Library Hi Tech News

Libraries, data and the fourth industrial revolution (Data Deluge Column)
Donna Ellen Frederick

Libraries, data and the fourth industrial revolution (Data Deluge Column)

Donna Ellen Frederick

The World Economic Forum held in Davos, Switzerland, in January 2016, brought together leaders from the areas of science and technology, business, health, education, government and other fields as well as representatives from the media. A key theme of the forum was what has come to be known as the "fourth industrial revolution". News reports and blog posts about the forum gave the impression that this new "revolution" would bring unprecedented advances in science and medicine as well as would hold the potential for a future dominated by intelligent robots and massive levels of unemployment. For example, on January 24, 2016, Elliot of The Guardian reported that the "Fourth Industrial Revolution brings promise and peril for humanity" (Elliot, 2016). Sensational headlines and sound bites are good at attracting attention but they are not very effective with regard to communicating what this revolution is about and what it could mean for our lives, communities, governments and our workplaces in the near and distant future. The snippets of information reported here and there give the impression that robots, artificial intelligence, cloud-based computing, big data and a combination of other technologies are gradually merging to create a new reality which has the potential for revolutionizing our way of life. This installment of the Data Deluge consists of an exploration of the fourth industrial revolution, what role libraries might play in this revolution and how our information environment could be forever changed.

The phrase "industrial revolution" is one that many readers undoubtedly encountered in their school history lessons and likely associate with a time-period lasting from the late eighteenth century to the mid-nineteenth century where small home-based industries gradually gave way to larger scale production in industrial shops. The type of technological change which set the revolution in motion was the invention of machines which could manufacture products faster and more efficiently than the home-based craftsperson. The revolution was strengthened by the increasing use of steam power in the operation of the new machinery. Some historians argue that there has been a succession of industrial revolutions since the mid-nineteenth century. Chandler (1990) introduced a distinct second stage of industrial development which eventually came to be known as the "second industrial revolution". The second revolution began around the middle of the nineteenth century and was characterized by full use of steam power in industrial operations and the building of railroads. These developments, both supported and required, increased production of steel. The resulting demand for energy sources lead to a gradual shift from reliance on traditional steam power to oil- and electricity-based industrial activities. It is also the age of the second industrial revolution which saw the development of "electrical communications". These communications originally took the form of the telegraph and eventually developed into the breadth of twentieth century telecommunication technologies. It is this telecommunications shift which is strongly associated with the "third industrial revolution". The idea of a third revolution being a key event of the twentieth century was made popular by Jeremy Rifkin in his 2011 book on the topic (Rifkin, 2011). Long before it was characterized as a third phase in the progress of world industrial development, the author of this column suggests that Western society recognized that there was an information technology, communications and energy-based revolution which began during the mid-twentieth century and hit a critical mass in the 1970s. This period of change and development led to today's "digital age".

As for the phrase "fourth industrial revolution", it appears to be a synonym for "Industry 4.0" which a web search reveals is most often associated with industrial developments in Germany and discussions at Hannover Messe (Hannover Industrial Fair) events as early as 2011 as well as the related high-tech strategies of the German government. In Germany, it appears that the most common term used is Industrie 4.0. Searches also reveal that the phrase "the fourth industrial revolution" is found on websites hosted by industry, governments outside of Germany, media and universities. Therefore, perhaps the growing use of this terminology is evidence that there is a growing belief that the world has begun to transition to a new phase of industrial development. A search of YouTube videos alone reveals over 20,000 hits for recent videos about the "fourth industrial revolution" or "Industry 4.0" and some of the video titles reveal the same sort of sensationalism as found in the Davos news reports. Titles include "Will humans become obsolete after the new industrial revolution?" and "Fourth Industrial Revolution tsunami warning [. . .] ". Given that the current patterns in industrial development are being described as an "industrial revolution" and given that industrial revolutions of

the past were times of economic and social upheaval, it is not surprising to find extreme-sounding predictions and warnings about the near future. While the previous three industrial revolutions were times where both occupations and "ways of life" became obsolete, the actual revolutions themselves are often viewed today as times of great opportunity and progress. It is not surprising to read or hear expressions of marvel at the possibilities of what the near future may bring, mixed in with anxiety about disruptive change and loss.

