledge Graph for Personalized Medicine". In: *The Semantic Web*. Ed. by Paul Groth et al. Lecture Notes in Computer Science. Springer International Publishing, 2022, pp. 461–477. DOI: 10.1007/978-3-031-06981-9_27.

[496] James Skelton. *NeDRex*. URL: https://github.com/repotrial/nedrex.

[497] snap-stanford. *ogb*. URL: https://github.com/snap-stanford/ogb.

[498] Sorger Lab. *INDRA*. URL: http://www.indra.bio.

[499] Sorger Lab. *INDRA*. URL: https://github.com/sorgerlab/indra.

[500] Sorger Lab. *INDRA REST API*. URL: http://api.indra.bio:8000.

[501] Sorger Lab. *indra_assembly_paper/data at master · sorgerlab/indra_assembly_paper · GitHub*. URL: https://github.com/sorgerlab/indra_assembly_paper/tree/master/data.

[502] Diana Sousa and Francisco M. Couto. "Biomedical Relation Extraction With Knowledge Graph-Based Recommendations". In: *IEEE Journal of Biomedical and Health Informatics* 26.8 (Aug. 2022), pp. 4207–4217. DOI: 10.1109/JBHI.2022.3173558.

[503] Lincoln D. Stein. "Integrating biological databases". In: *Nature Reviews Genetics* 4.5 (May 2003), pp. 337–345. DOI: 10.1038/nrg1065.

[504] tasosnent. *iASiS Open Data Graph*. URL: https://github.com/tasosnent/Biomedical-Knowledge-Integration.

[505] CROssBAR Project Team. *CROssBAR Knowledge Graphs*. URL: https://crossbar.kansil.org.

[506] OGB Team. *ogb · PyPI*. URL: https://pypi.org/project/ogb.

[507] PyKEEN Project Team. *Datasets — pykeen 1.10.2-dev documentation*. URL: https://pykeen.readthedocs.io/en/latest/reference/datasets.html.

[508] PyKEEN Project Team. *PyKEEN*. URL: https://pykeen.readthedocs.io.

[509] Translational Research Informatics Team. *PheKnowLator*. URL: https://github.com/callahantiff/PheKnowLator/wiki.

[510] Dalian University of Technology. *Dalian University of Technology*. URL: https://www.dlut.edu.cn.

[511] The Regents of the University of Colorado. *Computational Bioscience Home — School of Medicine*. URL: https://www.cuanschutz.edu/graduate-programs/computational-bioscience/home.

[512] The Regents of the University of Colorado. *Department of Biomedical Informatics*. URL: https://medschool.cuanschutz.edu/dbmi.

[513] The Tagkopoulos Lab. *Integrative Biology and Predictive Analytics – @ UC DAVIS*. URL: http://tagkopouloslab.ucdavis.edu.

[514] Anne E. Thessen et al. "Transforming the study of organisms: Phenomic data models and knowledge bases". In: *PLOS Computational Biology* 16.11 (Nov. 24, 2020), e1008376. DOI: 10.1371/journal.pcbi.1008376.

[515] William A. Baumgartner Tiffany J. Callahan. *pkt-kg · PyPI*. URL: https://pypi.org/project/pkt-kg.

[516] Santiago Timón-Reina et al. "A Knowledge Graph Framework for Dementia Research Data". In: *Applied Sciences* 13.18 (Jan. 2023), p. 10497. DOI: 10.3390/app131810497.

[517] togogenome. *togogenome*. URL: https://github.com/togogenome/togogenome.

[518] University of Tokyo / DBCLS. *OReFiL: an Online Resource Finder for Lifesciences*. URL: https://orefil.dbcls.jp.

[519] The Board of Trustees of Leland Stanford Junior University. *BioPortal — Biomedical Ontology Repository*. URL: https://www.bioontology.org.

[520] MRC Integrative Epidemiology Unit. *epigraphdb*. URL: https://github.com/MRCIEU/epigraphdb.

[521] MRC Integrative Epidemiology Unit. *EpiGraphDB API - Swagger UI*. URL: https://api.epigraphdb.org.

