Presented to the Interdisciplinary Studies Program:

UNIVERSITY OF OREGON
APPLIED INFORMATION MANAGEMENT

Applied Information Management
and the Graduate School of the
University of Oregon
in partial fulfillment of the
requirement for the degree of
Master of Science

Mitigating the Risks of Insider Threat on Unstructured Data through Data Governance

CAPSTONE REPORT

Michael C. Egli
AIM Program
University of Oregon

University of Oregon
Applied Information
Management
Program

May 2016

Academic Extension
1277 University of Oregon
Eugene, OR 97403-1277
(800) 824-2714

Approved by

Dr. Kara McFall

Director, AIM Program

Mitigating the Risks of Insider Threat through Data Governance

Michael C. Egli

Varonis Systems

Abstract

This paper examines the growing risk of insider threat on unstructured data (Gartner Research, 2015). This review of selected literature identifies the risks and challenges in managing unstructured data, and presents best practices for risk mitigation, security and audit controls, compliance implications, and technical processes. With implementation of these practices, it is feasible that organizations can reduce the probability and impact of an insider breach on their unstructured data.

Keywords: data governance, unstructured data, insider threat, risk management, compliance, information security

Table of Contents

List of Tables and Figures

Introduction

Problem

Inmon and Nosavich (2008) define unstructured data as an unstructured system that has form but is rich with text such as emails, contracts, reports, transcriptions, and other documents. The information security industry currently faces exponential growth in unstructured data. Gartner (2013) estimates that the amount of this type of data is growing between 40-60% each year, and it is estimated to increase by 800% from 2012 to 2022 (Berry, 2012). As a result, organizations are struggling with not just how to manage the storage of unstructured data, but also how to continue to make it useful and secure in light of insider threat breaches (Mearian, 2010). Recent examples of high-profile data breaches that included high volumes of unstructured data loss include those at Sony and Target and the unauthorized leaking of National Security Administration (NSA) confidential data by former Central Intelligence Agency (CIA) employee Edward Snowden (Ponemon Institute, 2015a).

Gartner Research (2015) has coined the term “dark data” and describes it as “the information assets organizations collect, process, and store during regular business activities, but generally fail to use for other purposes (for example, analytics, business relationships and direct monetizing)” (p.1). Unstructured data now accounts for nearly 95% of a typical organization’s data (Tanwar, Duggal, & Khatri, 2015). In addition, the International Data Corporation (IDC) indicates that over 90% of unstructured data is never analyzed (Pornain, 2014) for the purposes of data classification. The lack of analysis and rapid growth of unstructured data have created challenges for organizations, including difficulties in selecting technology to assist with the management, storage, and security of the data; the lack of insights from the data; and the mitigation of associated risks such as insider threats (Taylor, 2013).

Recent breaches of unstructured data at organizations such as Sony, the National Security Agency, eBay, JPMorgan Chase and Co., and Home Depot have "raised senior managements" level of concern about how cybercrimes might impact their organizations" across the United States (Ponemon Institute, 2015b, p.1). In a study by the Ponemon Institute (2015b), organizations that were surveyed in 2015 indicated that 67% of them saw an increase in budget over the last year aimed at defending data breaches. However, in the same study, the Ponemon Institute found that of the organizations with breach detection technology, 45% of those who had discovered a breach caught it by accident.

Research suggests that although many organizations have implemented technology to improve security controls, the methodologies to manage the unstructured data and the security controls surrounding them require improved practices (Ponemon Institute, 2015a). In a survey performed by the Ponemon Institute (2015b) of 1,006 senior-level leaders in information technology and information technology security in the US, UK, Europe, Middle East, and North Africa, respondents indicated that technology spending to improve security controls through the application of technology increased an average of 34% from 2014 to 2015 (Ponemon Institute, 2015b). In addition, 35% of these organizations acknowledge that a principle cause of unstructured data security issues is the lack of in-house expertise, 15% cite poor leadership, and 12% identify an incomplete knowledge of where sensitive data exists. A study by Vormetric (2015) consisting of 800 senior business managers and IT professionals indicates that while 49% of the respondents indicate that sensitive data for organizations resides within databases and 39% within file servers, 70% of decision makers were more focused on mobile device protection than addressing the volume of at-risk data on these inside storage devices. Although executives have increasing concerns about insider threats as a result of recent breaches, and these concerns have dramatically increased spending on technology, systems administrators and information security practitioners are concerned that the methods in place for securing unstructured data are not working (Ponemon Institute, 2015a).

Organizations that are facing challenges from unstructured data will benefit from data governance strategies. However, industry experts such as Christophe Marcant (2015), Vice President of Product Strategy at Stibo Systems, indicate that while there are many tools available that offer governance of unstructured data, they only offer solutions to execute data governance activities rather than aiding in the definition of the organizational goals and methods inherent within a data governance strategy. Organizations that do not currently manage their unstructured data will benefit from the documentation of industry best practices that address the need for data governance strategies, thus enabling the organizations to begin to turn dark data into managed unstructured data. Doing so can greatly reduce the associated risk of exfiltration (Gartner Research, 2011) and the resulting expense and loss of customers and company reputation (Filkins, 2015).

Purpose Statement

The purpose of this annotated bibliography is to present literature that addresses the need of organizations for a vendor-neutral data governance strategy that concentrates on unstructured data and the risk of insider threats such as exfiltration. This document endeavors to provide resources on best practices for data governance of unstructured data so that organizations can implement methodologies to leverage, secure, and manage human-generated data to meet increasing demands. In addition, this study provides sources that define unstructured data and provide examples of threats that organizations have faced with regard to the securing of this key organizational asset.

Research Question

What security *best practices* can organizations implement as part of their data governance plans to reduce the risk of exfiltration and the exposure of unstructured data?

Audience

Unstructured data, mitigating the risks of insider threat, and methods to improve data governance strategies are of interest to Chief Information Officers (CIOs) and Chief Information Security Officers (CISOs) according to a study performed by the Ponemon Institute (2015a). This interest has spawned an increase in technology spending in the area of improving security controls an average of 34% since 2014, yet many information technology workers are concerned that this spending is insufficient. In a 2015 study conducted by the Ponemon Institute, 66% of the 1006 respondents indicated that their organizations need more knowledge to effectively manage the threats (p. 16). Therefore, it is crucial that Chief Financial Officers (CFOs) are also informed on these issues as they have an interest in ensuring the budgets allocated to security spending are sufficient to address the continued effectiveness and efficiency of unstructured data, while also reducing the risk of loss.

The collection and reporting of best practices for unstructured data governance assists in the creation of effective information security policies and the successful application of technology (Marcant, 2015). For example, a study performed by the Ponemon Institute (2015b) identified that of the organizations that had experienced a data breach, 65% indicated the attack evaded existing security controls, yet only 9% of organizations indicate they will spend budget on sensitive data management as a result. CIOs, CISOs, and CFOs will benefit from a resource that documents data governance best practices to improve upon the implementation of strategies for ongoing management and security of unstructured data that are now the target of many exfiltration attempts.

