An Overview of Covid-19 Dedicated Scientific Databases

Meisam Dastani1

1- PhD in Knowledge and Information Sciences, Gonabad University of Medical Sciences, Gonabad, Iran. (Corresponding author) meisam.dastani@gmail.com

ABSTRACT

The rapid and challenging spread of the COVID-19 virus disease has become a significant threat in all countries of the world, which has provoked immediate reactions from the scientific and medical society and led to scientific publications on various aspects of the disease. Therefore, quick and easy access to these publications' results and sharing scientific data and findings to understand the disease control and create treatments and vaccines is one of the biggest ways to quickly and usefully transmit research results. The purpose of the present narrative review is to introduce the dedicated scientific databases of COVID-19 disease. For this purpose, an appropriate search was used to extract the studies conducted in the Google Scholar and PubMed databases. In the results, 16 databases of COVID-19 disease have been identified and introduced. Researchers are able to use these resources for their science and research purposes in accordance with the tools and capabilities available in the COVID-19 databases provided in this report.

Paper Type: Review Article

Keywords: COVID-19, Database, Scientific Publications.

▶ **Citation:** Dastani M. An Overview of Covid-19 Dedicated Scientific Databases. *Journal of Health Literacy*. Winter 2021; 4(5): 56-62.

Introduction

The outbreak of the new coronavirus (COVID-19) in late 2019 and early 2020 has affected the lives of people all over the world. It has led to the infection and death of many people. The rapid and unknown spread of the virus had become a significant challenge in all countries of the world. The outbreak of COVID-19 has provoked immediate reactions from the scientific and medical society, leading to an explosive growth of new data on possible treatments opportunities. Therefore, it has produced considerable knowledge in this regard in international databases (1, 2). Efforts and scientific productions in response to obscurity, danger, the rapid spread of coronavirus, its significant social and economic consequences have been part of the scientific community's conflicts to this epidemic, so the rapid dissemination of COVID-19 scientific results is critical to the rapid operation of successful clinical consequences (3). Besides, quick and easy access to the results of these studies is essential because researchers, medical staff, health providers, paramedical staff, and health support organizations, need to be aware of the results of recent vaccination developments, diagnostic kits, and the latest literature published in this field. Sharing scientific data and findings to understand the disease and to develop treatments and vaccines is one of the most significant ways to quickly and effectively transmit the results of COVID-19 research, which is essential.

Accordingly, all popular databases, including Oxford, Elsevier, BMJ, Wiley, Sage, Cambridge, Nature, and Emerald, provide free access to the COVID-19 scientific information (4).

Also, different health organizations and institutions have set up separate databases for quick and easy access of researchers to COVID-19 scientific publications and information.

Access to accurate knowledge and research services is also one of the benefits of these databases because lead to the rapid and quick access to data and scientific evidence of COVID-19. A medical science majors often take advantage of this technology with the advancement of Internet tools and the growth of online search in scientific and university study, because the field of science needs reliable, up-to-date and accurate information due to direct contact with human life as well as the amount of information in the medical sciences should be accurate, up-to-date and precise.

Likewise, the inability to locate health-related information, due to a lack of awareness and skills concerning the health-related tools and the location of information on the Internet (5), as well as increases the promotion of eHealth literature to access accurate and scientific information (6). Given the importance of the COVID-19 outbreak and the access of people to related knowledge and scientific services, this research explores the literature related to the implementation of specialized coronavirus databases and the COVID-19 study.

Method

This narrative review was designed to collect published articles and literature according to the subject and the aim of the investigation (7). In this review, all articles published in PubMed and Google scholar were searched by using keywords including COVID-19, CoronaVirus, academic databases, and scientific databases on January 2021. Then, articles that introduced new databases about coronavirus and Covid-19 were selected.

COVID-19 Open Research Dataset (CORD-19)

The development of the COVID-19 Open Research Dataset is one of the first open initiatives realized

with this portal's creation (8). In CORD-19, most of articles were sourced from PubMed Central (PMC) that expanded access to COVID-19 information by linking with scientific publishers. These publishers make COVID-19 -related articles accessible and discoverable through open-access license terms that allow for secondary and reuse analysis.

Other datasets are derived from the Medrxiv and BioRxiv preprint servers, the World Health Organization (WHO), and a collection of hand-curated articles related to COVID-19 information.

