

Content Curation for Research: A Framework for Building a “Data Museum”

San Cannon
Federal Reserve Bank of Kansas City

Abstract

In the current digital age, data are everywhere and are continually being created, collected and otherwise captured by a range of users for a variety of applications. Curating digital content is a growing concern both for business users and academic researchers. Selecting, collecting, preserving and archiving digital assets, especially research data sets, are important steps in the research life cycle and can help expand the boundaries of research by allowing data to be reused. Creating research data sets often starts with selecting input data sources; in this age of new or “big” data, that choice set keeps expanding, thereby making it more difficult and time consuming to discover and understand the vast data landscape when beginning an empirical research project.

This paper proposes an approach to make finding and learning about data easier and less time-consuming for researchers. While cognizant of the role of digital curation for research data sets, we focus on the traditional “museum” definition of curation to outline how data-oriented content curation can support research. The process of selecting, evaluating and presenting information about potential data inputs can help researchers more easily understand how certain data sets are used and better determine which data sources might be fit for their purposes. Although the paper draws on examples from economics citing U.S. data, the techniques could be used across disciplines and countries.

Received 9 February 2015 | Revision received 3 April 2015 | Accepted 30 April 2015

Correspondence should be addressed to San Cannon, Federal Reserve Bank of Kansas City, 1 Memorial Drive, Kansas City, MO 64198. Email: sandra.cannon@kc.frb.org

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see <http://creativecommons.org/licenses/by/2.0/uk/>



Introduction

Content curation is not new. As long as there have been museums, curators have arranged and organized physical artifacts to tell a cogent story. In the digital realm, content curation – defined as “the act of discovering, gathering, and presenting digital content that surrounds specific subject matter” – has been the basis of marketing strategies for years (Mullen, 2011). Entire e-publications and software platforms are devoted to “sorting through a large amount of web content to find the best, most meaningful bits and presenting these in an organized, valuable way” (Lee, 2014). And as the tidal wave of digital information threatens to overwhelm consumers of digital content, curation is getting more and more press (e.g., Collis, 2014).

But creating RSS feeds, blog posts, and newsletters on popular topics are not the only ways to curate content. Nor is popular media the only sphere in which we are drowning in information. The academic world, especially research areas focused on empirical or analytical work, is dealing with a similar issue: there is now more information, of varying quality and pedigree, than traditional approaches and techniques can readily process. In many academic realms, simply the increase in the number of data sources has been astounding and presents both benefits and challenges. Researchers regularly struggle to identify what data are available, which data are best to address a particular topic, and how to process or analyze these data when they are found. Although these challenges aren't new to the empirical research landscape, their solutions are vastly different in this world of terabytes, petabytes, and exabytes of digital information.

Researchers in economics and finance have slowly been adapting to this new landscape. Although theoretical papers dominated economic journals until the mid-1980s, a recent survey of articles (Hammermesh, 2013) found that empirical papers now account for 70% of those published in top economics journals. But how do economists learn about data sources and how to use them? Often, they sift through pages of documentation, websites, and other journal articles and try to make sense of “what the literature says” about not only their topic but the data sources used to research it.

In most instances, this work is done by each researcher, for each data source, each time the data are used, because no one has yet effectively curated the content for that topic or data set. A well-trained curator, however, could “filter a lot of the less important content and allow quality material to surface” (Flintoff et al., 2014). In this paper, we propose building a “data museum,” a framework for scholarly curation that staff at the Federal Reserve Bank of Kansas City are currently undertaking.