Search Report

Search strategy. The principal method for identifying relevant resources for this research is centered on the following three steps:

- Identifying key words and themes;
- Accessing libraries and research institute databases; and
- Correlating data between various topics to draw conclusions based on the intersections between discussions on insider threats, data breaches, big data, unstructured data as a subset of big data, data governance, IT spending, IT security policies, incident response, and risk management.

Evaluation criteria. Sources are evaluated using The University of Florida's Center for Public Issues Education (2014) evaluation criteria. This evaluation process includes identifying if the source demonstrates bias, is authoritative, if the content is reflective of quality standards, if the material is timely, and if the content of the source is relevant to the subject of this paper.

Bias. The topic of this paper is based on a growing issue with insider threats and the unstructured data (Gartner Research, 2015); at the time of this research study there are limited scholarly resources available on this new and narrow topic. Therefore, vendors who have stakes in the outcomes perform much of the research on unstructured data. To mitigate the risks associated with the use of potentially biased sources, content is evaluated for *authority* and *accuracy* through the following means: Content is included within this paper if there are statistics and knowledge gained directly from survey results rather than a presentation of the author's opinion. Resources used by research firms should be surveys that include respondents from multiple industries and positions, or the source must provide clear recommendations on resolving issues such as data leakage in a vendor-neutral manner without the requirement of specific software purchases. Portions of papers written by vendors who have a stake in selling a service must clearly mark which areas are written by the research firm and which are written by the vendor so that a clear distinction can be drawn between survey results written by an authority in comparison to material that reflects a strong potential bias.

Authority. Except for the exceptions stated above, academic sources are evaluated by analyzing their citations and verifying that the document has been peer reviewed and that source materials are relevant to the topic of this paper.

Quality. Evaluation of all sources includes the review of punctuation, grammar, flow, and presentation to validate a level of professionalism exists for the formal publication.

Timeliness. The topic of study is a relatively new subject; therefore, sources that have been published after 2010 are sought.

Relevancy. A source is selected if it has a key focus on either unstructured data as a main focus, or as a subset of Big Data.

Documentation approach. References are documented within three individual locations. Initially, they are stored using bookmarks for further review. Once a source is vetted and determined to meet the evaluation criteria for inclusion in this study, it is moved to Zotero alongside a description and the abstract for quick reference. Finally, the source is included within the main paper under references and if applicable as an entry in the annotated bibliography. The annotated bibliography and references section grow throughout the research process to build a solid foundation for the paper as a whole, which aids in writing other sections of the paper once the primary research phase is completed.

Key words. Key words, themes and topics include:

- Unstructured data;
- Data governance as it relates to unstructured data;
- Data breaches;
- Insider threat and unstructured data breaches;
- IT security policy and data governance;
- Incident response to data security and data breaches;
- Risk management and analysis approaches to data governance;
- Exfiltration, privilege abuse;
- Impacts of security breaches such as monetary loss and cost of recovery;
- Mitigating risks related to insider threat (exfiltration, abuse, etc.);
- Unstructured data management practices for common industry compliances (HIPAA, PCI-DSS, SOX); and
- IT spending and budgetary impacts.

Databases. The following databases are used to identify appropriate sources for this study:

- Gartner Research;
- Business Search Complete;
- Academic Search Premier;
- Sciencedirect.com;
- Google Scholar; and
- JSTOR.

Annotated Bibliography

The following Annotated Bibliography contains references that provide information on the impact of insider threat on unstructured data. Topics are focused on key areas including (a) challenges and risks with managing unstructured data, (b) best practices for mitigating unstructured data risks, and (c) best practices for data governance management. The selected references are intended to set a baseline of best practices that information security and information systems administrators can use to implement improved unstructured data management practices that are focused on security, as well as inform executive-level decision makers on the importance of implementing a corporate strategy for data governance.

Each annotation consists of three individual elements: a bibliographic citation, an abstract, and a summary. Abstracts are provided verbatim when available. Summaries are intended to highlight the key topics presented in the material that best align with the core intent of this paper: To provide insight on best practices on unstructured data governance, and how to reduce insider threat as a result.

Challenges and Risks with Managing Unstructured Data

Berry, D. (2012, April 9). Unstructured data: Challenge or asset? Retrieved March 30, 2016, from http://www.zdnet.com/article/unstructured-data-challenge-or-asset/

Abstract. Diane Berry presents many statistics backed by major research firms including IBM, Gartner and the IDC to instruct the reader that unstructured data is a challenge for most organizations. The major struggles experienced by study participants revolve around the inability to gain value from unstructured data. Coveo, the organization that the author works for, performed a study that indicated that 85% of executives felt that management of unstructured data would have an impact on their ability to serve customers (2012), yet a study performed by IBM of 1,500 CEO’s indicated that they have little insight and lack the ability to “transform available data into feasible action plans” (2012).

Summary. The article references a number of sources including studies by IBM, Gartner Research and Coveo that relate to the effects of growth in unstructured data. A study including 1,500 CEOs performed by IBM indicated frustration over how to utilize unstructured data, while Gartner Research predicts that there will be a growth of enterprise data of 800% between 2012-2017, of which 80% is unstructured. The author indicates that unstructured data is being semi-structured through the use of powerful engines that index the data, making it useful for enterprises to meet the need identified by the CEOs mentioned in the IBM study above. This indexing process is intended to assist organizations with leveraging the enormous growth in unstructured data so that the information within can become an asset. This has led to a paradigm shift where stakeholders are investing greatly in technology that is predicted will make this data useful and lead to huge economic value.

This article is useful for this study because it provides context for the value of unstructured data and the challenges that occur with rapid data growth and a general lack of visibility into what is contained within these documents.

Bertino, E. (2013). *Big Data - Opportunities and Challenges*. West Lafayette, Indiana: Purdue University.

Abstract. Recent technological advances and novel applications, such as sensors, cyber-physical systems, smart mobile devices, cloud systems, data analytics, and social networks, are making possible to capture, process, and share huge amounts of data –

referred to as big data - and to extract useful knowledge, such as patterns, from this data and predict trends and events. Big data is making possible tasks that before were impossible, like preventing disease spreading and crime, personalizing healthcare, quickly identifying business opportunities, managing emergencies, protecting the homeland, and so on [1]. As discussed by The Economist [2] "Managed well, the data can be used to unlock new sources of economic value, provide fresh insights into science and hold governments to accounts". Unlocking the potential of big data requires however addressing several major challenges. The goal of this panel is to identify and discuss research directions to address these challenges. In what follows, we first discuss the notion of big data and application domains where big data is relevant. We then outline relevant challenges and summarize questions addressed by the panel.

Summary. Bertino explains the importance of what big data provides such as the ability to gain knowledge and turn that knowledge into practice. Leveraging big data allows for the advancement in technology such as identifying trends in behavior that can lead to improved analysis or differences in application behavior to meet user needs. However, Bertino explains that this same volume of knowledge also poses an issue as the data itself can be used for other harmful purposes. For example, the data could provide the ability to gain knowledge on groups of people for the purposes of discrimination, which highlights the importance of population and personal privacy as a principle concern when managing large volumes of data. This critical component to the security of big data must include the resolution of related challenges, including the difficulties of securing access to unstructured data to only the users that require it through scalable security administration, inefficient management of unstructured data due to the sheer volume of files and folders, and integration of data security policies to govern the practices of security and management. Other important challenges that remain in the domain of data governance are how information can be cleaned when the information is no longer relevant and assuring that only relevant data is stored.

