Governments, like other large organizations, collect, produce, and distribute information. They generate statistical, demographic, economic, and geospatial information; and publish documents, such as laws, regulations, and executive orders, and other public records of legislative, judicial, and executive decisions and activities. During the print era, libraries acquired and organized these materials for use by researchers. CRL played an important role in this activity by preserving government publications and records not widely available elsewhere.

In the digital era, as governments expose more information directly to the web, the role of libraries in facilitating access is being re-examined. CRL recently convened two events to look at this role: a webinar on “Access to Government Information” that explored the implications for libraries of changes in the “supply chain”; and a Charleston Conference panel session, “From LexisNexis to WikiLeaks: the New Marketplace for Government Information.” This issue of Focus reports on those events, and the measures CRL is taking to address these new challenges.

—Bernard F. Reilly, Jr.
President
Access to Government Information

A webinar held by CRL on October 23, 2013

The open online distribution of government information is changing the ways that researchers access data and documentation produced by governmental agencies worldwide. This webinar explored:

- The changes in the “supply chain” for government information brought about by digital media;
- How libraries might continue to ensure access to this information for researchers; and
- What role CRL might play in supporting those library efforts.

Background: Current CRL Support for Access to Government Information

CRL has always played an important role in ensuring the survival and integrity of primary publications of U.S., Canadian, and foreign governments. CRL collections include historical documents from a great variety of jurisdictions and government agencies. CRL does not, however, hold extensive U.S. federal government publications, which are widely available through the Federal Depository Library Program.

CRL strengths include:

- **Canadian and U.S. state legislative journals**: CRL has extensive holdings of Canadian legislative journals and U.S. state legislative journals. Many of these materials have now been digitized and are accessible in the LLMC-Digital database.

- **U.S. state documents issued prior to 1951**: CRL holds more than a half-million volumes of monographic and serial publications of U.S. state government agencies and legislatures from the earliest periods through 1950. Of major historical value for various disciplines, these materials are notoriously difficult to access, since no comprehensive bibliographies or catalogs of them exist. Most are uncataloged, but are listed in CRL finding aids.

- **Publications of foreign governments**: CRL foreign documents holdings were initially formed of deposits from member libraries, with particular strengths in Western Europe and Latin America, dating from roughly 1800–1950. As CRL’s focus evolved from a depository to a center for cooperative acquisition, CRL initiated blanket orders and subscriptions to a limited number of foreign document series, over time concentrating on South Asia, Southeast Asia, Yugoslavia, Poland, and Israel. An important source of these was the PL-480 exchange program.

Other notable CRL holdings include an extensive set of central bank reports, dating from the late 19th century through the 1990s; official gazettes published by 161
countries from the 17th century through the 1990s; and a great variety of primary source materials and archives microfilmed and/or digitized by area studies interest groups affiliated with CRL. CRL has posted topic guides, listing key holdings of U.S. federal and state publications.

Presentation: Sustained Access to Government Information: the Changing Roles of Libraries

Annelise Sklar
Social Sciences Collections Coordinator
University of California San Diego Library

Even in the “straightforward” print-only Federal Depository Library Program (FDLP) collection of times past, up to a third of the collection was uncataloged and thus unfindable.¹ Today, this collection disorder is compounded by the individual agency quirksiness that accompanies distribution of state, local, and international documents, multiplied by the complexity of the online environment.

The infamous Ithaka S+R modeling initiative on the future direction of the FDLP (commissioned by the Government Printing Office) notes that libraries are moving away from the “just in case” model of purchasing tangible materials, to licensing electronic collections instead.² In fact, approximately one-fifth of FDLP libraries are almost exclusively collecting electronic-only new government publications.³

Libraries are adapting collection strategies to meet evolving user expectations and behaviors, of course. Similar to discovery practices with other types of resources, the GPO has found that 55% of self-acknowledged FDLP users list “Google or other search engine” as a frequent source for U.S. government information (a number that rises to 91% with the addition of “sometimes” users). Not surprisingly, FDLP users also list “access to more materials online” as their top most-desired improvement to government information access.⁴

Access to government information is also now part of the larger “Open” movements. The key driver behind Open Access is the idea that research results, especially those funded by taxpayers, should be freely shared for the greater good. Open Data, on the other hand, is about providing free access to data for the replication of scholarly results or repurposing into new research or new tools. Open Government, something of an extension of Open Access and Open Data, is about transparency and government accountability through public access to government-produced information.