[522] MRC Integrative Epidemiology Unit. *epigraphdb-r*. URL: https://github.com/MRCIEU/epigraphdb-r.

[523] MRC Integrative Epidemiology Unit. *neo4j-build-pipeline*. URL: https://github.com/MRCIEU/epigraphdb-graph.

[524] Bielefeld University. *Bioinformatics / Medical Informatics Department*. URL: https://www.techfak.uni-bielefeld.de/ags/bi.

[525] Columbia University. *Peter Rose — COVID Information Commons*. URL: https://covidinfocommons.datascience.columbia.edu/content/peter-rose.

[526] Murdoch University. *School of Information Technology — Murdoch University*. URL: https://murdoch.edu.au/schools/information-technology.

[527] Stanford University. *Computer Science — Stanford University School of Engineering*. URL: https://engineering.stanford.edu/faculty-research/departments/computer-science.

[528] Sun Yat-Sen University. *Sun Yat-Sen University: Research Institutes*. URL: https://www.sysu.edu.cn/sysuen/Research1/ResearchInstitutes/National_Level.htm.

[529] The Regents of the University of California. *Baranzini Lab*. URL: https://baranzinilab.ucsf.edu.

[530] The Regents of the University of California. *Neighborhood Explorer*. URL: https://spoke.rbvi.ucsf.edu.

[531] The Regents of the University of California. *People*. URL: https://spoke.ucsf.edu/SPOKE-team.

[532] The Regents of the University of California. *SPOKE*. URL: https://spoke.ucsf.edu.

[533] The Regents of the University of California. *SPOKE Nodes and Edges*. URL: https://spoke.rbvi.ucsf.edu/docs.

[534] The Regents of the University of California. *Swagger UI*. URL: https://spoke.rbvi.ucsf.edu/swagger.

[535] The Regents of the University of Colorado. *Computational Bioscience Home — School of Medicine*. URL: https://www.cuanschutz.edu/graduate-programs/computational-bioscience/home.

[536] University of Galway. *Data Science Institute - University of Galway*. URL: https://www.universityofgalway.ie/dsi.

[537] Information Sciences Institute University of Southern California. *KGTK: Knowledge Graph Toolkit*. URL: https://github.com/usc-isi-i2/kgtk.

[538] Deepak Unni. *kgx · PyPI*. URL: https://pypi.org/project/kgx.

[539] Deepak Unni. *Knowledge-Graph-Hub*. URL: https://github.com/Knowledge-Graph-Hub.

[540] Vaticle. *BioGrakn*. URL: https://github.com/vaticle/biograkn.

[541] Vaticle. *Releases · vaticle/biograkn · GitHub*. URL: https://github.com/vaticle/biograkn/releases.

[542] Vaticle. *Vaticle · GitHub*. URL: https://github.com/vaticle.

[543] Aleksandar Veljkovic. *BioGraph*. URL: https://github.com/aleksandar-veljkovic/biograph.

[544] Aleksandar Veljkovic. *biograph/biograph-backend/importers at master · aleksandar-veljkovic/biograph · GitHub*. URL: https://github.com/aleksandar-veljkovic/biograph/tree/master/biograph-backend/importers.

[545] Aleksandar N. Veljković. *BioGraph*. URL: http://andromeda.matf.bg.ac.rs:54321.

[546] Aleksandar N. Veljković, Yuriy L. Orlov, and Nenad S. Mitić. "BioGraph: Data Model for Linking and Querying Diverse Biological Metadata". In: *International Journal of Molecular Sciences* 24.8 (Apr. 9, 2023), p. 6954. DOI: 10.3390/ijms24086954.

[547] Medical University Vienna. *Section for Artificial Intelligence and Decision Support, CeMSIIS, MedUni Vienna*. URL: https://www.ai.meduniwien.ac.at/en.

[548] *Virtuoso SPARQL Query Editor*. URL: https://labs.tib.eu/sdm/covid19kg/sparql.

[549] W3C. *JSON-LD 1.1*. URL: https://www.w3.org/TR/json-ld.

[550] W3C. *Notation3 (N3): A readable RDF syntax*. URL: https://www.w3.org/TeamSubmission/n3.