Blanchard, R., & O'Sullivan, K. (October, 2015) Big data risk and opportunity: Having an action plan to address both can add tremendous value to the organization. *Internal Auditor*. Retrieved from http://go.galegroup.com/ps/anonymous?id=GALE%7CA434320765&sid=googleScholar&v=2.1&it=r&linkaccess=fulltext&issn=00205745&p=AONE&sw=w&authCount=1&isAnonymousEntry=true

Abstract. To an internal auditor, just the term big data can elicit a sinking feeling. The challenges associated with the volume, complexity, and variety of big data can be overwhelming. The good news is, with a solid action plan, internal auditors can do more than just mitigate the risks associated with big data. Internal audit also can help exploit big data to identify and mitigate existing risks. Big data is the collection of data sets that are so large and complex that they are difficult to process using conventional database tools. Big data comes in two flavors: Structured data (e.g., data in spreadsheets and databases) and unstructured data (e.g., social media posts, emails, audio, video, and GPS data). And, of course, big data can have multiple sources. Typically, working with big data requires new technologies to identify usable business insights, trends, and correlations — often in real time.

Summary. This document approaches big data and unstructured data from an auditor’s perspective and specifically focuses on identifiable risks that the data poses. The authors focus on personally identifiable information and the ownership of that data, the impacts of regulatory compliance on the potential sensitive content within documents, exposure to reputation risk if a data breach occurs, and data retention policies. The document provides clear guidance on how to address these issues by performing a combination of internal audits and creating organizational plans to specifically address governance over big data. At a high level, the approach involves a five phase plan which includes: (a) identifying where data resides and who owns it, (b) defining organizational goals regarding the data, (c) assessing critical data issues, (d) identifying key risk indicators (KRIs), and (e) identifying opportunities to add value. The intention is to drive an organization through this plan and perform routine audits to reduce the risk of the organization in the event of privacy breaches.

Miller, P. (2014). Applying big data analytics to human generated data. Austin, TX: Gigaom Research. Retrieved from https://gigaom.com/report/applying-big-data-analytics-to-human-generated-data/

Abstract. As the analytics industry moves to address an explosion in machine-generated data, another opportunity is already here. Emails, texts, documents, and other unstructured human-generated data -- and the metadata associated with them -- deliver significant insight to businesses with the resources and will to mine them. Taking control of human-generated data provides companies with a more complete understanding of their intellectual property, enables them to aggregate business intelligence for sharing with employees, and allows security professionals to identify and mitigate both casual and deliberate breaches of policy. However, the operational cost of normalizing and mining this data is significant and requires a sound strategic understanding of technologies and goals. This research report evaluates the opportunities and challenges associated with analyzing human-generated data. It examines early adoption in the risk management and governance use cases, and evaluates the potential impact of these analytics for other use cases and industries. Key findings include: Human-generated data in word-processed documents, presentations, spreadsheets, and emails typically comprises an organization’s most prized assets, including key intellectual property, operating procedures, and the plans and strategies that shape future development. Most organizations fail to adequately manage the creation, use, and dissemination of these key assets. As a result, they either introduce friction into collaboration through excessively strict access controls or risk serious data loss by sharing data too permissively. Tools and techniques from the big data sector offer the means to monitor human-generated data across an organization’s different IT environments, protecting key assets and ensuring that regulatory obligations are met in a cost-effective and timely manner. Data governance, audits, and other regulatory requirements are typically the initial drivers for deployment of these technologies, but other opportunities present themselves once systems and procedures are in place. The same tools, for example, can identify individuals and teams in different parts of a large organization who happen to be accessing similar resources without knowledge of one another, brokering introductions to teams that may be tackling complementary problems unwittingly.

Summary. The conflict between an organization’s desire for simplicity for file access rights, such as enabling every user to access any file, and the inherent security risks involved with this approach are discussed in this document by Gigaom Research. The research company indicates that challenges in managing data access are not only limited to understanding who should have access to which data, but that organizations are also challenged by inadequate resources to both identify who needs access and then implement the changes as these efforts are cumbersome. The various technologies and processes impacted by unstructured data are requiring more resources within the Information Technology (IT) organization, yet budgets are not available to address the need. Lastly, organizations struggle to mitigate the risks resulting from overly permissive access rights, a lack of visibility into who should and should not have access, and a lack of staff to make permissions changes and routinely audit them.

Ponemon Institute. (2015a). 2014: A year of mega breaches. Traverse City, MI. Retrieved from http://www.ponemon.org/local/upload/file/2014%20The%20Year%20of%20the%20Mega%20Breach%20FINAL3.pdf

Abstract. 2014 will long be remembered for a series of mega security breaches and attacks starting with the Target breach in late 2013 and ending with Sony Pictures Entertainment. In the case of Target breach, 40 million credit and debit cards were stolen and 70 million records stolen that included the name, address, email address and phone number of Target shoppers. Sony suffered a major online attack that resulted in employees' personal data and corporate correspondence being leaked. The financial consequences and reputation damage of both breaches have been widely reported.

Summary. The Ponemon Institute performed a study that included 735 IT and IT security practitioners, of which 2% were executives, 38% managers, and the remainder a mix of consultants and IT administrators. This study provides statistics on how the high profile security breaches in 2014 such as those experienced by eBay, JPMorgan Chase & Co., Home Depot, CHS community Health Systems, Target and Sony Pictures

Entertainment affected IT spending, decision making and the overall focus in managing Information Technology. A significant focus of this document is the technology investments and operational changes made after 2014. While data governance as a concept was not directly addressed in this document, related topics are discussed such as the impacts on an organization after data exfiltration occurs; loss in brand awareness, brand value and productivity due to a data breach; and the need to purchase new technologies to support better security. An important statistic from this study that relates directly to challenges and risks with managing unstructured data is that 65% of survey participants indicated that attacks evaded existing preventative security controls and 37% indicated that insufficient funding for technology to monitor and prevent breaches was a related cause (p.10). Following the 2014 breaches, additional budget was granted to IT to spend specifically on mitigations for the risk of data exposure; however, participants acknowledged that additional spending may not address the issues as they expect technology will not mitigate the risks and consequences of the breaches by itself (p.11).

Ponemon Institute. (2015b). *2015 Global Megatrends in Cybersecurity*. Traverse City, MI. Retrieved from http://www.raytheon.com/news/rtnwcm/groups/gallery/documents/content/rtn_233811.pdf

Abstract. We are pleased to present the findings of the 2015 Global Megatrends in Cybersecurity sponsored by Raytheon. The purpose of this research is to understand the big trends or changes that will impact the security posture of organizations in both the public and private sector in the next three years. Moreover, the study looks at the next generation of protocols and practices as the cybersecurity field evolves and matures. We surveyed 1,006 senior-level information technology and information technology security leaders (hereafter referred to as respondent) in the US, UK/Europe and Middle East/North Africa (MENA) who are familiar with their organizations' cybersecurity strategies. The research covered a range of trends related to an organization's ability to protect itself from cyber threats and attacks. Some of the areas addressed in this report are: The critical disconnect between CISOs and senior leadership, insider negligence, the Internet of Things, adoption of new technologies such as big data analytics, predictions of increases in nation state attacks and advanced persistent threats and the dearth of cyber talent.

