Big Data: What Is It and How Can Academic Libraries Use It?

Rachel Hooper
Head of Reference
Troy University Library
hooperr@troy.edu

Abstract

Numerous articles have been written about innovation and technological advances throughout the years. Starting in the 21st century, big data has become a popular topic throughout the innovation literature (Richey et al., 2016). However, the extant literature on big data in academic libraries is not as extensive. This review will examine the extant literature and discuss the broad scope of innovation. Next, the focus will narrow down to a frontier technology, big data (United Nations, 2018). What big data is, examples of it, advantages, and concerns will be discussed before delving into the relationship between libraries and big data.

Keywords: innovation, big data, libraries, academic library, global leadership, leadership, frontier technologies

Throughout the past few decades, technology has grown exponentially, causing shockwaves across the world. As technology advances, so does the ability to create and access data. In addition to that, barriers are lowered with the less expensive storage options now available. Even as advanced as the technology now is, it is expected to continue to grow (United Nations, 2018). To put things in perspective, former CEO of Google, Eric Schmidt, was quoted as saying "there were five exabytes of information created between the dawn of civilization through 2003, but that much information is now created every two days" (Walker, 2015, p. 3). With the advent of advanced technology and the growth in data, opportunities to mine and use that data are immeasurable.

The purpose of this study is to evaluate the current literature on innovation, specifically big data and its potential impact on academic libraries. The primary databases used for this research include Emerald Insight and EBSCO Databases combined. These were chosen due to the coverage of the topic, as well as the international coverage. However, this search could have been expanded by using additional resources that were not available at the time. While some of the extant big data research focuses on libraries, a majority of it focuses on businesses. However, this information may be able to transfer to the library world in certain cases, so it has been included. By collecting this information and evaluating the literature, more detailed information may be found on how big data is being used, or can potentially be used, in the academic library context. This information can help librarians determine the best course of action for their libraries and implement the needed innovations.

Literature Review

Innovation is defined as "the fulfillment of a new service/product, a new marketing method, or an organizational practice" (Hung et al., 2021, p. 3). Because innovation can vary by

industry, Hung et al. (2021) provided two classifications of innovations: technical, which includes “new products that demonstrate inventiveness in engineering, applied science, and pure science”, and non-technical, which includes changes to “marketing, management, or organizational structure” (p. 3). Regardless of whether it is a product to sell or a service to offer, innovations are typically created to serve the needs of the customers (Jalonen, 2012; Rampersad, 2020). Although Jalonen (2012) stated that innovation is “a process that is fraught with uncertainty” (p. 1), organizations must find ways to overcome the uncertainty in order to innovate and differentiate from others.

Innovation is crucial for an organization’s growth and survival. In order to compete with other companies, organizations must innovate to create a competitive advantage, and it is an on-going process (Barsh et al., 2008; Koziol-Nadolna, 2020; Lisak et al., 2016). As with any type of business, it is also important for academic libraries to innovate and attract patrons (Wojcik, 2019). Innovation can take place throughout different levels of the organization. According to Katarzyna Koziol-Nadolna (2020), “innovation in different areas of a company’s activity improves productivity, efficiency and quality of work, thereby increasing the quality of products and their competitiveness and improving the overall efficiency and productivity of the company” (p. 1). As the world continues to change, the library must adapt and change to meet the needs of the patrons. If libraries are able to innovate and advance, they can gain a competitive advantage and potentially increase the usage of the library (Wojcik, 2019). In order to have the innovative culture in which employees may thrive, the organization must also have innovative leaders.

Leaders and Innovation

One may ask what role leaders play in this new innovative context. Overall, leaders greatly influence how innovative an organization may be (Koziol-Nadolna, 2020; Overall, 2015).

The father of management, Peter Drucker, even stressed the need for an organization to innovate (Overall, 2015). However, the former mindset of leadership that focused more on inspiring and motivating the employees is no longer enough; leaders must expand their duties and have more of an innovative mindset to properly lead their companies to cultivate the proper innovative culture (Koziol-Nadolna, 2020). Since, by definition, innovation is synonymous with uncertainty (Jalonen, 2012; Mitchell, 2018), leaders must embrace the unknown, be flexible, be creative, and be open to risk-taking (Koziol-Nadolna, 2020; Mitchell, 2018). Additional descriptors for innovation-oriented leaders include having a "long-term perspective, vision…initiating changes, encouraging diversity, [and using the] employee as a strategic resource" (Koziol-Nadolna, 2020, p. 4), all of which create a culture in which innovation can thrive.

