Beyond LMGTFY*: Access to Government Information in a Networked World

James A. Jacobs
University of California–San Diego Library
and
James R. Jacobs
Stanford University Libraries

*LMGTFY, an acronym meaning Let Me Google That For You, is an internet meme that began as a sarcastic reply to someone asking a simple question. It now has its own website: lmgtfy.com.

The internet has exponentially expanded the public's access to government information, making more of it more accessible faster than it ever was when access was limited to paper-and-ink publications. Unfortunately, this rapid deluge of fast, short-term access has exacerbated long-standing problems that impede long-term access to this information. Access does not just happen all by itself; it is only the visible tip of the iceberg. Under the surface, there is a large network of individuals and organizations that select, organize, and preserve the information output of government agencies and provide access to and services for it. The

problems that have not gone away in the age of the internet are those that in many cases are actually aggravated and intensified by it: most notably, gaps in inventorying, cataloging, and tracking all that information; the passive neglect of preservation; and even the active deletion and alteration of information.

These problems are familiar to librarians at the more than 1,200 libraries in the Federal Depository Library Program (FDLP) that have historically filled gaps and actively preserved and provided access to paper publications and now struggle to do the same for digital information. That "struggle" is made much harder when the short-term benefits of digital access mask many of the long-term problems.

Viewing Value Through Service and Use

The value of government information itself can best be understood when viewed in the context of who uses the information, how it is used, and how real-life questions are answered by FDLP librarians. The internet age often leaves users on their own amid a sea of undifferentiated and under-described information that is stripped of its context, and often, even of its authorship and origins. The provision of human information services is therefore not only still important—it is often essential.

Library users sometimes ask questions that are quick and easy to answer ("I need a hearing"), but such questions often evolve into more complex problems requiring digging below the surface of easy access to locate hard-to-identify information hidden in government agencies, nongovernment archives, or information silos of digitization projects. Such research questions can be open-ended, last months, and involve following an information trail through secondary sources that may also require data analysis. As more and more government information is digitized and born digital, the nature of these research queries has actually become more complex, not less, and finding answers is often more complicated, not easier.

Government information is in high demand by library users. Reference statistics at the Stanford University Libraries show that government

document queries have accounted for approximately 10%–15% of reference statistics annually over the last 5 years for which data have been collected. In the following sections, several case studies display library research, which give context to how access to government information typically happens in an academic library and also show the value of that information to real users. These examples highlight *what* library patrons seek, *how* they want to use government information, and *why* government information librarians remain critical to the process of information access.

Use Case 1: Into the Archives

A researcher working on the evolution of cost-benefit analysis techniques at the EPA (United States Environmental Protection Agency) went to the library seeking historical *Federal Register* (*FR*) documents.[1] The *FR* is the official journal of the federal government of the U.S. and contains most of the routine publications and public notices of government agencies. It is the main source for proposed and final rules and regulations, changes to existing rules, and notices of meetings and adjudicatory proceedings. The researcher had compiled a list of a dozen *FR* documents, including citations to regulation titles and *FR* volumes and page numbers spanning the years 1991–2000. This seemed at first to be a straightforward, simple access request because the library has full-text online availability of the *FR* from several providers, both commercial and public.

But this researcher needed more than simple access to the final rules as published in the *FR*, which are easily identifiable online documents. She also wanted the EPA's Regulatory Impact Analysis (RIA) for each rule. RIAs contain descriptions of the potential social benefits and social costs of a regulation and may be completed by consultants under contract to the agency or by EPA staffers themselves. In the past, to access these RIA dockets would have required an in-person visit to one of the EPA's libraries or to the National Archives in Washington, D.C., but over the last decade, many of these RIAs have become available in different online collections, including on the EPA's website and the

National Technical Information Service's subscription database NTRL (National Technical Reports Library). Although some RIAs are more easily accessible than in the past, the fact that they are now in several different collections increases the amount of work it takes to find which collection has the needed documents. Some but not all of the RIAs this researcher wanted were readily available, but at the time of this writing, the researcher was still waiting to hear back from the EPA library for several of the needed RIAs.[2]

Use Case 2: Government Publications for Corpus Analysis

A legal researcher at Stanford Law School recently wanted to do a meta-analysis of the reports of inspectors general in several federal agencies in order to find similarities and differences among them and requested the "Inspector General Semiannual Reports" to Congress from 1995 to 2005 for six departments: Defense, Energy, Homeland Security, Justice, State, and the Treasury. Once again, what seemed on the surface to be a simple matter of document access was much more complex, which highlights several issues that government information librarians deal with regularly: online access, findability, collection management, and the completeness of regularly published agency reports.

