Data is everywhere these days.

Businesses, industries, governments, universities, scientists, consumers, and nonprofits are generating data at unprecedented levels and at an incredible pace.

We are not just talking about gigabytes of data. IBM reports that every day we create "2.5 quintillion bytes of data" (www-01.ibm.com/software/data/bigdata). In May 2012, Cisco forecast that "annual global IP traffic is forecast to be 1.3 zettabytes—(a zettabyte is equal to a sextillion bytes, or a trillion gigabytes)." Cisco also predicts that, between 2015 and 2016, global IP traffic will grow to more than 330 exabytes (http://investor.cisco.com/releasedetail.cfm?releaseid=678049). That's almost more data than the 2 previous years combined.

To put it another way, think about McKinsey Global Institute's statement, in its "Big Data: The Next Frontier for Innovation, Competition, and Productivity" report: "One exabyte of data is the equivalent of more than 4,000 times the information stored in the US Library of Congress" (www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation). Now that's a lot of data.

Big Data is a big deal, and it is a big opportunity for librarians and information professionals to play a role in the Big Data universe. Why? Librarians have the skills, the knowledge, and the service mentality to help our businesses, governments, universities, and nonprofits capitalize on all that Big Data has to offer.

DATA EXPLOSION

Why now for this data explosion? One important reason is the widespread accessibility, affordability, and availability of new digital devices that make access to the internet easy and relatively inexpensive. Billions of people and the billions of mobile phones, smartphones, tablets, computers, laptops, and other digital devices are able to access the internet and "contribute to the amount of big data available," as *The Economist* expressed in a 2010 special report, "Data, Data Everywhere" (www.economist.com/node/15557443).

The list of digital sources keeps growing. According to *The Economist*, the "trail of clicks that internet users leave behind from which value can be extracted—is becoming a mainstay of the internet economy." Consumers generate vast amounts of data every day through email, searching, browsing, blogging, tweeting, buying, sharing, and texting.

A Big Opportunity for Librarians

by Laura Gordon-Murnane

This "digital exhaust," data created as a byproduct of other activities, is contributing to the dramatic increase in digital data. Combine that data with the data from the growing number of embedded networked sensors in cars, highways, and transportation networks; green buildings and the smart grid; the retail and manufacturing of RFID tags that track inventory; and healthcare services; and you can see why data is exploding like never before.

It's not just increased amounts and types of data; it's also improved tools to store, aggregate, combine, analyze, and extract new insights. Put Big Data together with big analytics, and it becomes possible to spot business trends, uncover ways to prevent diseases, combat crime, add economic value, gain new insights in scientific research, and make government more transparent.

ENTERING THE ERA OF BIG DATA

Welcome to the era of Big Data. McKinsey refers to Big Data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze." Edd Dumbill from O'Reilly Media, Inc. defines it as "data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the strictures of your database architectures" (http://radar.oreilly.com/2012/01/what-is-big-data.html).

Public and private sectors are feeling the pressure to take advantage of all that Big Data has to offer—with the expectation that it will spur new innovations and new product opportunities, achieve cost savings and efficiencies, and use predictive analytics that will enable businesses to understand what their customers want now as well as in the future. In its report on Big Data, the McKinsey Global Institute states that mining big data for insights "can create significant value for the world economy, enhancing the productivity and competitiveness of companies and the public sector and creating substantial economic surplus for consumers."

Warnings have been issued—ignore Big Data at your peril. The days of "gut instinct" for executives in business, manufacturing, and government are in the decline—replaced by "data-driven" leadership. Economist Erik Brynjolfsson, the Schussel Family professor at MIT Sloan School of Management, reported that executives who used "data-driven decision-making" saw their businesses experience a

Recommended Reading on Big Data

Anderson, Janna and Lee Rainie, "The Future of Big Data," July 20, 2012 (http://pewinternet.org/Reports/2012/Future-of-Big-Data/Overview.aspx).

Avanade, "Global Survey: The Business Impact of Big Data," November 2010 (www.avanade.com/Documents/Research%20and%20Insights/Big%20Data%20Executive%20Summary%20FINAL%20SEOv.pdf).