CORD-19 open dataset also linked directly with several publishers, such as Nature, Springer, and Elsevier to provide full-text coverage of articles that available in their back catalog. These articles are usually published under the special COVID-19 open access licenses. CORD-19 is currently identified as the most extensive coronavirus literature corpus that publicly available (9).

LitCOVID databas[1]

LitCOVID is a scientific information base that was created to track and quickly access to research related to COVID-19 among medical research. LitCOVID allows users to track the most up-to-date research related to COVID-19. LitCOVID includes the latest research published in the field of COVID19 in the PubMed science database, which is updated daily and is classified into groups, Mechanism, Diagnosis, Transmission, General Information, Prevention, Treatment, "Case Report, and Epidemic Forecasting.

All these activities are done manually and with the help of a machine. Among this database's features are showing the trend of COVID-19 publications on a monthly basis and quick access to scientific publications related to each country individually (10, 11).

The COVID-19 data portal[2]

In April 2020, the European Commission established the COVID-19 data portal to facilitate the exchanging and sharing of COVID-19 information and data (8).

The COVID-19 Data Portal is synchronized with COVID-19-related data and scientific literature that held in EMBL-EBI's (The European Bioinformatics Institute) data resources, including ENA, to ensure the availability of the latest data sets in UniProt PDBe, EMDB, Expression Atlas, and Europe PMC. The data continue to grow in diversity and volume and include sequences, structures, expression data, compound screens, biochemistries, and scientific publications.

COVID-19 Drug Repository[3]

The COVID-19 Drug Repository traces relevant data from pharmacogenomics and chemo-genomics studies and provides human and viral proteomics and genomics information on approved drugs and other therapeutics. This Repository objective is to automatically collect relevant data on approved drugs that implemented against COVID-19 around the world and establish a structured repository to save available publications, drug descriptions, and side effects. Likewise, This repository consists medicine- and pharmacology-oriented information, including therapeutic agents (experimental drugs), annotated information on (FDA) approved drugs, and natural chemical substances or drug-like synthetic. Recent data on therapeutic combinations and approved drugs were considered as useful methods for the treatment of SARS-CoV-2 infection that has been reported at PubMed and ClinicalTrials.gov.

This database enables users to focus on different complexity, clinical trials and formulations, starting from general information and increasing the resolution to the molecular mechanisms of drug action. Thus, the database can serve as a recommendation and navigation tool for healthcare and research purposes (8).

---

1. www.ncbi.nlm.nih.gov/research/coronavirus
2. https://www.covid19dataportal.org
3. http://covid19.md.biu.ac.il/

COVID Scholar[1]

COVID Scholar differs in the breadth of literature collected. In addition to the biological and medical research were collected by other large-scale aggregation efforts such as CORD-19 and LitCOVID. COVID Scholar's collection includes the full breadth of COVID-19 research public health, behavioral science, physical sciences, economics, psychology, and humanities.

The COVID Scholar research corpus consists of research literature from different open-access and preprint services. COVID Scholar web portal provides an accessible user interface to various literature search tools and information retrieval algorithms explicitly tuned for the needs of COVID-19 researchers (12).

Visualizing COVID-19 Research[2]

This database categorizes Covid-19 publications by topic, and it is valuable to (a) recognize trends over time by visualizing research volume in specific topics; (b) identify connections between research areas, and (c) find available resources. The Visualizing COVID-19 Research uses text mining and topic modeling techniques based on the title and abstract of articles to categorize publications. The data used in this database is received from two sources, Dimensions' COVID-19 dataset and the CORD-19 Dataset (13).

Covidex[3]

The Covidex uses artificial intelligence (AI) techniques and state-of-the-art neural network models to answer questions using the CORD-19 Dataset (14).

COVID-SEE[4]

The COVID-SEE was designed to support medical experts based on literature discovery. COVID-SEE has been designed to facilitate more interactive exploration of the COVID-19 literature by integrating sub-collection thematic analysis, document-level visual concept summaries, and PICO-structured concept relations (15). This Database data is extracted from CORD 19 Dataset.