Stanford University Libraries' collection has most of these semiannual reports to Congress in paper or on microfiche, but there are some gaps in the collection. Because the reports are within the scope of the FDLP, Stanford was able to claim the missing reports from the U.S. Government Publishing Office (GPO). In addition, the individual departments each have some of the reports in digital form on their websites, but most do not go back to before 1998. All have gaps in coverage, and the digital files are in a variety of file formats, including Microsoft Word, PDF with digital text obtained by optical character recognition (OCR), and PDF without OCR'd text. One agency, the Office of Inspector General for the Department of Energy, had posted two files on its website that were corrupted. (To its credit, the agency replaced those files quickly with uncorrupted files after being informed about them.)

In the end, the best way to meet the researcher's needs and fulfill this seemingly simple access query was to request the bound, complete runs of the reports of all of the agencies and years via interlibrary loan (ILL) from the Southwestern Law School library and then work with the Stanford Digital Library Systems and Services (DLSS) staff to scan, perform OCR, catalog, and save all of the reports to Stanford's digital repository. What seemed to be a simple request for specific known publications took approximately 6 weeks to complete and involved staff from the library's ILL, DLSS, and Government Information services. It was made possible, not by the work of the issuing agencies, but by the diligence of the depository library at the Southwestern Law School.

Use Case 3: Blogging About Collections

Early in James R. Jacobs' tenure at the Stanford University Libraries, his responsibilities centered on international government information, intergovernmental organizations (IGOs), and nongovernmental organizations (NGOs). He noticed over the first few quarters that he was regularly receiving reference requests for average tariff levels, which is a fairly straightforward data question that is regularly asked in classes doing work on comparative political systems. The International Customs Tariff Bureau (ICTB), an IGO, has published customs tariffs since 1890 in the *International Customs Journal*, but there are several formats and resources to be used depending on the date and era of the data needed. In addition to the ICTB, UNCTAD (United Nations Conference on Trade and Development), the United Nations Statistics Division, the World Bank Group, and the World Trade Organization (WTO) have all collated trade and tariff data. While the data are available, it is difficult to know, without having gone through the iterative search process a few times, which organization or resource to use to satisfy the requester's need.

In an effort to collate that information and help the staffers in the Information Center (Stanford University Libraries' reference unit) who

might receive that query at the reference desk, Jacobs started a Q&A section on the library's blog (library.stanford.edu/blogs/topic/qa) and wrote a post in 2008 on how to find historic and current average tariff levels for countries around the world (library.stanford.edu/blogs/stan ford-libraries-blog/2008/08/qa-average-tariff-levels). That small effort at describing and collating library resources has paid off handsomely: Stanford's reference staffers have used the blog post numerous times to help students and researchers. The section also turned out to be one of the library's most popular pages; amazingly, a Google search for "average tariff levels" lists the blog post near the top of the first page of search results. It was also the impetus for a researcher from Carnegie Mellon University to request an ILL of the *International Customs Journal* (Stanford is one of the few libraries in the U.S. that owns it), which he needed for a research project.

Data: A Main Function of Government

Perhaps one of the most frequent and challenging problems researchers face is finding data and statistics. Statistical information can be published in standalone documents, but it also can be embedded in other government publications, collated in secondary research literature, or released as a dataset. Uses of data and statistics vary widely and may include the need for a single number, the use of a dataset for statistical analysis, or the merging of datasets from multiple sources.

