Demystifying Big Data: Value of Data Analysis Skills for Research Librarians

Tammy Ann Syrek-Marshall, MLS

Abstract

There was a time when librarians learned statistics solely as a management tool, a way of measuring library usage and materials circulation. More advanced users of statistics might use it in research papers as a metric for in-depth study of patron behavior and bibliographic analysis. At one time, Excel spreadsheets primary use was in managing library budgets or producing attractive graphs for library reports and publications. As for data, it's no longer enough to simply provide access; the value now is in understanding and interpreting this data. This begins with knowing the difference between small data and big data. Small data is static, a snapshot of a single moment in time. Businesses and other organizations, however, are now eagerly tapping into the predictive power of big data. To fully realize the potential of big data, people who possess certain skill sets are in high demand, making data analytics and data science two of the fastest growing career fields. Depending on the size of the organization, librarians will find themselves called upon to work closely with data analysts. Librarians and researchers who possess even a basic understanding of data analysis are now positioned to competently collaborate with these data professionals. Research librarians with a strong grasp of the methodology and processes of data analysis stand the best chance of expanding their roles within their organizations. To this end, for those who currently have had limited exposure to big data and analytics, it is time to become acquainted with this growing field; to begin the process of expanding research methodology with an understanding of predictive data analysis.

Introduction

There is an adage about librarianship being the smartest profession, not for what librarians know but for how quickly they can access what they do not.

The evolution of the role of the librarian began long ago. In the beginning, librarians were the keepers of the library, but that wasn't enough. As long ago as Callimachus of Alexandria, the process of keeping records of the contents of a library and organizing those contents by subject proved to be an invaluable tool for researchers. Moving up to the twentieth century, using metadata to catalog objects, concepts and information sources has become a science onto itself. Providing access to massive quantities of information has proven to be a turning point. The stepping stone for the explosion of data in the twenty-first century, the data revolution.

Careers in librarianship are evolving, changing with the ever-increasing role of technology. A data revolution is occurring and we as librarians stand alongside others on the ground floor. Do we join in and climb to the top? Or do we just stand there and watch while others take the first step?

In this paper I will attempt to offer some context and insights into a relatively new field of study, to demonstrate its relationship to library science, and offer a stepping stone for those who may wish to follow this new path.

A Four-Sided Coin, A Five Step Pyramid

A four-sided coin, a sisterhood of sciences each devoted to studying different aspects of what is essentially the same discipline. Another way of looking at it is through the DIKW Pyramid model [1] which shows Data, Information, Knowledge and Wisdom as having a hierarchal and transitional relationship with each other. Each of the first three concepts have been formalized as fields of study; Data Science, Information Science, and Knowledge Science. There is one additional concept that can be added to this model, Understanding. (Figure 1)

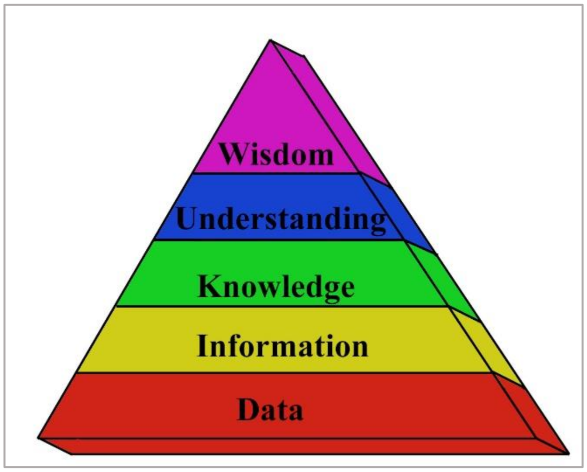

Figure 1: DIKUW Pyramid model

Defining, conceptualizing and interpreting this model is still subject to debate, but the basic ideas hold true. The idea that data are the building blocks of Information. Information is the foundation of Knowledge. Knowledge is the basis of Understanding. Understanding then leads to Wisdom.

As already stated, the first of these, Data, Information and Knowledge each have their own field of study. But what about Understanding? What field of study correlates with that concept? While Knowledge is a singular concept, Understanding is attained when Knowledge is made accessible. When it is shared and in turn looped back into the system in order to build upon that original Knowledge to generate new Knowledge. Sound familiar? It should be, this is a core element of Library Science.