One factor of innovative leadership is the creation of an innovative culture (Koziol-Nadolna, 2020; Overall, 2015). In this type of environment, employees are encouraged to innovate and are not punished for failure (Koziol-Nadolna, 2020). Rather, they are encouraged to learn from their failures and use that knowledge to improve future projects (Overall, 2015). Overall (2015) stated that "innovation is dependent on an organizational culture that incentivizes creativity" (p. 41). It has been suggested that leaders provide rewards in the form of raises or promotions for innovative employees. Another way is to allow the employee to take the reins of a project, providing an atmosphere of independence and creativity (Koziol-Nadolna, 2020). A study published in 2016 listed findings that revealed the top "three leadership dimensions that lead to an organization being considered innovative by its employees" (Dodge et al., 2017, p. 25). These included "providing organizational encouragement, ensuring challenging work, and fostering support within the work group" (Dodge et al., 2017, p. 25) These findings reemphasize the need for the former mindset of leaders, while also incorporating the new innovative aspects

(Koziol-Nadolna, 2020). In the academic library, this could include innovative programming ideas, finding ways to partner with other departments to improve certain aspects of the operation, or trying new ways to conduct instruction sessions or to reach students. If librarians are encouraged to try new ideas without the fear of punishment or failure, the innovative culture could flourish. While the creation of an innovative culture is imperative, other important factors influence organizational innovation.

Another area in which leaders should focus is on the creation of a multicultural team. The demands of this global environment have truly fast-tracked the essential need for global leaders (Hanges et al., 2016). According to Lisak et al. (2016), multicultural teams "benefit from the large pool of diverse knowledge, skills, and perspectives of their members" (p. 655) and that diversity "enables team members to gain new insights and create innovative solutions" (p. 655). Leaders of these global teams should embrace the diverse backgrounds and cultures, and also encourage their team members to use the variety of skills and knowledge to increase their creative potential (Lisak et al., 2016). This could include hosting more international-themed library programs or providing more international resources for the students. Another advantage of the multicultural team is the ability to better understand their customers (Lisak et al., 2016). If the team only consisted of people from one area, they would not necessarily have the knowledge needed to understand the needs of patrons from around the world. One may not consider an academic library in the United States from a global perspective, but it is something to consider. Large numbers of international students seek to study at higher education institutions in the United States. During the 2021-2022 academic year, almost 950,000 international students were studying in the United States (Open Doors, 2022). This shows that academic libraries could benefit from multicultural employees who may be able to relate well to those from different

cultures and help provide services or resources that may not have been used previously. In conclusion, many factors play into a successful, innovative library, all of which lead back to a successful innovative leader.

Big Data

Described by Carillo (2017) as the "next frontier for innovation" (p. 598), and a frontier technology (United Nations, 2018), big data has been a hot topic in the business world. Over the past decade, big data has created waves and is essentially changing the way the world works. Although there is not a single agreed-upon definition, and different sectors and industries conceptualize it and its usefulness in different terms, it is generally agreed that big data has the potential to positively impact organizations, as well as assist in global issues (Ahmed & Ameen, 2017; Richey et al., 2016; United Nations, 2018). Mayer-Schonberger and Cukier (2013) provided a definition that seemed to cover all areas: big data is "the ability of society to harness information in novel ways to produce useful insights or goods and services of significant value" (Mayer-Schonberger & Cukier, 2013, pp. 2-3). Then big data analysis, or BDA, occurs when data is evaluated to find connections that had not previously been discovered (Carillo, 2017; Richey et al., 2016). Although a common definition does not exist, many researchers agree on some of its characteristics.