[551] W3C. *OWL - Semantic Web Standards*. URL: https://www.w3.org/2001/sw/wiki/OWL.

[552] W3C. *RDF - Semantic Web Standards*. URL: https://www.w3.org/RDF.

[553] W3C. *RDF 1.1 JSON Alternate Serialization (RDF/JSON)*. URL: https://www.w3.org/TR/rdf-json.

[554] W3C. *RDF 1.1 N-Quads*. URL: https://www.w3.org/TR/n-quads.

[555] W3C. *RDF 1.1 N-Triples*. URL: https://www.w3.org/TR/n-triples.

[556] W3C. *RDF 1.1 TriG*. URL: https://www.w3.org/TR/trig.

[557] W3C. *RDF 1.1 Turtle*. URL: https://www.w3.org/TR/turtle.

[558] W3C. *RDF 1.1 XML Syntax*. URL: https://www.w3.org/TR/rdf-syntax-grammar.

[559] W3C. *RDFa 1.1 Primer - Third Edition*. URL: https://www.w3.org/TR/rdfa-primer.

[560] W3C. *W3C*. URL: https://www.w3.org.

[561] Andra Waagmeester et al. "Wikidata as a knowledge graph for the life sciences". In: *eLife* 9 (Mar. 17, 2020). Ed. by Peter Rodgers and Chris Mungall, e52614. DOI: 10.7554/eLife.52614.

[562] Brian Walsh, Sameh K. Mohamed, and Vít Nováček. "BioKG: A Knowledge Graph for Relational Learning On Biological Data". In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. CIKM '20. Association for Computing Machinery, Oct. 19, 2020, pp. 3173–3180. DOI: 10.1145/3340531.3412776.

[563] Qingyun Wang et al. "COVID-19 Literature Knowledge Graph Construction and Drug Repurposing Report Generation". In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations*. Ed. by Avi Sil and Xi Victoria Lin. Association for Computational Linguistics, June 2021, pp. 66–77. DOI: 10.18653/v1/2021.naacl-demos.8.

[564] Shike Wang et al. "KG4SL: knowledge graph neural network for synthetic lethality prediction in human cancers". In: *Bioinformatics* 37 (Supplement_1 July 12, 2021), pp. i418–i425. DOI: 10.1093/bioinformatics/btab271.

[565] Zhen Wang et al. "Knowledge Graph Embedding by Translating on Hyperplanes". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 28.1 (June 21, 2014). DOI: 10.1609/aaai.v28i1.8870.

[566] OBO Technical WG. *OBO Foundry*. URL: https://obofoundry.org.

[567] Patricia L. Whetzel et al. "BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications". In: *Nucleic Acids Research* 39 (suppl_2 July 1, 2011), W541–W545. DOI: 10.1093/nar/gkr469.

[568] Wikidata. *Wikidata Query Service*. URL: https://query.wikidata.org.

[569] Wikidata. *Wikidata:Database download - Wikidata*. URL: https://www.wikidata.org/wiki/Wikidata:Database_download.

[570] Wikipedia. *Comma-separated values*. URL: https://en.wikipedia.org/w/index.php?title=Comma-separated_values.

[571] Wikipedia. *GraphML*. URL: https://en.wikipedia.org/w/index.php?title=GraphML&oldid=1060832076.

[572] Wikipedia. *JSON-LD*. URL: https://en.wikipedia.org/w/index.php?title=JSON-LD.

[573] Wikipedia. *N-Triples*. URL: https://en.wikipedia.org/w/index.php?title=N-Triples.

[574] Wikipedia. *N-Triples: N-Quads*. URL: https://en.wikipedia.org/w/index.php?title=N-Triples#N-Quads.

[575] Wikipedia. *Neo4j*. URL: https://en.wikipedia.org/w/index.php?title=Neo4j.

[576] Wikipedia. *Notation3*. URL: https://en.wikipedia.org/w/index.php?title=Notation3.

[577] Wikipedia. *Ontotext GraphDB*. URL: https://en.wikipedia.org/w/index.php?title=Ontotext_GraphDB.