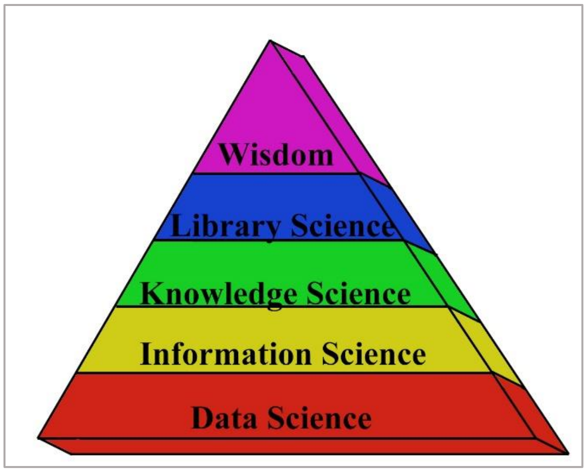

Figure 2: Relationship of Library Science to the other three sciences.

With the inherent background, training and experience of the majority of Librarians, it is a natural process to expand into the related fields of Data Science, Information Science, and Knowledge Science. With this, the DIKW Pyramid now becomes a DIKUW Pyramid.

Before we delve into the focus of this paper, we first need to take a closer look at both Knowledge Science and Information Science.

The Clover-Leaf on the Path to Wisdom

While the DIKUW Pyramid is a good model to begin with, a closer analysis on the interrelationships of each discipline shows a connection that is more linear, and more convoluted. Less like a stairway, more like a highway. As we move along this path from Wisdom to Data, first we must travel through Knowledge and Information.

The definition of Knowledge is fairly straightforward. Defining Knowledge Science, however, is a little more complex. In the book ‘‘Knowledge Science’’, Yoshiteru Nakomori breaks it down into four research fields:

- *Knowledge engineering*: Symbolizing (approximating) experts’ knowledge to develop artificial intelligence
- *Knowledge discovery*: Mining a large-scale data set to extract partial rules and adding their meanings based on domain knowledge
- *Knowledge construction*: Simulating complex phenomena based on some hypothesis and adding the meanings to emerged properties based on domain knowledge
- *Knowledge management*: Converting distributed (or tacit) knowledge into shared (or explicit) knowledge and using it effectively

In using this description, we can begin to see some commonality with Library Science.

Let’s now take a look at Information Science, an area of study that, while it roots extend back at least to the nineteenth century, has its origins in the 1950s with J.E.L. Farradane. Information Theory became Information Science and took on an expanded meaning with the post-WW II Computer/Technology Revolution.

In trying to define Information Science, you have to consider from which viewpoint the interpretation comes from. Is it from the viewpoint of Library Science seeking to show the interrelationships between the two fields? Or from the viewpoint of Computer and Data Scientists whose priorities affect their interpretation of Information and Information Science? Then there is another viewpoint that comes from a more philosophical perspective. So, how do we define Information Science?

"Information Science is that discipline that investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability." [2]

H. Borko continues by defining Information Science, "It is an interdisciplinary science derived from and related to such fields as mathematics, logic, linguistics, psychology, computer technology, operations research, the graphic arts, communications, library science, management, and other similar fields." [2]

As Information Science and Knowledge Science both interrelate with Library Science, and through that with each other a feedback loops are created. Library Science, understanding in this model, acts as the gatekeeper to Wisdom, feeding Knowledge, Information and Data back into the system in order to create even more knowledge. (Figure 3)