Within this context, there is an increasing desire to remix government information. Organizations like the Sunshine Foundation and projects like Followthemoney.org and Govtrack.us are mashing up government information into new tools, and projects like IPUMS and the American Presidency Project are reformatting government data for one-stop-shopping. Users who produce these projects prefer direct access through APIs to feed into their increasingly sophisticated projects rather than having to mine that data from unwieldy pdf documents.

But at the same time, it should be noted that costs of distributing government information have shifted with the move to online access. As a 2012 CRS report notes, in the print model, depository libraries bore the costs of managing tangible materials, staff, physical plant needs, providing public access, etc. However, in the electronic model, costs of providing digital materials fall to the GPO for FDsys, and to the other agencies that provide online content.⁵

In contrast to the Open Government movement is a push toward smaller government, which means fewer government employees to produce and care for government infor-

---

mation. The direct impact on librarians and government information users has been shown in the defunding of key resources like the *Statistical Abstract* and *Sourcebook of Criminal Justice Statistics*. In fact, during the 2013 (partial) federal government shut-down, and to the surprise of many people, a number of heavily used government websites—including the Bureau of Economic Analysis, the Census Bureau, Data.gov, the Department of Agriculture, the National Center for Education Statistics and the ERIC database, National Institute of Standards and Technology (NIST), and the National Science Foundation—went offline.

Likewise, content has been lost to a “temporary” removal of online ERIC document and NASA technical reports. Though this material was available for years in the microfiche collections of FDLP libraries across the country, the implications of widespread access once the full text documents were made available online led the issuing agencies to remove the full collections of those online documents pending review (in the case of ERIC for “privacy concerns” and at NASA “to ensure that it does not contain technical information subject to U.S. export control laws and regulations.” Much of the content has since been restored, but there are no target completion dates, and any library that had weeded its microfiche collection with the intent of relying on the online simply had to do without.

Libraries, of course, need to respond to these changes in access to information in whatever ways are best for our individual user communities.

One option to compensate for the items no longer available through the depository system is to buy commercially produced government information resources. This is not a new concept (for example, libraries have purchased commercially published case law for over a century), and sometimes it is worth the money to have a better interface or more (perceived, at least) stable access. Vendors typically have the resources that governments, nonprofits, and libraries may not have to carry out these big digitization projects, update and maintain interfaces, and provide value-added context and/or metadata. However, they still have to rely on the government to produce that data in the first place, and libraries then have to justify spending the money on something that is “free.”

Another option is to digitize and preserve the materials we want ourselves. One advantage to digitizing FDLP materials is the lack of copyright restrictions. But because one cannot simply guillotine a Regional Depository’s full collection, libraries still need to develop realistic workflows for these projects. Additionally, state, local, and regional collections are often scarcer in print, and may come with additional intellectual property restrictions.

Luckily, there are already a number of collaborative digitization projects underway for libraries with the time and resources to model or join. These include FDSys, the American Memory Project, HathiTrust, LLMC-Digital, TRAIL, the Internet Archive, University of North Texas’s A–Z digitization project, and CRS archive (Indiana Virtual CD-ROM Library), to name a few. Likewise, there are a slew of projects to model or join for capturing and preserving born-digital materials: the EndofTerm collaborative web archive, the Internet Archive, UNT’s CyberCemetery, the Web Arching Service, Archive-It, LOCKSS-USDOCS, and public.resource.org.

