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

Using Aspects of Data Governance Frameworks to Manage Big Data as an Asset

CAPSTONE REPORT

Allen Chamberlain
Senior Practice Consultant
EMC Corporation

University of Oregon
Applied Information
Management
Program

July 2013

Continuing Education
1277 University of Oregon
Eugene, OR 97403-1277
(800) 824-2714

Approved by

_____

Dr. Linda F. Ettinger
Senior Academic Director, AIM Program

Using Aspects of Data Governance Frameworks to Manage Big Data as an Asset

Allen Chamberlain

EMC Corporation

Abstract

The emergence of big data has transformed how organizations realize the benefits of large datasets. This annotated bibliography examines how data governance frameworks may enable organizations to more effectively implement big data systems and maximize the rewards. Literature published from 2004 to 2013 is reviewed and key aspects of big data governance systems are identified. Aspects are aligned within a set of five domains (Khatri & Brown, 2010) to guide effective data governance.

Keywords: data governance, big data analytics, data quality management, data lifecycle management

Table of Contents

List of Tables

Introduction

Problem Area

Gartner (2012) defines big data as **"**high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making" (para. 1). In this study, big data is viewed within the context of the need for organizations to recognize greater amounts of information than ever before, from multiple and often entirely new origins, consumable almost instantaneously upon creation (Rosenbush & Totty, 2013).

Violino (2013) argues that in order to effectively prepare for the transition to big data, organizations adhering to an established data governance framework are positioned to achieve greater success – both in the short term and long term – and are therefore better prepared to succeed. The Data Governance Institute (2012) defines *data governance* as "a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods" (Definitions section, data governance). Soares (2013) further expands this definition to include "policies relating to the optimization, privacy, and monetization of big data" (p. 4).

Soares (2012) states that "done correctly, a strong big data governance program can take the guesswork out of finding and using the right data to make business decisions—regardless of where the information comes from, what type it is, or how fast it is moving"(para.7). *Business intelligence*, or the process of turning raw data into useful information to support business decisions, is further improved by the practice of computational analysis of large datasets (Hua, Huang & Yen, 2012). Gopalkrishnan et al. (2012) note that big data analytics prompts

organizations to determine how data is utilized strategically to remain competitive and support business decisions.

The absence of data governance is a serious and often ignored issue that should receive careful consideration to reduce the risk of unsuccessful software implementations, safeguard information and realize the concept of *data as an asset* (Khatri & Brown, 2010). Violino (2013) notes that "on a typical big data project, data management is often 'deprioritized' by development staff and can go unresolved" (p. 3). As organizations strive to implement systems to analyze big data, it is critical that they develop well-formed data governance frameworks to support the management of big data systems (Myers, 2012).

Purpose

The purpose of this scholarly annotated bibliography is to identify literature that explores data governance frameworks that enable organizations to effectively manage big data systems. (Borkar et al., 2012). The goal of this research is to present how governance frameworks enable organizations to better implement systems that capture and analyze large data sets. In this study, a *data governance framework* is defined as "a set of processes that ensures that important data assets are formally managed throughout the enterprise" (Sarsfield, 2009, p. 23). According to Peyret (2013) the advent of big data systems enables organizations to dynamically analyze and subsequently utilize information to make advantageous business decisions (para. 4). Peyret further advises that as organizations transition to this "data economy" adopting an actionable data governance framework centered on the organization's business model is key to achieving success (2012, para. 5).

Data governance may be structured in multiple forms across different organizations. For example, Khatri and Brown (2010) present a data governance strategy designed to guide

organizational decision making and reduce risks to the organizations; they organize data governance into five interrelated *domains* constituting an infrastructure from which data governance is implemented. Khatri and Brown (2010) describe both (a) how organizations should adhere to these domains, and (b) provide examples of how these domains are applied (see Table 1). For the purpose of this study, this set of domains constitutes a basic *framework* to guide effective data governance and improve success of such big data initiatives; additional aspects identified in the literature are added to broaden the potential application in a variety of organizational settings.

Table 1

Five Domains to Guide Data Governance (from Khatri & Brown,2010)

Domain	Description	Application
Domain 1: Data Principles	**O**rganizations should establish data principles "to which data is an enterprise wide asset, and thus what specific policies, standards and guidelines are appropriate" (Khatri & Brown, 2010).	Principles provide a way for organizations to recognize the significance of data as an asset.
Domain 2: Data Quality	**O**rganizations should establish standards to ensure data is up-to-date, complete and credible (Khatri & Brown, 2010)	Quality standards reduce financial impacts of poor data quality. These costs are estimated as 8% to 12% of revenue (Parssian et al., 2004)
Domain 3: Metadata	**M**etadata "provides a mechanism for a concise and consistent description of the representation of data, thereby helping interpret the meaning" (Khatri & Brown, 2010).	Metadata provides a consistent means for organizations to realize efficient searching and analysis of their data.
Domain 4: Data Access	**O**rganizations should effectively control access to data assets to manage risk and compliance (Khatri & Brown, 2010).	Access standards establish rules for organizations to control what data shall be accessed and by whom.
Domain 5: Data Lifecycle	**O**rganizations should understand and document how data progresses through its business	The lifecycle provides organizations with the means to understand how data is sourced from various contributors – both inside and outside of the

Domain	Description	Application
	processes. (Khatri & Brown, 2010).	organization – and its progression from active use to its eventual archival is a critical piece of an effective data governance program.

Significance

This annotated bibliography explores the potential relationships between data governance frameworks and big data systems initiatives. According to a study commissioned by Hitachi Data Systems in 2011, managing big data systems is new territory for many information technology leaders (Yap, 2011). A 2010 survey of 17,000 executives representing 19 industries worldwide found that 71 percent indicated they were uncertain that their companies were adequately prepared for the transition to big data (Rajan, McMillan & O'Connor, 2011).

Soares, Deutsh, Hanna and Malik (2012) caution "as big data technologies become operational—as opposed to exploratory—they need the same governance disciplines as traditional approaches to data management" (para. 4). These decision-makers require a thorough understanding of how data governance frameworks may (a) reduce the risks inherent in big data system implementations, and (b) realize the potential successes of big data system implementations.

Audience

The intended audience for this annotated bibliography includes information architects, chief information officers, chief information security officers and information technology managers responsible for managing big data systems and realizing the benefits of big data system initiatives. According to Borkar et al., (2012) the goal of realizing value from big data impacts

"virtually everyone, ranging from big Web companies to traditional enterprises to physical science researchers to social scientists" (p.1).

Research Questions

Main question. What are the most effective aspects of a data governance framework for managing big data system initiatives?

Sub-questions.

What is a *big data system*?

What is *data governance* and why is it important?

What is the role of selected *data governance frameworks* in improving the effectiveness of business processes?

Delimitations

Topic focus. This annotated bibliography explores the potential relationships between data governance frameworks and the management of big data systems initiatives. Information technology leaders face challenges in transitioning from their current level of data governance maturity to the level necessary to recognize the benefits of big data analytics (Manyika et al., 2011, p. 13). Soares (2012) states that "done correctly, a strong big data governance program can take the guesswork out of finding and using the right data to make business decisions—regardless of where the information comes from, what type it is, or how fast it is moving"(para.7). Data governance represents a foundational aspect of information management and constitutes a supportive infrastructure upon which transformational initiatives such as big data systems are built (Hsu & O'Neal, 2009, p. 3)

Definition of *data governance framework*. For the purposes of this study, a data governance framework is defined as "a cross-functional set of roles, policies and enabling

technologies that work together to ensure that an organization is getting the maximum net benefit out of its data assets (Stanford, 2013, para 1). Khatri and Brown (2010) present a data governance strategy designed to guide organizational decision making and reduce risks to the organizations; they organize data governance into five interrelated *domains* constituting an infrastructure from which data governance is implemented. Khatri and Brown (2010) describe both (a) how organizations should adhere to these domains, and (b) provide examples of how these domains are applied (see Table 1). For the purpose of this study, this set of domains is pre-selected as a basic *framework* to guide effective data governance and improve success of such big initiatives.

Time frame. Data governance is considered an emerging concept, with the first industry-wide conference on the subject occurring in 2006 (Thomas, 2012, p. 19). Therefore, the time frame for collection of references articles included in this study is limited to literature published between 2004 to 2013.

Audience. The audience for this study includes information managers responsible for driving big data initiatives within organizations. Stakeholders include CIO or CTO level executives, information technology directors and managers and other interested parties who are concerned with how to implement a data governance foundation. Forrester Research Inc. analyst Boris Evelson notes "big data is such a new area that nobody has developed governance procedures and policies…there are more questions than answers" (cited in Du Mars, 2012, p. 1).

Reading and Organization Plan Previews

Reading plan preview. Busch et al. (2013) provide a framework commonly used in research identified as *conceptual analysis*. Within this framework, the concepts embedded in the research questions proposed in this study are directly mapped to key terms and phrases that are

coded in the selected references. Busch et al. (2013) identify the following steps for conducting such an analysis to code and categorize relevant data:

- Determine the level of analysis.
- Determine what categories to code.
- Decide if concepts will be coded for existence or frequency
- Determine how concepts are distinguished
- Establish consistent coding rules
- Determine how to handle irrelevant information within references
- Complete the coding process.
- Analyze results and determine conclusions.