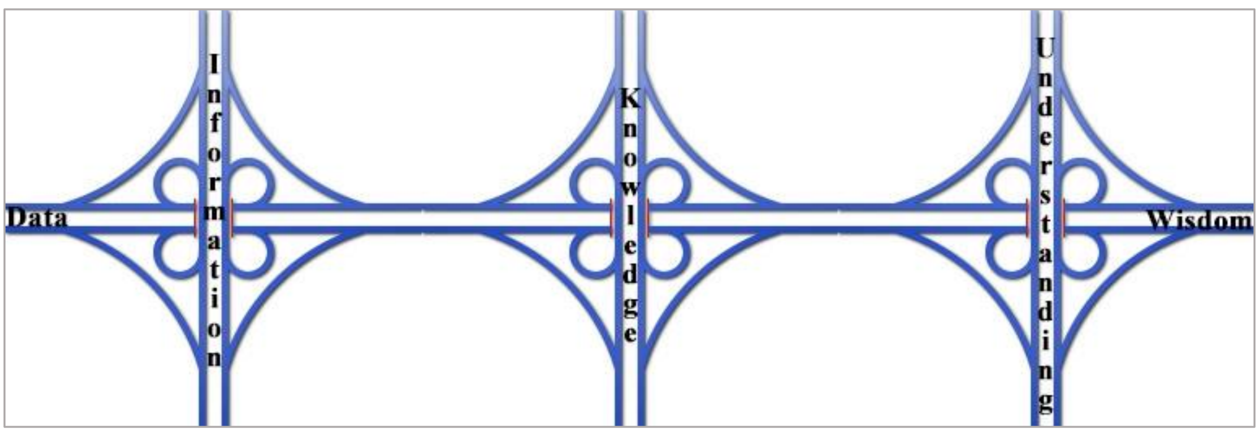

Figure 3: The Information Interchange.

The Many Flavors of Data

We have discussed the relationship between Library Science, Knowledge Science and Information Science. Before we can tell the story of Data Science, we first must analyze the history and nature of Data. The definition of the word data comes from the Latin word datum and means "Things known or assumed; facts or figures from which conclusions can be inferred; information" [3] With that in mind, the history of data, even before it was defined as such, is as old as the human record. Hunters who find a watering hole where prey come to drink might record that information on cave wall drawings.

Counting may have been the first form of numeric data. Tally sticks, from as early as ca. 18,000 BCE, were used to keep count of supplies, of belongings or merchandise for trade. This would evolve over human history into mathematics and eventually statistics. The scientists of ancient Greece used data to record their observations and develop theories based on them. The Library of Alexandria, ca. 350-150 BC, is believed to have possessed the largest collection of printed knowledge in the world at that time. Callimachus, scholar and librarian, is credited for being the author of the Pinakes considered to be the first library catalog. Recording data on each of the manuscripts held within the great library. While the origins of data usage are an integral part of the evolution of human civilization, it wasn't until the twentieth and twenty-first centuries for data to reach its own evolutionary leap.

To begin with, let us start with the question, what is data? One way to look at data is by analogy. Subatomic particles are the building blocks of matter, Physicists are still trying to determine the exact number and types of these particles. Data are the building blocks of Information. The simplest definition of data is that data are facts that are neither subjective, nor subject to immanent change. For example, if I were to stand on a scale and measure my height, it would tell me that I am 5 feet, 5 inches or 165.1 centimeters tall. That would be an example of a single data point. Data points can be numbers or symbols; concepts or locations; objects or even people. Pretty much anything can be treated as a data point, so long as it meets the basic criteria. This seems simple enough, right?

It wasn't so long ago that data was simply that, data, a concept as plain as a vanilla ice cream. one of the earliest variations on data is metadata. As it relates to cataloging, the basic idea of metadata is as old as the Library of Alexandria. However, the term itself, in its modern usage, only dates back to around the 1980s. As most librarians and catalogers know, the standard definition of metadata is that it is simply 'Data about Data'. Yet, like with most aspects of data, this simplicity is only skin deep.

Another related concept that has recently become popular is small data. In its more conventional interpretation, small data is limited in size and in scope. It is the data used, not for predictions or modeling but for static snapshots of a specific event. How many students asked questions at the reference desk today? What is the average number of videos that circulated on a specific weekend?

The more recent, redefined concept of Small Data was initially developed as a response to big data. The first mentions of small data began as early as 2013 as a Marketing tool that focused on the individual, not on the masses. "In a telephone interview, [Allen] Bonde said big data is about machines and small data is about people." [4] An idea further popularized by Martin Lindstrom in his 2016 book, ''Small Data''.

"The small data approach Lindstrom offers is simple, at least in concept. As a marketer, he says, you should be spending time with real people in their own environments. That, combined with careful observation, can lead to powerful marketing insights.

This approach is the human-centric alternative to big data. In each case, one is collecting information to gain insights into behavior, interests, and so on. But, Lindstrom's approach relies on a mix of keen observation of small samples and applied intuition." [5]

Like the Three Bear's Chairs, we have gone from small to medium and now we have reached the largest of them all.

