



# Open Data: What It Is and Why You Should Care

Lori Bowen Ayre & Jim Craner

To cite this article: Lori Bowen Ayre & Jim Craner (2017) Open Data: What It Is and Why You Should Care, Public Library Quarterly, 36:2, 173-184, DOI: [10.1080/01616846.2017.1313045](https://doi.org/10.1080/01616846.2017.1313045)

To link to this article: <https://doi.org/10.1080/01616846.2017.1313045>



Published online: 19 Jun 2017.



Submit your article to this journal [↗](#)



Article views: 259



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)



## Open Data: What It Is and Why You Should Care

Lori Bowen Ayre and Jim Craner

The Galecia Group, Petaluma, California, USA

### ABSTRACT

Open data is information that is provided by public entities to be accessed and reused. Publishing open data is an excellent way to improve an organization's transparency and provide insight into the value of the organization. Libraries are uniquely positioned to assist their patrons in the use of open data by making them aware of it and helping them access and use it. This article provides a short history of open data and explores ways that library trailblazers are already using open data as well as contributing open data for others to use. Getting started with open data involves identifying key open data resources such as data.gov, identifying library information that would be beneficial to publish, and creating programs that provide digital literacy training and create opportunities for patrons to engage with open data in new and creative ways.

### ARTICLE HISTORY

Received April 2017

Accepted April 2017

### KEYWORDS

Government data;  
hackathons; library data;  
open data; transparency

“Open Data” refers to data collected and shared with others to use as they wish, without restrictions on copyright or usage. Traditional examples of open data include government-collected data (e.g., weather reports, crime incident reports, postal or Zip codes) as well as some academic sources (e.g., open-access journals, raw polling and survey data, scientific experiment results). Businesses are also realizing the benefits of sharing data and making it available for use.

What's being done with all of this data? Local, state, and national government agencies—including libraries—use shared data to lower costs and improve services they deliver to their constituents. Non-profit and community groups use valuable government data to better target services. Businesses rely on government economic statistics, maps, and commercial reports to structure their operations.

Two important qualities make data “open.” Data must be “legally” open, which means it is unrestricted by restrictive copyright terms and can be legally shared and used by individuals, businesses, academia, nonprofits, and others. Data must also be open in the “technical” sense of the word, which

**CONTACT** Lori Bowen Ayre  [lori.ayre@galecia.com](mailto:lori.ayre@galecia.com)  The Galecia Group, 15 Old Creek Road, Petaluma, CA 94952, USA.

Color versions of one or more of the figures in the article can be found online at [www.tandfonline.com/wplq](http://www.tandfonline.com/wplq).

Published with license by Taylor & Francis. © Lori Bowen Ayre and Jim Craner

refers to the delivery of the data in a standard, well-defined format, usually via the Internet.

A current definition of open data is provided by Open Definition:

Open data that can be freely used, re-used and redistributed by anyone—subject only, at most, to the requirement to attribute and sharealike.<sup>1</sup>

## A short history

In 2007, Larry Lessig and 29 other “open data pioneers” established a set of values articulating how the government could provide data in a manner that would empower others to review, analyze, and act on the data; the ultimate goal was to further improve government operations. They argued that data should be complete, primary, timely, accessible, machine-processable, non-discriminatory, non-proprietary, and license-free (although now an explicitly open license is preferred to unlicensed releases).<sup>3</sup>

In 2013, President Obama signed an executive order which resulted in the establishment of an Open Data Policy<sup>2</sup> and the data.gov website, the nation’s first federal government data portal. The Order stated as follows:

Openness in government strengthens our democracy, promotes the delivery of efficient and effective services to the public, and contributes to economic growth. As one vital benefit of open government, making information resources easy to find, accessible, and usable can fuel entrepreneurship, innovation, and scientific discovery that improves Americans’ lives and contributes significantly to job creation.<sup>3</sup>

This order led to the adoption of the Digital Accountability and Transparency Act of 2014 (DATA Act). The DATA Act expanded upon the Federal Funding Accountability and Transparency Act of 2006 which had establish government-wide data standards for financial data, to improve transparency about government spending thereby improving the quality of data submitted to USASpending.gov.

