Daniel G. Alemneh
University of North Texas, USA
Daniel.Alemneh@unt.edu
Kris S. Helge
Texas Woman's University, USA
khelge@twu.edu

Chapter

Providing Open Access to Heterogeneous Information Resources without Compromising Privacy and Data Confidentiality

ABSTRACT

Digital technologies and innovation can be a double-edged sword. Recognizing that cyberspace is an integral component of all facets of national interests (including economy and defense), increasingly nations are articulating their national cyber strategies and outlining efforts to harden national cybersecurity and deter malicious actors from launching digital attacks. The escalating cyber attacks can come from insider threats for data breaches or other malevolent tactics by cyber criminals (such as ransomware, viruses, spyware, phishing, and other malicious endeavors). There are other types of cyber attacks resulting from well-meant initiatives such as open access or globally shared data, information, and knowledge. As cultural heritage institutions embrace digital environments, they are facing unprecedented pressures to ensure privacy and reduce the exposure of their institutions to all kinds of data-related risks. This chapter briefly reviews the current status of data security and argues in favor of balancing the open access aspirations of cultural heritage institutions with the need for privacy and data confidentiality.

Keywords: Data Confidentiality, Privacy, Data Protection, Open Access, Cybersecurity

INTRODUCTION

It is a well-accepted fact that emerging trends in technology (such as the rapid growth of mobile devices and cloud computing solutions) are changing the landscape and the way business is conducted in many organizations, including in cultural heritage institutions. Digital technologies provide scholars with access to diverse and previously unavailable contents that span various formats and myriad technologies across institutions and nations. As noted by Janes (2013), the digital shift has been upon us all for some time now, and the issues and realities are getting deeper and more complex as library service continues to be transformed by the multifaceted changes already in place and others on the horizon. Although technologies such as automation, artificial intelligence, and machine learning already help to facilitate access and interactions with big data, such innovations can also increase risks. Data growth has reached explosive levels. As the legacy system simply cannot keep up with the pace of the digital transformation. Some of the concerns with big data applications relate to:

- system security (e.g., protecting digital preservation and networked systems / services from exposure to external/internal threats);
- collection security (e.g., protecting content from loss or change, the authorization and audit of repository processes);
- legal and regulatory aspects (e.g. personal or confidential information in the digital material, secure access, redaction).

This chapter will discuss challenges raised by concerns about ensuring long-term access to digital resources verses data confidentiality and balancing the right level of data security that addresses compliance requirements in the context of libraries and cultural heritage institutions.

BACKGROUND OF OPEN ACCESS MOVEMENT

The digital shift has challenged the status quo and existing values, and as a result data security is a critical imperative for all institutions. According to the 2018 Thales Global Data Security and Threat Report, the rate of enterprises that are encountering data breaches grew from 21% in 2016 to 26% in 2017 and now to 36% in 2018. Digital transformation requires new data security approaches. In fact, increasingly many nations are articulating and releasing their national cyber strategies. Accordingly, the White House published a comprehensive National Cyber Strategy in September 2018 detailing how the United States current administration aims to improve cybersecurity in government, critical infrastructure and the private sector, as well as tackling cybercrime and international issues.

The open access (OA) movement is part of the broader "open knowledge" or "open content" movement that transforms scholarly communication. In reviewing the literature of the past few years, there is no shortage of views on the role of digital libraries and open access in facilitating digital access to knowledge by reducing barriers. Many researchers articulate a vision of a digital library environment that resonates with possibilities to create a knowledge management system that will enable scholars to navigate through these resources in a standard, intuitive, and consistent way. Many researchers including Alemneh and Hastings (2006), Muir (2017), and Verma (2018), agree that the new scholarly communication systems will inevitably be based on capabilities of interoperable network technology.

As cultural heritage institutions embrace such digital environments, they are facing unprecedented pressures to ensure privacy and reduce the exposure of their institutions to all kinds of data-related risks. Escalating cyber attacks, together with the insider threats for data breaches, make balancing the open access aspirations of cultural heritage institutions without

compromising privacy and data confidentiality challenging. In July 2018, the U.S. National Academies of Sciences (NAS) released a consensus report titled "*Open Science by Design: Realizing a Vision for 21st Century Research*," which lays out a vision for a fully-open global science environment, and provides the following five specific recommendations for moving from vision to implementation:

1. Research Institutions and Funders to work to create a culture that actively supports Open Science by better rewarding and supporting researchers engaged in Open Science
2. Research Institutions and other entities to support the development of educational and training programs to support students and researchers in adopting Open Science practices
3. Research Funders and institutions to develop policies/procedures to identify research outputs for long term preservation and public access, and funding to be made available to support these activities
4. Funders and institutions to ensure that research archives are designed and implemented according to the FAIR (Findable, Accessible, Interoperable and Reusable) principles
5. The Research Community to work together to advance Open Science By Design in order to advance science and help science better serve the needs of society.

THE PROMISE AND SECURITY CHALLENGES OF OPEN ACCESS BIG DATA

The term "big data" increasingly refers to the use of advanced data analytics methods that extract value from data. According to the 2018 Thales Data Threat Report, compared to traditional relational databases, the data generated and stored within big data environments can be orders of magnitude larger, less homogeneous, and change rapidly. There are a number of concepts associated with big data, including the three top attributes what are often referred to as the

"Three 'V's: Volume, Variety and Velocity." Some experts (including Jain, 2016 and Rijmenam, 2013), go on to add two more Vs to the list, Variability and Value.