Organizational plan preview. The Writing Center at the University of North Carolina at Chapel Hill (Literature Reviews, n.d.) advises researchers to use predefined methods to provide organization and focus. This annotated bibliography utilizes a *thematic* method (instead of, for example, a chronological order). The key references selected for inclusion in the Annotated Bibliography section of this study are organized based on relevance to each theme. Themes are framed by the three research sub-questions:

Theme 1. What is a big data system?

Theme 2. What is data governance and why is it important?

Theme 3. What is the role of selected data governance frameworks in improving the effectiveness of business processes?

Definitions

This section provides definitions of terms appearing in the literature concerning big data systems and data governance. All definitions are supported by authoritative sources considered reliable for the purposes of this study and intend to clarify terms for the intended audience and decrease ambiguity.

Access Control – The Committee on National Security Systems defines access control as "the process of granting or denying specific requests…for obtaining and using information and related information processing services" (CNSS, 2010, p.1)

Big Data – Dumbill (2012) states that big data "is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the structures of your database architectures" (O'Reilly, Chapter 2, para. 1). Gartner further defines this as "high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making" (Gartner, 2012, para 1).

Business Intelligence – Hua, Huang and Yen (2012) define business intelligence as "solutions applying information technologies to retrieve heterogeneous and distributed resources in order to interpret, categorise, and integrate them, and then to formulate any potentially usable knowledge by employing analysis mechanisms" (Introduction, para. 1)

Data Analytics – Defined as **"**qualitative and quantitative techniques and processes used to enhance productivity and business gain. Data is extracted and categorized to identify and analyze behavioral data and patterns, and techniques vary according to organizational requirements" (Janssen, 2013, para. 1)

Data Governance – The Data Governance Institute offers the following definition of data governance: "a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods" (DGI, 2012, p. 3).

Data Governance Framework - "a set of processes that ensures that important data assets are formally managed throughout the enterprise" (Sarsfield, 2009, p. 23).

Data Lifecycle Management (DLM) – DLM is defined as **"**a policy-based approach to managing the flow of an information system's data throughout its life cycle: from creation and initial storage to the time when it becomes obsolete and is deleted. DLM products automate the processes involved, typically organizing data into separate tiers according to specified policies, and automating data migration from one tier to another based on those criteria. As a rule, newer data, and data that must be accessed more frequently, is stored on faster, but more expensive storage media, while less critical data is stored on cheaper, but slower media" (Rouse, 2010, para. 1).

Data Quality Management – Larry English, co-founder of the International Association for Information and Data Quality defines data quality management as "the application of sound quality principles and processes to information as a product of business and manufacturing processes" (English, 2005, para. 3)

Hadoop – Hadoop is a software platform consisting of two primary modules: the Hadoop Distributed File System (HDFS) and Hadoop MapReduce. The HDFS component enables big data analytics by separating large datasets into smaller sets stored across inexpensive computational hardware. The Hadoop MapReduce component leverages the HDFS structure to

disperse the processing of each subset of a larger dataset across separate computers working independently (Mone, 2013).

Metadata – The National Information Standards Organizations defines metadata as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information" (NISO, 2004, p. 1).

Research Parameters

This section details the methods and tools utilized to conduct this study. Research parameters are presented to document the approach used to gather, classify, analyze, and evaluate literature concerning big data systems and data governance within the context of the research question and associated sub-questions.

Search Strategy

References presented in this annotated bibliography are identified via online search engines and tools provided by the University of Oregon Libraries. While the primary focus of the search is *big data governance*, additional keywords correlate the focus to big data: *analytics, big data, compliance, considerations, data governance frameworks, governance, implementations, policies, practices, and risk*. References should be from peer-reviewed journals and identify relevant information regarding either the primary research question or sub questions for inclusion. The initial database search queried the following:

- Academic Search Premier
- JSTOR
- Project Muse
- UO Local Catalog

Searching the Computer Science "Search by Subject" indexed search for 'Big Data Governance' as well as the Business "Search by Subject" option yields the highest degree of success. Relevant articles are returned from the following databases:

- Academic Search Premier
- ACM Digital Library
- Business Search Complete

- EconLit
- Factiva
- IEEE Computer Science Digital Library
- JSTOR
- New Palgrave Dictionary of Economics Online
- Newspaper Source
- Regional Business News
- ArXiv.org
- Web of Science
- Computer Source
- UO Local Catalog

Search key words are intentionally limited as indicated above. Outside of the UO Libraries site Dogpile.com, Bing.com and Google.com are also utilized.

Evaluation Criteria

Literature is selected based on the works of those identified as key contributors to data governance (Wisker, 2005). References are cited from peer-reviewed publications and (a) address the importance of data governance in big data system implementations, (b) provide guidance to information technology leaders in implementing data governance frameworks, and/or (c) discuss data governance frameworks as a foundational aspect of information technology.

To ensure information sources are credible an evaluation framework by Bell and Frantz (2012) is utilized to determine (a) the authority and objectivity of the author and (b) the objectivity, currency and relevancy of the work, based on the following criteria:

- Is an author provided?
- Is the author a credible and reputable source?
- Is the publisher an objective entity free of bias?
- Is the citation adequately researched and cited by other works?
- Is the information presented in an organized fashion?
- Is the information current?

Documentation Approach

Machi and McEvoy (2009) suggest the following systematic approach to discovering and documenting relevant references:

1. Successfully develop a specific scanning strategy that connects the preliminary topic statement, its focus and vantage point to key ideas;

2. Frame key ideas as descriptors for the search;

3. Build Boolean frameworks to query the appropriate databases;

4. Develop cataloging tools to document what works you plan to review for potential inclusion in the study;

5. Define the sequence and purpose for each scan, and identify appropriate search data based and their accessibility (p. 45).

To aid in capturing relevant reference information, Zotero is also utilized. Zotero provides a plugin compatible with the Firefox web browser that eases data collection by capturing website information such as the URL into a database and includes options to record reference metadata to assist in producing bibliographical records. All references are finally captured as individual Adobe Portable Document Format (.PDF) files either directly from the publisher when available or via an electronic printing when a suitable output is not available.

Reading Plan

This study applies the conceptual analysis method of reviewing and categorizing content. According to Busch et al., the focus of the conceptual analysis method involves the following:

> . . . looking at the occurrence of selected terms within a text or texts, although the terms may be implicit as well as explicit. While explicit terms obviously are easy to identify, coding for implicit terms and deciding their level of implication is complicated by the need to base judgments on a somewhat subjective system. To attempt to limit the subjectivity, then (as well as to limit problems of reliability and validity), coding such implicit terms usually involves the use of either a specialized dictionary or contextual translation rules. (conceptual analysis, para. 1., 2013)

Busch et al. (2013) identify a coding process to identify pertinent concepts within selected works. The following steps are employed:

1. **Determine the level of analysis**. Will the single words or entire phrases suffice for the review? For the purposes of this literature review the primary coding term is *big data governance* with additional, secondary coding terms including *big data systems* and *big data management*.
2. **Determine what categories to code**. Should coding utilize an objective method of capturing every positive word match or a more subjective method of comparing search results with a pre-determined set of categories? This study uses the following pre-determined concepts to categorize and code references: (a) data governance with relation to big data systems, (b) big data initiatives, and (c) big data management.

3. **Decide if concepts will be coded for existence or frequency**. Selected references are coded solely based on the existence of the pre-determined concepts. Coding based on existence is preferred in the study as opposed to frequency to analyze the literature within the context of this conceptual framework, as a way to best determine meaning. Coding for existence is further utilized in response to the nominal volume of applicable research due to the emerging nature of the subject matter.
4. **Determine how concepts are distinguished.** This study allows for the use of synonyms or alternate keywords or phrases that appear in the selected literature. For example, the term *data management* may appear as written or as *information management* and both results are generalized according to the context within the reference.
5. **Establish consistent coding rules.** Concepts appearing in selected works are coded consistently and utilize translation rules when necessary based on the author's intent. In the case where a translation rule is utilized a notation is included and the entry is included in the pre-determined categories identified above.
6. **Determine how to handle irrelevant information within references.** This study ignores words deemed irrelevant such as *and* and *the*.
7. **Complete the coding process.** Selected works are reviewed in their entirety and keywords and key phrases are documented. Coding is conducted manually during the reading of each reference and the Adobe PDF "Find" function is utilized to facilitate keyword and key phase searching.
8. **Analyze results and determine conclusions.** Results are analyzed based on level of applicability to the research questions presented in this study. Information gathered

from the coding process is organized and presented by the Organization Plan illustrated below.

Organization Plan

This study utilizes a thematic method to organize and-present information. References within the Annotated Bibliography section relevant to data governance frameworks and their application to big data systems are grouped within four themes mapped directly to the research sub-questions, based on direction from the Writing Center at the University of North Carolina at Chapel Hill (Literature Reviews, N.D.). Organizational themes and their aspects are presented below:

Theme one: What is a big data system?