Data are Scholarly Digital Content

In the simplest sense, data are just another form of digital content. However, data are unlikely to be included in curation aimed at growing a brand or marketing a product. Instead, data support more rigorous academic analysis and thus deserve more scholarly treatment. Given that, museums and the museum definition of curation provide a helpful analogy. When a museum curator wants to put on an exhibit on a particular topic – say, King Tut – their goal is to provide relevant information and artifacts so that visitors to the exhibit can learn as much about that topic as possible. The artifacts may come from a variety of sources and take a range of forms: King Tut's tomb, hieroglyphics from the area and the era, poems, portraits, and scholarly work. To build this exhibit or room in

the museum, the curator must consult with foremost experts in this area, as it is unlikely he or she would be an expert in all areas of interest to the museum. Some exhibits are permanent, and the curator would need to be continually on the lookout for new or updated information or artifacts to add. Others may be temporary, and after a set time period or after initial interest has waned, the exhibit may be appropriately cataloged and archived or dispersed in order to make room for a new exhibit on an emerging topic.

Curating content for researchers is not unlike curating a museum exhibit. Although the “rooms” or “exhibits” might be digital, they could showcase premier research on a topic, highlight the development of thinking over history in a particular area, emphasize data sets typically or traditionally used for research in an area, identify methodologies used in that research, and provide code or tools useful for researchers in that field. A room dedicated to labor economics, for example, could highlight papers on wage inequality, unemployment and search, or any other area of the field. It could provide information on both the macro- and microdata available in this area and show how other researchers have used them. Furthermore, the room could provide algorithms, logic, or code to assist with common data problems, such as the challenges of working with complicated data collections. As new research evolves, the room could highlight new methodologies, theories, or data sources. For researchers looking for cutting edge opportunities, the exhibit could also provide insight into what types of research have *not* yet been carried out.

The Museum Floor Plan

A vast compendium of work, scholarly and otherwise, has focused on how people find information. Library science and information architecture are just two of the academic disciplines devoted to the endeavor. Indeed, the explosion of digital content has changed the way that libraries define themselves (Borgman, 1999; Fuhr et al., 2007) and has influenced the direction of developments in information architecture (Resmini, 2014). Data and information are more abundant but arguably less organized than in the days before the ubiquity of the World Wide Web and search engines. Luker (2008) likens the state of information today to “Europe after the fall of the Roman Empire. It’s entirely in disarray: every little principality mints its own money, passes its own laws, speaks its own language, and has its own rituals.” A wealth of research into taxonomies, ontologies, search engine optimization and other areas is needed to help users find content.

The issue we face is how to find, evaluate, and disseminate information for which we already have an organizational structure. Institutional repositories and content aggregators generally have or locate a wealth of content that they need to organize around topics, concepts, or categories. The framework we are proposing collects relevant information for a topic that has already been identified. However, the collection mechanism needs to go beyond simple harvesting of digital content. After all, the museum curator does more than merely amass artifacts about King Tut. They consider the educational value, provenance, and informational worth of each potential contribution before deciding whether and how to best display that artifact. We are interested in applying those same considerations to data sources. To that end, our endeavor is more akin to the collection development task faced by libraries, which Johnson (2014) defines as “the *thoughtful* [emphasis added] process of developing or building a library collection in response to institutional priorities and community or user needs and interests.”

To extend the data museum analogy, a “floor plan” may be a helpful way to describe how content might be presented to users. In one respect, the floor plan merely provides the information architecture around the organization of data sets by topic: the economics building on the museum campus, for example, is likely to have a labor wing, a macroeconomics wing, and a public economics wing, among others¹. In each wing, a room or exhibit would be dedicated to the major data sets used for research in that area. For those who prefer to browse rather than search, this approach gives a good overview of how topics relate to each other and how data sets relate to those topics. When a particular data set supports research in multiple areas, some deliberation may be necessary to determine where that exhibit will be housed, making sure to provide links to that room from related parts of the museum.

In addition, a well-tuned search algorithm should be available throughout the site to provide assistance with finding specific information based on user-provided words or phrases.

The Guide to Exhibits

While there are guidelines to curating content for commercial purposes (Collis, 2014), we propose a new framework for making information about data available in an organized manner for research and academic purposes. The following information should be included in each exhibit.