It would be difficult to define what these 5Vs mean in ways that can work in various contexts. When it comes to handling big data, different disciplines or organizations might use the same tools for collecting and manipulating the data at their disposal, but there are significant differences in how they use technologies to organize, analyze, interpret and put the output data to work in general. The following brief description provide some points about the five Vs and their impacts on information professionals:

1. Volume: Data is being produced at astronomical rates, and size in this case is measured as volume. As Cano (2014) noted, with the Internet of Things (IOT) and all kinds of smart devices that feed smart living, the sheer volume of the data continues to grow every second. No wonder 90% of all data ever created was created in the past two years
2. Velocity: In the context of big data, velocity refers to the speed at which huge amounts of new data are being created, collected and analyzed in near real-time using various technological tools. Big data technology helps to cope with the enormous speed the data is created and used in near real-time.
3. Variety: With increasing volume and velocity comes increasing variety. Big data technology allows structured and unstructured diverse data to be harvested, stored, and used simultaneously (George 2017).
4. Variability: It refers to the inconsistency, which is the quality or trustworthiness of the data. According to Rijmenam, (2013), Variability is the variance in meaning, or the meaning is changing (rapidly). In indexing the same term or word can have a different meaning. In the same way, to perform proper sentiment analysis, algorithms need to be

able to understand the context and be able to decipher the exact meaning of a word in that context.

5. Value: is referring to the worth of the data being extracted. Big data can create enormous value for the global economy, driving innovation, productivity, efficiency, and growth. Despite the size, unless big data can be turned into value, it is useless (cost-benefit). In other words, the value is in the transformation and how the data is turned into information and then into knowledge.

Firican (2017) emphasized the importance of understanding the characteristics and properties of big data to prepare for both the challenges and advantages of big data initiatives. Some used the term complexity to refer to the complex process in which, where large volumes of data from multiple sources is collected, linked, connected and correlated to be reliable in order to grasp the information that is supposed to be conveyed by in original data.

Unintended Consequences of Open Access

Good intentions of open access may results in deleterious consequences. As mentioned above, most modern information institutions attempt to offer hosted data, information, and knowledge as openly as possible. Yet, such open access can result in data and information descending into the possession of sinister individuals. For example, copious libraries now offer data repositories for their researchers, faculty, and students (University Libraries, Data Repository Services 2018). Faculty place their raw data, both quantitative and qualitative, into such a repository. Uploading their data benefits faculty by assuring that it will be accessible to colleagues who may comment, utilize, question, and otherwise implement their raw data; their data will be preserved and safe from corrupt jump drives or personal drives; and their data will be harvested and visible globally.

However, since these researchers' data is globally accessible, danger of a data parasite obtaining and misusing this data is also possible.

Data parasites are individuals who through little or no achievement of their own obtain other people's data and use it maliciously to ultimately publish articles or other written documents, and fail to give attribution to the original data gatherers or creators. For example, data parasites will troll several different data repositories from universities, colleges, and other research institutions in hopes of gathering specific data about new technologies that could promote cleaner forms of energy for automobiles. They will then piece this data together and attempt to publish a paper or offer a conference presentation using the fragmentary data, while offering no credit to the original data gatherers (Longo and Drazen 2016). Thus, they take credit for proposing some form of this new technology and convey it as their idea and research - a form of plagiarism (Helge and McKinnon 2013). Such parasitic pseudo-research harms the original creators of the data and scientific research as a whole.

Plagiarism of others' data, information, and knowledge occurs frequently and for various reasons. Sometimes, researchers accidentally use another's research and data without giving proper attribution. Other times, such as with data parasites, plagiarism is intentional. Such malevolent intent can occur because a student researcher simply believes he or she will not get caught in such a malicious act. Other reasons for plagiarism include not taking an academic course seriously; not understanding self-plagiarism, improper conceptualization of what common knowledge is; and not knowing how to accurately cite scholarship, research, and data (Helge 2017). Dissertations and theses also often become the target of cyber criminals preying on academic informational institutions.

Intellectual Property Rights and Pirated Theses and Dissertations

Many benefits arise when students and faculty place their dissertations or theses into an open access scholarship repository. Their research is instantly accessible to anyone around the globe; sharing their research globally results in personal and professional benefits; they have a permanent and convenient hyperlink with which to refer prospective employers, research collaborators, and other research entities; they may receive invaluable constructive criticism from many researchers globally; and other altruistic researchers have perpetual and efficient access to this invaluable scholarship (Abrizah et al. 2015). Despite these benefits, as with open data, negative ramifications may manifest with open access to dissertations and theses as well. Serving as a Scholarly Communications Librarian at the University of North Texas, I was approached by a faculty member who had just obtained her Ph.D. from North American's university. She deposited her recently completed dissertation into her university's digital scholarship repository and was excited about the potential benefits of such a deposit. However, she discovered her dissertation had been pirated and was being sold in China. She queried whether anything could be done to stop the scholarship bootlegging. The response given to her explained that unfortunately, legally not much could be offered. In the United States, one may be sued for copyright infringement and other intellectual property crimes when a dissertation is improperly reproduced, distributed, displayed, or when illegal derivatives are created within the borders of the United States of America (17 U.S.C. 2018). However, when such intellectual property crimes occur outside of the United States of America, such as in China, the United States courts do not have legal personal jurisdiction to allow a prosecution to proceed, without proper extradition. Pennoyer v. Neff , 95 U.S. 714 (1878). Obtaining proper extradition from China is very cumbersome, especially for a stolen dissertation or thesis. So, at best for faculty or

students whose dissertations or theses are stolen and sold, they should be happy someone is actually reading their scholarship.

Internet Crimes Complaint Center (IC3) Roles in Protecting IP Rights

Besides being elated someone is actually reading and paying money for their dissertation or thesis, victims of scholarship piracy may also file a complaint with the Internet Crimes Complaint Center (IC3) https://www.ic3.gov/default.aspx. IC3 is a branch of the Federal Bureau of Investigation and examines Internet facilitated criminal activity (Federal Bureau of Investigation, Internet Crime Complaint Center 2018). There is no guarantee victims of scholarship theft will receive any equitable or monetary relief from filing a complaint with the IC3, however, filing a complaint with this federal entity could help in such recovery. Although, it is difficult to legally punish cyber criminals who steal and misuse intellectual property outside of specific legal jurisdictions, some countries such as the European Union (EU) formed alliances and passed legislation that protects the use of certain individual data.

