

Description

Quite simply, a data repository is a place where data is stored. If you are a researcher, you already know the importance of your research data that has taken years to compile. You don't want to lose it now! In addition, you also want access to others' data in your specific discipline. Where can you find it and what can you do with it? Researchers are now encouraged to store their research datasets in repositories for safe keeping. Repositories might be located at the institution where the researcher conducts his or her studies, or one that is discipline specific, such as GenBank, which is specific to only DNA sequences. Here, we discuss some of the major research repositories and how to access them.

Finding and Using Data

We presume here that a researcher would most likely look for data in their own discipline or field of study. We also presume that the researcher intends to either analyze that data or reuse it in his or her research. This data is known as "secondary data" because it is gathered and compiled by someone for use by others. In most cases, data centers and repositories are indexed, which makes it easier to search. In addition, there are searchable registries of repositories. With the above presumptions, there are a few strategies for finding the right data. The University of Sheffield suggests the following points to begin your data search:

- Search "an appropriate disciplinary data repository or data center,"
- Search "a data portal or indexing service," and
- Find data that is "referred to or presented as supplementary material by journal papers."

Searchable Registries of Databases

There are several registries available for online searches that help find the appropriate database for your needs. For example:

- Research Pipeline: Provides a list of free data and includes multiple categories, such as bioinformatics and astrological date.
- Biosharing.org: Provides data locations for subjects in the life, environmental, and biomedical sciences.

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license

- DataCite: Indexed by the University of Sheffield Library catalog StarPlus.
- European Union Open Data Portal: Allows browsing by subject or group.
- Research Data Australia: Under development as a UK research data registry and discovery service.
- Data Citation Index (DCI): A single point of access to quality research data from repositories across disciplines and around the world. See also Web of Science.
- EMBL-EBI: Most up-to-date molecular databases.
- Registry of Research Data Repositories (re3data): Provides access to more than 1,000 repositories.
- The National Institutes of Health (NIH): Provides its own list of repositories specific to the biomedical field of research.

In addition, there is a tool entitled "OpenRefine," a former brainchild of Google, that teaches you how to find datasets, and how to clean up and transform your own datasets into various formats to post in public registries. This is by no means an exhaustive list of searchable registries, but they can help you find what you need or point you to other directories.

Data Use, Licensing, and Attribution

Researchers are finding that making their data available free of charge to other researchers not only helps to promote their own professionalism and findings, but also contributes to the overall dissemination of information to others. In the fields of science and technology, this information is valuable to both researchers and the general public who want to learn more about the latest scientific developments. Open access publishing and free databases have greatly contributed to this desire to learn, but you should remember that all data comes with licenses.

A license to use secondary data is simply a legal document that clearly states how the data can be reused. This user information should always be provided with the database itself or by the custodian of the database that you've accessed. But licensing doesn't stop there. When secondary data is combined with your own new data to create a new dataset, you must also have your own license that will apply to the new information. The terms of this new or revised license should also be listed in the original license so you should become familiar with it.

You should also be aware that when reusing data that you've pulled from a repository, you must cite its source in much the same way as any citation and reference in your research paper. The license under which you have been granted access to the secondary data will provide information on how to cite the source.

Make the Most of the Information

With the Internet, the capabilities for disseminating information are endless. Reusing data compiled by others is a great help to any researcher because it saves time, quite possibly hundreds of hours, to enable them to focus on a project. Take the time to browse through some of the links provided here to familiarize yourself with what is available. Once you have mastered the art of searching for what you need, you will be on your way to being able to use others' data in your own research, and this will be

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license

invaluable to you.

Category

1. Career Corner
2. PhDs & Postdocs

Date Created
2017/05/24
Author
daveishan

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license