[578] Wikipedia. *RDF/XML*. URL: https://en.wikipedia.org/w/index.php?title=RDF/XML.

[579] Wikipedia. *RDFa*. URL: https://en.wikipedia.org/w/index.php?title=RDFa.

[580] Wikipedia. *TriG (syntax)*. URL: https://en.wikipedia.org/w/index.php?title=TriG_(syntax).

[581] Wikipedia. *TriX (serialization format)*. URL: https://en.wikipedia.org/w/index.php?title=TriX_(serialization_format).

[582] Wikipedia. *Turtle (syntax)*. URL: https://en.wikipedia.org/w/index.php?title=Turtle_(syntax).

[583] E. C. Wood et al. "RTX-KG2: a system for building a semantically standardized knowledge graph for translational biomedicine". In: *BMC Bioinformatics* 23.1 (Sept. 29, 2022), p. 400. DOI: 10.1186/s12859-022-04932-3.

[584] Zexu Wu et al. *HALD, a human aging and longevity knowledge graph for precision gerontology and geroscience analyses*. May 16, 2023. DOI: 10.6084/m9.figshare.22828196.v4.

[585] Zexu Wu et al. "HALD, a human aging and longevity knowledge graph for precision gerontology and geroscience analyses". In: *Scientific Data* 10.1 (Dec. 1, 2023), p. 851. DOI: 10.1038/s41597-023-02781-0.

[586] Jun Xia. *BIOS EntityClassification*. URL: https://github.com/xiaj1011/bios_entity_classification.

[587] Jun Xia. *BIOS NER*. URL: https://github.com/xiaj1011/bios_ner.

[588] Yasunori Yamamoto and Toshihisa Takagi. "OReFiL: an online resource finder for life sciences". In: *BMC Bioinformatics* 8.1 (Dec. 2007), pp. 1–8. DOI: 10.1186/1471-2105-8-287.

[589] Peiru Yang et al. *LMKG*. URL: https://github.com/ypr17/LMKG.

[590] Peiru Yang et al. *LMKG – Google Drive*. URL: https://drive.google.com/drive/folders/13oDTHUnkgaN__pw_XKGEotmu97odmUpF.

[591] Peiru Yang et al. "LMKG: A large-scale and multi-source medical knowledge graph for intelligent medicine applications". In: *Knowledge-Based Systems* 284 (Jan. 25, 2024), p. 111323. DOI: 10.1016/j.knosys.2023.111323.

[592] Xi Yang. *Neo4j Browser*. URL: http://114.115.208.144:7474/browser.

[593] Xi Yang. *Stroke*. URL: https://github.com/yangxi1016/Stroke.

[594] Xi Yang et al. "Mining a stroke knowledge graph from literature". In: *BMC Bioinformatics* 22.10 (July 29, 2021), p. 387. DOI: 10.1186/s12859-021-04292-4.

[595] Jason Youn, Navneet Rai, and Ilias Tagkopoulos. "Knowledge integration and decision support for accelerated discovery of antibiotic resistance genes". In: *Nature Communications* 13.1 (Apr. 29, 2022), p. 2360. DOI: 10.1038/s41467-022-29993-z.

[596] Sheng Yu et al. *BIOS: An Algorithmically Generated Biomedical Knowledge Graph*. Apr. 24, 2022. DOI: 10.48550/arXiv.2203.09975.

[597] Gergely Zahoránszky-Kőhalmi et al. "A Workflow of Integrated Resources to Catalyze Network Pharmacology Driven COVID-19 Research". In: *Journal of Chemical Information and Modeling* 62.3 (Feb. 14, 2022), pp. 718–729. DOI: 10.1021/acs.jcim.1c00431.

[598] Biyu Zhang and Qi Liu. *Synthetic Lethality Knowledge Graph*. URL: https://www.slkg.net.

[599] Biyu Zhang and Qi Liu. *Synthetic Lethality Knowledge Graph*. URL: https://www.slkg.net/query.

[600] Biyu Zhang et al. "The tumor therapy landscape of synthetic lethality". In: *Nature Communications* 12.1 (Feb. 24, 2021), p. 1275. DOI: 10.1038/s41467-021-21544-2.