CURRENT DATA AND INFORMATION LAWS

To help assuage individuals' fears of their data being misused in sinister guises, the EU recently passed legislation that directly addresses data misuse. In 2018, the European Union passed into law the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation (GDPR)) (Intersoft Consulting, 2018). The GDPR regulates how specific individuals' (e.g., student, medical patient…) data (e.g., grade point average, medical diagnosis…) is

processed or utilized by an individual, or an organization in a professional or commercial guise (European Commission 2018). For example, the GDPR does not apply to a private individual utilizing home addresses and phone numbers of other individuals who live in the same neighborhood in order to organize a block party. However, a commercial organization that is collecting that same data along with data about the shirt sizes of all who live on that block, and that plans on selling that data to another commercial company, is regulated by the GDPR. The difference is that the individuals or entity whose motivation is to utilize the personal data for monetary purposes is regulated by the GDPR, whereas the private individuals using the data for personal non-commercial purposes is not regulated by the GDPR. Newly passed GDPR also grants to EU citizens many protections and opportunities to become aware of how their data is being utilized.

Impact of GDPR in Protecting European Union Citizens

The GDPR grants to European Union (EU) citizens many rights, which include to discover and have access to personal data other entities hold, be aware of the processing of one's personal data, have incorrect data about an individual be corrected, have obsolete personal data deleted, object to the processing of one's data for commercial purposes, request the restriction of some personal data, and to obtain personal data in a machine-readable format (European Commission 2018). Such regulation allows EU denizens the opportunity to be more aware of where their data is being utilized, how it is used, who is using it, what it is being used for; and for EU citizens to rightfully object, correct, and have more control over the use of their personal data. Ultimately, in information warehouses, such as libraries, this could help researchers, students, and other patrons become more aware of how their research data is being utilized, by whom, where; and

also allot them more legal protections to reverse the use of data usage when their data is used for malevolent purposes. In fact, GDPR impact companies beyond user privacy. Article 33 of GDPR specifies that organizations must report a breach to the supervisory authority within 72 hours of detection, detailing the nature of the breach, the approximate number of data subjects and personal data records impacted, the likely consequences of the breach, and measures taken or proposed to address the breach and its negative effects (Wood, 2019). In the past few decades, not many countries, including the United States have passed similar broad sweeping legislation that offers as much macro-level protection.

Proposed United States Model Statute for Data and Information Privacy

In the United States, legislation passed offers more directed, micro-level protections, targeting specific industries. Laws such as the 1996 Health Insurance Portability and Accountability Act (HIPPA), the 1999 Gramm-Leach-Bliley Act, and the 2002 Homeland Security Act direct healthcare entities, financial institutions, and federal agencies to ensure data and information systems are protected with a reasonable level of security. Such reasonable levels of security are usually satisfied via an entity tangibly displaying it has created and documented specific protocols, policies, principles, standards, and guidelines that reasonably protect and secure healthcare, financial, and/or other data, information, and knowledge. Such vague statutory language does not afford private citizens an opportunity to edit incorrect data, to know how one's data is being utilized, by whom it is being utilized or to whom it is being sold, where one's data is being transferred, and other micro-level uses of one's data. Perhaps it is time for the United States Congress to follow the example of the EU and create legislation that better empowers the

citizens of the United States to have more control over their personal data, and that fosters them more knowledge about who, what, when, and how their personal data is being utilized.

A proposed model statute that could offer denizens of the United States more protection of the use of and more awareness of how their data is being utilized is conveyed below.

1. Citizens may request and be given information and access to information regarding who, what, when, why, and where their personal data is being utilized. They further may be given access to when, where, and why their data was transferred, sold, or otherwise disseminated.
2. Any organization using data for monetary purposes (whether a for-profit or not-for-profit entity such as a school, doctors office, law firm, charity, church…) must deliver a tangible response to a request for data within 30 days of the request for personal data, and the use of personal data. The response may detail how much of the personal data was sold, to whom it was sold, how the sold data may be utilized, the date the data was sold, and other possible pertinent requested information.
3. Misuse of personal data occurs when a for-profit or not-for-profit entity collects a person's personal data, and then uses it in a deceitful manner to create false digital personas, false digital likenesses, or otherwise uses the data in a malicious manner.
4. Any individual may have the right to force a for-profit or not-for-profit entity to cease the use of his or her personal data if he or she was not properly notified of such sale and use of data in a timely manner.
5. Any individual may force any for-profit or not-for-profit entity to cease the use of an individual's data for commercial purposes, to cease using an individual's data in an

incorrect or deceitful manner, to cease the use of obsolete data, or to cease the use of a deceased person's data.

6. Penalties for using a person's data in a malicious manner; or for noncompliance or disclosure of delivery of data, correction of inaccurate data, delivery of the location, specific use, and identification of the persons or function of such use of data may result in a fine of at least $250,000 and two years in prison.

Due to the ubiquitous threat of cyber security breaches and the misuse of a person's data, information, or knowledge; such a statute needs to be proposed and enrolled into law at the state and federal levels. Such a statute could provide more legal guidance and general protection to information entity users' privacy, which is a cornerstone value to libraries and other information entities. By ensuring such privacy and data protection, patrons of these information entities can confidently and comfortably use various information warehouses without fear of having their data, information, and knowledge compromised. Another consideration, beyond proposed legislation, is horizon technologies that will be adopted by cultural heritage institutions, such as libraries.

EFFECT OF HORIZON TECHNOLOGIES ON DATA PRIVACY AND CONFIDENTIALITY

Many horizon technologies will soon affect library patron data privacy and confidentiality. One such technology currently being experimented with is termed sixth-sense-technology. This technology interacts with digital world phenomena as a person gestures with his or her arms, hands, or other parts of the body in the physical world (Nuistry 2009). Such a gesture could involve motioning with one's hand and fingers to swipe from right to left to turn a page, pointing

to an object to retrieve more information about it, or setting an object on a tablet to discover where similar items may be located. For example, a person may place a soccer ball on a tablet, and utilizing the sixth-sense-technology, he or she could then theoretically locate all nearby physical locations where a soccer match was being played within the next two weeks.