- Characteristics and qualifications of a big data system (Borkar, Carey, & Li, 2012a; Cukier & Mayer-Schoenberger, 2013).
- Business intelligence, analytics, and big data (Chen, Chiang, & Storey, January 01, 2012).
- Innovative aspects of big data systems (Gartner, 2012).

Theme two: What is data governance and why is it important?

- Aspects and intent of data governance (Rosenbush & Totty, 2013).
- Data governance as a framework for managing information (Borkar, Carey, & Li, 2012b; Khatri and Brown, 2010; Österle, Otto, & Weber, 2009).

Theme Three: What is the role of selected data governance frameworks in improving the effectiveness of business processes?

- Risk management strategies inherent in data governance frameworks (Khatri & Brown, 2010).
- Data governance as a means to drive improvements within organizations (McGlinchey, 2013)
- Business practices most likely to require improvement (McGlinchey, 2013); Franks (2012).

Annotated Bibliography

References selected for annotation in this annotated bibliography are presented in a consistent manner and include (a) complete APA citation; (b) author or indexer-provided abstracts; (c) verification of the credibility and authority of the authors; and (d) a summary of the relevance of the reference to the respective research question or theme. Abstracts are taken directly from the reference; summaries represent the ideas of the author(s).

The selected references relevant to data governance frameworks and their application to big data systems are grouped within three themes mapped directly to the research sub-questions.

What is a big data system?

Borkar, V., Carey, M., & Li, C., 2012a. Big data platforms: What's next?. *XRDS* 19, 1 (September 2012), 44-49. doi:10.1145/2331042.2331057. Retrieved from http://doi.acm.org/10.1145/2331042.2331057

Abstract. Three computer scientists from UC Irvine address the question "What's next for big data?" by summarizing the current state of the big data platform space and then describing ASTERIX, their next-generation big data management system.

Credibility. Vinayak R. Borkar holds a Master of Science degree in computer science and at the time of publication was a doctoral student in computer science at The University of California, Irvine. Michael J. Carey and Chen Li, both hold PhD degrees in computer science and are professors in the Bren School of Information and Computer Sciences at The University of California, Irvine. This article appears in a peer-reviewed journal and cites additional leaders in the field of big data systems.

Summary. The authors discuss the expanding definition of what constitutes a big data system from its appearance in the 1970s as data sets measured in megabytes to the

current state wherein data sets are comprised of terabytes and higher. The article also explores the origins of big data systems and the underlying architectures invented to support analytical processing. One such architecture, known as the ASTERIX data model (ADM) is explored in detail including a description of the model's key components, query language and capabilities. The authors conclude the article by discussing the future-state of big data systems and how they will differ from the current state solutions available for implementation. The article cites eleven sources from a variety of respected technical journals and one article from *The New York Times*.

Chen, H., Chiang, R. H. L., & Storey, V. C. (January 01, 2012). Business intelligence and analytics: From big data to big impact. *Management Information Systems Quarterly, 36*, 4, 1165-1188. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=cph&AN=83466038&site=ehost-live&scope=site

Abstract. Business intelligence and analytics (BI&A) has emerged as an important area of study for both practitioners and researchers, reflecting the magnitude and impact of data-related problems to be solved in contemporary business organizations. This introduction to the MIS Quarterly Special Issue on Business Intelligence Research first provides a framework that identifies the evolution, applications, and emerging research areas of BI&A. BI&A 1.0, BI&A 2.0, and BI&A 3.0 are defined and described in terms of their key characteristics and capabilities. Current research in BI&A is analyzed and challenges and opportunities associated with BI&A research and education are identified.

Credibility. Dr. Hsinchun Chen holds a PhD in information systems from New York University and is a Regents' Professor at the University of Arizona, where he was

awarded the Thomas R. Brown Chair in Management and Technology and directs the Artificial Intelligence Laboratory. Dr. Chen is also the Founder of the Caduceus Intelligence Corporation, a healthcare information systems company. Dr. Chen has authored 28 journal articles as of 2011, concerning business intelligence systems, artificial intelligence and computer networking. Dr. Roger Chiang holds a PhD in computer and information systems from the University of Rochester and is an associate professor at the University of Cincinnati's Carl H. Lindner College of Business. Dr. Chiang's research and teaching interests include but are not limited to business intelligence, data management, data mining and knowledge discovery. Dr. Chiang has authored 56 scholarly publications and supervises undergraduate students. Veda C. Storey holds a PhD from the University of British Columbia and is a professor of computer information systems and computer science at J. Mack Robinson College of Business at Georgia State University. Dr. Storey's research appears in a number of peer-reviewed, scholarly journals including *CM Transactions on Database Systems, IEEE Transactions on Knowledge and Data Engineering, Information Systems Research,* and *Information & Management*. Dr. Storey has also served on the editorial board of multiple information technology journals including *Information Systems Research, MIS Quarterly, DataBase,* and *Decision Support Systems*.

Summary. The authors discuss the history of business intelligence and analytics (BI&A) from both an applied and research-oriented perspective. This article describes how BI&A frameworks have developed to support: (a) BI&A 1.0, characterized by a data-centric approach towards large data sets, prevalent in the 1990's; (b) BI&A 2.0, in which Internet search engines enabled Web 2.0 applications to dynamically collect user-profile

information and prompt organizations to expand storage and analytical capabilities; to the future-state (c) BI&A 3.0 wherein big data analytics will focus primarily on mobile devices to align with emerging Web 3.0 (i.e., semantic web) technologies.

The authors also provide relevant research concerning BI&A and identify subject matter experts leading research into big data analytics and big data systems development. Additionally, the article includes how big data analytics, enabled by big data systems, are applied in:

- Commerce and Market Intelligence
- Government
- Science and Technology
- Healthcare
- Information and Physical Security

Cukier, K., & Mayer-Schoenberger, V. (2013). The rise of big data. *Foreign Affairs*, *92*(3), 27-40. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=87000329&site=ehost-live&scope=site

Abstract. The article discusses the effect of increasing quantities of digital information, or big data, on the way humans interact, communicate, and learn, with information on how it has altered the popular perception of data. Topics include the determination of correlative rather than causative relationships in statistical research using large quantities of data, the lack of accuracy and precision of data created through resources such as the Internet, and the ability of technology to produce larger statistical samples.

Credibility. This article appears in *Foreign Affairs*, a peer-reviewed journal founded in 1922. Kenneth Cukier is the Data Editor of *The Economist* whose writings primarily focus on information technology and internet governance. Mr. Cukier was a research fellow at the Kennedy School of Government at Harvard University where he focused on the internet and foreign relations and is a frequent contributor to respected publications such as *The New York Times, The Washington Post, The Financial Times* and news outlets including CBS, CNN, NPR, and the BBC. Viktor Mayer-Schonberger is Professor of Internet Governance and Regulation at the Oxford Internet Institute at the University of Oxford. Prior to Oxford, Mayer-Schonberger was a faculty member at the Kennedy School of Government at Harvard University. Mayer-Schonberger holds graduate degrees from the University of Salzburg, Harvard Law School, the London School of Economics and the University of Graz. Mayer-Schonberger is considered a leading expert on laws and regulations pertaining to information.

Summary. This article is adapted from the authors' book, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* published by Houghton Mifflin Harcourt in 2013. The authors identify the term *datafication*: as the ability to collect and analyze interactions and concepts in technological ways for the first time. For example, global positioning systems (GPS) used to capture location and online user preferences correlated with *likes* on social media applications such as Facebook. The article also discusses how big data transforms healthcare, marketing, politics, education and social work. The authors conclude that decisions based on big data analytics require human interaction and judgment to provide context within purely statistical data sets.

Dysart, J. (2013). The dawn of big data. *ABA Journal, 99*(5), 42-47. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=87445429&site=ehost-live&scope=site

Abstract. In this article, the author provides a brief insight of data mining techniques and its relevance in data management of corporate laws and legislations. It informs that data allows companies and lawyers to understand the real opportunities and risk in a lawsuit or claim. It also emphasizes the role of digital libraries in predicting the outcome of a court case and also offers information on a content fabrication system of the company LexisNexis Group.

Credibility. This article appears in the ABA Journal, a monthly professional publication of the American Bar Association. The ABA Journal is a well-respected journal first published as the Annual Bulletin in 1908. The author, Joe Dysart is a freelance writer unaffiliated with any publisher or organization and has published in over 40 publications including *The New York Times* and *The Financial Times of London*. Within the article, Mr. Dysart cites credible subject matter experts such as Mark A. Lemley, a law school professor at Stanford University; Ian Koenig, the chief architect of LexisNexis; and software executives of several applications providing big data analytics.

Summary. The author explores the impacts of big data systems within the context of practicing law within the United States. The author begins by defining big data and proceeds to discuss how applying big data analytics within law firms and legal publishers is transforming how attorneys manage their practices:

- **Big data and predictive analytics.** Software developers are marketing applications that provide predictions of case outcomes based on precedents established by similar cases for malpractice and patent attorneys.
- **Big data and changing fee schedules.** The author indicates law firms may now utilize data warehouses such as the TyMetrix LegalView data service to compare their rates for services against what corporate legal departments pay for similar services.
- **Hiring practices and big data.** Hiring practices for law firms usually consists of hiring recent law graduates based on law school rankings and test scores; however, big data analytics allow for law firms to benchmark hiring practices based on key success metrics inherent in the performance of their most successful lawyers.