Summary. The Ponemon Institute performs a study each year on the previous year's Information Security landscape and identifies where organizations are focusing projects and spending for the upcoming year. This 2015 study covers a variety of information security topics, yet a majority of questions relate to insider threat, storage, and other topics that relate or fall under the topic of data governance. Important information from this study includes an increase for the majority of organizations in spending on technology related to data governance such as Identity & Access Management (IAM), automated tools to facilitate forensics, and data analytics and encryption for data at rest (p.10). However, a reduction in spending was seen in other areas of security such as network edge and end-point security spending, showing a trend in focusing additional security at the core of the network and not the perimeter.

Survey respondents provided feedback on where governance practices will need to be focused within their organizations. They identify existing governance concerns that require improvement to meet requirements over the next three years, including secure access rights to data, unstructured data growth, lack of cybersecurity leadership, and the inability to integrate data sources for actionable cyber intelligence to protect assets. Results also indicated that only 37% of participants believe their organizations are prepared to tackle cybersecurity risks, 37% are investing in big data analytics for the purposes of cyber defense, and 47% believe they have the resources to tackle cybersecurity requirements.

Ponemon Institute. (2013). *The Post Breach Boom*. Traverse City, MI. Retrieved from http://www.ponemon.org/blog/the-post-breach-boom

Abstract. Data breaches have become a fact of life for organizations of all sizes, in every industry and in many parts of the globe. While many organizations anticipate that at some point a non-malicious or malicious data breach1 will occur, the focus of this study is to understand the steps organizations are taking—or not taking--to deal with the aftermath of a breach or what we call the Post Breach Boom.

Sponsored by Solera Networks, The Post Breach Boom study was conducted by Ponemon Institute to understand the differences between non-malicious and malicious data breaches and what lessons are to be learned from the investigation and forensic activities organizations conduct following the loss or theft of sensitive and confidential information. The majority of respondents in this study believe it is critical that a thorough post-breach analysis and forensic investigation be conducted following either a non-malicious or malicious security breach.

Understanding the differences between these two types of breaches can help organizations anticipate the financial consequences. In Ponemon Institute’s most recent Cost of Data Breach study published in March 2012, the average cost of a data breach per compromised record is $194. However, if the root cause is the result of a malicious insider or attack the average per record cost climbs to $222. While breaches attributed to a negligent insider averages far less at $174 per compromised record.

In this study we surveyed 3,529 IT and IT security practitioners in the following eight countries: United States, Canada, United Kingdom, Australia, Brazil, Japan, Singapore and United Arab Emirates. Most of the respondents (54 percent) report directly to the chief information officer or head of corporate IT. Fifty percent of respondents are at the supervisor level or higher.

To ensure quality responses, all participants in this study are in organizations that had one or more data or security breaches in the past 24 months. They also have significant or at least some level of understanding about the data or security breach incidents experienced by their organizations.

Summary. The Ponemon Institute performed a study to identify the causes of data breaches globally. The study was performed with 3,529 respondents in the IT field, of which 54% report directly to the CIO or hold a similar role. The survey provides important results to identify the highest risks and related challenges as they relate to the root cause of data breaches. Data breaches in 2012 and 2013 were the results of employee negligence 47% of the time, due to system errors and malfunctions 32% of the time, and were discovered by accident 34% of the time. Other relevant results from the study were reports of why the breaches were not prevented, including 64% of participants indicating that a breach was a result of a lack of expertise, 37% pointing to inadequate security processes, and 36% noting inadequate technology. This study covered all forms of corporate data including structured and unstructured data. 39% of breaches were focused at unstructured data. These statistics could help security practitioners and leadership identify areas of improvement in their own environments, as well as provide insight on where data governance practices are most applicable.

Sarkar, K. (2010). Assessing insider threats to information security using technical, behavioural and organisational measures. *Computer Crime, 15*(3), 112–133. http://doi.org/10.1016/j.istr.2010.11.002

Abstract. The UK government took a bruising in the headlines (Sep 2008) after a Home Office contractor lost a USB stick containing unencrypted data on all 84,000 prisoners in England and Wales. As a result, the Home Office terminated the £1.5 million contract with the management consultancy firm. The world woke up to the largest attempted bank fraud ever when the UK's National HiTech Crime Unit foiled the world's largest potential bank robbery in March 2005. With the help of the security supervisor, thieves masquerading as cleaning staff installed hardware keystroke loggers on computers within the London branch of a Japanese bank, to steal £220m. It is indeed sobering to imagine that any organisation could fall victim to such events and the damage an insider can do. The consulting firm lost the contract worth £1.5 million due to a small mistake by an employee. The London branch of the Japanese Bank would have lost £220 million had not the crime been foiled.

Insider threat is a reality. Insiders commit fraud or steal sensitive information when motivated by money or revenge. Well-meaning employees can compromise the security of an organisation with their overzealousness in getting their job done. Every organisation has a varied mix of employees, consultants, management, partners and complex infrastructure and that makes handling insider threats a daunting challenge. With insider attacks, organisations face potential damage through loss of revenue, loss of reputation, loss of intellectual property or even loss of human life.

The insider threat problem is more elusive and perplexing than any other threat. Assessing the insider threat is the first step to determine the likelihood of any insider attack. Technical solutions do not suffice since insider threats are fundamentally a people issue. Therefore, a three-pronged approach - technological, behavioural and organisational assessment is essential in facilitating the prediction of insider threats and pre-empt any insider attack thus improving the organization's security, survivability, and resiliency in light of insider threats.

Summary. Sarkar discusses the approach of protecting sensitive organizational data through the use of a three-pronged approach: Technological, behavioral, and organizational assessments. Discussions surrounding technology include potential methods of access to sensitive data for an insider and the types of potential breaches including social engineering and exfiltration. Behavioral assessments include the anatomy of insider attacks, the psychology of why an insider would take corporate assets such as a feeling of entitlement or ethical flexibility, and also why organizations themselves often ignore the risks associated with insider threats. Finally, Sarkar explores what insider threat assessments are and how they can be performed to help identify issues with both the technology and the staff to identify areas of risk so they can be targeted for mitigation efforts.

This article provides important information on the challenges and risks that accompany the management of unstructured data, including the criticality of considering the human factor. Understanding why individuals within the organization may deliberately perform a data breach, how to mitigate the risk through management practices, and how to provide protection through monitoring staff using relevant technology can assist in securing structured and unstructured data.

Best Practices for Mitigating Unstructured Data Risks

Balakrishnan, B., & Northcutt, S. (2015). *Insider Threat Mitigation Guidance* (SANS Institute InfoSec Reading Room). SANS Institute.