Identification

The most fundamental information about a data set or data collection should be succinct text telling the user what the data are. Exhibits should include some core metadata elements to identify and describe each data set (e.g., Cannon, 2013). These metadata elements allow users to readily understand the basic information about a dataset and compare it to other data sets. These items are often used as the metadata for entries in a data catalog or inventory and where possible should be repurposed from such entries. Most standards for core elements would include items like data set name, creator, description, subject area or key words, and access information.

Beyond basic descriptive items, the exhibit should include more detailed information about the data set, its construction, and user experience descriptions. This part of the data exhibit should include brief, plain language explanations of some of the following points:

- **Purpose of the data set:** What was the reasoning or intent for which these data were captured, compiled, collected, or created? Transaction-level credit card data, for example, may seem useful for personal consumption research, but the inferences one can draw from those data may be limited since they were not collected with that purpose in mind. Understanding the questions the data were meant to answer can help researchers identify whether or not the data might be applied meaningfully to a different question. This will also help inform researchers about the suitability of particular data sets to be combined with other data sets.

¹ In fact, a logical organization for the economics museum would be to follow the Journal of Economic Literature (JEL) codes that are already widely used in the profession. See the American Economic Association’s JEL classification codes guide at: <https://www.aeaweb.org/jel/guide/jel.php>

- **Description of the collection mechanism:** Are the data recorded responses to a survey instrument, administrative entries from an information system, or sensor data from scientific equipment? Knowing the collection mechanism can help researchers understand the data quality, likelihood of revision, and other dimensions of the data to help decide if they are fit for a particular purpose.
- **Terms and conditions of use:** What restrictions have been placed on the data? As data continue to become commoditized, contractual obligations or conditions may apply even to data for which there is no fee. “Free” data are rarely without cost, and any restrictions on their use should be clearly expressed so that researchers can consider whether the restrictions affect the data’s fitness for their purpose (Cannon, 2010).

Besides the information outlined above, any data exhibit must provide links to data documentation such as the study description, survey codebook, collection template, variable definitions, and data dictionary. These need not reside on the servers that house the museum, but links to both official documentation and secondary usage guides should be provided and maintained to make sure that researchers can make the most complete and appropriate use of the data.

Beyond the official documentation and curated user guides, a good room in the data museum will have a place for users to add their own notes, tips, and tricks for using these data. In our social and collaborative digital environment, these items are likely to already exist in blog posts, Twitter feeds, LinkedIn boards, and other social media sites. A good curator can find this information and include it in a wiki or other editable site where collaborative curation can readily occur. For example, there may be good reasons why a particular data series or official statistic is not available for a particular time period, and this information may or may not be readily available in official documentation. Longtime users of those data can help new users understand these issues so there are fewer questions about the quality of the data series. Users might think of these entries as a “guest book” for the room: as is often done for exhibits in a museum of physical artifacts, the data museum can collect suggestions and hints from visitors.

Indeed, a “dummies guide” for each data set may spare a great deal of pain for new researchers and may be relatively easy to compile from existing sources; curators can provide easy, one-page “Guide to the exhibit” summaries to make it simple for first-time visitors to navigate the collection.

Relevance

Beyond the fundamental information about what a dataset is and how it came to be, the data museum can present information about how the data are used. Most new users will investigate an exhibit about a data set because they saw it referenced in other research, received a recommendation to use that particular set from a knowledgeable librarian or colleague, or found it in an Internet search. In the first two scenarios, users will likely have some notion of at least one or two applications for the data in question. For the uninitiated, however, general search results may not provide examples relevant to their particular case. If users know the name of the data set, a simple internet search should bring up pages of links to documentation on the data set and, hopefully, information on how to access the data before presenting results identifying papers or projects that use it. Of course, there are ways to search for papers directly and these methods will turn up tens of thousands of links to papers for major data sets. Continuing with the labor economics research example, it is not difficult to find papers that mention the Current

Population Survey (CPS – 80,000+), the Panel Study on Income Dynamics (PSID – 20,000+), the Survey of Income and Program Participation (SIPP – 12,000+), and the National Longitudinal Survey of Youth (NLSY – 19,000+). A novice user, however, would have to sift through potentially thousands of papers written *about* the data set to get to ones that *use* the data set in the right subject area. Although these are very commonly used data sets for labor economics research, they are also widely used for research in demography (CPS), medicine (CPS), child development (PSID, SIPP, NLSY), poverty (PSID, SIPP), and education (NLSY, SIPP) among others.