Today, the data.gov data portal contains nearly 200,000 datasets<sup>4</sup> contributed by virtually every major department of the United States government, including the Departments of Agriculture, Commerce, Health and Human Services, Transportation, and Homeland Security as well as other agencies such as the Environmental Protection Agency, Federal Communications Commission, Federal Election Commission, and the Consumer Product Safety Commission. In addition, hundreds of cities, counties, and almost all of the U.S. states have established their own data portals<sup>5</sup>.

## Benefits of open data

Open data offers numerous benefits to public agencies. Project Open Data,<sup>6</sup> another outcome from the 2013 Executive Order, is an “online, public repository intended to foster collaboration and promote the continual improvement of the Open Data Policy.” They describe the benefits as follows:

- Save time and money when responding to Freedom of Information Act requests
- Avoid duplicative internal research
- Use complementary datasets held by other agencies
- Empower employees to make better-informed, data-driven decisions
- Attract positive attention from the public, media and other agencies
- Generate revenue and create new jobs in the private sector<sup>7</sup>
- The power of open data—like any other use of data—is its ability to drive and support intelligent decision-making, from individuals to organizations to entire communities and nations. The “openness” of open data reduces friction and increases the ability of people and organizations to use that data effectively.

So what does the “use” of open data look like? The potential applications are limitless and countless examples exist already in many sectors of the economy. Let us consider a few examples tangentially related to a visit to a local public library:

- (1) Our patron—a preschool teacher on her day off—begins the journey to the library by opening her city’s “BusFinder” transit app on her phone. Because her city’s transit agency publishes its routes and schedules online, using the open GTFS transit data format, it is easy for application providers to build useful local transit apps that help her plan an effective journey. Her app even integrates weather warnings from the National Weather Service’s open data feed, notifying her of potential delays.
- (2) Upon arrival at the library, our patron—an astronomy buff—heads for the library’s new 3D printer and downloads a free printable model of the Webb Space Telescope from NASA’s public domain data repository. The plastic model, once printed, will make a neat item for her daycare’s “Show and Tell.”
- (3) Our patron is finishing a degree in early childhood development. She asks a reference librarian for some statistics required for a thesis paper. Previously, that information was locked in a proprietary academic database, but thanks to federal open data initiatives, the information

can now be easily retrieved from the Office of Head Start, accessed via [data.gov](http://data.gov).

- (4) Finally, our patron uses a public access computer to visit her local chamber of commerce's website. Each month, the site provides a spreadsheet of new business registrations in the city; our patron prints this data for use in her daycare's annual fundraising drive.

## **The role of public libraries in the open data ecosystem**

Public libraries are uniquely positioned to provide multiple important roles in the open data world: helping patrons find open data and applications that meet their needs, helping their community partners publish open data within the community, and publishing open data themselves.

As trusted sources of information, public librarians should be aware of local, state, and federal open data assets as well as related applications that could benefit patrons. As in our example above, librarians are often the first point of contact for members of the public seeking information. The existing information retrieval and digital literacy skills of librarians can be put to use helping patrons navigate open data repositories, retrieve needed data, and directing them to other tools and resources for analysis.

As relatively neutral authorities on information management, public libraries are excellent partners for local and county governments that wish to publish their own open data for the benefit of their communities. Some libraries are already the de facto or official archivist of municipal/county information; helping operate an open data repository can be a logical extension of that role.

Your library also has data that would be useful to publish including as follows:

- (1) Statistics about visitors, computer usage, materials borrowed, etc., which illustrate the value of the services you provide (see [Figure 1](#));
- (2) Budget data that can be accessed by journalists, academics, good-government groups, and citizens interested in how taxpayer dollars are spent;
- (3) Information published on your existing website—such as event schedules or operating hours—when offered in a “machine-readable” format become available for “mash-ups” with other data or local applications (see [Figure 2](#)).

## **Library open data trailblazers**

The Newcastle (UK) Libraries embraced open data at all levels and in all aspects—philosophically, strategically, technologically, and more—to great success. The article “Engaging Citizens with Data that Belongs to Them”<sup>8</sup> tells their story. It is instructional and inspiring.