Abstract. Insider threats are complex and require planning to create multi-year mitigation strategies. Each organization should tailor its approach to meet its unique needs. The goal of this paper is to provide relevant best practices, policies, frameworks and tools available for implementing a comprehensive insider threat mitigation program. Security practitioners can use this paper as a reference and customize their mitigation plans according to their organizations' goals.

The first section provides reference frameworks for implementing an insider threat mitigation program with the Intelligence and National Security Alliance (INSA) Insider Threat roadmap, Carnegie Mellon University's Computer Emergency Response Team (CERT) insider threat best practices, CERT insider threat program components, National Institute of Standards and Technology (NIST) Cybersecurity Framework, and other relevant guidance. This section provides an implementation case study of an insider threat mitigation program for an hypothetical organization.

The second section of this paper will present example use cases on implementing operational insider threat detection indicators by using a risk scoring methodology and Splunk. A single event might not be considered anomalous, whereas a combination of events assigned a high-risk score by the methodology might be considered anomalous and require further review. A risk scoring method can assign a risk score for each user/identity for each anomalous event. These risk scores are aggregated daily to identify username/identity pairs associated with a high risk score. Further investigation can determine if any insider threat activity was involved. This section explains how to implement a statistical model using standard deviation to find anomalous insider threat events. The goal is to provide implementation examples of different use cases using a risk scoring methodology to implement insider threat monitoring.

Summary. The SANS Institute provides a document to guide organizations on how to mitigate insider threat through a combination of threat assessments, roadmaps, and frameworks based on best practices. While the practices are, in general, high level and do not apply directly to unstructured data governance, they can be applied to any form of insider threat mitigation techniques. For example, the document provides guidance in relevant areas such as identifying how the kill chain, the common procedure used in breaches, leads to the exfiltration of data and how detection indicators can help when threat activities occur by triggering notifications of a breach.

The document provides details on performing risk assessments to identify where vulnerabilities exist across internal monitoring platforms such as Security Information and Event Management (SIEM) logging platforms and their viability as methods for performing forensics on a breach. The document also includes information on the ability to baseline user activity to determine when there are deviations in behavior that could indicate suspicious behavior and the importance of user training in the areas of security awareness to ensure that the reality and consequence of insider threat are understood within an organization. Finally, it is crucial to identify stakeholders within the

organization for the development of an insider threat program, ensure that these stakeholders understand that they each own unstructured data, and subsequently secure leadership buy-in and recognition that the responsibility of data governance is an organizational goal.

Gartner Research. (2015). Turning security inside out to protect the most valuable data. Stamford, Connecticut, USA. Retrieved from http://www.gartner.com/imagesrv/media-products/pdf/varonis/Varonis-1-2ME5EK5.pdf

Abstract. The prevailing security model is not sustainable. A recent report from the Identity Theft Resource Center (ITRC) documents the number of known (reported) U.S. data breaches in 2015 are on pace to exceed the record of 783 set in 2014. According to an IBM study, the average data breach results in $3.79 million worth of damage. Most breaches involve unstructured data: The emails, word documents, spreadsheets, and presentations that employees are generating every day. Organizations store more unstructured data than any other type of data, and typically know very little about it. Gartner Research goes as far as to call this "dark data", and most CISOs know this is their biggest vulnerability. Employees and contractors have too much access to it, and nobody is watching their use of it. These over-privileged insiders represent a growing security threat with the rise of BYOD and cloud environments and the increasing sophistication of techniques that co-opt user's identities, like phishing. At many organizations, sensitive or valuable information can be extracted without anyone noticing. Simultaneously, hackers are getting more effective, constantly uncovering new ways around even the most advanced security perimeters. And to make matters worse, once these attackers are inside a network they frequently appear as an employee to an IT professional. Often the only

way they can be identified once they breach a perimeter is based on their behavior. By using software that can analyze file content, access rights and the actions users normally perform -- emails sent/received, files accessed, logins, and compare their actions to a baseline of normal behaviors for each user (also known as User Behavior Analytics), companies can and will be notified if hackers are in their networks, and fast.

Summary. This report is written in combination with Varonis Systems. Clearly marked areas written by Gartner Research indicate that user behavior analytics is a new use of technology aimed at identifying breaches in progress to help mitigate the exfiltration of unstructured data. The intention of this technology is help supplement data governance processes as an additional preventative measure to ensure that an active intrusion can be detected and subsequently prevented before a complete exfiltration of data occurs. Gartner Research indicates that the process of user behavior analytics also assists in the process of understanding baseline user behavior to help drive decisions on managing unstructured data by understanding the nature of how users and groups interact with said data. Case studies and examples are included to help provide real-world authenticity to the report detail.

Best Practices for Data Governance Management and Strategy

Bertino, E. (2012). Data protection from insider threats. *Synthesis Lectures on Data Management, 4*(4), 1–91. http://doi.org/10.2200/S00431ED1V01Y201207DTM028

Abstract. As data represent a key asset for today’s organizations, the problem of how to protect this data from theft and misuse is at the forefront of these organizations’ minds. Even though today several data security techniques are available to protect data and computing infrastructures, many such techniques—such as firewalls and network security tools—are unable to protect data from attacks posed by those working on an organization’s “inside.” These “insiders” usually have authorized access to relevant information systems, making it extremely challenging to block them is use of information while still allowing them to do their jobs. This book discusses several techniques that can provide effective protection against attacks posed by people working on the inside of an organization. Chapter 1 introduces the notion of insider threat and reports some data about data breaches due to insider threats. Chapter 2 covers authentication and access control techniques, and Chapter 3 shows how these general security techniques can be extended and used in the context of protection from insider threats. Chapter 4 addresses anomaly detection techniques that are used to determine anomalies in data accesses by insiders. These anomalies are often indicative of potential insider data attacks and therefore play an important role in protection from these attacks. Security information and event management (SIEM) tools and fine-grained auditing are discussed in Chapter 5. These tools aim at collecting, analyzing, and correlating—in real-time—any information and event that may be relevant for the security of an organization. As such, they can be a key element in finding a solution to such undesirable insider threats. Chapter 6 goes on to provide a survey of techniques for separation-of-duty (SoD). SoD is an important principle that, when implemented in systems and tools, can strengthen data protection from malicious insiders. However, to date, very few approaches have been proposed for implementing SoD in systems. In Chapter7, a short survey of a commercial product is presented, which provides different techniques for protection from malicious users with system privileges—such as a DBA in database management systems. Finally, in Chapter 8, the book concludes with a few remarks and additional research directions.

Summary. In this book, Bertino discusses the overall issue and impact of insider threats on corporate data. The general principle of the book is to elaborate on how common attacks have been performed, the kinds of data that are leaked, and practices that can be implemented to reduce the risk of a data breach on unstructured and structured data. Key discussions include the implementation of authentication and proper control lists for accessing data, the concepts of Separation of Duty (SoD), and how these approaches can help mitigate overly permissive access that leads to malicious insider breaches. The primary recommendation made in this publication is to combine the concepts of Access Control Lists for document access and Separation of Duty to create a model where authentication is required to access corporate data and access is limited to only those individuals that require the data by their job functions.