Searching for data by topic instead of by data set name has similar issues. A search for “labor market data” or “employment data” will lead to millions of hits; the primary listings are likely to be relevant for some questions but not for others. Need time series data from the Bureau of Labor Statistics? How about state-level data for New York? Many search algorithms will return those results first, but they aren’t very helpful for researchers looking for micro- or individual-level data. In fact, the aforementioned well-used, well-documented, individual- and household-level data sets are very frequently used for research in labor economics but none have any of the typical search terms (e.g., labor, employment, unemployment) in their names or the documentation that a search engine indexes. A researcher might need to page very far down the results before landing on anything that references one of these sources, if they are found at all.

New users visiting the data museum, however, could see a room for the CPS, a room for the SIPP, a room for the PSID, and a room for the NLSY. If they are strictly interested in the data set itself, locating the room should be sufficient to access the information they need. Topic browsers, on the other hand, could look over the floor plan to see that these rooms are in the labor economics wing. A content curator would be able to identify and appropriately arrange both scholarly papers about the data set and scholarly articles that use the data set.

An important part of the curation service indicating how data are used is the development and maintenance of a “research paper index” or “data bibliography” database. Entries in such an index would link the input data sets with a paper and any published output data set created for or from that paper. Some commercial services purport to do this, but the results are far less than satisfactory even for the large data sets used as examples here. Other outlets also try to provide this service with some mixed results: both the Minnesota Population Center (MPC) and the Inter-university Consortium for Political and Social Research (ICPSR) provide bibliographies of research using their data. The MPC bibliography relies on author self-identification; the ICPSR describes their version as a “continuously-updated database of thousands of citations of works using data held in the ICPSR archive” (2014). These are both rich resources that can serve as a model for others or, as will be argued later, an important input to a collaborative service.

As with any such endeavor, there will be perceived shortcomings as the research paper index will not be able to fulfill all possible user wants. First, the breadth of content that holdings of any repository must cover means that sifting through the results for any particular search may still be overwhelming. For example, nearly 300 CPS files are in the ICPSR holdings. The content curator for the CPS room would understand what the files are and how they relate to each other, making the information in the ICPSR bibliography even more useful. Even providing a faceted search by topic could help users discern between economics and education papers that use the same data set. Second, the ICPSR bibliography is limited to research using the data sets in their repository. While this coverage may be extensive, it leaves out potential data sources not available through the ICPSR repository. This issue may seem to contradict the previous

concern about too much data – in the realm of economics and finance, however, the number of private or contracted data sets continues to grow and not all will be candidates for formal repository holdings. Indeed, a well-curated data museum could provide information and documentation about possible data sources that are contractually acquired without providing access to the data itself or the marketing materials. This would provide researchers less biased information on how to best spend scarce funds.

One major challenge is compiling and maintaining the information for the research paper index. When all research papers directly cite their data inputs, including digital object identifiers, then scraping the information from bibliographies will make the compilation of index inputs less onerous. Until then, it still takes a human with some expertise in a particular subject area or data source to parse through the papers identified by a cursory search and determine if a particular reference to a data set actually indicates that the data were used for that paper's analysis. Indeed, many references may simply mention why a particular data set was *not* used rather than discuss how it *was* used. In addition, the data descriptions in many academic articles are woefully inadequate for easily identifying input data sources. There may be only a sentence or two saying the author was using data from a large, multi-program agency like the Census Bureau or a commercial data service like Bloomberg which aggregates over an even larger number of data sources. A curator familiar with the data and/or literature that use them would contribute importantly to this body of knowledge.