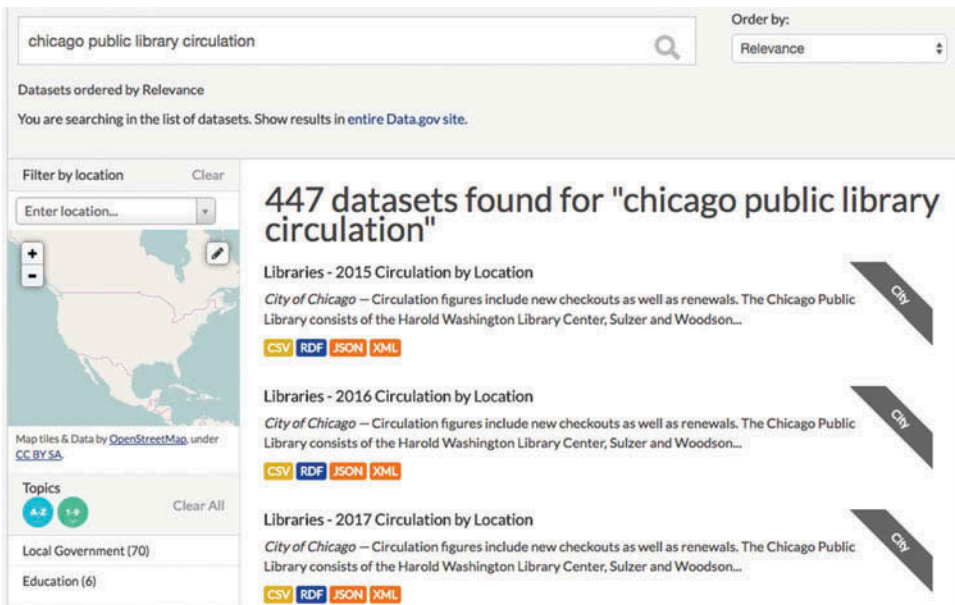


Figure 1. Datasets contributed by Chicago Public Library and available on Data.gov.

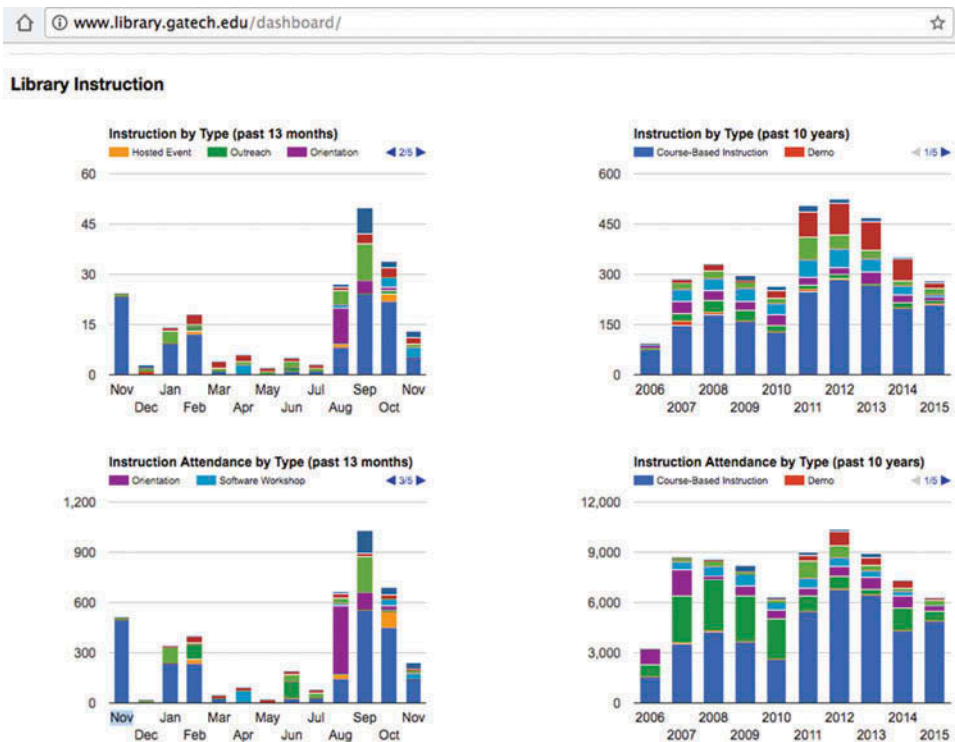


Figure 2. Example of a publicly available dashboard providing library program statistics.