Some common characteristics associated with big data include volume, velocity, variety, and variability. The volume refers to the vast amount of information, or data, that is available now, or the information that is able to be stored. As technology advances and storage increases, the volume of data that is produced has grown exponentially. The velocity refers to the speed at which data is being created, while the variety refers to what type of data is being created. Lastly, the variability of data refers to the inconsistency of that information. The amount of information,

as well as what type of information, cannot be guaranteed, so this is where the variability comes in, making big data more complex (Ahmed & Ameen, 2017; Anna & Mannan, 2020; Richey et al., 2016). A fifth characteristic is the value of big data, which encompasses the expenses related to storing the information as well as the profit from investing in big data (Ahmed & Ameen, 2017; Anna & Mannan, 2020).

Big data has transitioned the way in which we do business and has actually created new jobs (Carillo, 2017; United Nations, 2018). In the past, information was collected in small amounts, due to a lack of storage and limited capabilities. Now, with increasing technological abilities, data has grown on a much larger scale. Traditionally, research has been conducted using sampling (Mayer-Schonberger & Cukier, 2013). With the advent of big data, research can be much more exhaustive and comprehensive that sampling originally allowed (United Nations, 2018). “Translating the strategic potential of big data into actional insights takes many of the skills LIS professionals either have or can master” (Dority, 2022). Since data can be evaluated and used in libraries, data librarian positions are now becoming more prevalent. Some of the tasks that data librarians handle include collecting, processing, analyzing, communicating, and preserving the data (Dority, 2022). This next section will provide advantages and concerns with big data, and its impact on leadership. To conclude, the last section will focus on big data in libraries and its impact in the library environment.

Advantages of Big Data

Big data is present in various facets of everyday life, many of which people may not be aware (Klous & Wielaard, 2016; Mayer-Schonberger & Cukier, 2013). Some examples of these areas include trading on the stock market and insuring customers at banks or at insurance agencies (Klous & Wielaard, 2016). Big data provides numerous advantages for organizations,

including the opportunity for advanced innovative opportunities (United Nations, 2018). Big data is not specific to one industry or sector, but available to any area in which data can be mined, including the academic library.

An advantage of big data is to help managers and leaders by its ability to help make predictions, improving the decision-making area of libraries. By having data to back up choices, managers can make more informed decisions, hence reducing the associated risks (Richey et al., 2016; Simovic, 2018). In the past, choices may have been based upon emotions or relationships, which could have horrific results. Big data can change and improve that process. BDA can be particularly beneficial in the efficiency of a library's supply chain operations, as well as inventory management (Richey et al., 2016). For example, by using BDA, libraries have the potential to waste less money and eliminate excess (Richey et al., 2016).Although this is a good advantage of big data, even more exist.

Customizing the products for the customers is another way that big data can be advantageous. By "assess[ing] customer needs and perspectives" (Richey et al., 2016, p. 728), libraries can be better informed and help provide the wanted and needed items for the customers/patrons. This makes for a more patron-centered approach and helps eliminate some guessing. This use of big data "increases the ability to predict and provide for customer needs" (Richey et al., 2016, p. 729). As with any positives, negatives also exist.

Concerns with Big Data

"Big data erodes privacy and threatens freedom" (Mayer-Schonberger & Cukier, 2013, p. 163). While big data provides many advantages, many concerns also exist. Ahmed and Ameen (2017) created a mind map for big data, and some of the resulting terms, starting with the most relevant, included privacy, "risk, industry, market, and creditworthiness" (p. 23), all of which

deal with issues, concerns, or are facets of big data. One concern is that of errors, or creditworthiness: once the data has been saved, will it be analyzed correctly (Ahmed & Ameen, 2017; Klous & Wielaard, 2016)? Klous and Wielaard (2016) quoted big data critic Taleb as saying "Big Data means Big Errors" (p. 108). Will his prediction hold true that incorrect predictions or conclusions will be made with the data, making it less advantageous than previously stated? Data analysts must be cautious and remember "the difference between correlation and causation" (Klous & Wielaard, 2016, p. 109) when examining the data. Not only are there concerns with the analysis of the data, but also prior to that stage of analysis, will people "know how to use it and how to translate it, or even if it is usable" (Richey et al., 2016, p. 719)? Other aspects related to errors with data include that of imperfect information, inaccurate data (Graves et al., 2016) and unreliable information (Klous & Wielaard, 2016).