Access and Usage

Once prospective researchers determine if a particular data source is fit for their purpose, they need to actually understand how to gain access to those data. For public data, this may seem very easy: just go to a web page and download the data. If things were so easy, there would be no room for the dozens of commercial data vendors whose entire business model relies on the disparate and inadequate methods of accessing even data in the public domain.

Ideally, the data museum would provide easy access to the data themselves from within the exhibit. However, this may not be feasible or even desirable in many cases. If curators can't provide a front door to a data store, they can easily help identify the various options available. The PSID, for example, has a fairly centralized distribution: it is run by the Institute for Social Research at the University of Michigan, and nearly all links to download the data point to the website there. For data sets like the decennial Census or the CPS, a user can download the data from a multitude of sites. Some organizations, such as the National Bureau of Economic Research (NBER), host the raw data files as well as supporting documentation and some computer code to help users handle and parse the files. Others, such as the MPC, create databases of the CPS variables to allow for custom downloads. In the latter case, the MPC states that the data are "integrated over time and across samples by assigning uniform codes to variables" which means the data may not be exactly as reported and assumptions may have been made to bridge across survey changes (2014). While this may be desirable in many instances, a novice user may find it challenging to understand. Even if the CPS room in the museum could not provide access to the data, it should, at a minimum, be able to explain the pros and cons of using data retrieved from a particular site, should multiple copies be available.

Supplementary Information

Researchers may also wish to understand the relationship between data sets. When a researcher is trying to decide which data set to use to answer a particular question, understanding how two data sets are similar or different can be incredibly useful. Placing a room within a certain topic wing of the museum can help highlight which data sets might be related but not what those relationships are. For example, two data sets might have a derivative or lineage relationship. The microdata collected for the CPS are used by the Bureau of Labor Statistics to calculate the official unemployment statistics. It may be helpful for new users to know that one is used to create the other.

In some cases, multiple measures may be available for the same concept. Researchers looking for data on income, for example, would find different measures of income available from government sources. They would find that the Bureau of Economic Analysis publishes a measure of personal income, the Census Bureau publishes data on income from several different sources, the Statistics of Income division of the Internal Revenue Service publishes income data from tax returns, and the SIPP, CPS, and PSID provide income data at the micro level.

Experienced researchers who have worked in a particular area for some time may understand these relationships and have less trouble figuring out which source fits their purpose best, but not everyone has the necessary experience for all domains. This is where a curator could provide the most help: for researchers who are new to a particular data source or concept. Maybe the research question at hand is not actually about income but needs an income measure as part of the analysis. The expert in one field who needs information from another would benefit greatly from understanding the relationships between measures and sources. In many cases, scholarly articles make bilateral comparisons; finding and presenting those for users may be a first step, possibly followed by additional documentation and commentary to tie bilateral comparisons together.

Breaking Ground

The aforementioned framework is currently being followed by the Center for the Advancement of Data and Research in Economics (CADRE) at the Federal Reserve Bank of Kansas City. The Center comprises research economists, library and curation specialists, and technology experts whose mission is to support, enhance, and advance data- and/or computationally-intensive research in economics. Laying the foundation for a data museum is part of that mission.

As a start for our museum, we are building an exhibit based on the Current Population Survey². The CPS room contains metadata identifying the data set by describing its core attributes. Additional information on the data set's purpose, collection mechanism, and terms and conditions are also provided, even though this particular data set is fairly well-documented and generally has no restrictions.

The exhibit compiles and presents links to various sources, both official information and third-party user guides. In addition, CADRE staff are collecting suggestions, tips, and tricks from their own work as well as what others have noted to help new users get started with this data set. Library staff have begun to build the research paper index for papers using the CPS – capturing metadata about the paper, data, methodology and links to other scholarly work in a searchable database. Building on the information available

² Current Population Survey: <http://www.kansascityfed.org/research/cps>

from other similar directories, the index extends previous work in this area and allows researchers to search links between papers and data sets in a way not previously possible. A program to expand the coverage of the data set for this and related data sets is currently being developed.