They began their open data journey philosophically—understanding that their role includes providing and curating local government data that their constituents technically “own” and they recognized that constituents also had a right to library data as well.

They began by releasing data about library services to the public—locations, inquiries, material borrowing and circulation, and more—and then set out to inspire the local application developer community to harness this data by hosting “hackathons.” The hackathons were librarian-hosted one- or two-day events where software developers converged to focus on a single community goal. The first event helped them forge connections with local software developers and identify data-savvy allies. The librarians were able to tie the goals of their second open data hackathon to general digital literacy skill development. Likewise, they were able to establish connections with other librarians and data professionals interested in engaging data and data products for the improvement of libraries around the world.

Some of the hackathon projects were of direct relevance to library management, such as an interactive dashboard displaying data and trends about important library metrics like visitors and circulation. Many of the projects, however, were relevant to the general public or were simply the result of a developer with a particular technological “itch that needed scratching,” to borrow a phrase from the open source software community. For instance:

- (1) Developers worked on a project that would allow a patron to scan a book’s barcode at a kiosk and immediately hear an audio description of the book, based on its Google Books record.
- (2) “Georeferencing maps”—patrons were enlisted to help digitally “superimpose” historical maps over modern digital maps, vividly illustrating the degree of change in their community over the years, and generating a new combined data source for use by other amateur historians.
- (3) Statistical analysis of library data—this project highlighted the promise of releasing raw machine-readable data for use by others, in addition to human-readable, formatted summary reports.

One of the most promising examples of open data in the U.S. library sector is the “Open Data to Open Knowledge” project in the city of Boston, MA.<sup>9</sup> Supplemented by a Knight Foundation grant, Boston is aggressively moving to open civic data in all departments, with strategic guidance and on-the-ground assistance coordinated by the city’s public libraries. Dozens of staff throughout the city’s many departments were named departmental “data coordinators” and then convened at a local library for a civic data summit. This project also aims to jointly develop

training for librarians that will allow them to “democratize access to data”—that is, to make data available to the general public to use freely, instead of keeping it in “siloes.” Although this project is just getting started, examples of local open data use are already being shared, such as a professor who analyzes “311” government service call data for important research, and a local map app that incorporated newly-released city bike lane maps into their online GPS system.

Other interesting library-based open data repositories or open-data-based applications are as follows:

- (1) The New York Public Library sponsors numerous applications and tools involving the use of open data, as well as publishing their own data. One of their most interesting tools in development is the NYC Space/Time Directory,<sup>10</sup> a combination of online maps (historical and modern), historical collections and multimedia, and discovery tools that integrate with NYPL’s existing collections. The tool is intended to provide a complete picture of the entire city’s history, visible through the lens of the library’s incredible historical collections.
- (2) The Newberry Library in Chicago worked with historians and geographers to create the Atlas of Historical County Borders,<sup>11</sup> an online atlas and data repository containing the maps of every U.S. state, territory, county, and parish border change from the colonial era to today. While visitors use the site’s online maps to see the history of their region, historians and genealogists can download the raw data to perform their own research and generate their own custom maps. (Disclosure: the authors assisted with the technical implementation of this site.)
- (3) The Digital Public Library of America is a collaborative project aiming, among other goals, to serve as a portal to all of the digitized works collected by U.S. libraries. With such a focus on digital collaboration, it is no surprise that DPLA maintains an extensive set of open data downloads, system integration interfaces (also known as Application Programming Interfaces, or “APIs”), and other tools aimed at helping developers and librarians collaborate on digital projects.<sup>12</sup>

## Getting started with open data

If we have piqued your interest in open data, in what follows we present some ideas for how to get started.

- (1) Familiarize your staff with data.gov and your state’s open data portal, if any. Visit <https://www.data.gov/open-gov/to> find a list of state,



county, and local open data portals. If your county or local is not listed, try searching for them (e.g., “‘your city’ open data”); newer portals may not have been added to the federal list.

