The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
ASEE
Paper ID #42047

Gray Goldmine: Charting the Course to Engineering Literature's Treasures

Jamie M. Niehof, University of Michigan

Engineering Librarian Nuclear Engineering & Radiological Sciences, Engineering Education, Robotics, Integrated Systems & Design University of Michigan â€” Ann Arbor

Sarah Barbrow, University of Michigan

Sarah Barbrow is a librarian and the Assistant Director of the Engineering Library at the University of Michigan. She is a liaison to three departments: Computer Science and Engineering, Mechanical Engineering, and Naval Architecture and Marine Engineering.

Sarah graduated with an MSI in Library and Information Services from the University of Michigan School of Information. She also has an MSc in Ecology and Evolutionary Biology also from the University of Michigan.

Mr. Paul Grochowski, University of Michigan

Paul Grochowski is an engineering librarian at the University of Michigan. He holds a bachelor's degree from the University of Michigan and a MLIS degree from Wayne State University.

Luesoni Kuck, University of Michigan

Luesoni Kuck is the Biomedical Engineering Librarian at the University of Michigan and provides instructional and research support for the students, faculty, and staff within the areas of Biomedical Engineering, Civil and Environmental Engineering, and the Center for Entrepreneurship. Luesoni has a Master's of Library and Information Science from the University of Illinois, Urbana-Champaign. She also has a Master's in Biology from the University of Missouri-Columbia.

Gray Goldmine: Charting the Course to Engineering Literature's Treasures

Abstract

The landscape of gray literature in the field of engineering has undergone significant evolution over the last two decades, driven by the proliferation of the internet and diligent efforts in digitization and cataloging. It is noteworthy that the last comprehensive investigation of gray literature in engineering was published in 2001. Gray literature encompasses a diverse array of document types, including but not limited to technical reports, patents, standards, clinical trials, dissertations, white papers, case studies, preprints, and university departmental documents. These resources are a vital part of the engineering research landscape, and scholars often lack the expertise to search for them, or even the knowledge they exist. This paper is designed to demystify the search process and facilitate the discovery and utilization of these elusive, valuable, and often under-appreciated resources, emphasizing their relevance to research and academic endeavors. This paper will focus on several gray literature formats: preprints, technical reports, conference papers and proceedings, datasets, clinical trials, syllabi, and a unique collection of nuclear reactor logbooks. These formats were chosen based on practical experience with engineering student and faculty requests at an R1 institution and as a means to highlight diverse forms of gray literature. For each of these gray literature content types, this paper seeks to provide a practical roadmap for librarians and researchers to navigate the often complex ways of finding and accessing gray literature.

Introduction

A number of different working groups have offered definitions of gray literature (GL), the most notable being the Luxembourg Definition from the Third International Grey Literature Conference in 1997 and amended in 2004. At this conference, GL was defined as "information produced on all levels of government, academia, business and industry in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body" [1]. The key element in this and in most definitions of GL is that materials are not produced by commercial publishers. One implication of this is that there can be many different types of GL. Indeed, Schöpfel and Farace reproduce a list of 131 different document types originally compiled by GreyNet International, all of which can "contain unique and significant scientific and technical information that is often never published elsewhere" [1]. While readers might quibble with some of the types of GL compiled by GreyNet International, this list of 131 types serves to illustrate that librarians face a challenging task in supporting researchers who need to access such diverse formats as blogs, datasets, fact sheets, memoranda, research proposals, and myriad others. Some, though certainly not all, of the most common formats of GL used by engineering scholars and practitioners include standards, patents, technical reports, conference proceedings, case studies, preprints, datasets, and dissertations and theses.

Historically, because these resources are not produced by commercial publishers with standardized distribution networks and because their formats are so diverse and numerous, GL has been difficult to identify, locate, and access. Lawrence summarizes some of the issues driven by major technological advances surrounding GL production, dissemination, and retrieval. Problems often surface around methods of access, collection, description, and long-term preservation, though these problems manifest in different ways before and after the implementation of machine-readable catalogs and the rise of the internet [2], [3]. Savić explains how in the late 1990s, as libraries digitized print collections onto new digital formats like CD-ROMs, they encountered challenges related to the quality of the scanned content as well as long-term preservation issues [4]. The internet then allowed for an explosion in the amount and type of born-digital GL content produced by an increasing number of entities. This new production and dissemination landscape, when paired with a lack of comprehensive and standardized bibliographic information attached to these resources, continues to make it challenging to identify, find, and access GL through traditional catalogs and databases [1], [2], [3].

As Schöpfel and Farace as well as Lawrence note, scholars and practitioners benefit from a number of key features of GL [1], [2]. GL often contains unique information not found in other published sources. This information can be incredibly detailed and extensive, and not beholden to space constraints imposed by traditional scholarly publishing. Relatedly, information disseminated in GL formats may be published on an accelerated timeline relative to scholarly works, for which the peer review process can take quite a bit of time. In the case of preprint articles, access to the public can happen in a matter of hours.

Because GL represents an important swath of sources engineering scholars and practitioners need, it is critical for our patrons to have the skills to find, access, evaluate, and use it. The literature demonstrates both the need for and best practices in information literacy instruction around GL. One study by two academic engineering librarians at McGill University and a librarian working at a nearby engineering company offers concrete evidence that engineers in the workforce would benefit from more information literacy instruction at the undergraduate level around finding and evaluating GL resources [5]. Leachman and Leachman noted that engineering students at Washington State University who had trouble incorporating GL and standards effectively into their senior capstone projects benefited greatly from a particular pedagogical intervention in their library instruction session [6]. A common theme from these articles is that students find it tricky to identify and access GL, and they are eager to learn these skills. Librarians need to be equipped to help our patrons and therefore need to be well-versed in how to find and access these resources as well.

In 2001, at a moment when the internet was drastically revolutionizing how people produced and disseminated GL materials, Thompson provided a set of key strategies for finding a number of different formats of commonly used GL, such as technical reports, standards, military specifications and more [7]. In his paper, Thompson outlines how these types of GL were produced and distributed, and how those production and distribution methods could impact where one needed to look to find physical or electronic copies of the desired materials. In each section, he makes particular note of how the internet was actively changing modes of access for the different GL formats.