In addition, technology staff have loaded the monthly CPS data into a database and developed an interface to allow customized downloads. While variations of this service are available elsewhere (King et al., 2010; ICPSR, 2014), a unique facet of the CADRE implementation is the ability to use these data to replicate previously published research findings. Replication is not something researchers regularly undertake in empirical economics (Ioannidis and Doucouliagos, 2013) and the increase in the size and complexity of data is likely to exacerbate that shortcoming. Nevertheless, researchers using these data are invited to provide their code as part of the exhibit, highlighting certain examples to allow users to replicate their calculations. Not all calculations can be done interactively, but CADRE staff will try to make as many examples available as is practical.

Finally, the CPS room allows users of its data to connect and share their own tips and tricks. Suggestions for additional documentation or other information can be submitted to the curation staff who will verify the accuracy of the information and make it available to other visitors to the exhibit as appropriate. An interactive “guest book” is also being considered. All of the content and functionality of the CPS room, and the museum in general, will be publicly available at no cost to the user, although registration for some features may be required.

Completing the Museum

As with a physical museum, a single curator is unlikely to have the expertise to maintain all the exhibits personally. CADRE staff, therefore, are looking to encourage collaboration among subject matter experts to broaden the knowledge available through the museum. To provide comprehensive coverage, additional topic rooms will need contributions from additional experts who may be scattered across various academic institutions. Indeed, for some topics, the exhibit itself can be created from information collected for a particular need at a university or research institute where it is used for their regular work. This “adjunct collection” can be digitally linked to other like content and made more broadly discoverable as a room in the data museum. CADRE staff can then provide a place in the floor plan to allow access to this additional content.

The model of having different parts of the collection produced and maintained by different institutions or individuals is common for social media; adding the expertise of a curator, or group of curators, in content selection, maintenance and presentation is less so. “Content curation communities” exist in some spheres (Rotman et al., 2012) but generally not with an academic focus. Furthermore, we have seen none focused on data for research.

Building the more comprehensive research paper index can also be a community effort. While CADRE staff will continue to catalog information about the data for which they curate exhibits, the coverage will not be complete. To be fully comprehensive, the index would need contributions like those that might be done by other repositories (ICPSR, MPC, etc.) for their collections. Then this broader index could be augmented by additional entries found for data that are not stored in any of the participating repositories. Researchers can collect, and contribute, additional information on input

data sets and methodologies each time they do the obligatory “literature review” for a scholarly article. These notations may add a small amount of time to their preparatory work but could have great benefits for the research community.

Such a collaborative endeavor requires not only a central location in which to make these contributions but infrastructure to harvest entries stored elsewhere. CADRE has a suitable infrastructure and could easily provide a centralized storage location for entries created elsewhere. Some existing resources provide a central location for related content and offer access to a variety of self-provided and electronically harvested information about economic research³. These collections are created by aggregation rather than curation, and data are not included. That said, a review akin to the collection analysis librarians undertake (Johnson, 2014) could help transform a collaborative aggregation into a curated collection.

In the collaborative digital world around us, it would be efficient and effective to limit the number of platforms that host the floor plan or map of the museum, possibly to a single one, and to “crowd source” the curation of the various rooms to experts in that domain. This is not unlike the Encyclopedia of Life model where “citizen scientists” and subject matter experts contribute from across the globe to a central infrastructure (Rotman et al., 2012). There would need to be common agreement on the minimum information provided in an exhibit, and users would benefit from some common architecture for the content of each room. Governance and maintenance roles and schedules would need to be arranged as well. CADRE staff are working to develop the business processes and other workflow details to make collaboration straightforward.