One of the main functions of government is to gather and compile data about society, some of which—such as economic data and censuses—is gathered specifically for its own inherent information value. Other information, such as birth and death records, is gathered for administrative reasons and is collected as a byproduct of legislatively required functions and services. Typically, governments compile the information they gather (anonymizing it to protect the privacy of the citizens who have contributed the information) and make it available to the public. Until the mid-20th century, most dissemination of

this kind of government information consisted of reports analyzing and aggregating large data collections. Perhaps the largest and most prominent information publication of this type is the Census of Population and Housing, which includes anonymous, aggregated, and detailed demographic information in tables of multiple geographies down to the block level. But many other smaller, one-time reports fill government documents libraries. One example is the 1915 100-page Bureau of Labor Statistics report, "The Boot and Shoe Industry in Massachusetts as a Vocation for Women." All such reports, big and small, old and new, provide the raw facts for citizens, scholars, journalists, and governments to better understand society and culture as well as the tools for improving it. The collection, creation, and publication of such information are hallmarks of civilized society, providing an official, factual, objective perspective on the state of the state. Such information is useful in making federal, state, and local governments more effective and efficient, but it is notably used by citizens and voters to better understand their communities and by businesses to plan and grow successfully.

From Paper to Digital Data

The migration from paper publications to digital distribution of data has been steady and significant over the last 60 years. By the mid-20th century, data collection and analyses were increasingly becoming computerized. The release of raw (but anonymized) data to the public happened slowly following the development of technologies for government distribution and consumer use of data. Distribution of large census datasets on computer tape became common in the 1970s, and university computer centers and private companies developed procedures for acquiring, storing, and using the data. Private sector businesses found a lucrative market for repackaging such data. In the 1980s, governments started distributing data to libraries on CD-ROM and later on DVDs. By 1990, government agencies were going online, using dial-up computer "bulletin board" software to release time-sensitive economic statistics more quickly and in a digital form that businesses, economists,

and students were finding increasingly easy to use. By the mid-to-late 1990s, university libraries were increasingly becoming a part of the information life cycle of such government data by acquiring government datasets and providing public services for them. As the World Wide Web became popular, government agencies started posting data and reports on the internet, and during the Obama administration, the open government and open data movements led to the development of Data.gov, a dedicated website for the distribution of government data.

The *use* of government data has, over time, followed a development curve parallel to the *distribution* development curve. Data usage increases every year as the quantity of data available increases and as computer tools for using data become easier to use and acquire.

How Are Government Datasets Used?

One of the most important aspects of government datasets is that they can be used for so many different kinds of analyses. Demographic information collected by the U.S. Census Bureau is, of course, useful to economists, demographers, city planners, corporations, state and local governments, students, researchers, and academics. But the uses and repurposing of this information are limited only by the imagination. Clever computer programmers have, for example, incorporated the Census Bureau's database of the top 1,000 male, female, and last names from the 1990 census into spell-checkers.[3] Social scientists have examined the race classification categories used by the bureau over time to track changes in attitudes toward race and immigration.[4]

Census data are particularly flexible. One study combined U.S. census data with local data about public schools to study school closings.[5] Administrative data (such as "vital statistics" of births, deaths, and marriages) are particularly useful for analyzing societal events, trends, and patterns. Two researchers combined cause-of-death data by date with nongovernment information sources to study the causes of suicide among teenagers.[6]

Some government-sponsored data collections are so rich that they warrant their own bibliographies of studies that have used them. The

Survey of Income and Program Participation (SIPP) maintains its own bibliography of nearly 2,000 reports, conference papers, working papers, journal articles, dissertations, books, and book sections that use its data.[7]

With the release of data in computerized formats, it has become easier to combine data from several sources (which may have been gathered for very disparate reasons) in order to address new questions. One study combined economic gross national product (GNP) data, city-level mortality rates, and weather and air pollution data to study urban vulnerability to climate change in 12 cities on five continents.[8]

Indeed, the richness of data collected by the government, combined with new technologies that allow social participation and collection of data, are together creating a whole new category of "hybrid data." A news article from January 2014 captured the significance of these changes in its headline, "Government Data Saves Lives."[9] That story describes ongoing work inside the government to facilitate combining aerial pictures taken by the United States Air Force Auxiliary's Civil Air Patrol following a natural disaster with geotagged pictures taken by disaster victims on the ground that were culled from Twitter, Flickr, and Instagram. This two-way sharing of all kinds of data brings with it many promises of new and significant uses of government data.