Another burgeoning technology that will be adopted by cultural heritage institutions is the Internet of Things (IoT). IoT digitizes the physical world, and allows an information warehouse to digitally connect all of its electronic items and simultaneously collect, share, analyze, and project data and information (Geng 2016). Therefore, a library could digitally connect its data visualization screen with student checkout records, student grade and attendance data in the registrar's office, and student financial aid records to determine whether any of these factors correlates to students' academic success, and ultimately to high student retention and enrollment. Both of these above-mentioned horizon technologies allow for efficient access to various types of information beneficial to students, library staff, faculty, and possibly members of other cultural heritage institutions. They also can assist in generating synergized data that can help predict what types of behavior, financial assistance, information retrieval and study habits may correlate highly with student success, re-enrollment, and retention. However, each of these technologies also potentially expose students, staff, faculty, and other members of cultural heritage institutions to breaches of confidential data and information.

Each of these above-mentioned technologies gathers sensitive and private information pertinent to users. If cybercriminals implement one of the types of cyber attacks mentioned at the beginning of this chapter (e.g., unleashing ransomware, hacking into the servers holding such private information and data, or some other type of cyber attack); then student, faculty, staff, or other uses of these new technologies, may place their financial, medical, academic, or other

personal data and information at risk of being stolen by cyber criminals. Such misappropriated data and information may eventually be sold or used in another malevolent manner against the will of the original data owner.

Data Confidentiality

Digital tools hold a lot of promise in terms of empowering individuals to take control over their personal data. However, there is significant gap in terms of practices around different groups. For example, collecting data about vulnerable populations by humanitarian organizations may not adequately address the possible implications of collecting and using data about such populations (Vannini et al. 2019). Depending upon the importance and sensitivity of the data being shared, this may be especially critical for marginalized individuals, such as students, undocumented or irregular immigrant. Similarly, use of the digital tools increases the probability that a patron's data confidentiality may be violated as well. Anytime a patron utilizes any technology in a cultural heritage institution, he or she also should feel confident his or her data (research trails; pertinent information retrieved will also be kept confidential. Such confidentiality is vital so that patrons medical history, financial research, and religious preference are not exposed and coupled with his or her name, lest the public discover personally sensitive data about specific patrons. If such confidentiality is breached, basic tenants of all cultural heritage research institutions are eroded and patrons are likely not to return and utilize technology that helps them learn, discover, and synergize new information and data. Unfortunately, because cyber criminals persist, some degree of risk of breach of confidentiality is ever present. Thus, information warehouses must remain cognizant of such risks and perform every possible action to assure patron confidentiality.

Although a foolproof manner of preventing all types of cyber attacks will probably never exist, state and federal legislatures should exercise due diligence in ensuring laws are current to address the expeditious changes in technology and the sinister ways in which cybercriminals exploit such abrupt changes. Further, staff of cultural heritage institutions should ensure they utilize the most current cyber securities, both in software and hardware, to protect their patrons' privacy and confidentiality in data, information, and knowledge.

Data, Information, and Knowledge Privacy

Horizon technologies will affect data, information, and knowledge privacy. Although the tools to manipulate big data (including capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, preserving) are improving by the hour, the top big data challenges include information privacy. Jail (2016) noted that as more and more medical devices are designed to monitor patients and collect data, there is great demand to be able to analyze that data and then to transmit it back to clinicians and others. With increasing adoption of population health and big data analytics, we are seeing greater variety of data by combining traditional clinical and administrative data with unstructured notes, socioeconomic data, and even social media data. All these will only lead to increasing velocity of big data and presents data security challenges, as sensitive data can be anywhere–and therefore everywhere.

Tene and Polonetsky (2012) among other privacy advocates and data regulators, call for the development of a model where the benefits of data for businesses and researchers are balanced against individual privacy rights. This presents us with what Janes (2013) calls a classic balancing act: for instance, taking advantage of what libraries could learn from their

communities' habits, tastes and activities without crossing the line into what would be perceived as misuse of personal data.

A cornerstone value that library personnel perpetually promote is patron privacy. This value stems from the Association for Research and College Libraries (ACRL) and the American Library Association (ALA) – promoted fundamental right to have one's research, checkouts, search history, and other library conduct remain confidential, unless otherwise consented to by a patron. ACRL and ALA's privacy perspective supports current United States law opined in the United States Supreme Court case, Olmstead v. U.S., 277 U.S. 438 (1928), which conveys one of the most comprehensive rights a free people have is the right to be left alone. The ALA Bill of Rights (ALA Library Bill of Rights 2018) additionally suggests a lack of privacy for library patrons chills patrons' research choices and access to information, and further it undermines basic tenets of a democratic society (ALA Privacy and Confidentiality 2018). Thus, a library patron's right to privacy is a preeminent value for all information institutions. This right to privacy extends to all types of patrons' data such as checkout records, home addresses, phone numbers, gender, digital research trials, cached images on utilized library computers, and other digital trails. Ensuring privacy of such data is increasingly a challenge in information warehouses such as libraries as more and more data, information, and knowledge is offered globally via open access.

According to Verma (2018) and Raisaroa (2018), the top choices to secure big data were stronger authentication and access controls, monitoring and encryption. In most modern libraries, privacy is balanced with an effort to ensure global access to information to as many individuals as possible. Such global access, which relies upon intentions of open access, often necessitates the uploading of digital data, information, and knowledge which may overtly or covertly convey

health records, academic records, financial records, or other personal information or data to a university or college server, institutional repository, or other digital storage medium. However, some digital data, information, or knowledge that library or other information warehouses may store might not be intended to be accessed openly by the public. This data or information may concern student grades, patron research history, or other data or information normally considered to be private. Yet such data or information may still be uploaded to a library server. Uploading this data and information, whether it is intended to be open to the public or not, may result in privacy breaches via cyber criminals' malevolent conduct, data parasites, or other inadvertent breaches.