Over 20 years have passed since Thompson's paper was published, and given the paradigm shift precipitated by the internet and the challenges that remain in the GL landscape, we feel it is timely to revisit this topic and to illuminate strategies for finding and accessing these important resources. The number and varieties of GL formats is enormous, and we do not have space to cover them all. Instead, we focus on preprints, technical reports, conference papers and proceedings, datasets, clinical trials, course syllabi, and a unique collection of nuclear reactor logbooks. This range was selected based on practical experiences with engineering student requests at our R1 institution and encompasses several of the most commonly sought formats of GL. Though they are considered a major part of the GL corpus, both standards and patents are well-covered elsewhere in the literature and on any number of institutional Springshare LibGuides [8], [9].

Preprints

A preprint is a scholarly manuscript posted on an open access platform before it undergoes peer-review and formal publication (Fig. 1). Typically, preprints have already been accepted to a journal when they are posted, and posting an open access copy speeds the dissemination of the work. Preprints are often hosted on preprint servers, which tend to be broadly subject specific [10].

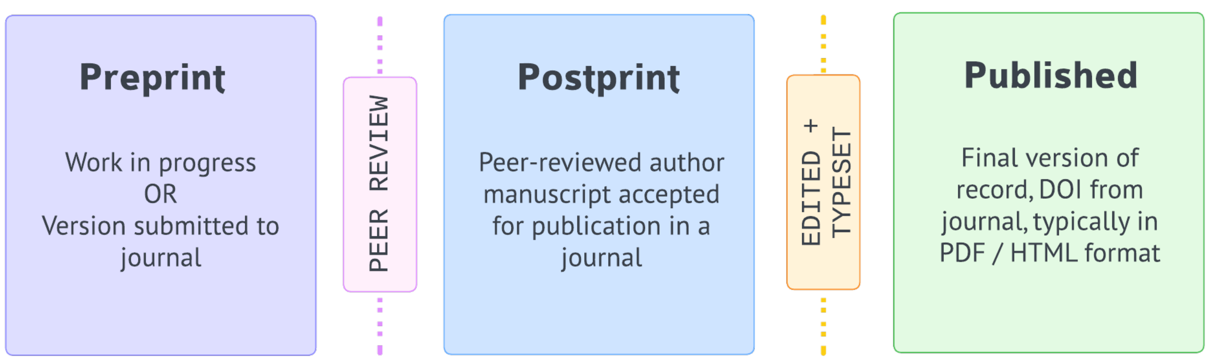

Fig. 1. Preprint lifecycle diagram.

Although preprints have not undergone the peer review process, they are of value to the scientific community and "deserving of being easily discovered and accessed" [11]. This value was especially evident at the beginning of 2020. Preprint dissemination skyrocketed immediately after the onset of the COVID-19 pandemic, with 30,000 scientific articles related to the disease posted on preprint servers within 10 months of the first confirmed case [12].
Unpublished manuscripts have long been exchanged between researchers, although they have not been easily searchable or available to the public until the early days of the internet.

First established at Los Alamos National Laboratory in 1991, the pioneering arXiv repository enabled the sharing of high-energy physics preprints [13]. This open access platform is now housed at Cornell University. arXiv paved the way for the proliferation of other preprint repositories across disciplines over the subsequent three decades, such as bioRxiv and medRxiv [14].

These repositories are excellent at featuring recently-posted articles but lack optimization for searching with specific syntax and facet limiting, which engineering librarians are accustomed to doing.

In the last five years, traditional subscription databases like Scopus, Web of Science, and Compendex have begun indexing preprints from various repositories. This discovery option allows librarians to take advantage of familiar database features like complex search strings and limiting by subject or keyword, therefore making databases excellent resources to use when working with students.

A. Subscription Library Databases

A.1. Scopus

Preprints have been included in Scopus since early 2021. As of spring 2024, the repositories indexed are arXiv, bioRxiv, ChemRxiv, medRxiv, Research Square, Social Science Research Network (SSRN), and IEEE's TechRxiv. Coverage includes preprints posted from 2017 onwards.

To find preprints in Scopus, after running a search, select "Preprints," which exists as a top-level category on the results page, between Documents and Patents. Citation metrics are not tracked for preprints, so they can only be Sorted By "Date (newest)" and "Relevance."

Preprint results in Scopus also have fewer filtering options than a typical Document search, with no Subject Area, Keyword, or Funding Sponsor options. Instead, there is a filter for "Repository," allowing users to limit to arXiv (or other repositories) if desired. The same limiting behavior can be accomplished in the Scopus Advanced Query by using:

```
LIMIT-TO ( EXACTSRCTITLE , "arXiv" )
```

A.2. Web of Science

In early 2023, the "Preprint Citation Index" was added to the suite of available database options in Clarivate's Web of Science (WoS). It is a stand-alone product and is not included in the WoS Core Collection. As of spring 2024, the repositories indexed are arXiv, bioRxiv, medRxiv, ChemRxiv, and Preprints.org, with coverage dating back to the establishment of each repository.

These repositories can be searched inside WoS by choosing the “Preprint Citation Index” or the “All Databases” dropdown menu (Fig. 2).

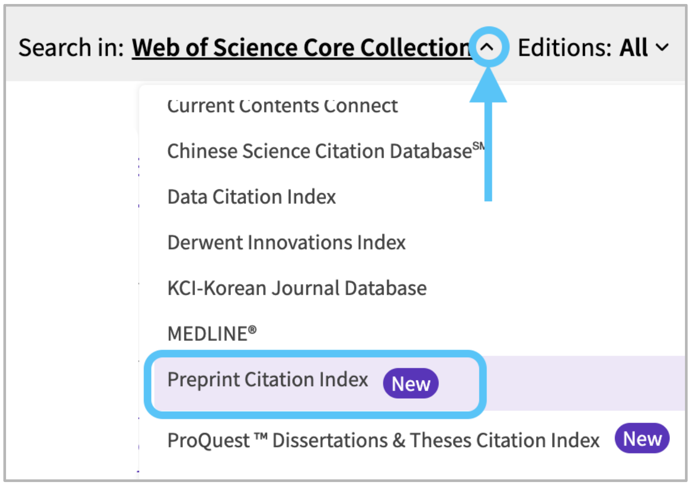

Fig. 2. Web of Science - Preprint Citation Index.

In order to make preprints visible while using the “All Databases” option, the user needs to X out the button labeled “NOT Database: Preprint Citation Index” after running a search (Fig. 3). The button indicates preprints are excluded by default. This behavior can be confusing for new users, especially because the breadcrumbs at the top of the page seem to indicate the Preprint Citation Index is being searched.