COVID-19 LOVE[5]

Covid-19 Love serves from the Epistemonikos project that organizes the literature according to study type (systematic reviews, broad syntheses, and primary studies) as well as PICO categories and question types (e.g. diagnosis and prognosis). This database is not designed based on automatic analysis or on the CORD collection, and does not provide visual exploration (15). To complement the searches in the ten sources routinely performed in Epistemonikos Database such as PubMed, EMBASE, Cochrane Database of Systematic Reviews (CDSR), PsycINFO, Database of Abstracts of Reviews of Effects (DARE), CINAHL (The Cumulative Index to Nursing and Allied Health Literature), LILACS (Literatura Latinoamericana y del Caribe en Ciencias de la Salud), JBI Database of Systematic Reviews, the Campbell Collaboration online library, and Implementation, EPPI-Centre Evidence Library.

Trial streamer[6]

The Trial streamer is an artificial intelligence system, which finds and summarizes new trial publications, registrations, and preprints in COVID-19. This database continuously scan critical sources of research (PubMed and the WHO International Clinical Trials Registry Platform [ICTRP]), and use a machine learning system to automatically find the RCTs and extract critical pieces of information from them (16).

1. https://covidscholar.org/

2. https://strategicfutures.org/TopicMaps/COVID-19/dimensions.html and https://strategicfutures.org/TopicMaps/COVID-19/cord19.html

3. https://covidex.ai

4. https://covid-see.com

5. app.iloveevidence.com/loves/5e6fdb9669c00e4ac072701d

6. https://trialstreamer.robotreviewer.net/

SciSight[1]

A SciSight Faceted search is a tool for exploring how authors and topics interact over time. The user can select the desired topics from different PICO-like categories, which act as a filter for the shown articles. It also provides a co-mentions view with chord diagrams, displaying associations between diseases and chemicals or between proteins, genes, and cells. This is conceptually similar to our goal of representing concept relations (17). Its goal is to help accelerate scientific research and visualize the emerging literature network around COVID-19.

WellAI[2]

The WellAI has developed a analytics tool and Machine Learning (ML) search according to 4 neural networks and incorporating the complete list of NIH medical categories semantic types, for interrogation of the CORD-19 Dataset. At a practical level, searching the CORD-19 Dataset using the WellAI tool starts with the initial analysis results based on SARS and COVID-19 Coronavirus as the preloaded issues. This provides a list of 69 categories of COVID-19 concepts. Each category has associated with list of concepts that ranked according to their significance concerning regarding COVID-19 based on the WellAI neural networks (18).

COVID-19 Intelligent Insight[3]

The COVID Intelligent Search, which offers filters for various medical categories (e.g., drugs and indication) and suggests textual extracts as queries (15). Sinequa's COVID-19 Intelligent Insight portal is the repository of curated scientific publications focused on COVID-19 and SARS-CoV-2, with more than 128,000 papers (including preprints) and clinical trials, with new content added daily. These sources include scientific papers from CORD-19 (COVID Open Research Dataset), published research from Elsevier, clinical trials information from the WHO's ICTRP database, additional preprints from medrxiv, biorxiv, and arxiv, COVID-19 papers from the BMJ, official guidance from the World Health Organization, and official guidance from the Centers for Disease Control.

Doctor Evidence[4]

The DOC Search is a specialized medical search engine that is built with advanced artificial intelligence technology. This database currently contains biomedical citations and medical concepts. Doctor Evidence database includes PubMed, ClinicalTrials.gov, COVID-19 Open Research Dataset, DailyMed, EPAR, WHO ICTRP, and RSS feeds (15).

Covid Evidence[5]

COVID Evidence is continuously updated available database and evidence on interventions for COVID-19 in the world. Likewise, it monitors the present decision-makers and researchers with latest available clinical trial research agenda on COVID-19 to examine how to diagnose, prevent, manage, and treat COVID-19. This database provides information about ongoing, and completed randomized controlled trials and worldwide planned on any intervention to prevent and treat SARS-CoV-2-infections. COVID-evidence includes trials from preprint servers (medRxiv, bioRxiv), PubMed, international registries (ClinicalTrials.gov, WHO International Clinical Trials Registry Platform [ICTRP]), listing of all trials with ethical approval in Switzerland, and WHO COVID-19 literature database (19,20).

COVID NMA[6]

COVID NMA is planned as an international research initiative. It provides a living mapping of

1. http://scisight.apps.allenai.org/
2. http://wellai.health/covid/
3. https://covidsearch.sinequa.com/
4. https://covid-search.doctorevidence.com/
5. https://covid-evidence.org
6. https://covid-nma.com

COVID-19 trials that available through interactive data visualizations and Its database are obtained from the WHO International Clinical Trials Registry Platform (ICTRP) (https://www.who.int/ictrp/en/), and an international registry that assembles information on clinical trials registered in 17 primary registries. The database is weekly updated and is publicly available (21).