CASE STUDIES OF SECURITY AND PRIVACY ISSUES

Ransomware in St. Louis

Some case studies may further convey the possible pitfalls of digital access to data, information, and knowledge. Some data, information and knowledge stored in information institutions, such as libraries, may be susceptible to cyber breaches even though it is not proffered to the public via open access. In 2017, cyber criminals breached the information systems of seventeen separate public libraries in St. Louis, Missouri. As a result of this digital trespass, all seventeen libraries computer systems were infected with ransomware that encrypted most of the library system's digital files, and to unencrypt these files the cyber criminals demanded a $35,000 payment in the electronic currency Bitcoin. The cyber criminals effectively shut down the entire St Louis library system and destroyed the library staff's email system in a matter of minutes (Pagliery 2017). Until the sabotage was rectified, this caused patron and library staff angst due to them not having access to their normal information retrieval channels, and due to not knowing if confidential

information regarding their search history, personal data, and checkout histories would be made public.

Malware in Singaporean Health Records Database

In a separate unexpected attack in 2018, cybercriminals broke into a Singaporean health records database, SingHealth, and took the names and addresses of approximately 1.5 million medical patients, and the names of medicines dispensed to some of these patients. This breach occurred via malware through which the cyber criminals gained access to the personal health data. Just this year alone, numerous other nation-states claimed some of their governmental agencies have been hacked by cyber criminals such as Germany's government IT network and the United Kingdom's National Health Service (BBC Asia 2018).

These unanticipated criminal breaches of information entities exemplify that despite the best efforts of any security protocol, no information depository is impervious to cyber hackings. As a result, for libraries in the digital age, no patron can ever be completely confident his or her research history, checkouts, social media posts, grades, and even university digital health records are inpentrateable. For university library patrons in particular, student grades, courses taken, information literacy classes attended, entrance into a library, and other metrics are often collected by libraries to show correlations between library usage and student success in college or graduate school (LeMaistre et al. 2018). Due to these metrics being utilized and stored on library servers, cyber criminals could breach security protocols and obtain student health records, student grades, student or faculty research history, and utilize this data in malevolent guises. If such private student information and data is stolen, irreparable damage could be caused to students. Along with uploading students' data to information warehouse servers, promoting some

data and information via open access initiatives may also lead to malicious uses of data and information. These cyber criminals also seem to troll and assail universities and other related research institutions.

Cybersecurity Breaches Affect Universities and the Military

Cyberattacks also affect Universities and the military. In February of 2019, numerous universities' information technology systems were allegedly breached via Chinese hackers known as Temp.Periscope, Leviathan, and Mudcarp. These hackers targeted military defense information housed at various Universities such as the University of Hawaii, University of Washington, Duke University, Massachusetts Institute of Technology, Penn State University, and other Universities and Colleges around the United States, Canada, and Southeast Asia. During this cyberattack, hackers aimed to locate and steal United States, Canadian, and Southeast Asian military and economic secrets. The Universities and Colleges assailed in this cyberattack house key research institutes focused on undersea technology. Thus, it appears this is the focus of this cyberattack. Temp.Periscope, Leviathan, and Mudcarp are linked to previous cyberattacks where the hackers were seeking highly secure military information and data related to submarine missile creation and ship maintenance data (Volz 2019).

Allegedly, Temp,Periscope has targeted many United States Universities in the past because they tend to partner with military branches and usually have the digital infrastructure to house and preserve valuable military research. Hackers leverage the natural trust and desire to share information that most researchers display and promote

while working at a university or college. Hackers know because of this desire to share and build upon stored information and data, some researchers at colleges and universities may be more likely to click on a well produced spear phishing email. The universities and colleges hacked in this particular case probably had their information technology infrastructure hacked due to researchers clicking on spear phishing emails. Hackers have become well aware that researchers are willing to share their research, and they use this as a conduit in which to hack, steal, and illegally utilize sensitive military information (Volz 2019).

Academic libraries are often the location in which such sensitive data, information, and knowledge is digitally preserved. Scholarly communication departments often maintain digital scholarship repositories, digital data repositories, and other digital repositories that may house some of this highly classified military information. Since libraries personnel also naturally gravitates toward sharing information, this can be problematic when housing highly sensitive military data and information. Any data, information, or knowledge stored in a type of digital repository may easily be safely stored behind a dark archive perpetually. Such a dark archive theoretically ensures that only specific entities or individuals from those entities are able to access any deposited data, information, or knowledge located in that dark archive. That is partially why military institutions choose libraries in which to store sensitive information. However, as is mentioned in this chapter, cyber hackers constantly create new means to breach information technology security safeguards. Thus, no digital repositories are impenetrable. Knowing this, cyber criminals often target academic libraries to hack into, steal, and maliciously utilized highly

sensitive military data, information, and knowledge. Despite such technological vulnerabilities, academic libraries should continue to keep abreast of the best security measures that may prevent any data, information, and knowledge from being pirated, and then sold on the dark web, used for malicious purposes in other countries, or utilized in some other malevolent manner. Along with breaking into the digital infrastructure of academic libraries, cyber criminals also attempt to pillage data from small businesses too.

Small Businesses are Easy Targets for Cyber Assailers

Small businesses might assume they are non-targets from cyber criminals because larger corporations might have more data for criminals to quickly pillage and sell. This is an incorrect assumption however. Small businesses are frequently targets of sinister hackers because most cyber criminals are cognizant that most small businesses either cannot afford adequate antivirus protection, or simply do not have the awareness to keep such security software current.