When Data Gets Big

The late 1990's gave birth to the term, "Big Data" though the attribution is not quite as clear. One potential candidate is John Mashey, a Silicon Valley scientist at that time. When asked about his role in the conceptual development of the phrase, he replied "I was using one label for a range of issues, and I wanted the simplest, shortest phrase to convey that the boundaries of computing keep advancing." [6]

Since then, the concept of big data has evolved, making its biggest leaps from 2009 to the present, correlating with the increased interest in big data analytics. The basis of big data is born out of complexity. Not only as a factor of population or sampling size, but also of diversity and immediacy. Though this alone might not be sufficient to properly describe big data. A popular shorthand description of big data among professionals in the field is referred to as “The Three V’s”, "volume, velocity and variety," as originally described by (Laney 2001). Taking into consideration the nature and current uses of Big Data, (van Rijmenam 2013) has proposed expanding this definition to include Veracity, Variability, Visualization, and Value.

To simplify this complexity, let’s go back to my original analogy. I know what my height is, so what is the height of everyone in the Baltimore Convention Center? If we measured everyone, we would now have a large sample size, volume, though it is one that does not yet have variety or variability.

For variety let us add the qualifier of gender, then make it more complex by adding another qualifier, Job Title. For variability we can ask if we took multiple measurement over time, would each height measurement remain the same? Veracity takes into account accuracy. Were the measurements performed using the same equipment? Were those measured wearing shoes or standing in bare feet?

That leaves us with velocity, visualization and value. Velocity relates to the speed with which data is created and processed. Like a loaf a bread, one fresh from the oven tastes better than one that has been sitting around for a long time getting stale. If the height measurement data is not acquired and used quickly, it may lose its relevance, it might become stale. Visualization makes data easier to comprehend, simpler to relate to the purpose it was created for and is usually more pleasing to the eye. Value, the final concept, should be the most obvious. After doing all the work of gathering the data, analyzing it and visualizing it, was it worth it? Do you have new insight into the relationship between height and librarian’s job titles?

Now that we have big data, now that we understand it a little better, the next step is to decide what to do with it. How do we turn a massively complex quantity of data points and numbers into a model that can be used to help predict which patients are the most likely to have a recurrence of cancer, or which teaching methods produce the best educational outcomes, or which product distribution model is the most efficient and the most profitable?

From Descriptive Statistics to Predictive Analysis

The groundwork for statistics had been laid throughout the history of mathematics, yet it didn’t start taking on an identity of its own until the eighteenth and nineteenth centuries. This is in large part to the works of Carl Gauss, Thomas Bayes and Pierre-Simon Laplace. Statistics, as we will see, lies in the heart of Data Analysis. But its soul belongs to Pascal and Fermat’s development of Probability Theory in the 1650s.

Modern statistics of the twentieth century, commonly subdivided into Descriptive Statistics and Inferential Statistics, has become a mainstay for researchers in many fields, including Library Science. For the purposes of library management or research on bibliographic analysis or similar topics, knowledge of modern statistical methodology was going from useful to critically important. Yet, these statistical studies were often limited in size and in scope.

The twenty-first century data revolution not only brought with it a massive increase in accessible data, it also introduced new tools and techniques for dealing with this data. If you are interested in the programing aspect of big data analysis, then you need to familiarize yourself with at least one of these languages: R, Python, SQL, SAS, Java, Apache Hadoop, or C/C++. Knowing the language does little unless you also know the tools needed to perform data analysis. Some of the popular of these tools include Microsoft Excel, Tableau, RapidMiner, KNIME, BigML, Jupyter, Orange, or Anaconda.

From start to finish, working with big data is a multi-step process. Once you have a handle of the language and the tools, you would next need to understand this process.

Data Creation: if the data does not already exist, would be the first step in the process. Examples of data creation methodology include surveys, scientific research, or observation. Whichever method is used, the data must be accurate and verifiable in order to produce valuable results.

Data Warehousing: for organizations which generate their own data; organizing, storing and making accessible this data is essential precursor to big data analysis. Doing this systematically while using the proper data governance in order to make certain that the data will always be accurate and available as needed.

Data Retrieval: if the data already exists, it needs to be identified, verified and extracted. There are many sources of data, either internal to an organization or external from the internet or from a third part source. Locating and identifying the correct data to use in certain projects requires both the right tools and the right skills.

Data Mining/Data Analysis: two terms that are different yet closely related. Data mining is defined as a means to work with raw data, to process it in such a way as to derive information or gain insights from or see patterns within the data. Data analysis aims for a similar goal but it is more statistically intensive. Its main goal is in working towards creating models that can be used for informed decision making.