Although the private sector market for repackaging and redistributing data is probably larger than the market for printed publications, the same trends that make data more easily available and usable are increasing the opportunities for libraries and nonprofit companies to create data services, which is particularly important for those services that are not profitable enough to attract for-profit companies and for users who cannot afford the often quite high prices those companies charge.

Conclusion: Many Challenges Ahead

The long-term preservation of and free access to government information has many challenges, one of which is the egovernment movement. Egovernment is a service that uses government information as part of a

two-way transaction with individuals. Although egovernment services can speed up delivery of small amounts of information to individuals, their focus is on short-term and current information. Society as a whole requires a different focus on long-term access, comprehensiveness, and maintenance of the historical record. As the government develops more digital services, it will become increasingly important for libraries and others interested in cultural memory in all its guises to draw a distinction between egovernment, a *service* that *uses* government information, and the information itself. The governments that provide dynamic information services must also provide the information behind those services in fixed, stable, and preservable formats, and distribute that information outside the government to libraries and others that have different service goals from the short-term goals of government agencies. Too many librarians, overwhelmed by the speed of digital information services, overlook the long-term challenges of the preservation of information. The availability of egovernment information services should be seen not as an excuse for reducing library involvement in government information but as evidence of the need for even more involvement in the long-term preservation of and access to information.

In an age where governments provide fast access to current information through services driven by governmental priorities, there is a growing need for nongovernment organizations, which have different priorities, to provide a long-term perspective and to provide alternative information services. While governments may neglect or even discard noncurrent information that does not meet their short-term service missions, libraries and data archives can help insure the collection, description, preservation, and accessibility of information produced by our governments for the long term. There is a clear need for networks of libraries and skilled government information professionals to build and sustain the critical information infrastructures—the 90% of the "iceberg" below the surface—to assure that information is not lost or forgotten.

Endnotes

1. Office of the Federal Register, *Federal Register*. federalregister.gov (accessed Aug. 11, 2015).

2. This is just one example of the continuing need for federal agency libraries. Federal libraries are often the only repositories that collect obscure agency memoranda, working papers, technical reports, and the like. The slow-motion demise of federal libraries over the past decade is a topic outside the scope of this chapter.

3. Cooper, Peter, "whatlanguage," GitHub. github.com/peterc/whatlanguage (accessed Aug. 11, 2015).

4. Thomas, Wendy L., "Politics in Race/Ethnicity/Ancestry: Mining the Metadata for Answers," presented at the 2004 APDU (Association of Public Data Users) Annual Conference.

5. Butler, Asiaha, "Vacant Schools in Englewood Open Doors for Economic Growth Opportunities—White Paper," Dec. 12, 2013. ragenglewood.files.wordpress.com/2013/12/white-paper-vacant-schools-in-englewood.pdf (accessed Aug. 11, 2015).

6. Phillips, David P. and Lundie L. Carstensen, "Clustering of Teenage Suicides After Television News Stories About Suicide," *New England Journal of Medicine*, vol. 315, no. 11 (September 1986): pp. 685–689. doi:10.1056/NEJM198609113151106 (accessed Aug. 11, 2015).

7. U.S. Census Bureau, "About the SIPP Bibliography." census.gov/programs-surveys/sipp/publications/aboutbib.html (accessed Aug. 11, 2015).

8. McMichael, Anthony J., et al., "International Study of Temperature, Heat and Urban Mortality: The 'ISOTHURM' Project," *International Journal of Epidemiology*, vol. 37, no. 5 (October 2008): pp. 1121–1131. doi:10.1093/ije/dyn086 (accessed Aug. 11, 2015).

9. Marks, Joseph, "Government Data Saves Lives," NextGov. nextgov.com/emerging-tech/2014/01/saving-lives-government-data/76826 (accessed Aug. 11, 2015).