Gopalkrishnan, V., Guszcza, J., Lewis, H., & Steier, D. (2012). Big data, big business: Bridging the gap. In *Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications* (BigMine '12). ACM, New York, NY, USA, 7-11. doi:10.1145/2351316.2351318. Retrieved from http://doi.acm.org/10.1145/2351316.2351318

Abstract. Business analytics, occupying the intersection of the worlds of management science, computer science and statistical science, is a potent force for innovation in both the private and public sectors. The successes of business analytics in strategy, process optimization and competitive advantage has led to data being increasingly recognized as a valuable asset in many organizations. In recent years, thanks to a dramatic increase in

the volume, variety and velocity of data, the loosely defined concept of "Big Data" has emerged as a topic of discussion in its own right -- with different viewpoints in both the business and technical worlds. From our perspective, it is important for discussions of "Big Data" to start from a well-defined business goal, and remain moored to fundamental principles of both cost/benefit analysis as well as core statistical science. This note discusses some business case considerations for analytics projects involving "Big Data", and proposes key questions that businesses should ask. With practical lessons from Big Data deployments in business, we also pose a number of research challenges that may be addressed to enable the business analytics community to bring the best data analytic practices when confronted with massive data sets.

Credibility. This article is authored by four members of the consulting firm of Deloitte and Touche Financial Advisory Services. Deloitte and Touche created the Deloitte Analytics Institute to offer advisory services for big data initiatives in a range of industries including banking, insurance, media and healthcare. The authors include propriety research from Deloitte and Touche and reference peer-reviewed, scholarly journals including the *Harvard Business Review* and several proceedings from *IEEE* conferences.

Summary. The authors begin by discussing recent developments in how organizations use big data to make decisions. They conclude with a prediction that this trend will become the norm in the near future. The authors indicate that investing in big data analytics requires a business case approach to (a) determine return on investment; (b) assess the feasibility of successful implementation; and (c) identify business problems this level of analysis is intended to solve. The authors also provide an overview of

potential research opportunities seeking to improve big data systems implementations and lower the financial overhead of transitioning to big data analytics.

Gorton, I. & Gracio, D. (2012). *Data-Intensive Computing : Architectures, Algorithms, and Applications*. Cambridge University Press. Retrieved from http://orbis.eblib.com.libproxy.uoregon.edu/patron/FullRecord.aspx?p=1042395&userid=8r16fsFJMWM%3d&tstamp=1369014206&id=78740BE31F39B1357B660549FB3F1EC170A5BCB2&conl=uoregon

Abstract. The world is awash with digital data from social networks, blogs, business, science and engineering. Data-intensive computing facilitates understanding of complex problems that must process massive amounts of data. Through the development of new classes of software, algorithms and hardware, data-intensive applications can provide timely and meaningful analytical results in response to exponentially growing data complexity and associated analysis requirements. This emerging area brings many challenges that are different from traditional high-performance computing. This reference for computing professionals and researchers describes the dimensions of the field, the key challenges, the state of the art and the characteristics of likely approaches that future data-intensive problems will require.

Credibility. Ian Gorton directed the Data Intensive Scientific Computing group for the Pacific Northwest National Laboratory (PNNL). Mr. Gorton holds a PhD in computer science also served as the Chief Architect for the Data Intensive Computing Initiative at PNNL. Deborah Gracio holds a Master of Science degree in electrical engineering and currently serves as the Director of Computational and Statistical Analysis at PNNL.

Summary. Gorton and Gracio provide a detailed inquiry into the evolution of big data analytic processing systems and how computational technology is changing to meet the challenges large datasets pose. Gorton and Gracio discuss how systems are adapting in terms of (a) scalability, (b) modifiability, and (c) complexity to facilitate the demands of big data analytics within the exabyte range and beyond. The authors explore the current architectures utilized to design big data analytic systems and discuss how large, cloud-based data sets are accessed and analyzed.

Madsen, L. (2012). *Healthcare business intelligence: A guide to empowering successful data reporting and analytics + website* [Kindle version]. New York: John Wiley & Sons.

Abstract. Solid business intelligence guidance uniquely designed for healthcare organizations. Increasing regulatory pressures on healthcare organizations have created a national conversation on data, reporting and analytics in healthcare. Behind the scenes, business intelligence (BI) and data warehousing (DW) capabilities are key drivers that empower these functions. Healthcare Business Intelligence is designed as a guidebook for healthcare organizations dipping their toes into the areas of business intelligence and data warehousing.

Credibility. Laura Madsen holds a Master of Science degree from the University of Wisconsin, Stout. Ms. Madsen founded the Healthcare Business Intelligence Summit and leads the healthcare practice for Lancet, a leading BI consulting firm. At Lancet, Ms. Madsen leads the overall strategy and product development initiatives for the healthcare sector and works with key accounts across the country in the provider, payer, and healthcare manufacturing markets. Prior to joining Lancet, Laura held senior positions

with several leading healthcare technology companies, including UnitedHealth Group, and a national pharmacy benefit management company.

Summary. The author discusses business intelligence within the context of the healthcare field. Within the text, the author devotes a chapter to data quality including a guide to establishing data governance. The author notes that data governance is especially important in healthcare information technology as a means to achieve regulatory compliance. Two key layers of data governance are presented: (a) a data governance committee comprised of executives to enforce governance policies, and (b) data stewards, subject matter experts associated with business units to determine governance policies and drive compliance. Within the author's governance model, the data stewards represent the tactical, or process, element while the governance committee represents the strategic element. The author provides guidance on staffing the two elements of the governance model, templates for writing governance policies and a case study of how a healthcare organization implemented a customized data governance program.

Tamm-Daniels, R. (2013). The key to smart big data: Know thy technology. *Information Today, 30*(4), 19. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=86640870&site=ehost-live&scope=site

Abstract. The author discusses the concept of Big Data, or extremely large data sets which require specific technologies for its storage, management, and retrieval. Particular focus is given to the types of unstructured data, including machine or system-generated data and content such as online social media data, e-mail, and documents. Other topics

include real time and batch data processing, the effects of data volume and value on storage needs, and user access to relevant data.

Credibility. Rik Tamm-Daniels is the co-founder and vice president of Attivio, Inc. Attivio develops data analytics software for the purposes of gaining insights into big data sets including structured data such as information from enterprise applications and unstructured data such as information from emails and blogs. Rik Tamm-Daniels holds multiple Master's degrees in computer systems engineering from Boston University.

Summary. The author states that although big data is a commonly discussed topic within the information technology community, what he refers to as the *nature of data* is widely overlooked. Tamm-Daniels states that the nature of data refers to aspects distinguishing data from big data. These aspects include (a) volume of data in the dataset, (b) data location, (c) speed of creation, (d) accessibility requirements, and (e) insights provided by the dataset? The author notes that these aspects must first be realized to determine how to best implement big data systems. The author describes the differences between structured and unstructured data and how data type prompts decision-making concerning the type of big data technologies to implement. The author also discusses the importance of not only the nature of data but the specific problems that big data analytics seek to solve.

What is a data governance framework and why is it important?

Aiken, P., Gillenson, M., Rafner, D., & Zhang, X. (2011). Data management and data administration: Assessing 25 years of practice. *Journal of Database Management. 22, 3* (July 2011), 24-45. doi:10.4018/jdm.2011070102. Retrieved from http://go.galegroup.com/ps/retrieve.do?sgHitCountType=None&sort=DA-SORT&inPS=true&prodId=AONE&userGroupName=s8492775&tabID=T002&searchId

=R1&resultListType=RESULT_LIST&contentSegment=&searchType=AdvancedSearchForm&currentPosition=2&contentSet=GALE%7CA267609846&&docId=GALE|A267609846&docType=GALE&role=

Abstract. Data management (DM) has existed in conjunction with software development and the management of the full set of information technology (IT)-related components. However, it has been more than two decades since research into DM as it is practiced has been published. In this paper, the authors compare aspects of DM across a quarter-century timeline, obtaining data using comparable sets of subject matter experts. Using this information to observe the profession's evolution, the authors have updated the understanding of DM as it is practiced, giving additional insight into DM, including its current responsibilities, reporting structures, and perceptions of success, among other factors. The analysis indicates that successfully investing in DM presents current, real challenges to IT and organizations. Although DM is evolving away from purely operational responsibilities toward higher-level responsibilities, perceptions of success have fallen. This paper details the quarter-century comparison of DM practices, analyzes them, and draws conclusions.