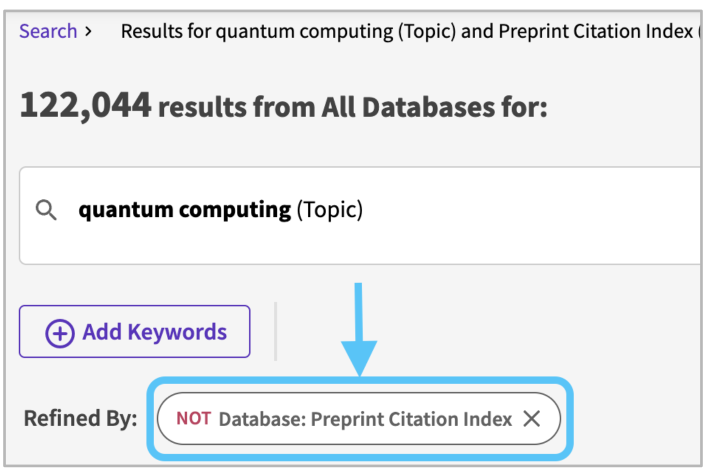

Fig 3. Preprint Citation Index button.

Once the default is toggled off and the preprints are included in the main search, the Preprint Citation Index will appear in the Database filter and can be included or excluded as needed. Individual repositories can also be included or excluded using the “Repositories” filter. Metrics are tracked for preprints in WoS, with citations and a citation alert available for each preprint. Filters are robust, including WoS Categories, Funding Agencies, and Grant Number.

These options appear whether a user is using the “All Databases” option or the “Preprint Citation Index.”

A.3. Compendex (on Engineering Village)

Preprints in Compendex are currently sourced from arXiv, Research Square, IEEE's TechRxiv, and the “Applied Sciences” and “Physical Sciences” collections from inside the SSRN. Coverage includes preprints posted from 2017 onwards.

In contrast to WoS, the Engineering Village platform includes preprints by default in search results. They can be excluded by using the Document Type filter dropdown before or after running a search. To search only preprints from the beginning, the “pp” designator can be used with “DT” document type in Expert Search:

```
("solar energy" wn TI) AND (pp wn DT)
```

Compendex allows limiting by 62 different physical properties (such as bit rate, luminance, and surface tension), and this unique feature can also be applied to preprint searches. Individual repositories can be included or excluded using either the Source Title or Repository filter.

A.4. Inspec

Preprints in Inspec come from arXiv only. Inspec began indexing preprints in 2016 and their coverage extends at least as far back as the year 2000. Inspec is available on eight different platforms, including Engineering Village, Web of Science, and ProQuest and search techniques vary by platform.

A.5. Dimensions Analytics

Preprints in Digital Science’s Dimensions Analytics come from arXiv, bioRxiv, ChemRxiv, EartharXiv, medRxiv, OSF Preprints and 50 other preprint repositories from around the world. Dimensions indexes more preprint repositories by an order of magnitude than any other subscription database. Preprints have been included in Dimensions since at least 2019, and coverage includes preprints posted from 2015 onwards.

In Dimensions, limiting to preprints is possible by using the Publication Type filter and selecting Preprint. Individual repositories can be selected or omitted by using the Source Title filter.

Dimensions features a dynamic landing page, with a feed of recent documents ingested into the platform instead of only a search box. This means from the landing page a user can select Publication Type > Preprint and see all currently-indexed preprints, which allows comprehensive analysis and visualization of preprints as a whole.

Citation metrics and Altmetrics are tracked for preprints within Dimensions, and sorting of results is available by Relative Citation Ratios and Field Citation Ratios.

B. Subject Repositories

B.1. arXiv

arXiv (pronounced "archive"), one of the oldest and most well-known preprint repositories, began in 1991 and focused mostly on high-energy physics. It is a vital resource for researchers, scholars, and graduate students across various academic disciplines. arXiv has since expanded into eight main categories and numerous subcategories (Table I).

TABLE I
arXiv Main Categories

Computer Science	Physics
Economics	Quantitative Biology
Electrical Engineering + Systems Science	Quantitative Finance
Mathematics	Statistics

Each main category and subcategory has an arXiv code, e.g.:

- Computer Science = .cs
- Artificial Intelligence = cs.AI
- Nuclear Theory = .nucl-th

The codes enable easy subject-limiting during the search process, thus:

```
chatGPT cs.AI
```

in the main search bar will look for arXiv preprints on chatGPT that fall inside the Computer Science subcategory of Artificial Intelligence.

arXiv search is limited and non-intuitive compared to search in subscription library databases. To search the entire arXiv and not a single main category, the small search box in the upper right corner must be used. Open Advanced Search from here, which has more search options.

Few filters exist in Advanced Search — before executing the search, a date range can be set. After executing the search, users can change how many results exist on each page and can sort results by Announcement Date, Submission Date, or Relevance. In arXiv, documents are “announced” first and “submitted” days or months later.

B.2. OSF Preprints

The Open Science Framework (OSF) established OSF Preprints in 2017. OSF partners with other repository service providers to help create and host niche subject repositories like EdArXiv and PaleorXiv. The search function pulls preprints from its own repository and partner repositories. Unique to OSF Preprints is a filter for "License," showing which Creative Commons or other license is applied to an individual preprint. Of interest to engineering librarians — OSF previously indexed engrXiv and still includes its preprints in their search results, but some engrXiv preprints are listed as "Withdrawn." Many of the withdrawn preprints are available directly on engrxiv.org instead of OSF.

B.3. IEEE's TechRxiv

TechRxiv was established as a preprint repository for engineering and computer science in 2020 by the Institute of Electrical and Electronics Engineers (IEEE). Its fields of interest are those that IEEE typically covers, primarily in electrical engineering, computer science and technology. IEEE Xplore does not include preprints, so searching TechRxiv can be useful for finding preprints in this domain. Results can be sorted by Recent - Most Viewed - Most Cited, and nothing else. Preprint results are presented as a grid and not a list.

B.4. Social Science Research Network (SSRN)

SSRN is a preprint repository established in 1994 and purchased by Elsevier in 2016 that focuses on social sciences. Relevant to engineering librarians are their "Applied Sciences" and "Physical Sciences" collections.