Bailey, Charles W. Jr., *Digital Curation Bibliography: Preservation and Stewardship of Scholarly Works*, CreateSpace Independent Publishing Platform, June 7, 2012.

Brynjolfsson, Erik, Lorin M. Hitt, and Heekyung Hellen Kim, "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?" April 22, 2011 (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486).

"Data, Data Everywhere," A Special Report on Managing Information, *The Economist*, Feb. 25, 2010 (www.economist.com/node/15557443).

Dumbill, Edd, "What Is Big Data? An Introduction to the Big Data Landscape," O'Reilly Radar, Jan. 11, 2012 (http://radar.oreilly.com/2012/01/what-is-big-data.html).

Gold, Anna K., "Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians," *D-Lib Magazine*, Vol. 13, September/October 2007 (http://works.bepress.com/agold01/6) and Anna K. Gold, "Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries," *D-Lib Magazine*, Vol. 13, September/October 2007 (http://works.bepress.com/agold01/4).

Gold, Anna K., "Data Curation and Libraries: Short-Term Developments, Long-Term Prospects," *Office of the Dean (Library)*, 2010 (http://works.bepress.com/agold01/9).

Haendel, Melissa A., Nicole A. Vasilevsky, and Jacqueline A. Wirz, "Dealing With Data: A Case Study on Information and Data Management Literacy," *PLoS Biology* Vol. 10, Issue 5, 2012 (www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1001339).

Hswe, Patricia and Ann Holt, "Transforming Research Libraries: E-Science Guide for Research Libraries: The NSF Data Sharing Policy," Association of Research Libraries (www.arl.org/rtl/eresearch/escien/nsf/index.shtml).

Manyika, James, et al., "Big Data: The Next Frontier for Innovation, Competition, and Productivity," May 2011, McKinsey Global Institute (www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation).

Miles, Doug, "Big Data–Extracting Value From Your Digital Landfills," AIIM Market Intelligence, 2012 (www.aiim.org/pdfdocuments/IW_Big-Data_2012.pdf).

Reed, Daniel, "My Scientific Big Data Are Lonely," Communications of the ACM blog, June 4, 2012 (http://cacm.acm.org/blogs/blog-cacm/150102-my-scientific-big-data-are-lonely/fulltext).

Shah, Shvetank, Andrew Horne, and Jaime Capellá, "Good Data Won't Guarantee Good Decisions," *Harvard Business Review*, April 2012 (http://hbr.org/2012/04/good-data-wont-guarantee-good-decisions/ar/1).

Stanton, Jeffrey, Teach Data Science, companion website to *Introduction to Data Science* (http://jsresearch.net/groups/teachdatascience).

Stuart, David, *Facilitating Access to the Web of Data: A Guide for Librarians*, Facet Publishing, 2011.

Swan, Alma and Sheridan Brown, "The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs: Report to the JISC," July 2008 (www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataskillscareersfinalreport.pdf).

"What Is Big Data?" IBM, Bringing Big Data to the Enterprise (www-01.ibm.com/software/data/bigdata).

5%–6% "increase in their output and productivity" (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486).

LIBRARIAN ENGAGEMENT WITH BIG DATA

However, recognizing that you need to make "data-driven decision-making" the foundation of your company and actually making it happen are two different things. That's the second reason why librarians need to be engaged in Big Data.

Many businesses are finding it difficult to exploit fully the data they are collecting. In several recently published surveys, executives were asked what problems they faced in extracting the most value from the data they were collecting. Four key areas jump out that are relevant to librarians—poorly organized collections; poor search tools and lack of accessibility and findability of internal data sets; lack of awareness of available third-party data sets; and copyright and intellectual property issues.

Consulting company Avanade, Inc. reported that business executives felt overwhelmed by the amount of data their company was managing. In fact, more than one-

quarter of executives reported that they had "lost business because they couldn't access the right information" (www.avanade.com/Documents/Research%20and%20Insights/Big%20Data%20Executive%20Summary%20FINAL%20SEOv.pdf).

In another survey, AIIM (Association for Information and Image Management) reported that executives recognized that their data collections were poorly organized, and that made it much harder for analysts and knowledge workers to access, query, and extract insights from the data that sits on their own servers. Many executives felt it easier to "do things on the web than in the office" (www.aiim.org/pdfdocuments/IW_Big-Data_2012.pdf).