Credibility. Dr. Peter Aiken is an associate professor in the information systems department at the Virginia Commonwealth University. Dr. Aiken holds a PhD degree in information technology from George Mason University and is the founding director of the Institute for Data Research. Dr. Aiken has published in peer-reviewed journals including *IEEE Computer* and *Computational and Mathematical Organization Theory*. Dr. Mark Gillenson is a professor of management information systems and director of MBA programs in the Fogelman College of Business and Economics at the University of

Memphis. Dr. Gillenson holds a PhD degree from The Ohio State University and has published in *Information & Management, European Journal of Information Systems* and *Information Systems Management*. Dr. Xihui Zhang is an assistant professor of computer information systems at the University of North Alabama. Dr. Zhang holds a PhD degree in business administration from the University of Memphis and has published in the *Journal of Strategic Information Systems, Journal of Database Management, Journal of Computer Information Systems, Journal of Organizational and End User Computing, e-Service Journal, Journal of Information Systems Education, Journal of Information Technology Management* and the *Journal of Information Technology Education*. David Rafner holds a degree from Virginia Commonwealth University and directs research and project development for the Richmond Group Fund Company.

Summary. The authors discuss data management as a means to comprehend not only the current needs of an organization to benefit from its information but also the future needs. Data Management requires the realization of an organization's current datasets via analytical systems and the formation of strategies to capture, distribute and benefit from internal and external data to the organization. The authors cite surveys conducted in 1981 and 2007 to the same community of data managers and determine that data management programs evolved into highly institutionalized governance efforts with beneficial outcomes for many of the respondents' organizations. The authors also note that a majority of respondents in both time periods initially invested in database technologies instead of data management and requisite data governance policies and procedures. As a means to improve data management, the authors identify findings from their 2007 survey indicating data management operations should be considered a management function and

those responsible for operating data governance models should be decentralized throughout the organization to enforce compliance throughout the enterprise.

Alhajj, R., Ridley, M., & Rifaie, M. (2009). Data governance strategy: A key issue in building enterprise data warehouse. *Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services* (iiWAS '09). ACM, New York, NY, USA, 587-591. doi:10.1145/1806338.1806449. Retrieved from http://doi.acm.org/10.1145/1806338.1806449

Abstract. This paper articulates data governance as one of the key issues in building Enterprise Data Warehouse. The key goals of this document are to: (a) define the strategy for Data Governance processes and procedures; (b) define the scope of and identify major components of the data governance processes; (c) adhere to enterprise Data Management standards, principles and guidelines; and (d) articulate a vision for building, managing and safeguarding enterprise data foundation. The client-centric focus of business organizations coupled with aggressive attention to the bottom line propelled initiatives such as Data Governance to the top of the list of IT and business executives. The recent financial crisis which spawned the worldwide economic meltdown has been to a great extent blamed on non-trustworthy and non-transparent data. It is becoming progressively and patently evident that data MUST be managed like other assets such as financial and human resources. It has to have defined and mandated set of controls where compliance can be objectively measured and reported.

Credibility. Mohammad Rifaie holds a PhD from Bradford University in computer science. While at the School of Informatics at Bradford University, Dr. Rifaie researched and published in peer-reviewed journals on data governance and big data analytics. Reda

Alhajj holds a PhD degree from Bilkent University at Ankara and is a professor of computer science at the University of Calgary. Dr. Alhajj has published over 300 papers in journals and conferences. Dr. Alhajj serves on the editorial boards of the following journals: *Journal of Information Assurance and Security, Journal of Data Mining and Bioinformatics, Journal of Data Mining and Modeling and Management*. Mick Ridley is a lecturer in computer at the Bradford University School of Informatics. Mr. Ridley holds a Master of Science degree and researches database management, XML and bibliographic databases. Mr. Ridley's work appears in peer-reviewed journals including the *International Journal of Machine Learning and Computing*, the *International Journal of Business Intelligence and Data Mining* and the *Journal of Information Management*. Mr. Ridley has also published chapters in books regarding database management and information technology.

Summary. The authors contend that as organizations pursue big data analytics initiatives, having an established a data governance program is a prerequisite for achieving success. The authors refer to research in the field of big data management and data governance in describing data governance models that provide accountability regarding how data is collected, quantified and monitored. The authors note that a data governance framework requires: (a) determining how decisions regarding data management will be made, (b) establishing how decision-making processes will be made, and (c) determining how business intelligence derived from big data analytics will be communicated within the organization. The authors contend while data governance is mandated by organizational leaders, the effective management of data assets are the responsibility of all stakeholders

interacting with the data during its lifecycle. The authors identify and discuss specific governance objectives:

- **Managing risks**. The authors note that as data becomes an asset to the organization, risk management practices and controls should be applied to data to ensure security and recoverability.
- **Data accountability**. The authors compare data assets to financial assets in that executives must be held accountable for observing legal and regulatory compliance.
- **Measurement of Performance**. The authors suggest instruments such as a balanced scorecard approach to track performance of compliance, risk management and accountability efforts.

The authors provide a six-stage maturity model in which data governance moves from being nonexistent to completely optimal within the organization.

Borkar, V., Carey, M., & Li, C. (2012b). Inside "big data management": Ogres, onions or parfaits?. EDBT/ICDT 2012 Joint Conference Berlin, Germany. Retrieved from http://www.edbt.org/Proceedings/2012-Berlin/papers/keynotes/a3-carey.pdf

Abstract. In this paper the authors review the history of systems for managing "Big Data" as well as today's activities and architectures from the (perhaps biased) perspective of three "database guys" who have been watching this space for a number of years and are currently working together on "Big Data" problems. The focus is on architectural issues, and particularly on the components and layers that have been developed recently (in open source and elsewhere) and on how they are being used (or abused) to tackle challenges posed by today's notion of "Big Data". Also covered is the approach the

authors are taking in the ASTERIX project at UC Irvine, where they are developing their own set of answers to the questions of the "right" components and the "right" set of layers for taming the "Big Data" beast. The authors also sharing opinions on what some of the important open questions are in this area as well as thoughts on how the data intensive computing community might best seek out answers.

Credibility. Vinayak R. Borkar holds a Master of Science degree in computer science and at the time of publication was a doctoral student in computer science at The University of California, Irvine. Michael J. Carey and Chen Li, both hold PhD degrees in computer science and are professors in the Bren School of Information and Computer Sciences at The University of California, Irvine. This article appears in a peer-reviewed journal and cites additional leaders in the field of big data systems.

Summary. The authors begin by discussing the evolution of big data systems from the *database machines* of the 1970's to the Hadoop-powered platforms of today. The authors use an analogy from the film *Shrek* to illustrate their main point: Parallel Database Systems are like onions in that they have a number of layers and users are only allowed to see the outer layer; however, open-source big data stacks are like ogres since they are extremely strong and capable but difficult to work with. The parfait analogy is described as something everyone finds favorable and involves combining elements from parallel database systems and open-source big data stacks to produce new database architectures to support big data systems.

Breaux, T., & Alspaugh, T. (2011). Governance and accountability in the new data ecology. 2011 Fourth International Workshop of Requirements Engineering and Law (RELAW), 2011. 14, 30-30 Aug. 2011. doi:10.1109/RELAW.2011.6050267. Retrieved from

http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6050267&isnumber=6050260

Abstract. Electronic data licenses (EDLs) are data governance instruments that consist of legal rules (rights, obligations and prohibitions) governing an organization's data practices. These rules include data requirements, such as rights to collect, use, retain and transfer data to third parties and prohibitions preventing these practices. We introduce the EDL concept by describing the emerging data ecology, wherein information sharing will reach unprecedented scale, and by presenting legal foundations for the EDL concept. We conclude with a broad vision for the EDL framework by discussing the license management and composition strategies, criteria for evaluating solutions, and how EDLs should support data principles and standards, before concluding with a review of related work that supports this vision.

Credibility. Travis D. Breaux is an assistant professor of computer science at the Carnegie Mellon University Institute for Software Research and holds a PhD in computer science from North Carolina State University. Dr. Breaux has published in peer-reviewed journals including *Requirements Engineering Journal, Computers and Security, ACM Transactions on Software Engineering Methodology, I*EEE Transactions *on Software Engineering.* Thomas A. Alspaugh is an adjunct professor at Georgetown University Department of Computer Science and a visiting assistant researcher at the University of California, Irvine Institute for Software Research. Dr. Alspaugh holds a PhD in computer science from North Carolina State University and has published in peer-reviewed journals and conference proceedings including *Journal of Systems and Software, ACM Symposium on Software Visualization, First International Workshop on Software*

Ecosystems, 2nd International Workshop on Emerging Trends in FLOSS Research and Development and *Architecting Dependable Systems.*

Summary. The authors discuss the rise in large datasets and the harm poor data quality can cause individuals, especially when business decisions are made via automated data analytics based on incomplete or false information. The authors discuss electronic data licenses (EDLs) as a means to control the flow of information within a data supply chain. These EDLs serve as a means to enforce compliance to a larger data governance model. The authors define EDLs as contracts hard-coded into data sets that identify terms of use, privacy concerns and service terms. The EDL process represents a maturation of data governance in that policies and controls relevant in the governance model are made inherent in the datasets thereby encouraging compliance and programmatically controlling usage and transference. The authors advise EDLs may contain state laws and may also reference regulatory, authoritative sources. The authors present an EDL framework encouraging commonalities in how licenses are written, linked to specific sets of laws and utilized in much of the same way software licenses are distributed, activated and expired. Within this framework, data handling practices are easily identified and transparency in how datasets should be handled is immediately apparent and recognizable by data handlers and analytical systems.