Due to the nature of the internet and the ability of users to add any information regardless of accuracy, the user must be cautious when relying on information from the internet. Not only can the information be incorrect, but the results from a search engine may be filtered based upon prior search terms one has used, as well as one's online activity. Even if two people were to use the exact same search terms in a search engine, their results could vary drastically. Although this provides more personalized results, this is concerning because the relevant information for which one is searching may be "filtered" out from the results. This filter bubble in which users find themselves may not show the entire picture (Klous & Wielaard, 2016). Not only is the filter bubble of concern, but also that of privacy and security (United Nations, 2018). This is something that academic librarians must be aware of and be prepared to handle as students conduct research. Librarians may need to provide instruction on the impact of big data and the filter bubble, or at least be prepared to answer research questions pertaining to this topic.

A common concern throughout the literature related to big data is that of people's privacy and security in the online environment (Mayer-Schonberger & Cukier, 2013; Richey et al., 2016; United Nations, 2018). A fine line exists between policies needed to protect people and the freedom to use the data for innovation (United Nations, 2018). The information that is saved, shared, and used should not invade anyone's privacy. Those who harvest the data must determine who can access it, as well as how to use it (Richey et al., 2016). Should patrons' circulation records be saved for data mining, or should that data be destroyed? In light of the 2001 Patriot Act, some people may feel that that information should not be accessible at all (Martin, 2013).

Since so much data is now available, storage has become an issue of concern. Once the decision has been made on what data to store and save for later use, the next question may be how to actually save that data. Related to the storage is the price of that storage. Is the infrastructure in place to save it, and, if so, how much will it cost (Richey et al., 2016; Simovic, 2018). These concerns listed within the literature are all factors that leaders must acknowledge and handle.

Big Data and Leadership

Many of the concerns with big data which are listed above tie in with the managerial and leadership roles. What data should be saved? How will it be saved? What information can be shared? If certain information is shared, are there any legal or moral implications? Is the data useful now, or will it be useful in the future? How can the information be used to increase the library's competitive advantage? Is the cost of harvesting and storing the data worth the end-result (Richey et al., 2016)?

Leading during this time of great innovation, specifically with big data, can be fraught with uncertainties. If leaders have not had the training, they may not know what to do with the

data. This is where innovative leadership applies. Innovative leaders support an innovative culture, including the implementation of big data. Even if leaders do not have relevant background in this area, they need to be open to innovation. Supporting an innovative culture can include adding people to the team who are well-versed in big data, or hiring data librarians. This can help the library push forward in those endeavors, providing insights and opportunities that may not have been recognized previously (Richey et al., 2016).

Culture also plays a role in how data is handled. In regards to legal issues with big data, different countries may have diverse regulations for data security. While this can help leaders and managers, they must be aware of these differences when working in the global environment. Culture also influences the general morals people have, affecting their decisions on how much information to disclose (Richey et al., 2016). If a higher education institution has locations/campuses with libraries in different countries, leaders need to be aware of the differences and modify their policies, if needed.

Big Data in Libraries

Libraries are not immune to the allure of big data and all that it can offer: it can actually have many positive impacts on libraries. One area in which it could help is the previously mentioned use of data to assess the needs of the customer. In the library context, this could mean examining what the patrons have borrowed from the library to help predict future wants. If a system could determine common themes, subjects, or authors a patron has read, this could help forecast what the patron may want to read in the future. The system could potentially place that future item on hold for that specific patron once it is available in the library. Also, assessing the needs of the library's patrons by examining materials that have previously been checked out could assist in collection development by automatically preselecting items that match the specific

interests of the patrons (Richey et al., 2016; Simovic, 2018). Simovic (2018) stated that "library catalogs need to carry enough information about items and users' preferences to have the capacity to determine a potentially ideal result and to respond adequately to the given query" (p. 498). While this implementation of big data could be useful, implementing additional resources could be even more beneficial.

Although data collected by a library's integrated library system (ILS) or vendor reports may already be used for certain purposes, taking this idea of personalized patron assistance a step further is the idea of a smart library. Aleksandar Simovic (2018) proposed the idea of a "Big Data recommender system [which] enables data integration from various sources (e.g. Learning Management Systems (LMS), University online bookstore, Internet of Things (IoT), data from social media networks, and traditional library) into a smart library" (p. 499). Which courses the students have taken and are currently taking would factor into the recommendations, as well as bookstore purchases and book or eBook usage from the library's system. This big data system would compile the data from various departments across a university campus in order to determine personalized services or resources for the students, predicting which ones would meet their needs. This system would be reminiscent of those used in marketing and advertising, as well as on social media, linking data to match users' interests. Since little research exists on the integration of a smart library, additional research needs to be conducted to reduce this gap in the literature (Simovic, 2018).