Data Visualization/Storytelling: once the data has been processed, it needs to be transformed. You don’t need magic to create this transformation. What you do need is creativity and the ability to re-envision numeric results into visual graphics and text that tell a story. The story, of course, is not one of fiction but of facts. A story with a beginning, a middle and an end result told in a way that is clear and easy to understand.

In the end, for projects this big, you need teamwork.

Data Science is a Team Sport

The growth in demand for data scientists is dramatically increasing every year. Current predictions see an increase in demand of as much as 28% by the year 2020.[7] This is further emphasized by Glassdoor’s ranking of data scientists as the Number 1 best job in America for 2018.[8] It has even been called, “The Sexiest Job of the 21st Century” [9] If so, would that make data librarians truly ‘sexy librarians’?

Moving off the pedestal and back into reality, the work of data scientists is focused mostly in the central part of the process, the analysis of data. In creating a finished product, a team is needed to handle each step towards the completion of the goal. An online search of data related careers identified at least 28 different job titles relating to big data and data analysis. Including the position of data scientist, 8 of those job titles made it into Glassdoor’s top 50 ranking. (Figure 4)

Job Title/Job Skill	GD Ranking
Analytics Manager	18
Business Analyst	43
Business Intelligence Developer	42
Data Analyst	38
Data Architect	
Data Collector	
Data Consultant	
Data Driven Instruction	
Data Engineer	33
Data Entry	
Data Governance	
Data Integrity Specialist	
Data Integration Architect	
Data Integration Specialist	
Data Journalist	
Data Librarian	
Data Mining	
Data Mining Scientist	
Data Quality Analyst	
Data Science - Machine Learning	
Data Scientist	1
Data Storyteller	
Data Visualization	
Data Warehouse Specialist	
DevOps Engineer	2
Director Data Management	
Operations Research Analyst	
Statistician	
Systems Analyst	39
Data Manager	
Data Modeler	

Figure 4: List of common job titles in the data science field.

As the increase in demand for data professionals grows, so too will the need for librarians familiar with the process of big data analysis as well as having the skill set to handle data management. Though why stop there?

Casting Call

From stage hand to script writer, everyone on the team has a role to play. As a librarian with an interest in big data and data analysis, you may assume that the roles available to you are limited. Far from it, there are many ways and many roles to play in the production of a big data story.

A more traditional role for a librarian is in data curation and archiving. The expanded role of data librarian also opens up the possibility for the organization's library to serve as it's Data Warehouse. As a curator of data and the one who manages the Data Warehouse, another role opens up. As the one who understands the Data Warehouse, it is reasonable to be included as a member of the organization's Data Governance group. This position allows for the development of policies, procedures and dealing with issues relating to the storage, use and management of data.

There are other positions as well that are suited to a librarian with the right skills and expertise. Some of these positions may not require an extensive technical background. While others are suited more to those with a practical knowledge of computer coding, database, software and/or computer graphics. Mathematics and statistics skills would also be useful for some positions. Taken from the more extensive list already provided, here are a just few of many examples to consider.

Data Quality Analyst: responsibilities include reviewing and auditing the health and quality of data, investigating problems and remediating data issues, establishing standards and procedures, and reporting on issues found and actions taken.

Analytics Manager: responsibilities may include creating data presentations, reports, and descriptive analysis of post analysis data for client usage. Analytics Managers may also lead teams of data analysts in completing projects and achieving goals.

Data Storyteller: possess ability to transform data into written and visual stories, presentations and interactive graphics that convey the meaning within the data.

When deciding which path to follow, you first need to know what your strengths are and where your interests lie. Then you need to consider what additional training you might need, as well as the value of those skills in relation to the cost of the training.

Valuation of Big Data Skills

Valuation can be viewed in three ways; personal, professional, and organizational. This is a pretty straightforward concept, but one that should be viewed in context. Before considering the value of becoming familiar with, or even proficient in skills relating to big data, you need to ask who are you doing it for? Are you learning these sills for yourself, simply because you want to? Are you seeking to learn new skills with the goal of advancing your career? Or are you doing it for your organization, to fulfil a need and possibly secure or advance your position within it?