Dewan, R., Freimer, M., & Story, V. (2012). Data quality: Setting organizational policies. *Decision Support Systems. 54, 1* (December 2012), 434-442. doi:10.1016/j.dss.2012.06.004. Retrieved from http://dx.doi.org/10.1016/j.dss.2012.06.004

Abstract. The collection, representation, and effective use of organizational data are important to a firm because these activities facilitate the increasingly important analysis needed for business operations and business analytics. Poor data quality can be a major cause for damages or losses of organizational processes. The many tasks that individuals perform within an organization are linked and normally require access to shared data. These linkages are often documented as process flow diagrams that connect the data inputs and outputs of individuals. However, in such a connected setting, the differences among individuals in terms of their preferences for data attributes such as timeliness, accuracy, and others, can cause data quality problems. For example, individuals at the head of a process flow could bear all of the costs of capturing high quality data but not receive all of the benefits, even though the rest of the organization benefits from their diligence. Consequently, these individuals, in absence of any managerial intervention, might not invest enough in data quality. This research analyzes this problem and proposes a set of solutions to this, and similar, organizational data quality problems. The solutions focus on principles of employee empowerment, decentralization, and mechanisms to measure and reward individuals for their data quality efforts.

Credibility. Veda C. Storey holds a PhD from the University of British Columbia and is a professor of computer information systems and computer science at J. Mack Robinson College of Business at Georgia State University. Dr. Storey's research appears in a number of peer-reviewed, scholarly journals including *CM Transactions on Database Systems, IEEE Transactions on Knowledge and Data Engineering, Information Systems Research,* and *Information & Management*. Dr. Storey has also served on the editorial board of multiple information technology journals including *Information Systems*

Research, MIS Quarterly, DataBase, and *Decision Support Systems*. Rajiv M. Dewan is a senior associate dean and professor of computers and information systems at the William E. Simon Graduate School of Business Administration at the University of Rochester. Dr. Dewan holds a PhD in business from the University of Rochester. Dr. Dewan has published articles regarding information systems in journals including the *European Journal of Operations Research, Journal of Management Information Systems, Information Systems Research Journal and Communications of the ACM.* Marshall Freimer is a professor of management and information systems (retired) at the William E. Simon Graduate School of Business Administration at the University of Rochester. Dr. Freimer holds a PhD in mathematics from Harvard University and publishes research in journals concerning information systems, management, engineering and mathematics.

Summary. Authors identify data quality management as a component to an organization's overall data management model. The authors present a process-oriented organizational data quality model capturing the flow of data within an organization. The authors map the model to mathematical parameters, present assumptions and leverage economic analysis methods to construct two theorems:

- **Theorem 1.** A data quality problem exists within organizations in which individuals responsible for determining data quality commonly do not consider the needs of the entire organization and consequentially undervalue data.
- **Theorem 2.** To optimize data quality management the organization should properly incentivize and motivate employees to make decisions regarding data value with respect to the needs of the entire organization.

The authors present these theorems as a means to understand why data quality problems exist and offer a means to prompt compliance to a data governance model that mandates policies and controls in the transference of data inside and outside of organizations.

Gendron, M. (2012). Business intelligence applied: Implementing an effective information and communications technology infrastructure. New York: Wiley. Retrieved from http://onesearch.uoregon.edu/metasearch/record?group=2013-04-04-000383&resultSet=002350&startRecord=12

Abstract. Enterprise performance management (EPM) technology has been rapidly advancing, especially in the areas of predictive analysis and cloud-based solutions. Business intelligence caught on as a concept in the business world as the business strategy application of data warehousing in the early 2000s. With the recent surge in interest in data analytics and big data, it has seen a renewed level of interest as the ability of a business to find the valuable data in a timely—and competitive—fashion. Business Intelligence Applied reveals essential information for building an optimal and effective information and communication technology (ICT) infrastructure.

Credibility. Michael S. Gendron holds a PhD in information science from the State University of New York at Albany and is a professor of management information systems at Central Connecticut State University. Dr. Gendron has published over 30 articles and multiple books in information management and business intelligence.

Summary. Dr. Gendron discusses the paradigm shift in which information technology departments have moved from providing operational services to strategic services to enable the organization to realize their goals. The author notes that organizations are utilizing business intelligence gained from improvements of hardware and analytical

applications to derive insights from large sets of incoming data. The author presents a business-driven infrastructure design cycle in which the data lifecycle and governance model is driven primarily based on the needs of customers and other external stakeholders. Dr. Gendron discusses the need for governance and accountability in how organizations not only manage their data but also how they design systems to collect, analyze and report upon information assets.

Khatri, V., & Brown, C. (2010). Designing data governance. *Communications of the ACM, 53*(1), 148-152. Retrieved from http://cacm.acm.org/magazines/2010/1/55771-designing-data-governance/fulltext

Abstract. The article presents a discussion of data governance, as distinguished from data management. It is noted that increasing regulatory pressures have led many firms to highlight the importance of best practices regarding data by addressing such issues at the strategic level. Data governance is conceptualized as a set of big-picture executive decisions about where the responsibility for day-to-day data management should be located, and what goals should be set in that context. A framework for developing a data governance strategy is presented which emphasizes five main areas of decision making.

Credibility. Vijay Khatri holds a PhD in information systems from the University of Arizona and is an associate professor of information systems at Indiana University, Bloomington. Dr. Khatri is also the co-director of the Kelley Institute for Business Analytics and has authored peer-reviewed publications concerning data governance, predictive analytics and data management. Carol Brown holds a PhD in management of information systems from Indiana University and is a professor at the Howe School of Technology Management, Stevens Institute of Technology. Dr. Brown has authored nine

books regarding information systems and business, teaches graduate-level courses in information technology and has authored a significant number of articles for peer-reviewed, scholarly journals.

Summary. The authors begin by discussing how organizations are realizing the importance of the *data as an asset* concept and the significance of implementing a data governance framework to manage risk and compliance. The authors provide five *domains* that drive decision-making concerning data governance:

- Data Principles
- Data Quality
- Metadata
- Data Access
- Data Lifecycle

These domains include leading questions to guide stakeholders through a process to identify potential decision-making tasks within an organization. The authors also describe how data governance intersects with information technology governance to reduce risk to the organization and align with the information technology strategy of the organization.

Kooper, M., Lindgreen, E., & Maes, R. (2011). On the governance of information: Introducing a new concept of governance to support the management of information. *International Journal of Information Management 31*, 3 (June 2011), 195-200. doi:10.1016/j.ijinfomgt.2010.05.009 http://dx.doi.org/10.1016/j.ijinfomgt.2010.05.009

Abstract. Information governance as an approach to better govern the use of information within and outside an organization is rapidly gaining popularity. A common and scientific ground for this approach has not yet been formulated. In this article the authors

describe a definition for information governance, extending the common, one-dimensional approach into a more generic statement. Starting from the well-known principles of IT governance, the authors further explore the aspects of both information and governance. Four hypotheses are proposed to give ground to the use of information governance.

Credibility. Michiel Kooper is a research fellow and faculty member at the University of Amsterdam and holds a graduate degree from Delft University of Technology. Mr. Kooper publishes articles on information governance and is also a management consultant at KPMG Advisory. Rik Maes is a professor in Information and Communication Management at the University of Amsterdam. Professor Maes also serves as the current dean of the Executive Master in Information Management degree at the University of Amsterdam. Edo Roos Lindgreen is a professor at the University of Amsterdam and directs the Executive Master of IT-Auditing at the University of Amsterdam and is a management consultant with KPMG Advisory.

Summary. The authors discuss the role of data governance as a means to connect the needs of the business to the supportive direction of information technology departments. The authors also note that the data governance is important to organizations as a means of realizing the value of information – both internally and externally created – and easing compliance efforts of the larger corporate governance model. The authors note that research on data governance is lacking and offer four hypotheses to serve as a foundation for further research:

- Information should be optimized and its value either determined as a unique value or weighted across stakeholders

- The value of information directly relates to how those responsible for acting upon it consider it to be reliable, relevant and useful
- Information management is dependent upon the individuals governing the use and distribution of information
- The optimization of information value depends upon constraining factors placed on the information creator.

Myers, P. (2012). Information governance-control content chaos. *KM World, 21* (8), 9. Retrieved from http://search.ebscohost.com.libproxy.uoregon.edu/login.aspx?direct=true&db=aph&AN=79553476&site=ehost-live&scope=site

Abstract. The article discusses strategies on how business enterprises should deal with information governance for large data. It suggests understanding data topology, employing real-time indexing of the business enterprise, and creating an information intelligence service center. It says that active information governance can help address the problems with big data, including records management, storage optimization, and electronic (e)-discovery.

Credibility. Phil Myers is the Chief Executive Officer of StoredIQ, a company that specializes in big data analytics and business intelligence. Mr. Myers holds an MBA degree from Penn State University.