Applied Sciences includes preprints in: Agricultural Science, Computer Science, Energy, Engineering, Food Science, Forensic Science, Mathematics, and Transportation. The Physical Sciences collection includes preprints in Chemistry, Earth Science, Environmental Science, Geology, Materials Science, Physics, and Space & Planetary Science. Engineering librarians should not discount this repository because of its apparent social science focus.

B.5. Directory of Open Access Preprint Repositories

Many individual preprint repositories exist beyond what has been detailed here. A few portals allow researchers to determine if an individual preprint repository might exist that will serve their needs for posting and dissemination. It is recommended that librarians and researchers consider preprint repositories listed in the Directory of Open Access Preprint Repositories.

Technical Reports

Simply defined, technical reports "…communicate research progress in technology and science; they deliver information for technical development to industry and research institutions, contributing to the continued growth of science and technology" [15]. These reports are not typically indexed in library databases such as Scopus, Compendex, or Web of Science.

In his article from 2001, Thompson offered general observations on the publication and distribution of United States government technical reports, along with strategies for accessing the reports [7]. Since that time, we have seen important new initiatives that have aided the search for U.S. government-sponsored research, including the Technical Report Archive & Image Library (TRAIL) [15].

Thompson reported on the National Technical Information Service (NTIS) as a distributor of technical reports. In 2016, NTIS re-launched the National Technical Reports Library (NTRL) site and made it freely open to the public. Previously it was only available through a subscription. NTRL indexes over three million reports, and more than 800,000 reports are available in full text on the NTRL site [16], [17]. Although NTRL once offered a digitization-on-demand service, this is no longer an option.

NTRL allows users to search for technical reports by author, title, keyword, or other information. The search results page will include a PDF icon for reports that are available for download. Reports not available at the NTRL site may be available from the issuing agency's website. See Appendix A for a selected list of digital libraries from agencies whose reports are indexed in NTRL.

Many technical reports dating from the year 2000 and earlier might have been distributed to Federal Depository Library Program (FDLP) libraries in print or microform. At the University of Michigan, and likely at other FDLP locations, individual reports on microfiche were never cataloged. Instead, they are stored and organized by report number. For institutions that follow this practice, use the report number found in NTRL to locate the microfiche in your collection.

A second important initiative surrounding technical reports in the last two decades is TRAIL. TRAIL is a member organization that grew out of an initiative of the Greater Western Library Alliance in 2005. TRAIL "identifies, acquires, catalogs, digitizes and provides unrestricted access to U.S. government agency technical reports" [15]. As of 2023, TRAIL provides access to more than 94,000 reports [18]. TRAIL's digitization efforts have centered around mostly print reports. In 2019, TRAIL committed to the digitization of reports that were distributed in microcard format [19].

Thompson noted the difficulty accessing non-government technical reports, such as corporate documents [7]. Unfortunately, since Thompson's paper, the situation recently regarding access to these corporate documents appears to have worsened. We have witnessed engineering firms and corporations eliminating access to documents by shuttering their collections. Existing citations may allude to obsolete corporate library holdings or restricted intranet resources. With fewer inroads, establishing personal contacts internally can be an increasingly crucial avenue for prying reports loose when information cannot be discovered online or verified as disappeared altogether. As an example, we have attempted to source some reports on nuclear reactor research from the 1970s by the Electric Power Research Institute (EPRI), a nonprofit independent organization that researches electricity. A report written by EPRI researchers in 1979 does not exist anywhere on the open web or in technical report collections, and when EPRI was asked, they said the report was unavailable and archived "due to obsolete information." It can only be unarchived for "extenuating legal circumstances" [20]. Of course, while doing research into the history of nuclear reactor design, it doesn't matter that the information is obsolete. This example highlights the difficulty of unearthing GL and the importance of preserving it and archiving it in the librarian definition of the word.

Another type of technical report is one written internally within a university department. Where citations to these reports bear the university's name, researchers can try locating the reports in an institutional repository. For instance, the repository at the University of Illinois, IDEALS (Illinois Digital Environment for Access to Learning and Scholarship) includes some of these types of reports [21].

Conference Papers and Proceedings

Thompson [7] explained the critical factors that can make many conference papers and proceedings difficult to find, including:

- Not all meeting papers are officially presented
- Not all presentations given at a conference are available afterwards
- Meeting papers are sometimes published in multiple versions
- Meeting papers are sometimes published as only titles and abstracts

Although much has changed about meeting papers in the past twenty years, many of these difficulties remain. Major engineering societies including ASEE, ACM, AIAA, ASCE, ASME, IEEE, MRS, SAE, and SPIE now have robust digital libraries where they host and deliver papers from their sponsored meetings and conferences. Papers from these societies are also typically well-indexed in subscription databases such as Scopus, Web of Science, Compendex, and Inspec. However, some of these digital library collections are incomplete for papers that originated in print. Some societies with numerous satellite conferences under their umbrella might not have copyright permission to digitize back catalogs, and some societies have not digitized backfiles due to cost considerations. One hopes that societies will continue to digitize older papers and add them to their repositories to complete their holdings.

While the previous discussion centered on publications from major engineering societies that regularly convene sizable conferences, we would be remiss not to acknowledge valuable research content from smaller gatherings where sponsors do not have their own digital platforms. Many of these papers are challenging to find, for reasons stated by Thompson, but also because papers may have been made available in digital formats that are not ideal for library collections. Libraries may have purchased CD-ROMs, USB drives, or floppy disks that are no longer useful on modern computer hardware. Also, some of these electronic resources may have been distributed only to conference attendees and are unavailable for purchase for library collections.

Datasets

A dataset is a structured collection of data organized in a specific format. Datasets in engineering can come in various forms, including numerical data, experimental results, measurements, simulations, and more. Engineering students often request datasets with the goal of analyzing, manipulating, or visualizing the data instead of using the data to draw conclusions and make decisions.

Despite the fact that funding agencies now require that research data are discoverable, accessible, and preserved for further use, finding data for re-use and analysis can be time-consuming and complicated. This context motivates the following section, which is scoped to provide a starting point for helping librarians and students find sample datasets to practice their analytical, manipulation, or visualizing skills.

A. Subscription Library Databases

A.1. Web of Science (WoS)

Data Citation Index was launched inside Web of Science in 2012. It is a stand-alone product and is not included in the WoS Core Collection. It features research data from a wide range of international data repositories in the sciences, social sciences, and humanities. As of spring 2024, Data Citation Index houses more than 14 million datasets.

To access this resource, use the Web of Science dropdown menu to select “Data Citation Index” (Fig. 4).