Discussion and Conclusion

The COVID-19 pandemic has considerably affected lives of people around the globe. A significant amount of effort has been made to progress COVID-19 research since the beginning of the pandemic. Several hundred new papers on COVID-19 are now being published every day. With so many scientific papers on this problem together, it is clear that convenient access platform for scientific research such as databases or open repository can accelerate the research process.

Researchers have therefore used artificial intelligence and machine learning tools in the battle against COVID-19 to help researchers, governments and health professionals. In order to deliver better predictive results and reliable advisory systems, more information on COVID-19 is available. In the current research depots or science databases the use of data mining techniques will save time and effort and also provide good understanding to continuously improve research on this disease (22).

Due to the analysis and identification of research patterns from data available in the scientific research is a challenging and time-consuming activity and requires a special expertise, therefore, Data mining and Text Mining techniques were used in introduced databases in this study to analyze and explain data from scientific publication. For example, the visualizing COVID-19 research database and COVID-SEE utilizes methods for text data mining and topic modeling of research clustering algorithms (13, 15). The TF-IDF text mining algorithm is used by SciSight to rank key research issues (17). COVID scholar also uses natural language processing models for scientific publication classification (12). LitCOVID database also uses manual methods and expert approval to automate and computer methods (10, 11).

In contrast to traditional search engines, the benefits of the use of artificial intelligent software and machine learning include finding and discovering knowledge that is concealed in research (18).

Relevant characteristics also occur in each scientific database and repository. SciSight describes the connection between authors and themes and determines the relationship between diseases and chemicals and proteins, genes and cells (17). DOCSerch offers also a powerful method to access scientific biomedical publications (15). COVIDLove organizes research articles by study category such as systematic analysis, detailed syntheses and preliminary studies but fails to promote visualization. WellAI also offers document retrieval through associated concepts based on concept search. COVID-19 Intelligent Insight also includes special filters, including Drugs and indication (15). Publications in the COVID scholar are divided into five areas: biological, chemical, medical, physical, humanities and social and cultural sciences (12). Clinical research data bases also display trial outcomes and results by the drawing of proof of interventions (16, 19, 20, 21). LitCOVID also classifies goods according to expert checks (10, 11).

Also, COVID-19 Drug Repository enables the user to focus on different levels of complexity, starting from general information, clinical trials and formulations, and increasing the resolution to the level of molecular mechanisms of drug action(8).

Researchers are able to use these resources for their science and research purposes in accordance with the tools and capabilities available in the COVID-19 databases provided in this report. These instruments may be used for scientometric and bibliometric research by researchers in the field of information technology.

Conflict of interest statement

There is no conflict of interest in this study.

References

1. Meskarpour Amiri M, Nasiri T, Mahdizadeh P. Subjects Clustering Analysis and Science Mapping on COVID-19 Researches in Scopus database. Journal Mil Med. 2020;22(6):663-9.

2. Kousha K, Thelwall M. COVID-19 publications: Database coverage, citations, readers, tweets, news, Facebook walls, Reddit posts. Quantitative Science Studies. 2020:1-28. https://doi.org/10.1162/qss_a_00066

3. Song P, Karako T. COVID-19: Real-time dissemination of scientific information to fight a public health emergency of international concern. Bioscience trends. 2020. https://doi.org/10.5582/bst.2020.01056 PMid:32092748

4. Ali MY, Bhatti R. COVID-19 (Coronavirus) Pandemic: Information Sources Channels for the Public Health Awareness. Asia-Pacific Journal of Public Health. 2020. https://doi.org/10.1177/1010539520927261 PMid:32429681 PMCid:PMC7240310

5. Dastani M, Mokhtarzadeh M, Eydi M, Delshad A. Evaluating The Internet-Based Electronic Health Literacy Among Students of Gonabad University of Medical Sciences. Journal of Medical Education and Development. 2019;14(1). https://doi.org/10.18502/jmed.v14i1.683

6. Shoaei MD, Dastani M. The Role of Social Media During the COVID-19 Crisis: a Narrative Review. Health Technology Assessment in Action. 2020;4(1).