One example of how easily cybercriminal can infiltrate small business is from a case study of Quaint Bakeries. A couple of years ago, two recent graduates of California Polytechnic State University commenced a vegan bakery that took orders online. They even hired a third party vendor to install and maintain what they thought was adequate software to protect their online assets. This third party also set up a virtual personal network to keep the bakery's IP addresses confidential and encrypt various Internet connections. However, this was not quite enough protection to prevent all cyber attacks (Strauss 2018).

While participating in a demo with a Microsoft store (lucky for the bakery, they learned their site was not full-proof via a demo rather than really losing valuable data to malicious hackers), the two bakers learned that their website could be easily spoofed. In other words, Microsoft successfully created a derivative of their site with one slight change in the sites URL. The bakers did not notice this small derivation and logged into the fictitious site thinking it was their actual website. Luckily this was just a demo and the bakers learned a valuable lesson (Strauss 2018).

However, many real cybercriminals create fictitious websites and embed links with malware on these sites. When one clicks on one of these links the malware can secretly install key-logging software on the user's computer, and then cyber hackers can discover exactly what a user is typing (Strauss 2018). Or worse, what the user is doing via video, or saying via various sensors.

Small businesses need to remember they are just as vulnerable as any other entity to cybercrime. They are often a more desirable target due to cyber hooligans' knowledge that owners of small businesses might not have the financial means or appropriate digital security knowledge to protect themselves. These small business entrepreneurs should educate themselves as did the Bakers at Quaint Bakery about valid and reliable products and services that may protect their digital assets from cyber criminals. Cyber criminals also target the secure data and information located in larger companies as well.

Cyber Breaches Occur in the Larger Corporate Arena as Well

Cybercriminals attack corporate information technology venues as well. In January of 2014, Ukrainian cyber-hackers broke into Target's information technology system and

stole up to 110 million customers private data, including credit card and debit card accounts, names, phone numbers, and email address. A month prior to this attack, similar cyber criminals hacked into Neiman Marcus's information infrastructure and pirated customer credit card data. One huge concern for Target, is that when a customer at this store purchases alcohol, a store clerk scans his or her driver's license (Jayakumar 2014). Thus, during this hack, millions of customer's driver's license data was stolen. Today, with a person's driver's license number, much other data can quickly be garnered online such as: residence and business addresses and phone numbers, public civil and criminal records may be quickly tracked, and other personal data.

At least some of the perpetrators of this crime have been arrested and prosecuted. Rusland Bondars, a Latvian citizen was arrested, found guilty, and sentenced to 14 years of prison for designing a program that helped hackers improve malware. This malware was used by hackers later to breach Target's information infrastructure. Those criminals who breached Target's system first used their developed Scan4You malware to determine whether an antivirus program would recognized their software as malicious. Cybersecurity officers believe the other hacker responsible for this hacking of Target and possibly Neiman Marcus is "Profile 958," is likely a Ukrainian named Andrey Hodirevski (Weiner 2018). Why do hackers do when they steal personal and private data from corporate, educational, governmental, and public sector entities?

Cybercriminals Quickly Make Money Off of Stolen Data

After breaking into a business, educational, governmental, or other public sector digital database, cybercriminals quickly sell stolen private and personal data on the dark web. The dark web is a part of the Internet not discoverable by traditional search engines such as Google of Firefox, and

is only discoverable via special web browsers such as Tor. The dark web accounts for approximately less than .01% of the Internet. The entire Internet itself is known to be broken into three parts. The surface web, the deep web, and the dark web. The surface web makes up about 10% of the Internet. Information, data, and knowledge located in the surface web may easily be retrieved via simple search from most web browsers such as Google and Internet Explorer. An example of surface web content is a gaming company selling legally downloadable video games via their commercial website. The deep web consists of 90% of the Internet, and offers data, information, and knowledge to users via search engines such as scholarly books or articles located in databases protected by a paywall. An example of this would be when a student searches in EbscoHost and downloads a scholarly article via his or her university or college username and password. The dark web is a part of the deep web, consisting of about .01% of the Internet, where much illicit trading of black market items and services occurs. Using primarily the web browser called Tor, stolen data, information, and knowledge is often illegal traded and sold (Ablon 2014).

Cybercriminals have a complete logistics system setup on the dark web on which stolen data can be marketed, sold, delivered and lightning speed to buyer via a sophisticated logistics system, and implemented via buyers for their malicious purposes (Ablon 2014). Thus, when cyber criminals are able to successfully break into and private millions of customers' private data, they are highly likely to have a quick return of possibly hundreds of thousands of dollars to possibly millions of dollars by selling this data within minutes on the dark web.

What Recourse Do Victims of Cyber-Crime Have?

Of course, when one falls victim to cyber-crime, the natural response other than feeling violated and somewhat helpless is, is there any legal recourse, or anywhere one may turn to for help. While receiving legal help from cybercrime is somewhat cumbersome due to cybercriminals usually residing in legal jurisdiction from where the victim's data or information is stolen, there are some inter-jurisdictional agencies to where one may attempt to seek help. For example, one may contact the Internet Crime Complaint Center (IC3) https://www.ic3.gov/default.aspx and file a complaint regarding cybercrime. The complaint should include the victim's name, address, telephone, and email; financial transaction information; specific details regarding how one was victimized; and any other relevant information.

If one falls victim to a phishing email scam, or if one simply receives an email phishing scam attempt, one may forward the attempt to reportphishing@antiphising.org (the Anti-Phishing Working Group (APWG)) and to phishing-report@us-ert.gov where it will be received by the United States Computer Emergency Readiness Team (US-CERT). The APWG is a private, international work group that tracks various types of cybercrime, and attempts to deter such future cybercrime. The US-CERT group is an arm of the United States Department of Homeland Security, and attempts to garner information about cybercrimes and subsequently better educate the its denizens regarding awareness and potential dangers of certain Internet activity (Smith 2016).

