Jacquelyn Leger

LIS 545B

2023-02-12

Updated: 02-26-23

Term Project – Repository Profile

Data Set: [Mortality of women of fertile age between 2006 and 2019: causes and trends](https://figshare.com/articles/dataset/Mortality_of_women_of_fertile_age_between_2006_and_2019_causes_and_trends/21971443)

Repository: [Global Health Data Exchange](https://www.re3data.org/repository/r3d100010252)

Word Count: 882

Due to the nature of the data set that I chose for the data and metadata profile assignment, my option for repositories is limited to ones related to health and medicine. Although there are many focused on medicine, I am only able to find one that focused on women, birth, and fertility. There are other options that focus on world health, which includes mortality rates data, but none that are open for data submissions. This is why I am choosing the Global Health Data Exchange as a possible data repository for the chosen data set, even though it is not explicitly stated whether it is opened or closed for submissions.

Another key factor in this decision is how open the repository is to submissions of data sets. The Global Health Data Exchange is not only a repository for data and metadata, but it is also a “comprehensive catalog of surveys, censuses, vital statistics, and other health-related data” (*Global Health Data Exchange | GHDx*, n.d.-b). The website makes that note that they both house the data sets themselves and provide the links to where data sets are house. This is important to take into consideration when the repository does not exactly provide a defined collection scope, and re3data.org states that the type of data upload is restricted and that the restrictions are other and institutional memberships. Neither of which gives researchers a clear understanding of what can and can not be submitted, or even who can submit.

Upon further inspection of the website, there is a lack of information on what data is eligible to contribute and what data they accept. They also do not provide a guide for submission, and the only assistance one may receive is through the ‘Contact Us’ form, which isn’t guaranteed to provide the answers one may be seeking. Therefor, it stands to reason that this particular repository or catalog is not open to submissions. Its contents, however, do include data sets from a plethora of countries in many different languages, including Portuguese. Among those data sets are yearly ones through the Sistema de Informação sobre Mortalidade (SIM), which provides data on the mortality rate of women of fertile age (WFA) in Brazil. SIM was even a primary source of data for the data set, Mortality of women of fertile age between 2006 and 2019: causes and trends.

Of course, SIM is but one of many organizations whose data sets have been provided to and catalogued in the Global Health Data Exchange. With over 1,000 organizations listed as data set sources, it stands to reason that there are a wide variety of data sets accessible through the catalog. According to re3data.org the subjects of these data sets include the following, social sciences, public health, health services research, social medicine, epidemiology, medical biometry, medical information, social and behavioral sciences, and humanities. It also notes that the forms that these data sets appear in are standard office documents, databases, structured graphics, scientific and statistical data formats, plain text, and archived data. A lot can be found on this website, except for submission information.

From a user perspective, there are four level of access for microdata, which are as follows:

* **Microdata access: Download** - the microdata are available for download either directly or with a simple registration process that can be completed in one day or less.
* **Microdata access: Register** - the microdata are available for download with a registration process that takes longer than a day.
* **Microdata access: Request** - the microdata are available by special request to the providing organization.  There may be a fee for the data.
* **Tabulations only** - Microdata are not available (or not applicable).  Tabulations are available either in a standalone file or in a report. (*About Data Availability | GHDx*, n.d.)

This means that there is open access data and restricted access data. The open access data can be downloaded by anyone without the need to login to the website. Whereas the restricted access data requires registration through the ‘Contact Us’ form found in the Resources section of the website, which there are varying levels of registration that are not clearly laid out. These files whether open access or restricted access are available to be downloaded directly from the website, or through the link to the providers website, which is typically tabulations only. And as far as metadata standards are concerned, I believe that the Global Health Data Exchange uses the DCAT – Data Catalog Vocabulary standard for their catalog.

In conclusion, options for open data repositories for health, women’s health, or even mortality rates in women of fertility age, are few and far between. Many of which do not provide information for potential submitters, even if they do not allow submissions. This one in particular also does not seem to provide a dissemination information package (DIP), but rather basic guides on how to retrieve the data. It also lacks in a given explanation of what it means by “microdata”. After some further research through the website, I was unable to find a clear answer or definition of what they mean by “microdata” when it comes to the levels of user access. And although it is the ideal candidate in certain aspects, it is lacking in some necessary information for both users and potential submitters.

Resources:

Albert, S. B. Z., Martinelli, K. G., Zandonade, E., & Santos Neto, E. T. do. (2023). *Mortality of women of fertile age between 2006 and 2019: causes and trends* [Dataset]. SciElo. <https://figshare.com/articles/dataset/Mortality_of_women_of_fertile_age_between_2006_and_2019_causes_and_trends/21971443>

Albert, S. B. Z., Martinelli, K. G., Zandonade, E., & Santos Neto, E. T. do. (2023b). Mortality of women of childbearing age in Brazil from 2006 to 2019: causes and trends. *SciElo*. <https://www.scielo.br/j/rbepop/a/CP6YdG3RxTqzXDBmyCRJK9x/?lang=pt>

*Global Health Data Exchange | GHDx*. (n.d.). <https://ghdx.healthdata.org/>

re3data.org: Global Health Data Exchange; editing status 2022-04-27; re3data.org - Registry of Research Data Repositories. http://doi.org/10.17616/R3QK52 last accessed: 2023-02-12

About Data Availability | GHDx. (n.d.). https://ghdx.healthdata.org/about-ghdx/about-data-availability