However, born-digital materials—such as databases, websites, and publications—may be composed of dynamic content and formats that our current tools do not harvest well. Our biggest challenge is probably going to be figuring out how to preserve not just the data but the ability to interact with the data. This includes not only old CD-ROM collections, but also databases and database-driven
content that go offline. We can harvest the search page, but someone needs to capture and then maintain the back-end code if we want to provide continued access to the content.

Concrete solutions are still scarce, but the key probably ultimately lies in those things librarians are really good at—preservation, metadata standards, providing access and making it all findable—and in the thing we’re not so good at: working together to make all of these projects interoperable and seamlessly searchable.

Response: Sustained Access to Government Information
Paul Belloni
Business and Economics Reference Librarian and Selector for Psychology at the University of Chicago Library

As Annelise Sklar has indicated, government information includes so much more than it used to. But with numerous distribution sources, distribution roles are not clearly defined. Things have never been so messy and exciting for researchers and librarians.

Free software packages such as R, Google Drive, and Zotero help empower researchers from just about anywhere to do research on par with researchers from elite institutions. In addition to free tools, many organizations are taking free government information and repackaging it in helpful ways. (For example, see FollowTheMoney and Open City.)

The World Bank (WB) is one organization that works well with the new complexities of the supply chain. It looked at the landscape, saw what tools researchers are using and reacted to it. The World Bank’s Data site offers many options for gaining access to their data including microdata, an open knowledge repository, and access to WB data via external platforms such as Google Public Data and Quandl.
The U.S. Census Bureau has taken a different approach. At first glance it appears unorganized about informing researchers where to access its information. There are in-house products such as American FactFinder and DataFerrett as well as commercial products, such as Social Explorer, and organizations like MetroTrends and MetroPulse. Even though it is difficult to be sure of all of the possibilities for locating census data, there is something exciting about how so many different government agencies, nonprofit organizations, research institutions and civic-minded geeks have placed themselves in the Census Bureau supply chain. The government information ecosystem is fluid and dynamic. And it doesn’t get any more fluid and dynamic than working with U.S. Census information.

The best strategy for libraries to remain relevant within the supply chain of information is more collaboration and better service. In the digital era, our expertise as information professionals becomes as important to our libraries as the collections that we house. Due to the complexity of the government information ecosystem, we must collaborate more to improve service. We must continue to grow and cultivate collaborations with librarians both inside and outside our own institutions, maintain strong relationships with the organizations that supply the information, work closely with our own IT services, and continue to educate ourselves.

It is an exciting and overwhelming time to be working with government information and we need a supportive community in order to do the best job we can.
CRL's 2013 pre-conference session at the Charleston Conference brought together aggregators and republishers of government information with collection development and government information specialists. The participants discussed:

- How the packaging and marketing of government information has changed over the past five years;
- What value commercial publishers add to raw government information and data in the form of analysis and tools, and at what cost;
- How new media and methodologies drive novel approaches to delivery of government content; and
- What challenges the new supply chain for government information presents for researchers.

Bill Sudduth, Head of Government Information and Maps at the Thomas Cooper Library of the University of South Carolina, framed the issues from a government information librarian’s perspective. A significant change for academic libraries has been a shift toward more focused, user-centered documents collections, which is the antithesis of the regional depository library model. Collaborative consortia-based efforts—including ASERL’s Cooperative Federal Depository Program (CFDP) and the TRAIL Project for technical reports—have supported this move, improving access through digitization and shared collections.

At the same time, significant changes have affected government publishing itself, resulting in gains such as the reorganized and expanded FDsys database (formerly GPO Access). But the gains have been tempered by losses including the Census Bureau’s cessation of the Statistical Abstract of the U.S. in 2011, and the more recent announcement that NTIS will cease distributing translations from the World News Connection service at the end of 2013. Unfortunately these losses are happening at a time of expanding research interest in “big data.”