Summary. The author discusses *content chaos*, a scenario in which managing data requires new techniques due to issues of volume and complexity. The author notes that past approaches to data management involved moving data to internal repositories where it is indexed, analyzed and stored; however, due to the advent of big data the time

required to support this approach is proving prohibitive as information is quickly out-dated. The author presents strategies leveraging new indexing technologies that, when applied to large datasets, increase enterprise search capabilities in a real-time capacity. The author offers strategies to apply new information governance principles to big data:

- Document data topologies and determine where enterprise data exists, quantity, ownership and business value
- Utilize indexing in real-time to track data changes and metadata
- Leverage effective data storage solutions to understand data relevancy
- Establish an information intelligence service center to manage not only enterprise data but the value of the organization's data
- Utilize change management processes to track shifting business processes and new data formations
- Ensure data security, storage, distribution and compliance is done in a proactive, policy-drive way
- Realize the potential of data as an asset and implement a data lifecycle model

These strategies constitute a data governance framework from which policies and procedures are aligned.

Österle, H., Otto, B., & Weber, K. (2009). One size does not fit all---a contingency approach to data governance. *Journal of Data and Information Quality* 1, 1, Article 4 (June 2009), 27 pages. doi:10.1145/1515693.1515696. Retrieved from http://doi.acm.org/10.1145/1515693.1515696

Abstract. The article presents the first results of a community action research project on data governance comprising six international companies from various industries. It

outlines a data governance model that consists of three components (data quality roles, decision areas, and responsibilities), which together form a responsibility assignment matrix. The data governance model documents data quality roles and their type of interaction with DQM activities. In addition, the article describes a data governance contingency model and demonstrates the influence of performance strategy, diversification breadth, organization structure, competitive strategy, degree of process harmonization, degree of market regulation, and decision-making style on data governance. Based on these findings, companies can structure their specific data governance model.

Credibility. The authors are researchers from the University of St. Gallen, Switzerland and this article is published in the *ACM Journal of Data and Information Quality*, a peer-reviewed publication. The researchers reference multiple peer-reviewed articles and present their research in a formal and professional manner.

Summary. The authors focus on the practice of Data Quality Management (DQM) in which organizations establish controls over how data will be collected, organized, used and eventually discarded. DQM drives organizations to achieve a higher quality of data utilization and therefore realize more value from their information assets. The authors indicate DQM is entirely reliant on data governance and observe that there is a limited amount of research on the subject. The authors present research into multiple types of information technology governance frameworks including centralized, decentralized and a federated model and describe how data governance aligns with the overall information technology governance model of the organization. The authors also discuss the benefits

of data governance in aligning information technology goals to the goals of the business as a principle aspect of the governance framework.

Soares. S. (2013). *Big data governance: An emerging imperative*. Mc Press.

Abstract. Written by a leading expert in the field, this guide focuses on the convergence of two major trends in information management—big data and information governance—by taking a strategic approach oriented around business cases and industry imperatives. With the advent of new technologies, enterprises are expanding and handling very large volumes of data; this book, nontechnical in nature and geared toward business audiences, encourages the practice of establishing appropriate governance over big data initiatives and addresses how to manage and govern big data, highlighting the relevant processes, procedures, and policies. It teaches readers to understand how big data fits within an overall information governance program; quantify the business value of big data; apply information governance concepts such as stewardship, metadata, and organization structures to big data; appreciate the wide-ranging business benefits for various industries and job functions; sell the value of big data governance to businesses; and establish step-by-step processes to implement big data governance.

Credibility. Sunil Soares is the founder and managing partner at Information Asset, LLC. Mr. Soares was the director of information governance at IBM where he provided consulting to organizations regarding best practices in data governance and information architecture. Mr. Soares holds a Master of Business Administration from the University of Chicago Booth School of Business.

Summary. The author begins by providing an analysis of the types of big data and offers applicable scenarios where various data types are analyzed to solve problem across

multiple industries. Mr. Soares proceeds with a sequential guide to implementing the data governance process within an organization including required inputs such as (a) a maturity assessment to determine readiness for data governance, (b) a business case to justify implementing data governance, and (c) a roadmap to guide the data governance implementation. The author presents case studies within each chapter citing real-world data governance implementations and heuristics. The author devotes an entire chapter to organizational considerations around managing a data governance program. These considerations include staffing data stewards responsible for enacting and controlling data governance processes such as ensuring compliance to metadata tagging and aligning executives to the data governance program by gaining sponsorship by means of representation on a governance council. The author devotes additional chapters to metadata management, privacy controls and data quality measurements.

What is the role of selected data governance frameworks in improving the effectiveness of business processes?

Allen M., & Cervo, D. (2011). Master data management in practice: Achieving true customer MDM. Retrieved from http://orbis.eblib.com.libproxy.uoregon.edu/patron/FullRecord.aspx?p=693520

Abstract. *Master Data Management in Practice* provides a logical order toward planning, implementation, ongoing management, and advanced practices of Customer MDM with tables, graphs, and charts. Authors Dalton Cervo and Mark Allen show organizations how to implement Master Data Management (MDM) within their business model to create a more quality controlled approach for successfully managing and maintaining their customer master data. This book focuses on techniques that can

improve data quality management, lower data maintenance costs, reduce corporate and compliance risks, and drive increased efficiency in customer data management practices.

- Designed for data management professionals and data management consulting firms
- Addresses the aspects of defining the underlying scope, approach, architecture, and objectives necessary for the planning and execution of a Customer MDM initiative
- Provides the practical insight, guidance, questions, and examples related to the implementation of the four foundational Customer MDM disciplines: Data Governance, Data Stewardship, Data Quality Management, and Data Access Management

Credibility. Dalton Cervo is president and founder of Data Gap Consulting and has over 20 years of experience in software development and data management. Mr. Cervo holds a Master of Business Administration degree. Mark Allen has over 20 years of experience in data management and consults to organizations in data analytics.

Summary. The authors describe *Master Data Management* (MDM) as a means to organize data into domains such as external customer data and internal business data and provide a sole, objective repository to enable business intelligence. The authors note that data governance is a key dependency to MDM and devote a chapter regarding initiating a data governance program. The authors advise that a successful data governance model deflects common mistakes organizations make in deriving optimal value from datasets and ensures data is properly managed throughout its lifecycle. The authors emphasize the importance of data governance, especially regarding customer data as this data domain is the primary means of making strategic decisions and realizing its value as an asset. Standards and procedures must be agreed to, adopted across the organization and ratified

within a data governance charter. The authors also advise organizations form data governance committees to enforce compliance to data governance policies and guide data stewardship. A unique aspect of the authors' governance model is utilizing service level agreements (SLAs) between business stakeholders and the information technology department to further enforce compliance to the data governance model with measurable performance indicators.

Alter, A. (2006). Information governance's big payback. *CIO Insight*, (68), 69-76. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=cph&AN=21057431&site=ehost-live&scope=site

Abstract. The article presents a research study on the information governance in the United States. The study conducted by the journal "CIO Insight" covers various aspects of information governance. According to the study, 65% of the respondents say their companies are effectively processing and using information for business advantage. But there are many companies in the United States whose information resources still need better management. The article also discusses the benefits of information governance. Of the companies which have information governance process in operation, 59% have started to get profit from the system. The study elaborates on the responses of the financial firms, manufacturers and executives about the usability, quality and productivity of information governance and management.

Credibility. Allan Alter has published articles concerning information technology and management consistently since 2003 with over 158 articles in *CIO Insight* and has edited two books. Mr. Alter was the editor-in-chief and director of new content development for

the *MIT Sloan Management Review*. Mr. Alter is currently a research fellow at the Accenture Institute for High Performance and holds a master's degree from the University of Michigan.

Summary. The author presents findings from a 2006 *CIO Insight* survey of information technology executives indicating that while a majority of respondents reported that one of their company's greatest strengths is the creation and utilization of information, only 52% of larger companies and 45% of smaller companies actually have a governance program to establish controls and grant decision-making authority for data. The findings also indicate that when asked what types of data the governance model concerns, respondents indicated that internal structured data is more important than external data, both structured and unstructured. The author also presents findings indicating that companies that do not have a successful data governance process report that their internal and external data is of low quality.

Ashford, W. (2013). Embracing big data can lead to greater security. *Computer Weekly, 6-7.* Retrieved from http://web.ebscohost.com.libproxy.uoregon.edu/ehost/pdfviewer/pdfviewer?sid=7a48c664-6d0b-4153-a2d0-3b6c44bf0425%40sessionmgr113&vid=2&hid=125

Abstract. The article discusses security issues related to big data. The author notes that executive chairman of RSA, the security division of information technology (IT) storage hardware company EMC, Art Coviello, has consistently warned against emerging threats to big data. Comments from Coviello on the ability of intelligence-driven security and big data analytics to protect against attacks, the value of big data, and the scalability of big data.

Credibility. Warwick Ashford is security editor for *Computer Weekly* and a frequent contributor to professional journals on topics including information technology security, business continuity and disaster recovery and information technology governance, risk and compliance.