- (2) Get familiar with some actual data. We recommend starting with library data since it will be immediately relevant to many librarians. For instance, the federal data.gov site contains interesting data from the Chicago Public Library; the following are just a few examples:

- Circulation by Location (by year, 2011–2016)
- Computer Sessions by Location (by year, 2011–2015)
- Popular Fiction (Jan. 2014)
- Popular NonFiction (Jan. 2014)
- Popular Teen Titles (Jan. 2014)
- Popular Kid Titles (Jan. 2014)
- WiFi usage (by year, 2011–2016)
- Visits by location (by year, 2011–2016)
- FOIA Requests received by CPL as of May 1, 2010

Note that this is the kind of data your library could publish as well, as discussed in what follows.

Check out other types of data available at the open data portals you find. Usually you can search for data by geographic area or by specific topics so you can find data relevant to a patron’s interest. For example, you may find rainfall data for a student’s science project, map data for an amateur geographer, or regional economic data for a budding entrepreneur.

- (3) Evaluate what information about the library should be published as open data. As noted in the example above, libraries collect a wealth of data that could be published as open data to the community. This is a great way to immerse your library in the open data ecosystem, and helps provide familiarity for future data-related projects and collaborations. You can start small—even a few spreadsheets of circulation and anonymized borrowing data posted in your website’s “About” section can lead the way to more ambitious data publishing later.

You might be wondering, “Why would anyone be interested in our circulation numbers?” One of the most compelling reasons is that we simply never know what great idea our citizens, our government partners, fellow librarians, and our software developers can come up with until we publish the data for them to use. Contributing your data to the open data commons puts it out there to be mashed up, combined, or otherwise integrated into services and applications that may never have crossed your mind. It can also function as an excellent marketing tool.

As Alex Carruthers of the Edmonton Public Library stated:

Were a library to collect and analyze its internal data and integrate it with publicly available data, it could improve the efficiency of workflows and provide evidence-based support for program development. Sharing library data such as in-branch technology, usage, anonymized circulation statistics, and catalogue metadata improves the organization's transparency and can provide citizens with insight into the value of the library.<sup>13</sup>

- (4) Review your local and county websites, bookmarking any data that might be beneficial to patrons. Even if your local city or county authorities have not launched open data portals, there might be valuable open data lurking on departmental pages. For example, the King County (WA) government publishes almost 300 data sets and information documents on a wide array of topics, including the following ones:

- Election results by district (and lots of other election info)
- Animal control calls
- Percent with unmet need by ethnicity
- County Community Service Locations
- Public Health Clinic Locations
- King County History Timeline
- Metro Service and Community Shuttle Schedules
- Recycling Options
- Flu Clinic Schedules
- Maps and GIS Data
- Budget and contract data (several)
- Weather Related Closure Updates
- Metro Bus Stops

All of this data is useful to someone out there and likely someone in your library as well.

- (5) Consider adding a “finding government data” section to your existing digital literacy curriculum. Many libraries already provide a list of approved links or bookmarks available by topic. For instance, a prospective entrepreneur would be shown links to the local chamber of commerce, information about registering a business in your state, and a link to the federal Small Business Administration information website. Consider enhancing this material by helping your visitors utilize open data.

Take advantage of the work of other libraries are doing on this topic. For example, one of the City of Boston/Knight Foundation project deliverables includes developing materials for helping librarians foster open data in their communities.<sup>14</sup> The California State Library is leading a multistate initiative to develop open data curricula for librarians and libraries,<sup>15</sup> also funded by the Knight Foundation. When these projects are complete, every library will have the ability to integrate open data stewardship and literacy into their service portfolios.

- (6) Consider hosting a “hackathon” to engage local technologists. As mentioned earlier, Newcastle had great success engaging the community by hosting a hackathon centered around the development and use of local open data, including library data. Service dashboards, engaging digital maps, historical analysis, and more—the possibilities of a hackathon are as varied as the data you supply to your local hackers. For help, check out Joshua Tauberer’s “Hackathon Guide”<sup>16</sup> which is based on his experience running numerous successful open data hackathons.
- (7) Engage and encourage your library to transform into a platform. Now that you have created a nascent open data initiative, take it to the next level and transform your library into a “platform,” a new paradigm many see as the future of the modern library. Just as the Internet is a “platform” of data, services, and tools upon which applications and businesses are constructed, many librarians are positioning their libraries as “platforms” for the community. As David Weinberger writes as follows:

A library platform would be about developing knowledge and community, not primarily for developing software ... physical libraries already offer services, data, and tools, just as software platforms do. The data are the books, magazines, DVDs, etc. The tools include electronic catalogs for finding works, and step ladders for reaching the high shelves. The services include the expertise of reference librarians, and the work done behind the scenes by, for example, the collection development team and the cataloguers. On top of this “platform” are built itineraries for family trips, genealogies, homework assignments, and happy summer afternoons reading Elmore Leonard.<sup>17</sup>

What does this mean in the real world? The possibilities are endless but some inspiring ideas can be found in The Aspen Institute’s 2014 report on public libraries.<sup>18</sup>

## Conclusion

As librarians, we know that data is distilled into information; information is distilled into knowledge; and knowledge drives better decision-making. When librarians help patrons, students, business owners, community members, and public servants harness information, we are helping them make better decisions in their communities, business, and lives. And librarians can benefit from this data focus ourselves: whether we are analyzing our libraries’ data as part of our internal management processes or witnessing the insights made by enterprising data-savvy constituents, it is important to make sure that our data is driving our own success.

And most importantly, in a time with increasing threats to information and media literacy, where public data is being removed from the public

sphere,<sup>19</sup> it is more important than ever to ensure that government-funded and government-derived data remains data “of the people, by the people, and for the people.”

## Notes

1. Open Definition is a community-based project intended to put forward principles that define “openness” in relation to data and content. For more information, see <http://opendefinition.org/>.
2. Although the data.gov website has a link to the Open Data Policy, the link is currently dead. However, the original document establishing the U.S. Data Policy arising from that Executive Order is available at <https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>.
3. The White House, Office of Press Secretary (May 9, 2013). Executive Order-Making Open and Machine Readable the New Default for Government Information. Available from <https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government->. Accessed February 24, 2017.
4. There were almost twice as many datasets on data.gov in 2012. There are several efforts underway to download and preserve datasets that are considered at risk such as data related to climate change, social justice and the environment. See Government Data At Risk (<https://datapub.cdlib.org/2017/02/09/government-data-at-risk/>) for more info.
5. For a list of state open data portals, see <https://www.data.gov/open-gov/>.
6. <https://project-open-data.cio.gov/>
7. Fretwell, <https://fcw.com/articles/2014/06/09/exec-tech-brief-history-of-open-data.aspx>
8. <https://www.cilip.org.uk/blog/engaging-citizens-data-belongs-them>
9. <https://www.knightfoundation.org/articles/how-city-boston-making-its-data-accessible-everyone>
10. See <http://spacetime.nypl.org/>
11. See <http://publications.newberry.org/ahcbp>
12. See <https://dp.la/info/developers/>
13. Carruthers, A. 2014. “Open Data Day Hackathon 2014 at Edmonton Public Library.” *Partnership: The Canadian Journal of Library and Information Practice and Research* 9 (2). <https://journal.lib.uoguelph.ca/index.php/perj/article/view/3121>
14. <https://www.boston.gov/innovation-and-technology/open-data-open-knowledge>
15. <http://lj.libraryjournal.com/2016/02/digital-resources/knight-news-challenge-winners-to-develop-open-data-training-program/>
16. <https://hackathon.guide/>
17. [http://lj.libraryjournal.com/2012/09/future-of-libraries/by-david-weinberger/#\\_](http://lj.libraryjournal.com/2012/09/future-of-libraries/by-david-weinberger/#_)
18. <http://csreports.aspeninstitute.org/Dialogue-on-Public-Libraries/2014/report/details/0087/Libraries>
19. <http://thememoryhole2.org/about/>

## Notes on contributors

**Lori Bowen Ayre** is the Principal Consultant at The Galecia Group, a library technology consulting firm. She is nationally recognized as an expert in automated materials handling, RFID, library workflow optimization, and open source library system software. Recognizing

that technology is a catalyst of change, she works to help libraries harness the power of technology as they navigate the changing needs of their communities.

***Jim Craner*** is the Development & Operations Manager for The Galecia Group. A former Code for America Fellow and Americorps volunteer, he has extensive experience building websites and mapping applications for libraries and nonprofit organizations.