[601] Chunhong Zhang et al. "Knowledge graph embedding for hyper-relational data". In: *Tsinghua Science and Technology* 22.2 (Apr. 2017), pp. 185–197. DOI: 10.23919/TST.2017.7889640.

[602] Yong Zhang. *huadingkg*. URL: https://github.com/dagege/huadingkg.

[603] Yong Zhang et al. "HKGB: An Inclusive, Extensible, Intelligent, Semi-auto-constructed Knowledge Graph Framework for Healthcare with Clinicians' Expertise Incorporated". In: *Information Processing & Management* 57.6 (Nov. 1, 2020), p. 102324. DOI: 10.1016/j.ipm.2020.102324.

[604] Jie Zheng. *KG4SL*. URL: https://github.com/JieZheng-ShanghaiTech/KG4SL.

[605] Jie Zheng. *KG4SL/data at main · JieZheng-ShanghaiTech/KG4SL · GitHub*. URL: https://github.com/JieZheng-ShanghaiTech/KG4SL/tree/main/data.

[606] Jie Zheng. *SynLethDB*. URL: https://github.com/JieZheng-ShanghaiTech/SynLethDB.

[607] Shuangjia Zheng et al. *PharmKG: A Dedicated Knowledge Graph Benchmark for Biomedical Data Mining*. Oct. 10, 2020. DOI: 10.5281/zenodo.4525237.

[608] Shuangjia Zheng et al. "PharmKG: a dedicated knowledge graph benchmark for bomedical data mining". In: *Briefings in Bioinformatics* 22.4 (July 1, 2021), bbaa344. DOI: 10.1093/bib/bbaa344.

[609] Zheng Lab, ShanghaiTech University. *SynLethDB*. URL: https://synlethdb.sist.shanghaitech.edu.cn.

[610] Zheng Lab, ShanghaiTech University. *SynLethDB: Download*. URL: https://synlethdb.sist.shanghaitech.edu.cn/#/download.

[611] Chaoyu Zhu. *RDKG-115*. URL: https://github.com/ZhuChaoY/RDKG-115.

[612] Chaoyu Zhu. *RDKG-115/Model/KG at main · ZhuChaoY/RDKG-115 · GitHub*. URL: https://github.com/ZhuChaoY/RDKG-115/tree/main/Model/KG.

[613] Chaoyu Zhu. *RDKG-115/RDKG-115 at main · ZhuChaoY/RDKG-115 · GitHub*. URL: https://github.com/ZhuChaoY/RDKG-115/tree/main/RDKG-115.

[614] Chaoyu Zhu. *SDKG-11*. URL: https://github.com/ZhuChaoY/SDKG-11.

[615] Chaoyu Zhu. *SDKG-11/Dataset at main · ZhuChaoY/SDKG-11 · GitHub*. URL: https://github.com/ZhuChaoY/SDKG-11/tree/main/Dataset.

[616] Chaoyu Zhu et al. "Multimodal reasoning based on knowledge graph embedding for specific diseases". In: *Bioinformatics* 38.8 (Apr. 12, 2022), pp. 2235–2245. DOI: 10.1093/bioinformatics/btac085.

[617] Chaoyu Zhu et al. "RDKG-115: Assisting drug repurposing and discovery for rare diseases by trimodal knowledge graph embedding". In: *Computers in Biology and Medicine* 164 (Sept. 1, 2023), p. 107262. DOI: 10.1016/j.compbiomed.2023.107262.

[618] Qian Zhu et al. "An integrative knowledge graph for rare diseases, derived from the Genetic and Rare Diseases Information Center (GARD)". In: *Journal of Biomedical Semantics* 11.1 (Dec. 2020), pp. 1–13. DOI: 10.1186/s13326-020-00232-y.

[619] Marinka Zitnik. *Artificial Intelligence for Medicine and Science*. URL: https://zitniklab.hms.harvard.edu.

[620] Marinka Zitnik. *Precision Medicine Oriented Knowledge Graph*. URL: https://zitniklab.hms.harvard.edu/projects/PrimeKG.