This book is useful for this research study because it clearly identifies practices that organizations can implement to reduce the probability and impact of a breach. By implementing a security posture that has users limited to access only the data necessary to perform their job functions, a singular user does not have the ability to exfiltrate an entire organization's sensitive data.

Gartner Research. (2011). Preventing data leaks with automated data governance. Stamford, Connecticut, USA. Retrieved from http://www.ainfosys.com/wp-content/uploads/2015/08/Gartner-Newsletter-Preventing-Data-Leaks-Automated-Data-Governance.pdf

Abstract. As Wikileaks and other recent newsworthy breaches remind us, organizations now house sensitive information belonging not only to the organization itself, but to its partners, clients, and employees. Governing access and monitoring use of so much data is mandatory to optimize productivity and security, and to remain a viable business partner. A critical part of the solution to preventing unauthorized or inappropriate access to sensitive data is the ability to leverage metadata - data about data (or information about information).

Summary. Gartner Research provides an analysis of data governance and the importance of data security. In an effort to provide clarity on the important of data security, the report references the data gathered by WikiLeaks to highlight how damaging exfiltration of data can be to an organization. Topics include insider threat, data protection, data governance, user activity monitoring, and risk of exposure. This document has been co-authored with Varonis Systems and has clearly marked which sections of the document Gartner Research wrote versus Varonis. Gartner Research indicates that information leaks are related to the poor management of data access and a lack of visibility into the behaviors of individuals using the data. Drawing attention to inappropriate behaviors through activity monitoring provides the ability to generate alarms before it is too late.

Gartner Research discusses the importance of combining various technologies to reduce the risk of exfiltration. Technologies such as data loss prevention (DLP), denial of service (DoS) protection, and data encryption coupled with policies and role and entitlement management are offered as best practices for use with a data governance strategy.

Loshin, D. (2013). *Big Data Analytics: From Strategic Planning to Enterprise Integration with Tools, Techniques, NoSQL, and Graph*. Elsevier. Retrieved from http://www.chinastor.org/upload/2014-05/14051214396874.pdf

Abstract. The goal of this book is to provide a firm grounding in laying out a strategy for adopting big data techniques. It is meant to provide an overview of what big data is and why it can add value, what types of problems are suited to a big data approach, and how to properly plan to determine the need, align the right people in the organization, and develop a strategic plan for integration.

Summary. The book covers the overall concept of Big Data and quickly moves to topics such as implementation, techniques for harnessing data, and governance strategies. While many aspects of this book are out of the scope of this paper, there are many relevant aspects to building a Data Governance strategy for unstructured data. The author recommends that the first stage of implementing a governance plan for corporate data include developing a technology adoption plan, identifying staff that will manage the data, documenting requirements for the implementation of new technology and the ongoing management of data storage, and finally documenting the practices for securing corporate data. The author also provides ideas for developing a general adoption plan, the creation of data governance policies and procedures to cover important items such as providing alerts for issues that will have a negative business impact, ways to triage and prioritize issues, and how to handle remediation through the assignment of data stewards. The key relevant content from Loshin's book on big data is the detail provided on how to manage growing data from an organizational perspective rather than from a technology perspective. The description of how to implement policies, procedures and plans can help organizations better shape their future efforts to manage unstructured data.

Tankard, C. (2015). Data classification – the foundation of information security. *Network Security*, *2015*(5), 8–11. http://doi.org/10.1016/S1353-4858(15)30038-6

Abstract. Data is the lifeblood of any organisation and has enormous value. Such data includes intellectual property and trade secrets, financial and customer records, and information related to employees. Much of that data is sensitive, and it is also of enormous value to those who would like to steal it for financial gain or commercial advantage.

Data classification is a foundational technology for ensuring effective data security and information governance by applying protective markings to documents.

Combined with clear policies and processes, this bolsters the capabilities of other controls used by organisations to safeguard data, such as data loss prevention and encryption technologies, helping to improve the overall security posture and to protect an organisation. It's key to your organisation's security, says Colin Tankard of DigitalPathways.

Summary. Tankard indicates that data classification is a critical element to securing data as it allows the creation of additional security measures. When an organization is unaware of what resides within documents, it is unable to identify which documents are safe to expose outside of the organization and which must reside under strict controls. Identifying which documents contain sensitive content can enable technology to restrict the ability to exfiltrate documents. Data classification can lead to compliance with regulations, protection of data, improvement of data loss prevention (DLP) technology, increased user accountability, easier encryption, and enhanced governance. Technologies such as applying protective markings to documents, usage of DLP technology, and encryption can assist in the development of a stronger security posture. According to the Corporate Executive Board references in the paper, 84% of organizations have experienced an insider breach from a staff member. The author explains that because insider breaches are so common, implementation of security consoles specific to sensitive data, such as controlling who has access and the communications methods used, are as important as encrypting data. However, the most critical recommendation made by this document and its relevancy to insider threat and unstructured data is the level of detail the document provides on the importance of data classification as part of information security over corporate data.

Conclusion

According to the SANS Institute, 92% of IT leaders acknowledge that they are vulnerable to insider threats within their organization, while 49% indicate they are extremely vulnerable (2015). The majority of insider breaches target unstructured data; 67% of insider breaches either directly target documents or target documents in combination with structured data (Ponemon Institute, 2014). Because unstructured data represents such a large percentage of overall data for organizations – 80% of all data in an organization on average, and growing by an average of 50% per year (Gartner, 2015) – the threat posed by insider threats targeting unstructured data is serious.

Organizations who have had an unstructured data breach indicate that the largest resulting impacts are a loss of time and productivity, the resulting cost of new technology purchases to mitigate the risk of future breaches, loss of reputation and brand value, and loss of revenue (Ponemon Institute, 2014). One means organizations have to help mitigate the risk of unstructured data breaches is through data governance plans (Gartner, 2011). This study explores the role that data governance plans can play in mitigating the risk of insider data breaches of unstructured data. Themes include (a) challenges and risks with managing unstructured data, (b) best practices for mitigating unstructured data risks, and (c) best practices for data governance management and strategy.

Challenges and Risks with Managing Unstructured Data

Gartner Research has termed unstructured data as “dark data” due to the lack of visibility into what is contained within the data (Gartner, 2015, p. 2). While the topic of monetizing dark data is a frequent concern for CEOs (Gartner, 2015), CISOs are concerned that dark data poses one of their organizations’ greatest security vulnerabilities as a lack of insight into which data contains sensitive information can lead to unexpected exposure (Gartner, 2015). This exposure has the potential to be damaging to organizations; in a study by IBM the average cost of a data breach was $3.79 million (Gartner, 2015).