When considering your position within your organization, your place of employment, you need to find identify possible opportunities. Learning more about how your organization deals with internal data, or if it has a need for external data is your first objective. If there are already data scientists, data analysts on the staff, meeting with them, talking with them is a way to learn more about the data system already in place as well as any needs that are not being met. Needs that you, with the right training, can fulfill. This will almost certainly increase your value to your organization. In some cases, you may even be able to use this value analysis to present it to your organization. In doing so, they might even be willing to cover the cost of training if you can convince them of the benefit.

If you are doing this for professional reasons, then you have more freedom to choose your own path. Bear in mind the cost of training and the time you are willing to invest into it. Set a goal, then choose a method of study that meets your goal. You may not need to learn to do advanced data analysis, so start with an overview of data science and decide from there which path would be best for you.

Valuation can also be subject to a cost/benefit analysis where cost is measured not only in dollars but also in hours. Fortunately, there are ways to learn the basics at minimal financial cost.

Taking the Path Forward

Once you have decided to take the next step, there are several ways to approach it. There are many books available that I would recommend as introductions to the field of Data Science. The first book on this list is Amy Affelt's ''The Accidental Data Scientist''. If you choose to take a more technical approach, O'Reilly publishes a number of books worth checking out, for example Joel Grus' ''Data Science from Scratch'' or Peter and Andrew Bruce's ''Practical Statistics for Data Scientists''. If you are interested in data visualization, I would recommend Jorge Camoes' ''Data at Work'' or Erez Aiden and Jean-Baptiste Michel's ''Uncharted''. To find more recommendations, I have listed several relevant articles in the endnotes. [10-12]

A common method of learning new skills is to enroll in online courses, some of which can lead to professional certificates or degree credits. There are several sites that offer various programs in Big Data, Data Science and Data Analysis. I would recommend looking into one or more of the following; Coursera, DataCamp, EdX, Khan Academy, Lynda.com, and Udacity.

If you would prefer a more academic environment, there are an increasing number of Data Science Certificate programs being offered either on-site or online from many major colleges and universities. These programs tend to be shorter than a degree programs, running anywhere from six months to two years.

Sometimes the best source of information comes from interpersonal interactions. There are a number of online communities and professional organizations that could help you get started. The SLA Data Caucus is where you can interact with other librarians who share your interest in data. Another way to meet with librarians who share an interest in data science is through regional data librarian symposiums. There are also a number of professional organizations you can look into, The Data Science Association, American Statistical Association, or The International Institute for Analytics just to name a few. One more source of information that I would recommend is Quora – Data Science, especially if you have immediate questions for those who already have experience.

Whichever path you choose, it is best to travel down the road with eyes wide open.

The Dark Side of Big Data: A Cautionary Tale

September 20, 2017, Hurricane Maria makes landfall on Puerto Rico causing widespread damage, loss of power, loss of clean water and disruptions in basic public services. By the end of October, the official death toll was reported to be only 64. After being pressured to reevaluate the initial analysis, to reexamine the data used to make the death toll determination, the number increased to well over 1000. Without accurate information, the already difficult recovery process for the island will be much greater than it should be.

Entering into the field of Data Science does not necessitate entering into it blind. Just as there are many invaluable and beneficial uses for big data analysis, there are also many ways that the process can go wrong, many ways it can be misused. Issues that affect accuracy, value and ethics.

The problems may begin if big data turns out to be bad data. Bad data occurs when the data wasn't collected in a way that is unbiased, or when the data contains gaps or inaccuracies, or when the data wasn't properly cleaned and scrubbed, or when the source of the data is questionable. Issues of ethics and of privacy come into play when the data is acquired through inappropriate sources.

Even if the data is good, the analysis process might be flawed. There are a number of ways to process data, different methods or systems. Was the data analyzed using R, Python, Hadoop, or Excel? Which data set did you use? What type of regression? Depending on the type of analysis done, the results could be similar or they could be very different. Knowing which of the methods is to be used and then using them correctly could mean the difference between an accurate analysis and one that could be very costly to the client.

Interpretation and visualization can turn a good analysis using good data into bad information if it is done subjectively. When the presentation is the final product, it needs to be done with an understanding, not just of what the client requires, but also of the truth in the data. If the story about the data is more fiction than fact, that could lead to dangerously bad decision making. It could also negatively influence future research and data analysis projects that might depend on this particular outcome.