7. Paré G, Kitsiou S. Chapter 9　Methods for Literature Reviews. In: Lau F, Kuziemsky C, editors. Handbook of eHealth Evaluation: An Evidence-based Approach [Internet]. Victoria (BC): University of Victoria; 2017 Feb 27. Available from: https://www.ncbi.nlm.nih.gov/books/NBK481583/

8. Tworowski D, Gorohovski A, Mukherjee S, Carmi G, Levy E, Detroja R, et al. COVID19 Drug Repository: text-mining the literature in search of putative COVID19 therapeutics. Nucleic acids research. 2020:1. https://doi.org/10.1093/nar/gkaa969 PMid:33166390 PMCid:PMC7778969

9. Wang LL, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D, Funk K, Kinney R, Liu Z, Merrill W, Mooney P. CORD-19: The Covid-19 Open Research Dataset. ArXiv. 2020.

10. Chen Q, Allot A, Lu Z. LitCovid: an open database of COVID-19 literature. Nucleic Acids Research. 2020. https://doi.org/10.1093/nar/gkaa952 PMid:33166392 PMCid:PMC7778958

11. Chen Q, Allot A, Lu Z. Keep up with the latest coronavirus research. Natur. 2020;579(7798):193. https://doi.org/10.1038/d41586-020-00694-1 PMid:32157233

12. Trewartha A, Dagdelen J, Huo H, Cruse K, Wang Z, He T, et al. COVIDScholar: An automated COVID-19 research aggregation and analysis platform. arXiv preprint arXiv:201203891. 2020.

13. Le Bras P, Gharavi A, Robb DA, Vidal AF, Padilla S, Chantler MJ. Visualising COVID-19 Research. arXiv preprint arXiv:200506380. 2020.

14. Zhang E, Gupta N, Tang R, Han X, Pradeep R, Lu K, et al. Covidex: Neural ranking models and keyword search infrastructure for the covid-19 open research dataset. arXiv preprint arXiv:200707846. 2020. https://doi.org/10.18653/v1/2020.sdp-1.5

15. Verspoor K, Šuster S, Otmakhova Y, Mendis S, Zhai Z, Fang B, et al. COVID-SEE: Scientific Evidence Explorer for COVID-19 related research. arXiv preprint arXiv:200807880. 2020.

16. Nye BE, Nenkova A, Marshall IJ, Wallace BC. Trialstreamer: Mapping and Browsing Medical Evidence in Real-Time. arXiv preprint arXiv:200510865. 2020. https://doi.org/10.18653/v1/2020.acl-demos.9

17. Hope T, Portenoy J, Vasan K, Borchardt J, Horvitz E, Weld D, et al. SciSight: Combining faceted navigation and research group detection for COVID-19 exploratory scientific search. bioRxiv; 2020. https://doi.org/10.1101/2020.05.23.112284

18. Kricka LJ, Polevikov S, Park JY, Fortina P, Bernardini S, Satchkov D, et al. Artificial intelligence-powered search tools and resources in the fight against COVID-19. Ejifcc. 2020;31(2):106.

19. Janiaud P, Axfors C, Saccilotto R, et al. : COVID-evidence: a living database of trials on interventions for COVID-19.2020. 10.17605/OSF.IO/GEHFX

20. Janiaud P, Axfors C, Van't Hooft J, Saccilotto R, Agarwal A, Appenzeller-Herzog C, Contopoulos-Ioannidis DG, Danchev V, Dirnagl U, Ewald H, Gartlehner G. The worldwide clinical trial research response to the COVID-19 pandemic-the first 100 days. F1000Research. 2020;9. https://doi.org/10.12688/f1000research.26707.2 PMid:33082937 PMCid:PMC7539080

21. Rivière P, Ripoll P, Barnier J, Vuillemot R, Ferrand G, Cohen-Boulakia S, Ravaud P, Boutron I, Alawadhi S, Amer-Yahia S, Ávila C. Research response to coronavirus disease 2019 needed better coordination and collaboration: a living mapping of registered trials. Journal of Clinical Epidemiology.;130:107-16. https://doi.org/10.1016/j.jclinepi.2020.10.010 PMid:33096223 PMCid:PMC7575422

22. Nath, Hemanta and Gary, Todd and Shepard-Smith, Andrew, Artificial Intelligence Tools and Models Used by the Scientific Community to Address the COVID-19 Pandemic (July 21, 2020). Available at SSRN: https://ssrn.com/abstract=3657855 or http://dx.doi.org/10.2139/ssrn.3657855 https://doi.org/10.2139/ssrn.3657855