Gartner Research indicates that unstructured data is growing at a rate of 50% per year (2015). of processes reliant on the creation of documents by humans as well as an “explosion in machine-generated data” (Miller, 2014, p. 6). This enormous growth does not pose a problem simply in the area of storage capacity management, but also increases the possibility of insider threats as organizations struggle to manage the growing volume of sensitive and unstructured data that puts them at risk for exposure (Ponemon Institute, 2013). Bertino (2011) notes that “the proliferation of web-based applications and information systems, and recent trends such as cloud computing and outsourced data management, have further increased the exposure of data and complicated the security problem” (p. 1). In a 2013 study performed by the Ponemon Institute of organizations that had suffered data breaches, 64% of participants indicated that a data breach that occurred within their organizations was due to a lack of expertise by the information security team, while another 47% stated that breaches were also due to employee negligence. This discovery indicates the need for not just technology expertise, but also the need for improved training to reduce negligence. The inclusion of best practices for data governance is a necessary part of this training (Balakrishnan & Northcutt, 2015).

While data governance and risk management for unstructured data are both in the early adoption stage, there are tools and techniques available to help improve security on unstructured data (Miller, 2014). An example of a challenging process for many organizations is managing access permissions across file systems and Active Directory, as exemplified by GDF Suez, a Global Gas and Electricity provider, which typically takes five to six days to complete an individual employee request for access to documents on a file system (Miller, 2014). The challenges that insider threats bring to the management of unstructured data and increasing regulatory requirements from the Health Insurance Portability and Accountability Act (HIPAA) and Sarbanes-Oxley (SOX) bring increased needs for effective monitoring of the access controls and efficient management of permissions (Miller, 2014).

Organizations that wish to reduce the risk of exposure and better manage their unstructured data will need to strengthen cybersecurity leadership and determine how they intend to govern corporate data, implement security controls for documents, and utilize technology that provides security auditing on data assets (Ponemon Institute, 2015b). However, these steps often clash with corporate culture and pose challenges in implementation due to insufficient resources (Miller, 2014). Organizations have become accustomed to providing lax security controls to data, often termed as over-permissive access, due to the challenges experienced in identifying who the correct staff members are that require access and difficulties in maintaining those access controls on an ongoing basis due to time constraints (Miller, 2014).

The risk of data breaches has traditionally led to investment in edge security, such as firewalls, and has overlooked threats posed by the trusted insider (Bertino, 2011). Since the insider has the “necessary authorizations to access much of the organization’s proprietary or sensitive data” (Bertino, 2011, p. 17), it is crucial that audits occur on any areas where data resides to reduce the risk of data loss from insider threats (Blanchard & O’Sullivan, 2015).

A study performed by the Ponemon Institute in 2014 identified that out of 1,587 global IT and IT security practitioners, only 7% felt they knew where unstructured data resided within their organization and only 16% believed that they knew where sensitive data resided. Most of the respondents in the same study felt that the lack of insight into where the sensitive data resides, not knowing where most unstructured data resides across the network, and the resulting inability to know where to apply security controls are the factors that created the largest concerns surrounding management of unstructured data (Ponemon Institute, 2014). As a result, 68% of respondents indicated that a need to classify sensitive data across the organization was the most critical control for protecting the organization from breaches, followed by the implementation of access controls at 62% of respondents (Ponemon Institute, 2014).

Best Practices for Mitigating Unstructured Data Risks

A collaborative effort between the for-profit Escal Institute of Advanced Technologies (SANS Institute), an organization that specializes in information security and cybersecurity training; the Intelligence and National Security Alliance (INSA); Carnegie Mellon University's Computer Emergency Response Team (CERT); and National Institute of Standards and Technology (NIST) has resulted in a selection of best practices from which organizations can select to assist in reducing the risk of insider threats (Balakrishnan & Northcutt, 2015). These general practices specify the importance of creating an insider threat program that covers general topics such as the formal creation of a security program and security policies and education of these items to all employees, the implementation of technologies to perform real-time monitoring and capture of all interactions with data, and the implementation of technology to actively detect and prevent access to data when suspicious behavior occurs. Technology to assist in mitigating the risk of an unstructured data breach includes security monitoring tools to ensure that every step of a data breach is monitored, that technology solutions are implemented that correlate security events together in a chain to identify behavior rather than missing a single anomalous event, and the implementation of the least permissive access model on all access points to data (Balakrishnan & Northcutt, 2015). User Behavior Analytics involves the use of software to analyze file content, access rights, and actions users normally perform such as emails sent and received, files accessed, and logins against a baseline of typical behavior for each user in order to quickly identify and notify companies of the presence of hackers in their networks (Gartner, 2015).

Best Practices for Data Governance Management and Strategy

The importance of data, structured and unstructured, is followed closely by the importance of "its security, privacy, and proper use" (Bertino, 2012, p. 1). Growing threats from insiders that are categorized as sabotage, theft of intellectual property or fraud are concerning, as an insider by definition is trusted by the organization to have access to sensitive information (Bertino, 2012). Gartner Research (2011) indicates that mitigating the risk of insider threats must focus on a strategy that focuses on data protection, analysis and user activity monitoring. This approach for securing the data from potential insider threats includes recommendations to create a formal data governance approach inclusive of data encryption, data loss prevention technology, monitoring of user activity, and implementation of role and entitlement management (Gartner, 2011).

Data governance for mitigating the risks associated with insider threats is a combination of "process changes and technology based solutions" (Gartner, 2011, p. 2). Implementation of encryption and data loss prevention technology will reduce the risk of exposure by making it more challenging for the data to be accessed and exfiltrated (Gartner, 2011). However, this technology should be coupled with "fine-grained" entitlement reviews (Gartner, 2011), internal auditing to identify risks posed by holes in access management (Blanchard & O'Sullivan, 2015), and user activity monitoring to identify suspicious patterns of behavior (Gartner, 2011).

Automated tools that perform data classification provide insight into data and can assist in securing unstructured data (Tankard, 2015). Unstructured data may contain trade secrets, intellectual property, financial data, customer records, and employee information, which are targets for exfiltration (Tankard, 2015). Data classification provides the ability to identify the contents of data and then classify the data so that data classification can bolster other processes such as data loss prevention and encryption (Tankard, 2015). Tankard (2015) stresses that data classification is best performed centrally from a single technology across the organization to reduce complexity and provide the ability to tag documents and emails with markings so the end-users can see their data is being monitored and identified, and that this data should be tied into multiple security controls to improve effectiveness.

While the terminology used by experts varies, a theme across data governance best practices is the concept of Separation of Duty (SoD). SoD involves identifying data and assigning its access by the role an individual serves, which helps reduce the risk of others accessing data that is out of the scopes of their positions. Furthermore, access can be managed through a combination of data classification efforts and Access Control Lists to limit exposure beyond those that absolutely need to access the data (Bertino, 2011). Lastly, incorporating these recommended practices within a data governance strategy will help with the adoption of policies and practices to assist with compliance efforts (Loshin, 2013).

Another critical approach to managing insider threats within an organization recognizes the importance of organizational measures that go beyond the application of technology. Findings indicate that while malicious insiders are a concern, a much larger percentage of data loss occurs due to the non-malicious actions of an insider (Sarkar, 2010). These losses can occur when users lack knowledge of how to prevent data loss due to common hacking methods such as

Phishing, theft of a misplaced smartphone or mobile storage device, and malware (Sarkar, 2010). While technology can reduce the risk of these threats, an important series of steps to drastically improve the effectiveness of data governance as a mitigation for insider threats involves performing threat assessments and implementing security awareness programs for staff to help educate them on the risks associated with unstructured data loss and the methods to reduce the associated security threats (Sarkar, 2015).