It is natural and expected of a librarian to provide the information, or data, requested often without judgement or bias. For those librarians choosing step into the world of big data, a decision may have to be made between providing the best possible data available versus providing the data that is requested, even if that data is questionable.

Commencement

Not a conclusion, but a beginning. A way to ride the wave of technology into the 21st Century and beyond. Books and traditional media are the heart and soul of the library profession. From Public Libraries to Academic Libraries to Special Libraries, there will always be a need for those who can be there on the front lines of public service in the way that librarians have for thousands of years.

The leaps and bounds of technology, the expanding concepts of what constitutes that which falls under the librarian's purview, has opened up new career pathways, along with new challenges. Finding a place for ourselves in this changing world is critically important. Data Science does not redefine what a librarian is as much as it expands the scope of what a librarian can do. Once you have come to understand the true nature and value of data, in all its forms, you will no longer be mystified by it. Instead you can begin to master it and your role in the coming data revolution.

Find your path and own it, trust in yourself and your abilities, then reach for whatever goal you set for yourself.

Endnotes

1. *What is DIKW?* (Complexity Labs, 2017), https://www.youtube.com/watch?v=h9gYk66yz-0.
2. Borko, H. 1968. "Information Science: What Is It?" *American Documentation* 19 (1). Wiley-Blackwell:3–5. https://doi.org/10.1002/asi.5090190103.
3. Guralnik, David B. (David Bernard). 1982. *Webster's New World Dictionary of the American Language*. Simon and Schuster.
4. Lundquist, Eric. 2013. "'Small Data' Analysis the Next Big Thing, Advocates Assert." *EWeek,* September 2013. http://www.eweek.com/enterprise-apps/small-data-analysis-the-next-big-thing-advocates-assert
5. Dooley, Roger. 2016. "Small Data: The Next Big Thing." *Forbes,* February 2016. https://www.forbes.com/sites/rogerdooley/2016/02/16/small-data-lindstrom/#324ddf387870.
6. Lohr, Steve. 2013. "The Origins of 'Big Data': An Etymological Detective Story." *The New York Times*, February 1, 2013. https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/.
7. Columbus, Louis. 2017. "IBM Predicts Demand For Data Scientists Will Soar 28% By 2020." *Forbes*, May 2017. https://www.forbes.com/sites/louiscolumbus/2017/05/13/ibm-predicts-demand-for-data-scientists-
8. "50 Best Jobs in America." 2018. Glassdoor. 2018. https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm.
9. Davenport, Thomas H., and D.J. Patil. 2012. "Data Scientist: The Sexiest Job of the 21st Century." *Harvard Business Review*, October 2012. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-
10. King, Timothy. 2018. "Top 20 Best Data Science Books You Should Read." Solutions Review - Business Intelligence. 2018. https://solutionsreview.com/business-intelligence/top-20-best-data-science-books-you-should-read/.
11. Lebied, Mona. 2018. "The Top 8 Best Data Science Books You Need To Read." The Datapine Blog. 2018. https://www.datapine.com/blog/best-data-science-books/.
12. Mester, Tomi. 2016. "Aspiring Data Scientists! Learn the Basics with These 7 Books!" Hackernoon. 2016. https://hackernoon.com/wannabe-data-scientists-learn-the-basics-with-these-7-books-1a41cfbbdd34.

References

Ackoff. n.d. "From Data to Wisdom." *Journal of Applied Systems Analysis* 16.

Affelt, Amy L. n.d. *The Accidental Data Scientist : Big Data Applications and Opportunities for Librarians and Information Professionals*. Accessed May 16, 2018. https://books.google.com/books?id=aQH3oQEACAAJ&dq=The+Accidental+Data+Scientist&hl=en&sa=X&ved=0ahUKEwifou3o-orbAhWptVkKHUsiBa4Q6AEIKTAA.

Aiden, Erez, and Jean-Baptiste Michel. 2013. *Uncharted : Big Data as a Lens on Human Culture*. Riverhead Books. https://books.google.com/books?id=4NOKDQAAQBAJ&dq=Uncharted&source=gbs_navlinks_s.

Bawden, David., and Lyn Robinson. 2012. *Introduction to Information Science*. Facet. https://books.google.com/books?id=Nc5qDQAAQBAJ&dq=Bawden,+Robinson+2012&source=gbs_navlinks_s.

Gent, Edd. 2016. "Beware of the Gaps in Big Data." *E&T Engineering and Technology*, September 2016. https://eandt.theiet.org/content/articles/2016/09/beware-of-the-gaps-in-big-data/.