Executives recognize that data sets available via third parties (governments, nonprofits, research universities, and company APIs) could be linked to their own data sets to derive added value and insights, but only a fraction of businesses are taking advantage of this external data. In its Big Data report, McKinsey points to intellectual property and copyright issues that need to be addressed. The report argues that Big Data, to be effective in both creating value and enabling data sharing, needs a "more effective and efficient means of establishing intellectual property rights (e.g., copyright and patents) and adjudicating disputes."

In an April 2012 report, *Harvard Business Review* described these same findability problems of Big Data: "Reliable information exists, but it's hard to locate. Many organizations lack a coherent, accessible structure for the data they've collected. They're like libraries with no card catalog and no covers on their books. The rise of social media, new selling channels, and devices such as tablets and smartphones has made it even harder to manage analytic content. Fewer than 44% of employees say they know where to find the information they need for their day-to-day work" (http://hbr.org/2012/04/good-data-wont-guarantee-good-decisions/ar/1). Big Data is only useful if it can be identified and found, used, and reused today and tomorrow.

LIBRARIAN SKILLS AND TOOLS

Here's the third reason that librarians and information professionals need to be involved in Big Data. Librarians bring to the table a whole range of skills and tools that address the very problems executives have identified as those that keep them up at night.

What can we do? We facilitate and enable data discovery and retrieval; we maintain data quality; we add value to the data through cataloging, indexing, and metadata; and we "provide for re-use over time through activities including authentication, archiving, management, preservation, and representation." The iSchool at the University of Illinois–Urbana-Champaign defines these skills as data curation (www.lis.illinois.edu/academics/programs/ms/data_curation). Librarians have been providing these services for print data sources for decades, and I think we are well-positioned to show what we know and apply it to the present day challenges and issues facing companies in the age of Big Data.

David Stuart, Research Fellow at the Centre for e-Research at King's College London and author of *Facilitating Access to the Web of Data: A Guide for Librarians* (Facet Publishing, 2011), writes that Big Data has "huge economic, research, and social value, and library and information professionals across the different sectors have a pivotal role in making sure that this value is realized."

Stuart makes the point that "there has probably never been a better time to be a librarian or an information professional. We live in an information society with access to more information than ever before, and librarians have an important role to play if people are going to successfully avoid the much discussed information overload."

DATA CURATION AND MANAGEMENT

Academic and research libraries and librarians continue to provide the traditional services of warehousing journals, books, and other materials, but many have also expanded that role by providing scientists and engineers with data curation services. With the growth of digital data, beginning in the last decade of the 20th century and growing rapidly in the first decade of the 21st century, funding agencies such as the National Institutes of Health (NIH) and the National Science Foundation (NSF) recognized that their investment in scientific research was at risk if the digital data was not properly stored, organized, and made accessible now and for use in the future. This data could be lost if data management plans were not put in place and implemented.

To develop data sharing and management plans, since 2003 the NIH has required that all investigators who request more than $500,000 in direct costs must include a data-sharing plan (http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html; http://grants.nih.gov/grants/policy/data_sharing). In the spring of 2010, the National Science Foundation "announced that it would alter its data sharing policy to require data management plans (DMPs) in future grant proposals to the agency" (www.arl.org/rtl/eresearch/escien/nsf/index.shtml). To help scientists achieve this, libraries and librarians at research institutions across the country have been actively involved in creating a place for scientists who need help managing their data.

It should come as no surprise that "[l]ibrarians have increasingly become experts in data management because of their combined knowledge of new data sharing standards, information science, and the Semantic Web," as Melissa Haendel and her colleagues argue, in an article published in *PLoS Biology* (www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1001339).

Take a look at the data curation and data management services that are now being offered by Cornell University's Research Data Management Service Group (https://confluence.cornell.edu/display/rdmsgweb/Home); Georgia Tech's Data Curation Services (www.library.gatech.edu/about/

data_curation.php); the University of Oregon's Science Data Services (http://libweb.uoregon.edu/faculty/SciDataAudit.html?tab=4); and Purdue University's Distributed Data Curation Center (http://d2c2.lib.purdue.edu).