Typical WoS filters like Year, WoS Category, Affiliation, and Subject can be applied after running a search in Data Citation Index. Unique to this index is the filter Data Type, which features a wide array of types like Meteorological, Nucleotide Sequencing,

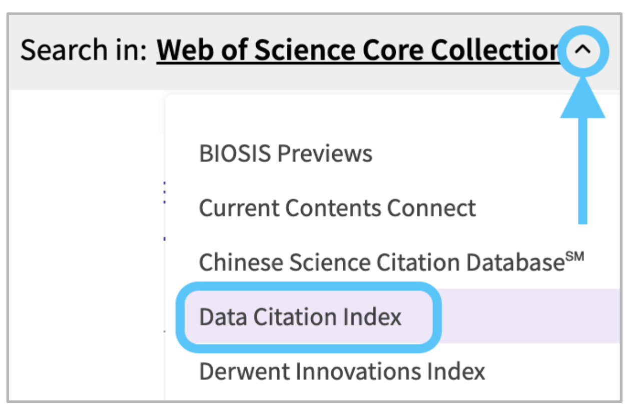

Fig. 4. Web of Science - Data Citation Index.

Molecular Structure, Images, Software, and a hundred more. When demonstrating Data Citation Index to students, the Data Type filter can be a useful tool.

The results for a given dataset will include a link to the data if available, plus standard metadata and a link in the sidebar to any WoS papers citing the data, which will go to a WoS Core Collection record.

A.2. Dimensions Analytics

Both Dimensions Analytics and the free option for accessing Dimensions have included datasets as a content type since 2020. Data are sourced from Figshare as well as Dryad, Zenodo, Pangaea, the National Institutes of Health and other repositories. "Datasets" is one of the featured document types on the landing page (Fig. 5), and selecting this will highlight the most recent datasets indexed by Dimensions.

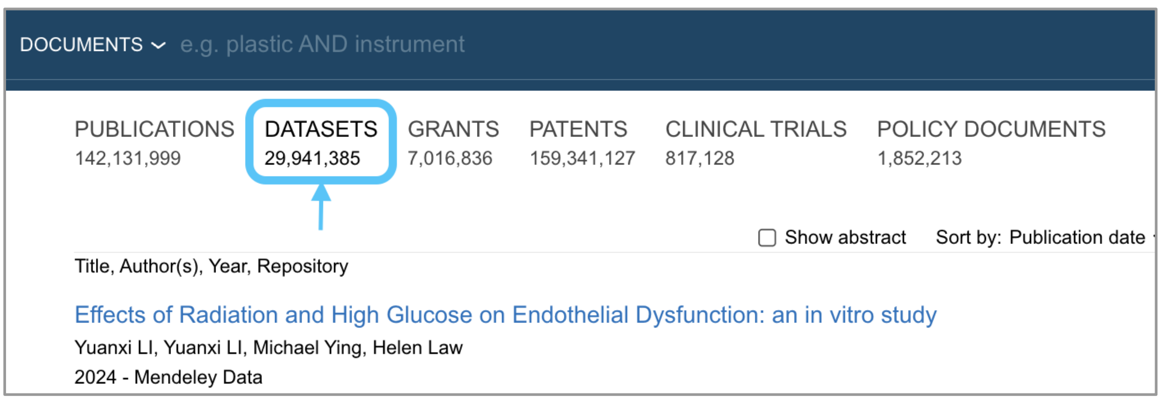

Fig. 5. Dimensions Analytics - Datasets.

Dimensions offers a traditional set of filters for datasets, including Funding Agency, Repository, and Source Title. The "Research Category" filter in Dimensions provides very granular health-related categorization, including a sub-filter called "Research, Condition and Disease Categorization," which includes options like Genetics, Vaccine Related, and Biotechnology.

A.3. Sage Research Methods Datasets

Sage Research Methods Datasets launched in 2015. This resource is excellent for students learning a new statistical technique because each dataset comes with an explanation of the analytical approach and an example to demonstrate the method. As of spring 2024 more than 600 datasets are available.

Datasets are separated into either Quantitative or Qualitative collections and include Engineering in their discipline filter. Engineering datasets teach students about simple and multiple regression using Stata and SPSS.

Sage Research Methods Datasets is useful for engineering students as a how-to guide for learning to manipulate data, instead of as a place to find datasets.

B. Open Web Data Sources

B.1. Google Dataset Search

Google launched Dataset Search in 2018. Dataset Search searches websites and data repositories for datasets using schema.org structured metadata. Subject coverage is multidisciplinary and includes government, scientific, and commercial datasets. As of early 2024 about 45 million datasets had been indexed.

Search is straightforward in Dataset Search, although the search results page is formatted differently than in Google or Google Scholar. Filters are along the top, and results are vertically aligned in a scrollable frame on the left. Datasets pop open on the right when selected from the results list.

Dataset Search filters include Download Format (tabular, image, etc.), Usage Rights, Topic, and original Provider. The Free filter can be toggled on and off. With Free toggled off, Dataset Search will pull data from inside library subscription databases like Statista and commercial sources like Global Data.

B.2. Inter-university Consortium for Political and Social Research (ICPSR)

ICPSR is an international consortium of more than 750 academic institutions and research organizations located in the Midwest. They maintain a data archive of more than 250,000 files of research in the social and behavioral sciences. Datasets can be searched by keyword and then limited by Data Format, allowing students who use specific tools like SPSS, Stata, or R to easily find relevant data. Not all datasets are available to the public, some are only available to ICPSR member institutions.

B.3. Government Data Sources

Governments around the world house myriad datasets on their websites. Many of these sites offer filtering options by Data Type or Format. See Appendix B for a non-comprehensive list of government and international data sources.

Clinical Trials

Accessing biomedical research, both peer-reviewed and GL alike, remains critical for those within or tangential to the medical community—serving to fill knowledge gaps and shed light on clinical outcomes. A vital component of biomedical GL resides in clinical trials. Defined as "any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes," clinical trials provide integral data to complement extant understanding on efficacy and procedures [22].

A. Clinical Trial Registries

Many important clinical trials may be ongoing or unpublished, so to expand a search, consider a variety of registries to identify relevant resources. Keep in mind that most databases, including subscription library databases, tend to source their clinical information from clinical trial registries.