Gold, Anna. 2010. "Data Curation and Libraries: Short-Term Developments, Long-Term Prospects." http://digitalcommons.calpoly.edu/lib_dean/27.

Haynes, David. n.d. *Metadata for Information Management and Retrieval : Understanding Metadata and Its Use*. Accessed May 7, 2018. https://books.google.com/books?id=muVFDwAAQBAJ&dq=metadata&sou

Laney, Doug. 2001. "3D Data Management: Controlling Data Volume, Velocity, and Variety." META Group. 2001. https://www.bibsonomy.org/bibtex/263868097d6e1998de3d88fcbb7670ca6/sb3000.

Lindström, Martin. 2016. *Small Data : The Tiny Clues That Uncover Huge Trends*. St. Martin's Press. https://books.google.com/books?id=hDJBCgAAQBAJ&dq=small+data&source=gbs_navlinks_s.

Lohr, Steve. 2015. *Data-Ism : The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else*. Harper Business. https://books.google.com/books?id=ett9BAAAQBAJ&dq=Dataism&source=gbs_navlinks_s.

Marr, Bernard. 2015. "A Brief History of Big Data Everyone Should Read." World Economic Forum. 2015. https://www.weforum.org/agenda/2015/02/a-brief-history-of-big-data-everyone-should-read/.

Matteson, Scott. 2013. "Big Data Basic Concepts and Benefits Explained." *TechRepublic*, September 2013. https://www.techrepublic.com/blog/big-data-analytics/big-data-basic-concepts-and-benefits-explained/.

Mayer-Schönberger, Viktor., and Kenneth. Cukier. 2013. *Big Data : A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt. https://books.google.com/books?id=uy4lh-WEhhIC&dq=Big+Data&source=gbs_navlinks_s.

Nakamori, Yoshiteru. 2012. *Knowledge Science : Modeling the Knowledge Creation Process*. CRC Press. https://books.google.com/books?id=vmbRBQAAQBAJ&dq=Knowledge+Science&source=gbs_navlinks_s.</p

Raber, Douglas. 2003. *The Problem of Information : An Introduction to Information Science*. Scarecrow Press. https://books.google.com/books?id=r6PVg479Sm8C&dq=Information+Science+introduction&s

Robles, Frances, Kenan Davis, Sheri Fink, and Sarah Almukhtar. 2017. "Official Toll in Puerto Rico: 64. Actual Deaths May Be 1,052." *New York Times*, December 9, 2017. https://www.nytimes.com/interactive/2017/12/08/us/puerto-rico-hurricane-maria-death-toll.html.

Shapiro, Fred R. 1995. "Brief Communication: Coinage of the Term Information Science." *Journal of the American Society for Information Science* 46 (5):321–97. https://doi.org/10.1002/(SICI)1097-4571(199506)46:5%3C384::AID-ASI8%3E3.0.CO;2-3

Spielkamp, Matthias. 2017. "Inspecting Algorithms for Bias." *MIT Technology Review*, June 2017. https://www.technologyreview.com/s/607955/inspecting-algorithms-for-bias/.

Stigler, Stephen M. 1986. *The History of Statistics : The Measurement of Uncertainty before 1900*. Belknap Press of Harvard University Press. https://books.google.com/books?id=M7yvkERHIIMC&dq=The+H

Stringfellow, Angela. 2018. "50 Best Data Science Tools: Visualization, Analysis, More – NGDATA – NGDATA." NGDATA. 2018. https://www.ngdata.com/top-tools-for-data-scientists/.

Tucker, Patrick. 2014. *The Naked Future : What Happens in a World That Anticipates Your Every Move?* Penguin Group (USA) LLC. https://books.google.com/books?id=rfIpAAAAQBAJ&dq=The+Naked+Future&source=gbs_navlinks_s.

Witty, Francis J. n.d. "The Pínakes of Callimachus." *The Library Quarterly: Information, Community, Policy*. The University of Chicago Press. Accessed May 9, 2018. https://doi.org/10.2307/4304755.

Penn State News. 2018. "Inaccurate Data Analysis May Affect Puerto Rico's Recovery," April 2, 2018. http://news.psu.edu/story/513137/2018/04/02/research/inaccurate-data-analysis-may-affect-puerto-ricos-recovery.