Sudduth argued the desirability of a rational and strategic national information policy rather than decisions based on short-term budget issues. He asked: In the digital age, how can a “complete” collection of government information (at the federal level) be defined, and what role can vendors play in assuring access? Furthermore, with the expanding role vendors are taking in publishing government-supplied information in online databases, what provisions are they making to ensure that what they publish is persistent, and maintains the same authority and authenticity as content supplied directly from the GPO, NARA, and government agencies? Will vendors view revised Title 44 legislative provisions for public printing as an opportunity or a threat?
Catherine Johnson, Publisher of Legislative Services at ProQuest, pointed out that the private sector has been involved in meeting government information needs since early publishing efforts in the 1970s; and that this activity has definitely accelerated over the past few years with the expansion of digitization. The private sector can fill an important role in providing deeper online collections and richer methods of access, as well as supporting innovative uses of statistical data. Researchers with varying interests and backgrounds look for different types of information in legislative and executive documents, from examining the legislative process itself, to understanding the social dynamics behind various initiatives. In recent years, libraries have presented vendors with expanding functional requirements for publishing government information online, including: advanced search and analytical tools; access to content for data mining; and depth of collections to ensure that full sub-sets of data are available to meet the needs of specific research projects. To ensure authenticity, ProQuest provides a GPO digitally signed PDF, but then creates “value-added” access versions of the document with a fully OCR’d copy accompanied by indexing and inserted citations. Moving forward, Johnson sees a multiplicity of roles in providing sustained access to government information with potential partnerships and shared interests between libraries, not-for-profit online innovators, and for-profit vendors.

Stephen Stesney, Managing Editor of Online Publishing for SAGE Publications, cited the “flood of data” being released at all levels from multiple sources, including nonprofits and private companies as well as government sources. On the one hand, researchers are simply interested in getting more data released by the government, since complex data analysis tools are now cheaper and easier to use. On the other hand, consumers of published government information also have elevated expectations for these digital collections, going beyond basic tools and functionality. Aggregating publicly available data is no longer sufficient; products need to have unique data and create additional value. At the same time, more data does not necessarily equal good data. It is challenging for publishers to hunt, extract, and clean data from multiple government sources. SAGE’s plans for meeting social science research needs will focus on incorporating unique data, which is “contextualized with public data”. It plans to work with specialist “data editors” to find and analyze overlooked data sources, as well as commissioning some new data sets. SAGE anticipates that the future will bring “more data, more data, more data,” including increased use in classrooms. For publishers, this brings an opportunity to create data-learning environments for students. It also represents a challenge to ensure authority of data, through detailed metadata that consistently describes and links back to the original data sources.

Jeffrey Cross, Academic Sales Manager for Statista, echoed Stesney’s emphasis on the growing demand for data from multiple sources. In a “multi-polar data world,” not only is more public information openly accessible, but multiple sources for the same data exist. Like SAGE, Statista seeks to integrate data from industry and non-governmental organizations with government-sourced information. Cross also emphasized the importance of trust and transparency in documenting sources, to guarantee the authority of data. Nothing can replace human editorial oversight. And he emphasized that vendors have a responsibility to work with each other, their sources, and libraries on preservation as well as access.

Robert Lee, Director of Online Publishing and Strategic Relationships for East View Information Services, described the particular challenges of sourcing and distributing non-U.S. governmental information. To start with, how does one define what constitutes government data, which varies
significantly across different world areas (for example: Russia, China, and the Middle East) and over time? The extent of availability or censorship can be very revealing of a government’s position on access, depending on what is flagged for internal consumption, obfuscated, censored, or simply not approved for release. The U.S. saw the recent example of cutting off the distribution of the translations in the *World News Connection*. Even in the existing WNC content, the basis for selecting materials to translate and disseminate is not well understood.

Additionally, Lee noted, differences exist between traditional definitions of government documents and the ways in which new media can facilitate access today, allowing more immediacy but also potentially more instability of authentic and persistent sources. Funding for distribution of information from agencies is sometimes limited, forcing selectivity. Third parties, such as non-governmental agencies, may serve as a hedge to problems of consistency in government dissemination. In any case, the potential benefits to research, of improved aggregation, normalization, and preservation of data from various government sources leads one to hope for more private/public synergies in the future.