In all, a small investment in organizational time and curation could provide vast benefits to the economics community and serve as a blueprint for other disciplines that could benefit from curated scholarly content. As sharing digital information and opinions on social media becomes second nature, so, too, might sharing data and research in an academic context. The framework outlined here is the foundation for a collection of information being built at the Federal Reserve Bank of Kansas City. We believe the data museum being built by CADRE could be the start of a richer, more collaborative collection and improve the quantitative research practice.

References

- Borgman, C. L. (1999). What are digital libraries? Competing visions. *Information Processing and Management*, 35, 227–243. doi:10.1016/S0306-4573(98)00059-4
- Cannon, S. (2010, June). *Copyright and “facts”: Issues in licensing and redistribution for social science data professionals*. Paper presented at the International Association of Social Science Information Service and Technology conference, Ithaca, NY.
- Cannon, S. (2013, March). *Defining “core” metadata: What are needed to make data discoverable?* Paper presented at the Federal CASIC Workshops, Washington, D.C.
- Collis, M. (2014, September 29). 6 Powerful tips to effective content curation. *The Huffington Post*. Retrieved from http://www.huffingtonpost.com/matthew-collis/6-powerful-tips-to-effect_b_5901856.html

³ For example, see the Research Papers in Economics (RePEc) website series: <http://repec.org/>

- Flintoff, K., Mellow, P., Clark, K. P. (2014). Digital curation: Opportunities for learning, teaching, research and professional development. *Transformative, Innovative and Engaging. Proceedings of the 23rd Annual Teaching Learning Forum*. Perth: The University of Western Australia. Retrieved from http://ctl.curtin.edu.au/professional_development/conferences/tlf/tlf2014/refereed/flintoff.pdf
- Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., ... & Sølvsberg, I. (2007). Evaluation of digital libraries. *International Journal on Digital Libraries*, 8(1), 21–38. doi:10.1007/s00799-007-0011-z
- Hamermesh, D. S. (2013). Six decades of top economics publishing: Who and how? *Journal of Economic Literature*, 51(1), 162–172. doi:10.1257/jel.51.1.162
- Inter-university Consortium of Political and Social Research (ICPSR). (2014). Find publications. Retrieved from the ICPSR Web site: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/citations/>
- Ioannidis, J., & Doucouliagos, C. (2013). What's to know about the credibility of empirical economics? *Journal of Economic Surveys*, 27(5), 997–1004. doi:10.1111/joes.12032
- Johnson, P. (2014). *Fundamentals of collection development and management* (3rd ed.). Chicago, IL: American Library Association.
- King, M., Ruggles, S., Alexander, J. T., Flood, S., Genadek, K., Schroeder, M. B., Trampe, B., & Vick, R. (2010). *Integrated public use microdata series, current population survey: Version 3.0* [Machine-readable database]. Minneapolis: University of Minnesota.
- Lee, K. (2014, May 5). The busy person's guide to content curation: A 3-step process for your blog, newsletter, or timeline [Web log post]. Retrieved from Buffer Social web log: <https://blog.bufferapp.com/guide-to-content-curation>
- Luker, K. (2008). *Salsa dancing into the social sciences: Research in an age of info-glut*. Cambridge, MA: Harvard University Press.
- Minnesota Population Center (MPC). (2014). Frequently asked questions (FAQ). Retrieved from the IPUMS-CPS web site: <https://cps.ipums.org/cps-action/faq#ques2>
- Mullen, E. (2011, November 30). What is content curation? *EContent*. Retrieved from <http://www.econtentmag.com/Articles/Resources/Defining-Econtent/What-is-Content-Curation-79167.htm>
- Resmini, A. (Ed.). (2014). *Reframing Information Architecture*. New York, NY: Springer.
- Rotman, D., Procita, K., Hansen, D., Sims Parr, C., Preece, J. (2012). Supporting content curation communities: The case of the Encyclopedia of Life. *Journal of the American Society for Information Science and Technology*, 63(6), 1092–1107. doi:10.1002/asi.22633