While the experience of the first few revolutions has faded from the collective memory of society, chances are that at least a few readers of this column will recall a time before smart phones, the World Wide Web or the presence of personal computers in every home and office. Over the past 45 years or so we have gradually seen the disappearance of once commonplace entities such as the office typing pool, the printed card catalogue, phone booths and, in most regions, the rooftop television antenna, to name a few. These things have disappeared slowly, but, for those who remember the time when they were still ubiquitous, we can understand the significant changes in our work and personal lives they represent. Accepting that perhaps we have just come through a third industrial revolution, readers may be asking the question, "exactly what is this fourth industrial revolution and what does it mean for me?" A closer look at the changes which make up the new revolution and examining them as disruptive technological changes may help to address these questions.

The fourth industrial revolution is considered to be the outcome of the convergence of a number of technologies in industrial operations. These technologies include, but are not limited to robotics, artificial intelligence, cloud computing, big data, linked data, 3D printing, biotechnology and the integration of technology with the human body. The latter is often characterized as a convergence of human body and machines where, for example, wearable technology may actually become embedded technology. While many of the sensational reports focus on the human-machine aspect of the revolution, web searches reveal that "Industry 4.0" and the "fourth industrial revolution" are sometimes described as an "internet of things". In reflecting upon all the author has read or watched lately about this topic, a more fitting description appears to be an "internet of the physical world" or of "real things" where physical things become part of the internet whether those "real things" are machines in factories, automobiles or parts of the human body. In reality, the new revolution appears to have two main streams. The first stream is what these new technologies can do to improve and augment the lives and bodies of human beings while the second stream has to do with creating new hi-tech industrial processes. Despite the appearance of these two different streams, the heart of what makes many of the inventions and innovations possible is the cluster of technologies which define the industrial revolution. It is essentially the same technologies, although implemented in a different way, that allow for a remote customer to design a customized product which can be 3D-printed, manufactured and delivered directly to the customer or to provide blind individuals visual information about their physical environment. The latter information can be provided to the individual because of technology which can access data stored on the web and use it in analyzing and interpreting visual information.

In addition to the human and industrial streams of the revolution, there is also a distinction between open and proprietary data. Proprietary data, or data generated and owned by a firm, is very common in discussions about Industry 4.0. For example, there are a number of YouTube videos which describe Industry 4.0 manufacturing processes at German auto manufacturing plants. A common thread in the videos is that the manufacturing process begins with a customer selecting customized options for a vehicle they wish to purchase. The data about the desired vehicle is sent to the factory which uses that data and combines it with their existing proprietary data about the manufacturing process to essentially create a program which will build the desired car. Workers in these new factories no longer handle tools but configure and tweak the programming of the manufacturing process to address the needs of the customer's specific design. The actual processes embedded in the programming are then carried out by robots and monitored by humans. Even in libraries, we can find less complex but otherwise similar processes. For example, some large academic libraries use automated robots in combination with RFID technology and data from the library's bibliographic and holdings records to retrieve from storage and deliver books that have been requested by users. While this application of technologies is newer for libraries, it is not on the cutting-edge for industry. It is an example of how libraries are already making use of the processes which are arising out of developments which are increasingly part of the new industrial revolution. It seems reasonable to expect that in the future, we will see an increasing use of Industry 4.0-style technologies and processes applied to the execution of routine library work and services for patrons. These applications will use the proprietary data of the technology vendors and library data which likely will include a combination or proprietary and publicly accessible data.

The process of migrating manufacturing from one model to another in response to the development of new technologies is not unlike what happened in the previous industrial revolutions. However, the "fourth industrial revolution" has a new quality that makes it different from the three which preceded it. This difference is the growing mass of non-proprietary or freely accessible data on the web. The central theory is that the growing mass of open data on the web covering topics