Compared to much of the literature, Ahmed and Ameen (2017) pose a different perspective when researching big data and libraries and focus more on the librarian than on the library itself. Since big data analytics is predicted to create even more jobs (Ahmed & Ameen, 2017; United Nations, 2018), the focus for hiring committees will be on those professionals who

are able to analyze data effectively, so “library and information management professionals are suitable candidates for these positions” (p. 22), thus providing additional job opportunities for librarians (Ahmed et al., 2019; Anna & Mannan, 2020). In order to stay up-to-date with those potential employment options, as well as in current librarian positions, librarians must continue to learn new skills and gain the knowledge that goes along with big data software and implementations (Ahmed & Ameen, 2017; Ahmed et al., 2019; Anna & Mannan, 2020).

Anna and Mannan (2020) conducted a review of the literature and focused on how many journal articles have been published with the topics of “big data in the library” (p. 4). Their research concluded that research had been conducted on the topic, but several gaps in the literature were exposed, calling for more future research to help fill those gaps. A majority of the research focused on the academic library and ways to integrate big data in that context, with a very limited number focusing on the public library and big data. While the lack of research in this area may be disconcerting, it can also be exciting to see the areas in which additional research may be conducted and what needs to be done (Anna & Mannan, 2020).

Critique and Implications

The literature reveals that, while big data may have its drawbacks, many advantages exist. With the careful consideration of what data to use and how to use it, endless possibilities abound. It is up to the individual libraries to determine the best way to innovate using big data, which can then lead to a competitive advantage. While this may improve products in some industries, the implementation of big data in others can improve and increase services to the customers, as is the case in libraries.

Suggestions for Future Research

Overall, an abundance of literature exists on the topics of innovation and big data, but research specific to big data in academic libraries is less extensive (Anna & Mannan, 2020). To help fill the gaps, researchers need to focus on the academic library setting and the implementation of a smart library. This integration of resources across a university campus can help determine the factors that are beneficial to personalizing resources, as well as those that may not be as helpful. This type of research could be useful to librarians who are interested in the concept, but may be wary about the implementation process and results (Simovic, 2018).

Although the focus of this literature review was geared more towards the academic library, Anna and Mannan (2020) recommend that additional research should focus on the public library, which could also apply to the academic context. They also suggested a few additional research gaps, including the use of data processing tools and techniques to determine and implement new services; the incorporation of outside data to improve services within the library; and the role of librarians in the era of big data. Since the first record of published research on big data in libraries dates back to 2012, many gaps still exist, but these recommendations should help eliminate some of them (Anna & Mannan, 2020).

Conclusion

While this review listed but a few possibilities for libraries in the era of big data, the possibilities are numerous. Although libraries may face some of the same concerns which other organizations face, they have the benefit of trained information specialists, or librarians, who have an educational background of information sciences (Anna & Mannan, 2020). This focus on information can be beneficial to librarianship. Big data is a stimulating opportunity for libraries to embrace, making the future of librarianship uncertain and exciting.

References

Ahmed, W. & Ameen, K. (2017). Defining big data and measuring its associated trends in the field of information and library management. *Library Hi Tech News, 34*(9), 21-24. https://doi-org.libproxy.troy.edu/10.1108/LHTN-05-2017-0035

Ahmed, K., JianMing, Z., & Rafi, M. (2019). An analysis of academic librarians' competencies and skills for implementation of Big Data analytics in libraries: A correlational study. *Data Technologies and Applications, 53*(2), 201-216. https://doi.org/10.1108/DTA-09-2018-0085

Anna, N. E. V. & Mannan, E. F. (2020). Big data adoption in academic libraries: A literature review. *Library Hi Tech News, 37*(4), 1-5. https://doi-org.libproxy.troy.edu/10.1108/LHTN-11-2019-0079

Barsh, J., Capozzi, M. M., & Davison, J. (2008). Leadership and innovation. *McKinsey Quarterly, 1*, 36.