**Angela Carreño, Head of Collection Development for the Division of Libraries at New York University**, responded to the panel’s points from the perspective of a collections manager. She is called upon to respond to various research interests as well as balance spending requests. First and foremost, it is clear that the role of vendors will become increasingly important to help aggregate as well as to preserve data.

Vendors will be especially important in identifying and providing access to information from governments in other parts of the world, as demonstrated by Robert Lee.
A systematic process for capturing, preserving, and disseminating foreign born-digital government information is currently lacking. Could a routinized collaborative library service make it possible for specialists to identify and also acquire open access government documents from non-U.S. jurisdictions (such as the Justice Verma Committee report issued following the rape homicide case in New Delhi)? She cited as another example a professor seeking authoritative data on 2006 election results in the Dominican Republic. This information was previously published in the Gaceta Oficial. Now there are various sources including a website hosted by the Junta Central Electoral, but the spreadsheet results found there are unwieldy and inconsistent with other sources.

Vendors will also play an essential role in supplying the growing research demand for “big data.” Emerging at NYU is the field of “urban informatics.” There the Center for Urban Studies and Progress (CUSP) plans to bypass library acquisitions efforts, to collect data directly from city and state agencies.
The presentations and discussions at the CRL-sponsored webinar and panel session suggest that this is an important moment for research libraries, and potentially for the scholars they support. Digital technologies and the Internet have fundamentally altered the ways in which publications, documents, and data produced by governments are distributed and accessed. In the print era, distribution involved a relatively small number of actors. The supply chain then was relatively simple: government data and documents were published in print, microform, and other tangible, fixed media normally by a government printing office, and were acquired, stored, and made available by public libraries, research libraries, and other official depositories. Information considered to have commercial value was often republished and distributed to libraries by commercial aggregators.

Today, as documents and data are exposed directly to the web, more actors are involved in the supply chain. Aside from the agencies that produce official documents and information, these actors include the organizations responsible for the technologies used in production and distribution of government information in electronic form: producers of the software, operating systems, platforms, and other tools that enable the storage, display, manipulation, and analysis of the documents and information, and their interoperability with related documents and information from other sources. The new actors include also the nongovernmental entities that reaggregate and provide access to government-produced information, including the traditional publishing aggregators (LexisNexis, Sage, ProQuest) and newcomers like AccuWeather and Bloomberg Government, who build new products and services around government-produced information.

In the webinar and the panel session, several presenters referred to the proliferation of Open Access databases and sites, which potentially reduce the need for libraries to mediate access to current content. The digital publishing process itself, however, is fluid and dynamic: instead of producing information in the form of “documents” and other discrete units, agencies output information in the form of XML and ASCII text, video, datasets, and other forms that they can easily and continually update, edit and augment—and even withdraw entirely from the platforms on which they are published. Several presenters noted the failure of many U.S. government web sites during the October 2013 government shutdown.

It is clear that the commercial publishing sector will continue to be as indispensable to scholarly access in the future as it was in the print era. With the vast oceans of data flowing from governments today, organizations with robust technical and financial resources are needed to gather, authenticate, organize, and publish.
government-produced data and digital content, and to provide platforms and tools that facilitate discovery and interpretation.

What do these developments mean for the roles research libraries play in ensuring the long-term accessibility of government-produced digital content? The webinar and forum presenters suggested several ways in which research libraries, collectively and individually, might figure in this new environment, including collaborative digitization and archiving of legacy print collections. CRL has undertaken several initiatives:

- **Digitization and archiving of legal materials:** Through CRL’s partnership with the Law Library Microform Consortium, targeted content has included U.S. and Canadian legislative journals; African law reports (colonial and post-colonial); and materials from Latin America and Eastern Europe, drawing not only from CRL’s collection but also through arrangements with CRL member libraries.