If an individual finds him or herself in one of the mass data/information breaches similar to the ones mentioned above (e.g., Target), he or she may enroll in the online account monitoring software called WebWatcher. This system offers reimbursement for monetary loss due to cyber breaches and fraud up to one million dollars. When one enrolls in these services, he or she also

obtains fraud counseling services and reimbursement coverage for free (Smith 2016). So, though sometimes a labyrinth to navigate, some sites are prepared to attempt to help victims of cyber-crime when it affects people.

Lessons Learned

It is apparent that cyber criminals target any and all type of organization when seeking to quickly breach an information technology security barrier and steal private data and information. All sectors are at risk, government, academic, small business, large corporations, non-profit charities… The motivation behind pirating private data and information is that cyber criminals can quickly sell such stolen goods and make a large profit on the dark web. Cyber security simply does not have the technology nor adequate person power yet to effectively patrol the dark web. There are too many discrete locations and ways for cyber criminals to complete illegal sales on the dark web. There are government and private agencies that can lend some help to victims of cybercrime on the dark web. However, all cultural heritage institutions need to remain cognizant of the possible malware and cyber-attacks to which their information infrastructure and their patrons may be susceptible. Such awareness needs to prompt information professionals to ensure the utilization of the most valid and reliable anti-malware software and that its patrons are properly notified of all possible risk when using digital materials in their institution. Information professionals should also actively advocate for the passage of updated laws that offer strict punitive consequences for those who commit cyber-crimes, and advocate for adequate funding with which to invest in proper safety measures to prevent cyber-crimes.

SUMMARY AND CONCLUSION

Emerging trends and horizon technologies are allowing cultural heritage institutions to develop new ways of gathering, preserving, analyzing, and synergizing data, information, and knowledge. These new technological endeavors offer great benefits to global humanity, but also, as with any new development, open new opportunities to individuals with malicious intent. As many data, information, and knowledge warehouses adhere to the global open access movement; opportunities for data breaches are further apparent. Dichotomously opposed to such data, information, and knowledge breaches; most cultural heritage institutions, such as libraries, adhere to valuing the utmost guarantees of privacy and confidentiality for their patrons. Such strong dedication to privacy and confidentiality is exemplified in international library policy and is further reflected in international law. Despite such efforts of cultural policy makers and legislators, cyber criminals continue to implement malevolent tactics such as ransomware, viruses, worms, spyware, trojans, phishing, pharming, and other malicious endeavors.

Some of these types of cyber attacks result from well-meant initiatives to globally share data, information, and knowledge. Dissertations, theses, e-journals, data, art, and other types of scholarship are increasingly deposited into open access digital repositories to allow and promote universal access. Promoting such access, unfortunately, sometimes leads to data parasites stealing researchers' data, or other cyber criminals pirating scholarly works such as theses or dissertations. These cyber criminals then claim such stolen information and data as their own, plagiarize it, offer no credit to the original creator(s), and often utilize this pilfered information, data, and knowledge for illegal commercial purposes.

The current state of the dark web also presents a challenge to all types of cultural heritage institutions based in government, academic, corporate, and not-for-profit sectors. Cyber

Criminals may quickly pirate private information and data from these entities and turn a quick profit by selling these goods on the dark web. To help reduce the illegal behavior being carried out on the dark web, more research needs to be completed about how the dark web works, more effective technology needs to be implemented to help reduce dark web crime, and new laws need to be passed that can enable security to better monitor the dark web and prevent and reduce the percentage of cybercrime occurring on the dark web.

Increasingly, organizations are reassessing their operational readiness to detect and respond to a breach. The European Union has addressed some data theft and misuse via the General Data Protection Regulation (GDPR). In addition to privacy, thanks to GDPR's stringent breach notification regulation, organizations have revamped their incident response programs over the past year to meet the requirements. With the recent release of the National Cyber Strategy (White House, 2018), the United States now has its first fully articulated cyber strategy. However, the United States has not recently passed proposed legislation into law to address modern data and information misuse. A proposed model statute in this chapter could offer insight into how the United States federal government, and the 51 state and district legislative bodies could provide legal guidance to this issue.

In the current data-intensive environments, all global standard organizations and legislative bodies will be precipitously challenged in perpetuity to continuously update standards and legislation as horizon technologies arise that alter the way in which data is collected, preserved, utilized, shared, and synergized. No wonder most ISO27k standards, which include many aspects of information technologies, are under review on an ongoing basis, and the publication of ISO/IEC 27045 (the specific standard for big data security and privacy) may not even be expected until the year 2022.

In today’s cybersecurity landscape, realizing the promise of open access and big data may well depend on our ability to continue our quest to maintain our cyber-readiness in the face of ever evolving threats. With appropriate legal policies, guidelines, and a combination of right technology, right people and right processes in place, organizations and nations at large will be able to contain damage and minimize risk when (yes, it's a matter of "when", not "if") they are breached and digitally attacked.

References

Ablon, L., Martin C. Libicki, and Andrea A. Golay, "Markets for Cybercrime Tools and Stolen Data: Hackers' Bazaar" (Washington D.C.: Rand, 2014).

Abrizah, A., Mohd Hilmi, and Norliyha Ahmad Kassim, "Resource Sharing Through an Inter-Institutional Repository: Motivations and Resistance to Library and Information Science Scholars," *The Electronic Library* 33, no. 4 (2015) 730-748, https://doi.org/10.1108/EL-02-2014-0040

"BBC Asia," Singapore Personal Data Hack Hits 1.5m, Health Authority Says, last modified January 22, 2019, https://www.bbc.com/news/world-asia-44900507.

"European Commission," Policies, Information and Service, last modified July 22, 2018. https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en

"Federal Bureau of Investigation," Internet Crimes Complaint Center, last modified January 22, 2018, https://www.ic3.gov/about/default.aspx

Firican, G., "The 10 Vs of Big Data . A publication in Transforming Data With Intelligence (TDWI)," (2019), https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx

Geng, H., "Internet of Things and Data Analytics in the Cloud with Innovation and Sustainability." In *Internet of Things and Data Analytics Handbook,* ed. Hwaiyu Geng (Hoboken, NJ: John WIley & Sons, 2017), 3-128.