MAKING PROGRESS

Anna Gold, university librarian at California Polytechnic State University's Robert E. Kennedy Library, reports on the progress libraries and librarians have made in offering data curation tools and services to scientists at major academic institutions around the country. She concludes that "[t]he match between libraries' deep knowledge base and expertise in the areas of metadata specification and development, user education and outreach, and user needs analysis, is a good indicator that libraries are well-matched to undertake long term roles in supporting data curation" (http://works.bepress.com/agold01/9).

Haendel, et al., also state that "[l]ibraries have traditionally been the place to acquire information; now they have become the place to learn how to manage it. … Information literacy has always been a topic of interest to research librarians, and it is natural that their role is expanding to include topics surrounding data curation and access. "

The lessons that have been learned in understanding and working with scientists and engineers at research institutions can be applied to the needs of those organizations outside of academia. Companies that want to take advantage of Big Data have encountered problems that librarians are well-prepared to help resolve.

The *Harvard Business Review* article mentioned earlier says it best: "To create an environment in which employees get the help they need, companies must rethink the kinds of people they bring in as experts. … Instead of simply answering questions as they arise, people oriented data experts can provide informal, ongoing training to employees in departments outside their own, increasing the organization's overall Insight IQ."

Who better than librarians to fill this need? This is what we do. The article concluded that "companies that want to make better use of the data they gather should focus on two things: training workers to increase their data literacy and more efficiently incorporate information into decision making, and giving those workers the right tools." In their report to JISC on "The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs," Alma Swan and Sheridan Brown identify two important roles that libraries and librarians can play in data science "training researchers to be more data-aware [and] adopting a data archiving and preservation role" (www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataskillscareersfinalreport.pdf). Seems like a great fit.

HIRING DATA LIBRARIANS

So you want to be a data librarian? Industry is looking to hire "data scientists" who have the technical skills and expertise to manage huge, varied data sets. We also need data librarians who can help everyone who has data and wants it better organized, accessible, searchable, retrievable, and deliverable.

Several graduate programs offer data librarian tracks such as the University of Illinois–Urbana-Champaign, Syracuse University, University of Michigan, and UCLA. The Center for Informatics Research in Science and Scholarship in the Graduate School of Library and Information Science at the University of Illinois–Urbana-Champaign has created the Data Curation Curriculum Search database—a very useful database of library and iSchool programs that offer data curation courses, programs, and certificates (http://cirssweb.lis.illinois.edu/DCCourseScan1/index.html).

Professional associations such as the American Society for Information Science and Technology and the Science and Technology Section of the Association of College and Research Libraries have regular sessions on data curation. Another superb resource is Charles W. Bailey Jr.'s recently published *Digital Curation Bibliography,* which provides a literature review of the state of libraries' and librarians' participation in cyberscience and data science (http://digital-scholarship.org/rdcb/rdcb.htm). IASSIST (International Association for Social Science Information Services & Technology) has also put together a very helpful collection of data curation materials (www.iassistdata.org/resources/category/data-management-and-curation).

EMBRACING BIG DATA

Librarians are essential knowledge workers. To continue to demonstrate our value, we need to embrace all opportunities. Big Data, in all areas of business, manufacturing, industry, government, academia, and nonprofits, is the hot new sexy buzzword. However, given the importance of data, it will only continue to grow in importance, which means that librarians can help business leaders, government executives, and scientists make better decisions by providing the services that will ensure that their data is available today and in the future.

Librarians are like the super domestiques of competitive cycling. The winner of the yellow jersey receives all of the acclaim and press, but he knows that without his team members to protect him in the mountains, to keep him out of the wind, to bring him water and food, and to even give up their machines if he suffers a mechanical problem on the road, he would not be the winner of the Tour de France.

So too librarians. We will have your back; we will organize and add value to the data; and we will protect and safeguard the data now and tomorrow. Librarians are a big competitive advantage that no business should be without. Big Data is a big opportunity for librarians for sure.

Laura Gordon-Murnane *(lgmurnane@gmail.com) is an information professional and freelance writer.*

Comments? Email the editor (marydee@xmission.com).