- **Digitization of U.S. government technical reports through the Technical Report Archive & Image Library (TRAIL):** The collaborative project, established by the Greater Western Library Alliance (GWLA) and now based at CRL, identifies, digitizes, and provides open access to federal technical reports primarily issued before 1976. TRAIL has digitized and mounted on the open web more than 40,000 full text reports, which comprise a valuable record of technology research and applications in the mid-twentieth century.

- **Digitization and archiving of serial publications of U.S. agriculture experiment stations and extension services:** Project Ceres, a joint undertaking of CRL, the United States Agriculture Information Network and the National Agriculture Library, supports digitization of serial publications, many of them products of public universities and state agencies.

Preservation of born-digital government information will require a different set of strategies. One appropriate new role for libraries may be monitoring and evaluating the gaps and limitations of systems and platforms used to maintain and distribute government information such as the Government Printing Office’s digital repository, FDsys, and third-party archiving systems such as LOCKSS for Government Documents, the Internet Archive, and the University of North Texas Libraries’ CyberCemetery Archive of Government Websites.

It might also involve identifying and evaluating the key databases of government information maintained by commercial publishers, and negotiating favorable terms for purchase of and subscription to same for academic and other research libraries. Collective negotiations will strengthen libraries’ hands in ensuring that the databases deliver the kinds of benefits that best serve scholars, such as comprehensiveness, persistence, and provenance and authenticity of content.

Through its Global Resources Forum, CRL has begun to provide independent analysis and venues for information-sharing to support this kind of work by member libraries. In April 2014, CRL will convene its annual Collections Forum, which will further develop a collective agenda on government information for CRL and its member community. The forum, “Leviathan: Libraries and Government Information in the Era of Big Data,” will bring together representatives of government agencies, historians, and members of the research library community, to further determine what role libraries can and should play in ensuring the preservation and future accessibility of the records and publications of U.S., Canadian, and foreign governments. A report from that forum will be published in the summer issue of *FOCUS*. ❖
FOCUS on Global Resources, published quarterly, is compiled by CRL’s Communications Department. Gwen Ihnat, Virginia Kerr, Editors. Graphic design services provided by Molly O’Halloran, Inc. Special thanks to: Bernie Reilly, James Simon, Virginia Kerr, Mary Wilke, Annelise Sklar, Paul Belloni, Bill Sudduth, Catherine Johnson, Stephen Stesney, Jeffrey Cross, Robert Lee, and Angela Carreño.

ISSN #: 0275-4924

Center for Research Libraries Staff Contacts
(800) 621-6044

President
Bernard F. Reilly x 334
breilly@crl.edu

Director of Information Systems
Patricia Xia x 341
pxia@crl.edu

Administrative Services Specialist
Yvonne Jefferson x 319
yjefferso@crl.edu

Head, Access Services
Kevin Wilks x 314
kwilks@crl.edu

Member Liaison and Outreach Services Director
Mary Wilke x 351
mwilke@crl.edu

Head, Stack Management
Bethany Bates x 339
bbates@crl.edu

Director of International Resources
James Simon x 324
jsimon@crl.edu

Digital Program Manager
Virginia Kerr x 265
vkerr@crl.edu

Director of Technical Services
Amy Wood x 327
awood@crl.edu

Communications Specialist
Gwen Ihnat x 289
gihnat@crl.edu

Global Resources Program Contacts
(800) 621-6044

Director
James Simon x 324
jsimon@crl.edu

Project Coordinator—Global Resources Network
Judy Alspach x 323
jalspach@crl.edu

Contact for Information Regarding:
(800) 621-6044

Billing/Invoicing/Deposit Accounts
Accounting Department x 345

Loans/Photocopies/Demand Purchases
Access Services Department x 314

Membership/Communications/Public Relations/Marketing
Don Dyer x 317
ddyer@crl.edu

Microfilm/Microfiche Sales
Lorraine Farley x 336
lfarley@crl.edu

CRL Publications Orders (non-FOCUS)
Meeting/Schedules
Yvonne Jefferson x 319

Purchase Proposals/Electronic Licensing
Mary Wilke x 351

Collection Deposits
Mary Wilke x 351

FOCUS on Global Resources 13