Carillo, K. D. A. (2017). Let's stop trying to be "sexy" – preparing managers for the (big) data-driven business era. *Business Process Management Journal, 23*(3), 598-622. https://doi-org.libproxy.troy.edu/10.1108/BPMJ-09-2016-0188

Dodge, R., Dwyer, J., Witzeman, S., Neylon, S., & Taylor, S. (2017). The role of leadership in innovation. *Research Technology Management, 60*(3), 22-29. https://doi-org.libproxy.troy.edu/10.1080/08956308.2017.1301000

Dority, K. (2022). LIS skills in a data universe: Your career as a data librarian. *LibGig.* https://www.libgig.com/lis-career-data-librarian/

Graves, J. T., Acquisti, A., & Christin, N. (2016). Big data and bad data: On the sensitivity of security policy to imperfect information. *University of Chicago Law Review, 83*(1), 117-137.

Hanges, P. J., Aiken, J. R., Park, J., & Su, J. (2016). Cross-cultural leadership: Leading around the world. *Current Opinion in Psychology, 8*, 64-69. https://doi.org/10.1016/j.copsyc.2015.10.013

Hung, S., Cheng, M., Hou, C., & Chen, N. (2021). Inclusion in global virtual teams: Exploring non-spatial proximity and knowledge sharing on innovation. *Journal of Business Research, 128,* 599-610. https://doi.org/10.1016/j.jbusres.2020.11.022

Jalonen, H. (2012). The uncertainty of innovation: A systematic review of the literature. *Journal of Management Research*, *4*(1), 1-47. http://dx.doi.org/10.5296/jmr.v4i1.1039

Klous, S. & Wielaard, N. (2016). *We are big data: The future of the information society.* Atlantis Press.

Koziol-Nadolna, K. (2020). The role of a leader in stimulating innovation in an organization. *Administrative Sciences, 10*(3), 1-18. http://dx.doi.org.libproxy.troy.edu/10.3390/admsci10030059

Lisak, A., Erez, M., Sui, Y., & Lee, C. (2016). The positive role of global leaders in enhancing multicultural team innovation. *Journal of International Business Studies*, *47*(6), 655-673. http://dx.doi.org.libproxy.troy.edu/10.1057/s41267-016-0002-7

Martin, K. (2013). USA Patriot Act's application to library patron records. *Journal of Legislation*, *29*(2), 283-306. https://scholarship.law.nd.edu/cgi/viewcontent.cgi?article=1112&context=jleg

Mayer-Schonberger, V. & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think.* Houghton Mifflin Harcourt.

Mitchell, M. T. (2018). Into the unknown: Leading people to do the work of innovation requires a different mind- and skill set. There are three crucial behaviors for success. *Independent School*, *77*(4), 46-49.

Open Doors. (2022). *Enrollment trends*. https://opendoorsdata.org/data/international-students/enrollment-trends/

Overall, J. (2015). A conceptual framework of innovation and performance: The importance of leadership, relationship quality, and knowledge management. *Academy of Entrepreneurship Journal*, *21*(2), 41-54.

Rampersad, G. (2020). Robot will take your job: Innovation for an era of artificial intelligence. *Journal of Business Research, 116,* 68-74. https://doi.org/10.1016/j.jbusres.2020.05.019

Richey, Jr., R. G., Morgan, T. R., Lindsey-Hall, K., & Adams, F. G. (2016). A global exploration of big data in the supply chain. *International Journal of Physical Distribution & Logistics Management*, *46*(8), 710-739. https://doi-org.libproxy.troy.edu/10.1108/IJPDLM-05-2016-0134

Simovic, A. (2018). A big data smart library recommender system for an educational institution. *Library Hi Tech, 36*(3), 498-523. https://doi-org.libproxy.troy.edu/10.1108/LHT-06-2017-0131

United Nations. (2018). *Technology and innovation report 2018: Harnessing frontier technologies for sustainable development*. https://unctad.org/system/files/official-document/tir2018_en.pdf

Walker, R. (2015). *From big data to big profits: Success with data and analytics.* Oxford.

Wojcik, M. (2019). How to design innovative information services at the library? *Library Hi Tech*, *37*(2), 138-154. https://doi.org/10.1108/LHT-07-2018-0094