Helge, K., "Law Student Information Seeking, and Understanding of Citation, Common Knowledge, and Plagiarism." In *Knowledge Discovery and Data Design Innovation,* eds. Daniel G. Alemneh, Jeff Allen, and Suliman. Hawamdeh (London, England: World Scientific, 2017). 249-263.

Helge, K. and Laura McKinnon, The Teaching Librarian: Web 2.0, Technology, and Legal Aspects (Amsterdam: The Netherlands: Elsevier, 2013).

“Intersoft Consulting,” General Data Protection Regulation (GDPR), last modified June 2, 2018, https://gdpr-info.eu/

Iso/iec 27045 - Information Technology - Security Techniques - Big Data Security and Privacy Processes (draft), last modified April, 2018, http://www.iso27001security.com/html/27045.html.

Jain, A. "The 5 Vs of Big Data," last modified January 12, 2019, https://www.ibm.com/blogs/watson-health/the-5-vs-of-big-data/..

Janes, J., “Balancing Privacy & Innovation | Reinventing Libraries,” In *The Digital Shift*, last modified November 15, 2018, http://www.thedigitalshift.com/2013/08/uncategorized/balancing-privacy-innovation-reinventing-libraries/

Jayakumar, A., “Target Tries to Reassure Customers After Data Breach Revelations,” *The Washington Post Online Edition*, last modified January 2014, https://www.washingtonpost.com/business/economy/target-tries-to-reassure-customers-after-data-breach-revelations/2014/01/13/3c0323e0-7c7e-11e3-95c6-0a7aa80874bc_story.html?utm_term=.009b25a8e9fa

LeMaistre, T., Shi Qingmin, and Sandip Thanki. "Connecting Library Use to Student Success." *Libraries and the Academy* 18, no. 1 (2018) 117-140, https://muse.jhu.edu/

Longo, D. and Jeffrey Drazen, “Data Sharing.” *New England Journal of Medicine* 374 (2016) 276-277, doi: 10.1056/NEJMe1516564

Olmstead v. U.S., 277 U.S. 438 (1928).

Pagliery, J., “St. Louis' Public Library Computers Hacked For Ransom”, last modified February 1, 2017, https://money.cnn.com/2017/01/19/technology/st-louis-public-library-hack/index.html.

Pennoyer v. Neff , 95 U.S. 714 (1878). 17 U.S.C. 106 (2018).

Strauss, S.. “How do cyber criminals hack small business startups? Here’s what we learned from Microsoft,” *USA Today Online Edition,* last modified October 17, 2018.

Tene, O., and Jules Polonetsky, “Privacy in the Age of Big Data: A Time for Big Decisions,” last modified January 22, 2019, https://www.stanfordlawreview.org/online/privacy-paradox-privacy-and-big-data/.

Raisaroa, J. L.,, “Are Privacy-Enhancing Technologies For Genomic Data Ready For the Clinic? A Survey of Medical Experts of the Swiss HIV Cohort Study,” *Journal of Biomedical Informatics*, 79, (March 2018), 1-6, https://doi.org/10.1016/j.jbi.2017.12.013.

Smith, B. R., “Don’t Step in the Trap: How to Recognize and Avoid Email Phishing Scams” (CreateSpace Independent Publishing Platform: 2016)

“Thales Data Threat Report: Trends in Encryption and Data Security - Global Edition,” last modified January 20, 2019, http://go.thalesesecurity.com/rs/480-LWA-970/images/2018-Data-Threat-Report-Global-Edition-ar.pdf.

“University Libraries Data Repository Services,” last modified August 1, 2018, http://www.library.unt.edu/datamanagement/data-repository-services.

Vannini, S., Ricardo Gomez, and Bryce C. Newell, “Documenting the Undocumented: Privacy and Security Guidelines for Humanitarian Work with Irregular Migrants.” In *Lecture Notes in Computer Science: 14th International Conference, iConference 2019,* Washington DC, USA, March 31-April 3, 2019, last modified March 8, 2019, https://www.conftool.com/iConference2019/index.php?page=browseSessions&form_session=365

Van Rijmenam, M., “Why The 3V’s Are Not Sufficient To Describe Big Data,” last modified June 14, 2018, https://datafloq.com/read/3vs-sufficient-describe-big-data/166

Verma S., “Data Confidentiality in Public Contracts: Why Typical 'What Me Worry' Attitudes May Not Really Be an Acceptable Position for Government Contracting Professionals Anymore” *The 8th International Public Procurement Conference (IPPC)* , Arusha, Tanzania, (August 2018), http://dx.doi.org/10.2139/ssrn.3159135.

Volz, D., “Chinese Hackers Target Universities in Pursuit of Maritime Military Secrets: University of Hawaii, University of Washington. And MIT are Among Schools Hit by Cyberattacks” *The Wall Street Journal Online*, Politics, National Security, (March 2019).

Weiner, R., “Hacker Linked to Target Data Breach Gets 14 Years in Prison,” *The Washington Post Online Edition*, last modified September 21, 2018, https://www.washingtonpost.com/local/public-safety/hacker-linked-to-target-data-breach-gets-14-years-in-prison/2018/09/21/839fd6b0-bd17-11e8-b7d2-0773aa1e33da_story.html?noredirect=on&utm_term=.e3b5fd6574df

“White House National Cyber-Strategy of the United States of America,” last modified September 1, 2018, https://www.whitehouse.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf

Woods, T., “GDPR's Impact on Incident Response,” last modified April 25,, 2019. https://securitytoday.com/articles/2019/04/24/gdprs-impact-on-incident-response.aspx

“XSI,” The V's of Big Data: Velocity, Volume, Value, Variety, and Veracity, last modified February 1, 2019, https://www.xsnet.com/blog/bid/205405/the-v-s-of-big-data-velocity-volume-value-variety-and-veracity.