such as climate, health, traffic, library holdings and so on can be combined and reused as "big data" to uncover new patterns and connections which would otherwise go undetected by human beings. Researchers have been doing work along these lines with big data for a while so the idea that this is possible is more than hypothetical. As the mass of data grows and the standards for presenting and organizing it are developed and implemented, it appears that humanity may be presented with new opportunities to study and address long-standing problems or issues which were previously otherwise too large or complex to analyze and understand. There is a catch which readers may have already detected. The reality is that there can be a big difference between proprietary data and the mass of freely accessible data on the web. With proprietary data, the business or organization that has created and used it has an interest in making it usable for their purposes and keeping it up to date. Proprietary data sources tend to be small. It is easier to keep smaller data sets up-to-date and in order. Scalability is less likely to be an issue with small proprietary data sets as well. Considering a popular source of non-proprietary data, it is mind boggling to think of how many images must have been taken over the past fifteen years or so to create Google Earth data and how many must be taken each day to attempt to keep these images up-to-date. How will the world keep up with the growing mass of data? How will an entity like the web, which is essentially unsupervised, become the medium for at least some of the data which will largely control our industrial processes, our vehicles and even parts of our bodies? Concerns over the disorganization and lack of control on the web are not unlike concerns that the author heard nearly 25 years ago as the world's information gradually moved out of printed books and journals and onto the web, but the concern seems more urgent today because of the potential integration of data with the functioning of real-world entities. Could, for example, either a mistake or malicious change to data lead to car crashes or manufacturing defects? Given the complexity of systems and the increasing mass of data, will either intelligent machines or human beings be able to detect such mistakes or deviations and be able to differentiate between innocuous variances and those with serious implications? It seems reasonable that those who lived through the third industrial revolution will carry with them a certain level of distrust of information technology, and the possibility of the "computer glitch" will likely loom in the back of the minds of most. However, perhaps artificial intelligence of the future will provide humanity with some relief from these concerns. For the time being, science fiction stories about an out of control HAL 9000[1] computer or malfunctioning super-robots will undoubtedly play in the imagination of many. Regardless, it seems reasonable to assume that for the "fourth industrial revolution" to see full fruition, issues regarding data management and integrity on the open web will need to be addressed.

Moving from the topic of data to that of the role and function of libraries in the new revolution, it seems natural to wonder about what will happen to libraries in a new environment in which the distinction between the physical and cyber world is diminished or absent. As with the changes of the third industrial revolution, there will undoubtedly be those who will claim that in the future there will be no need for libraries because the information we need will be embedded in our environment and accessed through this new "internet of things". The last part of the sentence may have some validity. However, libraries have many functions and purposes. It is likely that in the future families will still take their children to story times and other events at public libraries, researchers will need to pour through archival materials which seldom see the light of day and will likely never be digitized and students will want a place to study or will seek research and writing assistance from librarians. These are just some of the functions which will likely continue into the future. However, much of what "libraries do" is not seen by the average person. As a metadata librarian, the author is very aware that few library workers are aware of the full extent and type of work she does. It is important to consider many of the data creation and information management roles which libraries have played over time and likely will play into the future. Recalling the concerns in the early 1990s about the disorganization of information on the web, libraries have proven to be a useful aid in overcoming some of the disorganization. For example, the Virtual International Authority File (www.viaf.org), which brings together authority data from national libraries and other organizations from around the world, has been critical in disambiguating persons, places and things with similar or identical names in Wikipedia. This is just one example of the many possibilities for what work the librarians do and the use of library data in the emerging information environment may provide real and powerful solutions for taming the "disorganization" factor for non-proprietary data on the web. Perhaps library data, with its mature and robust procedures and standards for creation, has the capacity for providing a baseline from which artificial intelligence can learn to detect critical errors in the growing mass of data on the web. If such a use for library data were to come into being, libraries and information science can be expected to grow in relevance and importance in the near future. Perhaps the wider population will still never see or understand what libraries are doing to create and maintain their data any more than the average person understands the electricity that flows into their homes or the WiFi signals that envelope the urban landscape. Yet, it is possible that in an "internet of things" library data could become a part of the infrastructure human beings rely on to carry out their everyday activities. This is just one way in which the role and function of libraries could change as part of a new industrial revolution. If the entirety of the roles and services libraries provide were

considered, there is no doubt that many more possibilities exist.