Final Thoughts

This document presents literature to provide guidance in identifying the challenges associated with securing unstructured data from insider threat and recommended practices to reduce the risk of exposure. The techniques discussed vary from the implementation of technology such as the encryption of documents, to the implementation of audits to validate that security controls are implemented to meet Separation of Duty (SoD) best practices.

Bertino (2012) describes insider threat as “a current or former employee, contractor, or business partner who has or had authorized access to an organization’s network, system or data and intentionally exceeded or misused that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization’s information or information systems” (p. 2). Insider threat is a growing risk for organizations as the enormous growth of unstructured data has become a target (Gartner, 2015) and the contents of these documents and emails are often highly sensitive (Tankard, 2015). Development of a governance approach for unstructured data that includes technologies such as data loss prevention, encryption, access controls, and user activity monitoring will limit the threat and impact of data loss (Gartner, 2015). The addition of routine auditing of practices and technology (Blanchard & O’Sullivan,

2015) and the understanding of the psychology of the insider threat (Sarkar, 2015) can help reduce the risk of exposure by highlighting areas that require improvement.

References

Amato, F., Casola, V., Mazzocca, N., & Romano, S. (2013). A semantic approach for fine-grain access control of e-health documents. *Logic Journal Of The IGPL, 21*(4), 692-701.

Balakrishnan, B., & Northcutt, S. (2015). *Insider Threat Mitigation Guidance* (SANS Institute InfoSec Reading Room). SANS Institute.

Berry, D. (2012, April 9). Unstructured data: Challenge or asset? Retrieved March 30, 2016, from http://www.zdnet.com/article/unstructured-data-challenge-or-asset/

Blanchard, R., & O'Sullivan, K. (October, 2015) Big data risk and opportunity: Having an action plan to address both can add tremendous value to the organization. *Internal Auditor*. Retrieved from http://go.galegroup.com/ps/anonymous?id=GALE%7CA434320765&sid=googleScholar&v=2.1&it=r&linkaccess=fulltext&issn=00205745&p=AONE&sw=w&authCount=1&isAnonymousEntry=true

Center for Public Issues Education. (2014). Evaluating information sources. Retrieved from http://www.piecenter.com/wp-content/uploads/2014/08/evaluateinfo.pdf

Dark Data - Gartner IT Glossary. (2015). Retrieved March 30, 2016, from http://www.gartner.com/it-glossary/dark-data

Dayley, A. (2013). Innovation insight: File analysis innovation delivers an understanding of unstructured dark data. Gartner Research. Retrieved from https://www.gartner.com/doc/2394915/innovation-insight-file-analysis-innovation

Filkins, B. (2015). Cleaning up after a breach post-breach impact: A cost compendium. SANS Institute. Retrieved from https://www.sans.org/reading-room/whitepapers/analyst/cleaning-breach-post-breach-impact-cost-compendium-36517

Gartner Research. (2011). Preventing data leaks with automated data governance.

Inmon, W., & Nesavich, A. (2008). Tapping into unstructured data: Integrating unstructured data and textual analytics into business intelligence. Upper Saddle River, NJ: Prentice Hall Press.

Kshetri, N. (2014). Big data′s impact on privacy, security and consumer welfare. *Telecommunications Policy*, *38*(11), 1134–1145. http://doi.org/10.1016/j.telpol.2014.10.002

Liu, X., & Murphy, D. (2014). Document explosion in the world of big data – curriculum considerations. *Information Systems Education Journal*, *12*(6) pp 83-91. http://isedj.org/2014-12/ ISSN: 1545-679X. (A preliminary version appears in The Proceedings of ISECON 2013)

Loshin, D. (2013). Big data analytics: From strategic planning to enterprise integration with tools, techniques, nosql, and graph. *Elsevier*. Retrieved from http://www.chinastor.org/upload/2014-05/14051214396874.pdf

Marcant, C. (2015). Demystifying data governance: What it is and what it's not. Retrieved March 30, 2016, from http://data-informed.com/demystifying-data-governance-what-it-is-and-what-its-not/

Mearian, L. (2010, November 1). Data growth remains IT's biggest challenge, Gartner says. Retrieved March 30, 2016, from http://www.computerworld.com/article/2513954/data-center/data-growth-remains-it-s-biggest-challenge--gartner-says.html

Morabito, V. (2015). Big data governance. In *Big Data and Analytics* (pp. 83–104). Switzerland: Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-10665-6_5

Ponemon Institute. (2015). 2014: A year of mega breaches. Retrieved from http://www.ponemon.org/local/upload/file/2014%20The%20Year%20of%20the%20Mega%20Breach%20FINAL3.pdf

Pornain, X. (2014). Rewiring to tackle unstructured data. Retrieved March 30, 2016, from http://www.wired.com/insights/2014/07/rewiring-tackle-unstructured-data/

Sanders, M. (2015). Unknown threat detection with Honeypot Ensemble Analsyis using big datasecurity architecture. *Theses and Dissertations*. Retrieved from http://ir.library.illinoisstate.edu/etd/360

Seng, J.-L., & Lai, J. T. (2010). An intelligent information segmentation approach to extract financial data for business valuation. *Expert Systems with Applications*, *37*(9), 6515–6530. http://doi.org/10.1016/j.eswa.2010.02.134

Tanwar, M., Duggal, R., & Khatri, S. K. (2015). Unravelling unstructured data: A wealth of information in big data. In *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)* (pp. 1–6). http://doi.org/10.1109/ICRITO.2015.7359270

Tankard, C. (2015). Data classification – the foundation of information security. *Network Security*, *2015*(5), 8–11. http://doi.org/10.1016/S1353-4858(15)30038-6

Taylor, C. (2013 9–19). What’s the big deal with unstructured data? Retrieved March 30, 2016, from http://www.wired.com/insights/2013/09/whats-the-big-deal-with-unstructured-data/

University of California, Chico. (2010). Evaluating information – Applying the CRAAP test. Chico, CA: Meriam Library.

Vormetric. (2015). Trends and future directions in data security (2015 Vormetric insider threat report). San Jose, CA. Retrieved from http://enterprise-encryption.vormetric.com/rs/vormetric/images/CW_GlobalReport_2015_Insider_threat_Vormetric_Single_Pages_010915.pdf

What is unstructured data? - Definition from Tech Target. (n.d.). Retrieved March 30, 2016, from http://searchbusinessanalytics.techtarget.com/definition/unstructured-data

Williams, P. (2012). Unstructured data and the enterprise. *Dataversity*. Retrieved March 29, 2016 from http://www.peerevaluation.org/data/8065d07da4a77621450aa84fee5656d9/PE_doc_30009.pdf

Your unstructured data is sexy – you just don’t know it. (n.d.). Retrieved March 30, 2016, from http://www.theheadwatersgroup.com/your-unstructured-data-is-sexy/