A.1. ClinicalTrials.gov

ClinicalTrials.gov is a US-based online database that was created in 2000 by the National Institutes of Health and the Food and Drug Administration. It consists of clinical research studies and the results from these studies and has trials from all 50 states and over 200 countries. Submissions of trial studies are entered by study sponsors and investigators. It is up to the study sponsors and investigators to accurately keep their submission entry up-to-date, such as entering results when a trial has ended.

A.2. WHO International Clinical Trials Registry Platform (ICTRP)

ICTRP from the World Health Organization is a search portal that enables users to search for clinical trials across multiple international clinical trial registries.

A.3. National Registries

Many countries have a national registry for clinical trials. Knowing the country of origin for a specific clinical trial can make it slightly easier to locate the needed information.

B. Databases

Most completed clinical trials are reported and discussed in academic papers and these papers can be discovered through several specific databases.

B.1. PubMed (no subscription needed)

If a journal with clinical trial information is indexed within PubMed, then it is possible to find those clinical trials. Once a search is conducted on PubMed, use these filter options to limit results:

Article Type → Clinical trial
Article Type → Randomized Controlled Trial

PubMed allows for further categorization of a clinical trial with options found under "Additional Filters." After finding an article with a clinical trial, PubMed will provide an abstract or the fulltext of the article.

B.2. Cochrane Central Register of Controlled Trials (CENTRAL) (subscription needed)

CENTRAL is a repository for clinical trials from PubMed, ClinicalTrials.gov, International Clinical Trials Registry Platform (ICTRP), and more. CENTRAL only provides users with bibliographic data about the clinical trial article, along with the abstract of an article written about the trial. To search for clinical trials, run a search within Cochrane Library and along the top banner of the search result screen, select "Trials."

B.3. Dimensions Analytics (subscription needed)

Clinical trials in Dimensions Analytics are relatively easy to find. After logging in there is a category for "Clinical Trials" across the top banner. By selecting this option, users can view all the clinical trials within Dimensions. There are also additional useful filter options which can be found on the left-hand side of the page. A few of the filter options include:

→ Clinical Trial Status
→ Phase
→ Registry
→ Condition
→ Researcher
→ Sponsor/ Collaborator
→ Collaborating Funder
→ Location - Sponsor/ Collaborator

Once a selection is made from the search results, users can view the summary and design of the clinical trial. Often, a clinical trial will be presented with a national registry identification number, this ID is a unique identifier for each trial found in a national registry. Dimensions also provides a direct link to access additional data from the national registry. Clinical trials are not available in the free version of Dimensions.

B.4. Embase (subscription needed)

To locate clinical trials in Embase, run a keyword search, look along the top banner of the search result screen, select "EBM" and then "Controlled Clinical Trials." After selecting an article, an abstract is visible. Embase also has a search field option to search by clinical trial number.

C. Patent Databases

Searching a patent database can sometimes lead to useful information about a clinical trial, such as new drug development. Of interest, Citrome wrote an article highlighting the unique ways of searching for clinical trials beyond using PubMed and lists using patent databases in one's search [23]. Use freely available sources like Google Patents or Espacenet when searching for patents.

D. Manufacturer's Websites

Searching a manufacturer's website for clinical trials may be challenging as one navigates the website, but sometimes using this route can be rewarding in the search process.

Syllabi

Syllabi are documents that offer roadmaps for courses, including what topics will be covered in each class period, assignment descriptions and due dates, and what to read, view, or engage with in advance of each class period. Syllabi regularly clarify course basics such as the course topic and instructors, logistics of class meetings, learning outcomes, and policies such as grading, absences, and remote learning, among others. Graduate students in the Engineering Education program at the University of Michigan frequently request engineering discipline syllabi, as they analyze course content at a granular level.

Finding syllabi can be extremely difficult. Searching for a discipline's name or course title and the word "syllabus" can yield some results on GitHub or by using Bing, Google, or another preferred search engine. When using Google, for example, try searching using an advanced search method to narrow down specific sites. For example, [discipline's name] "syllabus" site:.edu will yield academic websites. Particular institutions may host sites that allow a syllabus search from their courses (e.g. UT-Austin, [24]). MIT OpenCourseware and other open course platforms will post syllabi as well.

Additionally, Open Syllabus can be a useful resource with several tools to examine elements of the 18 million syllabi collected so far. The Open Syllabus Explorer and the Open Syllabus Galaxy are both free tools that allow people to find, access, and visualize connections between citations of resources listed in the syllabi in the Open Syllabus corpus. Figure 6 provides an example of an Open Syllabus Galaxy visualization. Open Syllabus Analytics, a tool with some more powerful analytical options, can be accessed with a subscription.

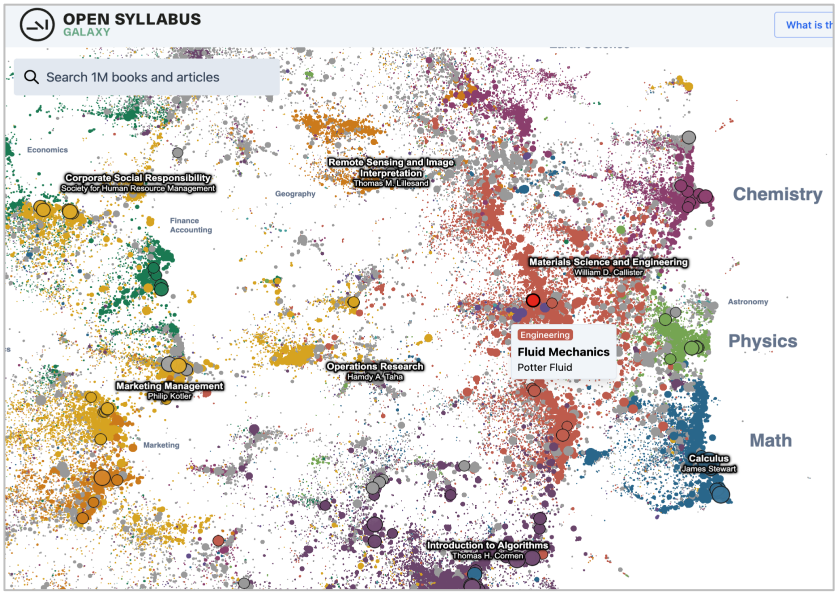

Fig. 6. Open Syllabus Galaxy visualization tool.

Note that Open Syllabus does not, as of spring 2024, provide access to full syllabi, but makes the citation data in them available and explorable. The Open Syllabus Explorer will have a small collection of syllabi to which they have full permission to index and provide access in a future version [25].

