

2016

# Research Data Management, Part Two: Data Reuse

Jane Kinkus Yacilla

*Purdue University Libraries*, [janeyat@purdue.edu](mailto:janeyat@purdue.edu)

Follow this and additional works at: [http://docs.lib.purdue.edu/lib\\_fsdocs](http://docs.lib.purdue.edu/lib_fsdocs)

 Part of the [Library and Information Science Commons](#)

---

## Recommended Citation

Yacilla, Jane Kinkus, "Research Data Management, Part Two: Data Reuse" (2016). *Libraries Faculty and Staff Scholarship and Research*. Paper 145.

[http://docs.lib.purdue.edu/lib\\_fsdocs/145](http://docs.lib.purdue.edu/lib_fsdocs/145)

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

## Internet Resources: Research Data Management, Part Two: Data Reuse

Submitted by Jane Kinkus Yacilla, Health & Life Sciences Information Specialist, Purdue University Libraries

As mentioned in September's column, the maintenance, sharing, and re-use of clinical data can be considered good stewardship of grant dollars by potentially increasing the amount of knowledge generated by the initial grant-funded data. However, clinical researchers cannot share HIPAA-sensitive data freely with the scientific community at large. A "data enclave" is an online research data center that offers a secure environment for storing sensitive data, such as clinical trial results, and protecting the privacy of study subjects by requiring that secondary use of the data meets criteria defined by the funding agency, the original researchers, or both. Data enclaves can be found at universities, government agencies, or research institutes, or may be created by an individual research group to house its own data. Perhaps due to the sensitive nature of the stored data, data enclaves are not easily discoverable with a search of Google or [re3data](#), a global registry of research data repositories. For this, researchers must rely upon their knowledge of data practices in their field, resources available at their home institutions, and conversations with staff in their funding agencies.

Here are a handful of sites through which researchers can either access public information or apply for access to proprietary or sensitive data.

### [HealthData.gov](#)

In response to the department of Health and Human Services' Health Data Initiative of 2010, HealthData.gov is a web site that aggregates datasets and other information resources created by several federal government agencies and several state and city governments. HealthData.gov makes data available to healthcare consumers, researchers, and entrepreneurs in formats that are easily downloaded and processed, while de-identifying data to ensure the privacy of individuals included in the data. HealthData.gov provides a description of the available datasets and the formats in which the data are available. Users can keyword search the site as well as browse by agency.

### [CDC WONDER](#)

WONDER stands for **Wide-ranging ONline Data for Epidemiological Research**, and is the CDC's open database of public health-related data sets. WONDER includes data on various public health topics, including AIDS, cancer statistics, tuberculosis, birth rates and infant mortality statistics, and sexually transmitted diseases, and more. Researchers can freely download these data sets and run them through a spreadsheet or statistical analysis program of choice, and they may access, use, and write about the data as long as they correctly cite the data providers, and WONDER simplifies the process by providing suggestions for citations.

### [CDC Research Data Center \(RDC\)](#)

The RDC provides an access point for restricted health data originating from the National Center for Health Statistics or the Department of Health and Human Services. The application process is in-depth, and all output from secondary research is similarly scrutinized and protected by the RDC, placing the highest emphasis on maintaining subject confidentiality throughout the

research process. While some datasets can be accessed remotely, approval may require the researcher to use some data on-site at NCHS or DHHS. Access to these datasets is fee-based.

#### [Yale Open Data Access \(YODA\) Project](#)

YODA promotes open science and research transparency by enabling researchers from Johnson & Johnson, Medtronic, and other data holders to responsibly share clinical research results. Applications for access to and reuse of data are subject to strict review to determine the appropriateness and scientific merit of the proposed reuse. Users can keyword search or browse by product brand name or generic name, therapeutic area, product class, or condition studied. There are currently 171 sets of clinical trial data in YODA.

#### [Clinical Study Data Request Site](#)

This site enables users to browse through a database of studies from several major study companies, such as Lilly, Takeda, GSK and others, and to submit a proposal for obtaining and reusing the associated datasets. Users can browse studies by sponsor, or filter all the studies by drug name, medical condition, or study phase. The site also serves as a mechanism for users to inquire about the availability of data from studies not already listed on the site. The site welcomes new clinical trials sponsors to include their study results in the database.

#### [Society of General Internal Medicine Dataset Compendium](#)

This site provides access to proprietary datasets being shared by SGIM members who are willing to collaborate with others on secondary uses of their research results, in addition to providing access to public datasets about health, health care, and medical education. Access to a proprietary dataset requires approval from the PI and a signed Data Use Agreement. At present there are seven proprietary datasets in the compendium.