Health Data Science Submission: A Science Partner Journal
Invited Perspective

Health Data Sharing Platforms – Serving Researchers through Provision of Access to High Quality Data for Re-use

Authors:
Rebecca Li
Nina Hill
Catherine D’Arcy
Amrutha Baskaran
Patricia Bradford
The objective of health-related data sharing platforms or repositories is to facilitate access to datasets to either researchers or the public. Data sharing platforms provide a service to researchers worldwide through providing access to high quality data. Oftentimes platforms aggregate disparate data sources that are siloed. In the current spectrum of data sharing platforms, there are a variety of implementation models – open, managed-access and closed. The selected model depends on the data type and governance required. “Open access” or “open data” approaches allow the data to be freely available via download if the user agrees to sign a simple legal data use agreement or terms of use agreement. “Managed access” or “gatekeeper” models require additional elements for data access – these may involve a research proposal, agreeing to data use agreements, and undergoing a review process. Closed or private repositories typically only allow individuals of a single institution or community to access data.

Some repositories are broad with a general remit whereas others are tailored to a more specific constituency based on the type of data they hold (genomics-only or data from a specific disease or therapeutic area, for example). The framework for how data may be shared is outlined within the FAIR principles — Findable, Accessible, Interoperable, Reusable. Many contemporary data sharing repositories and platforms have adopted the FAIR principles to ensure they are in-line with best-practices in sharing of scientific data [1].

We describe in this commentary two case studies - one of a generalist repository that offers researchers managed access to clinical trial data and a specialist repository that serves a specific constituency providing researchers with open access antimicrobial resistance (AMR) surveillance data.

**Generalist Repository for clinical trial data**

Generalist repositories do not have a disciplinary or therapeutic area focus and provide data for scientists to access and share scientific results, and methods and also to build upon these data to drive new scientific findings. Well-known “Generalist repositories” include Dataverse [2], Dryad [3], Figshare [4] and Vivli [5-6]. The Vivli platform launched in July 2018 and is the largest clinical trial repository globally. Vivli is an example of a managed-access repository acting as a trusted neutral entity balancing the interests of multiple stakeholders to achieve goals of transparency and access to anonymized human data. Researchers benefit from free access to individual participant-level data (IPD) to further science through accessing trial data in their field (over 20 therapeutic areas covered ranging from Alzheimer’s Disease to Oncology and Vaccines are available). Recently the most frequently searched terms in the last two months are Coronavirus, Breast Cancer, Diabetes, Non-Small Cell Lung Cancer and Crohn’s Disease. To support uptake, Vivli supports a spectrum of governance approaches from fully-downloadable data, to data that is restricted and only in a secure research environment. A specialized community portal focused on COVID-19 was launched in 2021. This portal enables in specific COVID-19 trials to be fast-tracked onto the Vivli platform with its own dedicated search and expedited request review. The Vivli platform currently provides researchers access to more than 6,600 trials representing 3.6 million participants from 40 data contributors (these are represented by 25 industry, 12 academic institution/non-profit foundations and 3 government platforms listed here [https://vivli.org/members/ourmembers/]). Through the use of the platform, researchers have been able to combine and access data from diverse sources and integrate these data in a manner not previously possible by connecting with existing platforms (such as Johnson & Johnson’s YODA Project (7) Project Datasphere (8), ImmPort (9) and NHLBI’s BioLINCC (10). The data held in Vivli’s clinical trial platform is individual participant-level data or IPD; researchers also typically receive the clinical protocol, statistical
analysis plan, clinical study report (for industry data) and data dictionary as part of the complete study package.

From the time Vivli was founded approximately 4 years ago, there have been over 550 proposals requesting data submitted. Secondary analysis of the data contributed in Vivli has resulted in 100 publications, many of these include important outcomes such as design of future clinical trials and trial protocols; preparation of funding applications; informing clinical guidelines; prediction of clinical responses to treatments and development of a variety of clinical tools and decision aides.

**Antimicrobial Resistance (AMR) Repository for Susceptibility Surveillance data**

Antimicrobial resistance has been recognized as a persistent and urgent global issue and a recent analysis showed that almost 5 M deaths were directly associated with infections caused by drug resistant bacteria [11]. Open and timely access to surveillance data is an important tool for public health projections of resistance trends, to fill the gaps in the prevalence map and most importantly to curb increasing antimicrobial resistance. Antimicrobial susceptibility surveillance programs conducted by industry are mandatory for approval of a new agents, and post-approval surveillance must take place over a period of years to monitor patterns of resistance. In June 2022 the AMR Register launched by Vivli provides researchers with high quality, open access biopharmaceutical surveillance data available at amr.vivli.org. This platform provides access to surveillance data from companies with major antimicrobial programs (including Pfizer-ATLAS, Merck-SMART, GSK-SOAR, Johnson and Johnson - DREAM, Paratek - KEYSTONE, Shionogi -SIDERO-WT and Venatorx-Global Surveillance program). The industry data contributors who have provided data to the Register monitor pathogens and changes in resistance patterns through centralized testing using internationally accepted reference antibiotic susceptibility methods. Datasets in the Register include:

- >925,000 isolates
- Collected 2004-2020
- From 81 countries (including 40 LMICs)
- Testing 69 antimicrobials
- On 434 organisms

The data hosted in the Register includes raw minimum inhibitory concentration (MIC) data from pathogen susceptibility studies. These can be used by researchers, policy makers and multi-lateral organizations to identify outbreaks, predict resistance trends and inform policy. The data is provided via download for most datasets after filling out a simple request form. Acknowledgement of the data contributor in public disclosures (publications / presentations) is required.

**Provision of data for re-use** – A key feature of the Vivli platform is that all researchers are welcome to submit their anonymized human data for long-term archive and re-use. Vivli adds incremental value as data contributed are able to be integrated with other data available in the repository. The NIH, the largest global funder of Biomedical research recently instituted a key change to its policy that affects its grantees. In that major policy change, the NIH issued an update to their Data Management and Sharing (DMS) Policy (12) that applies to “all research, funded or conducted in whole or in part by NIH, that results in the generation of scientific data,” which becomes effective on January 25, 2023. This policy requires researchers to include a data management plan along with their grant application and in most cases make their data available publicly. For sharing of data, NIH endorses the use of established data
repositories as this improves the FAIR aspects of the data. Vivli is cited as one of the acceptable-use repositories for human research data (13).

In conclusion, these shared data sources represent important opportunities to both drive forward scientific insights as well as to develop research careers across multiple disciplines. It is our hope that these precious data resources are utilized by researchers as a means to accelerate their scientific goals and open new avenues for intellectual pursuit.

Conflict of Interest statement:

Li, D’Arcy and Baskaran are employees of the non-profit entity Vivli
Hill and Bradford are consultants of the non-profit entity Vivli
References