Nuclear Reactor Logbooks: A Unique Case

One especially distinctive form of GL created by researchers at the University of Michigan is Nuclear Reactor Logbooks. A small swimming pool nuclear reactor was built and operated on our campus from 1957 to 2003 (Image 1). Powered by enriched Uranium-235 provided by and disposed of by the Department of Energy, the reactor was used to study medicine, cellular biology, chemistry, physics, nuclear science, and more [26].

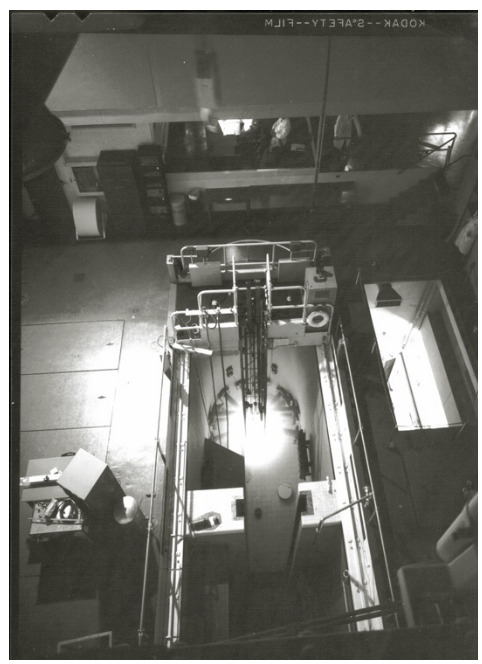

Image 1. Ford nuclear reactor [27].

During its operation the reactor was active and monitored 24 hours a day by reactor operators. Day-to-day operations were logged in 105 books of 200-300 pages each. Details include instrument calibration, shutdowns, bringing the reactor critical and up to full power, alarms, and more. This information is of historical importance to the home nuclear department, but also allows for an additional raw data source that can be processed by nuclear scientists to support validation of simulation tools [28].

In summer 2023, the University of Michigan library began the years-long process of digitizing the collection of nuclear reactor logbooks. Taken into account were copyright issues, export control issues, liability concerns with regard to the Nuclear Regulatory Commission, and compliance with our Research Ethics office. We created a custom metadata ingestion form, and the logbooks will be digitized in-house by our Digital Conversion Unit. Once the project is finished, the logbooks will be housed in our institutional repository and open to the public. Once posted, the logbooks will serve as an educational resource for the general public to learn more about how research reactors operate and could potentially be used to support the development of an open curriculum for reactor operator training [28].

Conclusions

In the future, gray literature will continue to fill knowledge gaps left by commercially published content and offer significant value in its depth and specificity of technical details that researchers use and value. Unique and rare GL resources, such as the nuclear reactor logbooks discussed above, will remain important and rich primary documentation that will inform both historical understanding and current or evolving research methods and models.

Just as the development of the internet vastly changed how GL was disseminated, sought, and accessed, artificial intelligence (AI) tools have vast potential to effect a similar revolution. AI tools may unearth connections and patterns within both structured and unstructured GL resources, allowing for powerful data mining at a breadth and depth not feasible manually. These tools may also begin to facilitate metadata generation, version control, predictive analytics, summarization, translation, and more robust information architecture. These capabilities will become increasingly valuable given the accelerating volume and forms of born-digital GL.

Truly optimizing the promise of GL in engineering scholarship and work requires shared commitments across many stakeholders, from engineers creating and documenting work, publishers and societies disseminating the work, librarians organizing and preserving the work, computer scientists leveraging AI to enrich it, etc. More cross-disciplinary partnerships and open resource initiatives will ensure critical gray information is fully utilized and leveraged for research and innovation now and in the future.

Acknowledgements

We would like to thank the following faculty and staff at the University of Michigan for their contributions to this paper: Dr. Brendan Kochunas, Leena Lalwani, Jennifer Nason, and Joanna Thielen. We are also grateful for the suggestions of our three reviewers.

References

[1] J. Schöpfel and D. J. Farace, "Grey literature [ELIS Classic]," *Encyclopedia of Library and Information Sciences*. CRC Press, Nov. 13, 2017. [Online]. Available: https://www.taylorfrancis.com/chapters/edit/10.1081/E-ELIS4-174.

[2] A. Lawrence, "Electronic documents in a print world: Grey literature and the internet," *Media Int. Aust.*, vol. 143, no. 1, pp. 122–131, May 2012, doi: 10.1177/1329878X1214300114.

[3] A. Lawrence, "From pamphlets to PDFs: The shadow histories of research publishing," *Pop Public Open Particip.*, no. 4, Oct. 2022. [Online]. Available: https://popjournal.ca/issue04/lawrence. [Accessed Dec. 11, 2023].

[4] D. Savić, "From digitization and digitalization to digital transformation: A case for grey literature management," in *21st International Conference on Grey Literature: Open Science Encompasses New Forms of Grey Literature, GL 2019, October 22, 2019 - October 23, 2019*, Hannover, Germany, 2020, pp. 13–19. [Online]. Available: http://greyguide.isti.cnr.it/index.php/greyguideportal/document-share/gl-proceedings-1993.

[5] N. Waters, E. Kasuto, and F. McNaughton, "Partnership between engineering libraries: Identifying information literacy skills for a successful transition from student to professional," *Sci. Technol. Libr.*, vol. 31, no. 1, pp. 124–132, Jan. 2012, doi: 10.1080/0194262X.2012.648104.

[6] C. Leachman and J. W. Leachman, "If the engineering literature fits, use it! Student application of grey literature and engineering standards," in *2015 ASEE Annual Conference & Exposition, Jun. 2015*, pp. 26.881.1-26.881.10. [Online]. Available: https://peer.asee.org/if-the-engineering-literature-fits-use-it-student-application-of-grey-literature-and-engineering-standards. [Accessed Nov. 1, 2023].

[7] L. A. Thompson, "Grey literature in engineering," *Sci. Technol. Libr.*, vol. 19, no. 3–4, pp. 57–73, 2001, doi: 10.1300/J122v19n03_05.

[8] M. Phillips, "Discovering and accessing standards," in *Teaching and Collecting Technical Standards: A Handbook for Librarians and Educators*, C. Leachman, E. M. Rowley, M. Phillips, and D. Solomon, Eds., West Lafayette, Indiana: Purdue University Press, 2023, pp. 53–65.