In conclusion to this discussion of the "fourth industrial revolution", it may be helpful to consider this revolution as part of a continuum of technological change and industrial development. With industrial revolutions, there is no place or day on which a battle is fought and a new regime assumes power. Essentially, way leads on to way, and soon people see patterns emerging which point to what some are now choosing to call a new industrial revolution. As with all predictions, we know that what is foreseen may or may not come to pass. We know that today's trends may grow in strength or they may fade away as unforeseen events or changes pushing progress in another direction. Understanding that we cannot predict the future does not invalidate the process of considering what may eventually become of the changes we observe and experience today. Just as history helps us to understand why things are as they are today, looking toward the future can help us to recognize opportunities when they present themselves to us. The author sees value in framing the progression of technological changes which have occurred over the past few centuries and impacted not only the industry but the nature of how humans live and interact with each other as a series of revolutions. If we consider our current age as the tail end of the third industrial revolution and that we are approaching a new era, we have a new lens through which we can analyze what is happening and could happen with library data. We can see that computing technology of the third industrial revolution allowed libraries to amass large amounts of data in the form of electronic bibliographic and authority records as well as collections of digitized resources. We can also see that limitations of technology and visions of purpose for that data, for most of the past 50 or so years, have largely kept the data siloed within the context of libraries. The past 25 years or so have seen a gradual breaking down of silos but the primary model upon which library data was created and used reinforced the presence of silos. Until relatively recently, library data were seen primarily as intended for use in library catalogues as a means for locating resources within library collections. Most librarians would still argue that this is the primary and ultimate purpose of library data. It was hard for many of us to envision bibliographic and authority data having any other purpose until recently. Most cataloguing and metadata librarians likely have stories about how their local MARC data had been "tweaked" or modified for local purposes over the years which, in turn, makes that data less interoperable than it needs to be for today's purposes. However, as library data or metadata gradually moves out of its silos, as the current linked data movement promotes that it should, are we starting to see the entrance of libraries and their data into this new "internet of things"? Is the use of library authority data in non-library environments such as Wikipedia just the tip of the iceberg for what we will see in the upcoming years? The author has already seen some experiments using bibliographic and holdings data from WorldCat as data sets for big data analyses such as OCLC's WorldCat Identities experiment (see: http://worldcat.org/identities/). In this experiment, data from 1xx and 7xx (author/creator) fields and 6xx (subject or topical) fields in MARC records is graphed to demonstrate the time periods during which a person or organization created content as opposed to when that person or organization was "written about". The graph also shows the relative rise and fall in the volume of publications during the active time period. This is not a purpose for which the MARC data were created but certainly this experiment is able to demonstrate that when MARC data are converted to linked data, it can successfully be used as big data to study a subject in a manner which may not have been otherwise possible. If we look at our current age as a time of transition and as the various types of data which libraries meticulously create and maintain as a treasure trove that is yet largely undiscovered within the wider world of information, we begin to see new possibilities and opportunities. Will the MARC record that a cataloguer dutifully struggled over in 2013 be part of the mass of information which drives our car or directs production in a factory in 2033? We do not know the answer to that question today, but it is reasonable to assume that we will be surprised by the many ways library data can and will be used in the future, regardless of whether or not Industry 4.0 or the fourth industrial revolution plays out in the way it is envisioned today. As with the previous revolution, we appear to be living in interesting times which are likely to get more interesting very soon. It does appear to be a good time to be a librarian.

NOTE

1 HAL 9000 or just HAL is a fictional computer from the movie *2001: A Space Odyssey* (1968). HAL had artificial intelligence and "malfunctioned". See a "biography" of HAL on the Internet Movie Database www.imdb.com/character/ch0002900/bio

REFERENCES

Chandler, A.D. Jr. (1990), *Scale and Scope: The Dynamics of Industrial Capitalism*, The Belknap Press of Harvard University Press, Harvard.

Elliot, J. (2016), "Fourth industrial revolution brings promise and peril for humanity", *The Guardian*, 24 January, available at: www.theguardian.com/business/economics-blog/2016/jan/24/4th-industrial-revolution-brings-promise-and-peril-for-humanity-technology-davos (accessed 20 April 2016).

Rifkin, J. (2011), *The Third Industrial Revolution; How Lateral Power is Transforming Energy, the Economy, and the World*, St Martin's Press, New York, NY.

Donna Ellen Frederick (*donna.frederick@usask.ca*) is Metadata Librarian at the University of Saskatchewan, Saskatoon, Canada.