[9] K. Kozak, "Patents: Patent Home." [Online]. Available: https://guides.lib.uiowa.edu/patents. [Accessed Jan. 15, 2024].

[10] J. Bhatt and L. Willems, "Preprints: Best practice tips librarians can share with researchers," Elsevier: Connect. [Online]. Available: https://www.elsevier.com/connect/preprints-best-practice-tips-librarians-can-share-with-researchers. [Accessed Jan. 17, 2024].

[11] P. E. Bourne, J. K. Polka, R. D. Vale, and R. Kiley, "Ten simple rules to consider regarding preprint submission," *PLOS Comput. Biol.*, vol. 13, no. 5, p. e1005473, May 2017, doi: 10.1371/journal.pcbi.1005473.

[12] N. Fraser *et al.*, "Preprinting the COVID-19 pandemic," Scientific Communication and Education, preprint, May 2020. doi: 10.1101/2020.05.22.111294.

[13] D. Garisto, "Preprints make inroads outside of physics," *APSNews*, vol. 28, no. 9, Oct. 2019. [Online]. Available: http://www.aps.org/publications/apsnews/201909/preprints.cfm. [Accessed Jan. 17, 2024].

[14] B. Björk, "Open access subject repositories: An overview," *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 4, pp. 698–706, 2014, doi: 10.1002/asi.23021.

[15] "TRAIL," Center for Research Libraries. [Online]. Available: https://www.crl.edu/programs/trail. [Accessed Jan. 30, 2024].
[16] NTIS, "About the National Technical Reports Library NTRL." [Online]. Available: https://ntrl.ntis.gov/NTRL/aboutus.xhtml. [Accessed Feb. 6, 2024].
[17] A. N. Link and C. S. Wagner, "The publicness of publicly funded research," *Sci. Public Policy*, vol. 48, no. 5, pp. 757–762, Oct. 2021, doi: 10.1093/scipol/scab050.
[18] "TRAIL Metrics." [Online]. Available: https://www.crl.edu/trail-metrics [Accessed Feb. 6, 2024].
[19] J. Kirk, S. Wood, and L. Sare, "Filling in the gaps, doing what we have always done in TRAIL," *DttP Doc. People*, vol. 51, no. 4, Dec. 2023, doi: 10.5860/dttp.v51i4.8151.
[20] EPRI Customer Assistance Center, "Personal Correspondence," Dec. 01, 2023.
[21] "Technical Report - Antenna Laboratory," IDEALS. [Online]. Available: https://www.ideals.illinois.edu/collections/1417. [Accessed: Feb 6, 2024].
[22] "Clinical trials (questions and answers)," World Health Organization. [Online]. Available: https://www.who.int/news-room/questions-and-answers/item/clinical-trials. [Accessed Jan. 31, 2024].
[23] L. Citrome, "Beyond PubMed: Searching the 'grey literature' for clinical trial results," *Innov Clin Neurosci*, vol. 11, no. 7–8, pp. 42–46, Jul-Aug. 2014.
[24] The University of Texas at Austin, "Access syllabi and CVs." [Online]. Available: https://utdirect.utexas.edu/apps/student/coursedocs/nlogon/. [Accessed Feb. 6, 2024].
[25] "What is Open Syllabus?," Open Syllabus Blog. [Online]. Available: https://blog.opensyllabus.org/about-os. [Accessed Feb. 6, 2024].
[26] J. J. Duderstadt, University of Michigan Department of Nuclear Engineering and Radiological Sciences: A History. Ann Arbor, MI: Millennium Project, The University of Michigan, 2018.
[27] Ford Nuclear Reactor Pool. 2007. [Online]. Available: https://en.wikipedia.org/wiki/File:Phoenix_fnr_pool_arial2.jpg#file
[28] B. Kochunas, "Interview," Jan. 15, 2024.

Appendix A: Technical Reports

A list of agency digital libraries where users can access reports that are not available at the NTRL site.

Agency	Website
DOD (DTIC): Department of Defense (Defense Technical Information Center)	https://discover.dtic.mil
DOE (OSTI): Department of Energy (Office of Scientific and Technical Information)	https://www.osti.gov/pages
DOT: U.S. Department of Transportation	https://rosap.ntl.bts.gov
EPA: U.S. Environmental Protection Agency	https://cfpub.epa.gov/ols/catalog/catalog_lookup.cfm
NASA (NTRS): National Aeronautics and Space Administration (NASA Technical Reports Server)	https://ntrs.nasa.gov
National Research Council Canada	https://science-libraries.canada.ca/eng/national-science-library
NOAA Central Library: National Oceanic and Atmospheric Administration	https://libguides.library.noaa.gov/c.php?g=853966
Nuclear IAEA (INIS): International Atomic Energy Agency (International Nuclear Information System) NRC (ADAMS): U.S Nuclear Regulatory Commission (Agencywide Documents Access and Management System)	 https://inis.iaea.org/search https://www.nrc.gov/reading-rm/adams.html
TRID: Transportation Research International Database	https://trid.trb.org
USGS (Publication Warehouse): United States Geological Survey	https://pubs.usgs.gov

Appendix B: Government Data Sources
A list of U.S. government and international data sources.

U.S. Government Data Sources	Website
BEA DATA Bureau of Economic Analysis	https://www.bea.gov/data
BEA GDF Bureau of Economic Analysis	https://www.bea.gov/data/gdp/gross-domestic-product
BLS Bureau of Labor Statistics	https://www.bls.gov/data
Census Bureau	https://data.census.gov/cedsci
U.S. data.gov	https://catalog.data.gov/dataset
FRED Federal Reserve Economic Data	https://fred.stlouisfed.org
NASA Open Data Catalog National Aeronautics and Space Administration	https://nasa.github.io/data-nasa-gov-frontpage
NOAA National Oceanic and Atmospheric Administration	https://www.ncdc.noaa.gov/cdo-web
NYC Open Data	https://data.cityofnewyork.us/browse
International Data Sources	**Website**
Canada Open Data Portal	https://open.canada.ca/en/open-data
EU Open Data Portal European Union	https://data.europa.eu/data/datasets
U.K. data.gov.uk	https://data.gov.uk
UNICEF data United Nations Children's Fund	https://data.unicef.org/open-data
WHO Open Data Repository World Health Organization	https://www.who.int/data/gho
World Bank	https://datacatalog.worldbank.org