Summary. The author discusses the proceedings of the RSA Conference 2013 in which Art Coviello, executive chairman of RSA presented a keynote address on big data and information security. RSA is the security division of EMC, a developer of big data analytics systems and storage devices for large data sets. The author provides direct quotations from Coviello's perspective on the relationship between big data systems and information security in which big data analytics, when integrated with information security initiatives, will predict threatening behaviors and provide real-time defensive controls. The author notes that when these future-state technologies are fully developed and deployed they must align with governance, risk and compliance systems and security strategies centered on big data analytics.

Coleman, D., Hughes, A., & Perry, W. (2009). The role of data governance to relieve information sharing impairments in the federal government, *Computer Science and Information Engineering, 2009 WRI World Congress on Computer Science and Information Engineering*, vol.6. pp.267 - 271, March 31 2009-April 2 2009. doi: 10.1109/CSIE.2009.630. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5171000&isnumber=5170943

Abstract. The technical mechanics of data sharing, particularly as related to metadata, has been reasonably well researched and to an extent is reaching technical maturity. The Federal Government Departments and Agencies have, to one degree or another, adopted

methodologies to organizationally frame their internal use of metadata. However there is evidence to suggest that the Federal Government has not been entirely successful in taking the necessary next steps to build a metadata governance model. One of the presumed impairments to this effort is that a mature metadata governance framework may be perceived as posing a threat to organizational and personal equities within government. This paper will examine from an organizational perspective this presumed impairment and provide a level of analysis as to relevant factors. Finally, it will examine potential data sharing governance policy changes that may be necessary to enable effective use of metadata in information sharing.

Credibility. David W. Coleman earned a PhD degree from George Mason University and is a lead engineer at The MITRE Corporation. Allen A. Hughes is a visiting professor of management information systems at George Mason University and holds a PhD degree from Carnegie-Mellon University. Wayne D. Perry holds a PhD degree in quantitative economics and public policy from Carnegie-Mellon University. Dr. Perry has published in peer-reviewed journals including *International Journal of Education Research, The Nonproliferation Review* and *Management Science Approaches to Manpower Planning and Organization Design.*

Summary. The authors discuss the importance of information technology systems to large organizations such as the United States government due to large data set reporting and manipulation capabilities. The authors note that while the effective use of large data sets within the private sector directly attributes to economically quantified gains, governmental organizations are not inclined to focus on profits; instead, these organizations concentrate efforts on the effective use of data and information sharing

amongst other government organizations and the public for which they serve. The authors discuss how organizations may utilize metadata, a key data governance component, to improve data quality and sharing efforts. Research on metadata to improve information sharing is cited and common metadata languages including XML and OWL are identified. A key impediment to such governmental organizations effectively implementing a shared metadata tagging procedure resides in the various organizational challenges inherent in these sometimes divergent units of governments. These impediments include issues concerning intra and extra-organizational cooperation, bureaucratic motivations and conflicting alignments due to rigid security and hierarchical structures. While these issues may appear to be specific to government organizations, private organizations may experience many of the same issues. The authors present a data governance model in which:

- Leadership is educated on the overall data governance model and engaged in the decision-making process
- Information technology goals are aligned with the needs of the customers
- A common information technology architecture is utilized resulting in a standardized portfolio of applications
- The governance model is institutionalized and its processes repeatable over time.

Franks, B. (2012). Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics. Retrieved from http://orbis.eblib.com.libproxy.uoregon.edu/patron/FullRecord.aspx?p=821898

Abstract. With its unique focus on the effective analysis of these increasingly popular data sources, Taming the Big Data Tidal Wave shows you step by step how to design,

incorporate, and profit from a world-class advanced analytics ecosystem in today's big data environment. Author and analytics expert Bill Franks shares his "from the trenches" perspective to show you how big data is changing the world of analytics. He maps out an easy-to-follow action plan to help your organization uncover new business opportunities, implement new business processes, and make better-informed decisions.

Credibility. Bill Franks is chief analytics officer for Teradata's global alliance programs and provides guidance to organizations in utilizing big data systems. Mr. Franks holds a Master of Science degree in applied statistics from North Carolina State University. Mr. Franks is also a faculty member of the International Institute for Analytics and is an active contributor to conferences and blogs concerning big data analytics.

Summary. The author advises that big data analytical processing will be a key component in services provided by information technology departments within organizations. Innovation, the author contends, requires risk and implementing big data analytical systems can be a challenging endeavor for organizations as they encounter new technologies and processes. The author provides a governance model that places much of the responsibility of ensuring compliance on the analytics professionals within the organization. The author also offers guidance on realizing the benefits of big data analytics by moving organizations away from the precedent of slow-moving, hierarchical information technology departments to faster, more agile teams focuses on immediate decision-making based on real-time data. Data analysts should be incentivized to follow governance policies and encouraged to leverage analytical outcomes in ways beneficial to the organization.

Goldston, D. (January 01, 2008). Big data: Data wrangling. *Nature, 455,* 7209. doi:10.1038/455015a. Retrieved from http://www.nature.com/news/2008/080903/full/455015a.html

Abstract. A former chief of staff of the House Committee on Science discusses applications of big data analytics for public benefit and how data governance manifests at the federal level.

Credibility. Dr. David Goldston is director of government affairs for the Natural Resources Defense Council. Dr. Goldston holds a PhD degree from the University of Pennsylvania and previously served as the chief of staff of the House Committee on Science from 2001 through 2006. Dr. Goldston is also a scholar in residence at Woodrow Wilson School at Princeton University. The journal *Nature* is peer-reviewed and presents research on science, medicine and technology.

Summary. The author discusses big data analytics within the context of environment policy. The author contends that the governance of big data analytics within the United States government requires statistical agencies completely objective and separated from political influence. The author cites the proposed formation of a Bureau of Environmental Statistics (BES) responsible for big data analytical processing and reporting to provide empirical research to guide federal policy decisions. The author discusses scenarios in which big data sets utilized for research purposes for the U.S. federal government are not freely released to the public due to concerns over privacy. The author argues that transparency and objectivity, key components in a data governance model, are essential for credible research and further inquiry.

Jacob, V., Parssian, A., & Sarkar, S. (2004) Assessing data quality for information products: Impact of selection, projection, and cartesian product. *Management Science*, (2004) 50:967-982; doi:10.1287/mnsc.1040.0237. Retrieved from http://mansci.journal.informs.org/content/50/7/967.full.pdf+html

Abstract. The cost associated with making decisions based on poor-quality data is quite high. Consequently, the management of data quality and the quality of associated data management processes has become critical for organizations. An important first step in managing data quality is the ability to measure the quality of information products (derived data) based on the quality of the source data and associated processes used to produce the information outputs. We present a methodology to determine two data quality characteristics—accuracy and completeness—that are of critical importance to decision makers. We examine how the quality metrics of source data affect the quality for information outputs produced using the relational algebra operations selection, projection, and Cartesian product. Our methodology is general, and can be used to determine how quality characteristics associated with diverse data sources affect the quality of the derived data.

Credibility. Amir Parssian is an assistant professor of management information systems at the University of Illinois at Springfield. Amir Parssian holds a PhD degree in management science from the University of Texas at Dallas. Dr. Parssian's research is cited in journals including the Journal *of Organizational Computing and Electronic Commerce, the Journal of Data and Information Quality and Decision Support Systems.* Sumit Sarkar is a professor of information systems at the Naveen Jindal School of Management at the University of Texas at Dallas and holds a PhD degree in computers

and information systems from the University of Rochester. Dr. Sarkar is an associate editor in several professional, peer-review journals including *Management Science, IEEE Transactions on Management Information Systems, Information Technology and Management*, and *Risk and Decision Analysis.* Varghese Jacob is a senior associate dean and professor of management science and information systems at the University of Texas at Dallas and earned a PhD in management information systems from Purdue University. Dr. Jacob's work has been published in the following journals: *Management Science, Information Systems Research, Decision Support Systems, IEEE Transactions on Systems, Man, and Cybernetics, European Journal of Operational Research, Psychometrika, Group Decision and Negotiation and International Journal of Man-Machine Studies.*

Summary. The authors contend that data management is important in ensuring an organization's information is of good quality in order to drive business decisions. The authors also suggest that organizations take a metric-based approach to evaluating information discerned from data sets. The author's selection methodology considers three types of errors commonly encountered in managing data sets concerning data accuracy, data completeness and dataset membership. The authors present their evaluation criteria for determining data quality by analyzing the attributes of two datasets. The author's analysis illustrates how algebra operations, when applied to the datasets, provide outcomes to determine data quality. The authors conclude by providing mathematical proofs of their evaluation criteria and discuss its application within data warehousing systems as a means of determining the effectiveness of the data governance framework.

McGlinchey, J. (2013). Putting the right processes in place to benefit from big data. *Computer Weekly*, 14. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=cph&AN=86374841&site=ehost-live&scope=site

Abstract. The author discusses data and information technology (IT) as of February 26, 2013, with a focus on tools and strategies for big data management. Topics include a study of executives' attitudes about big data in the survey "Big Data Insights and Opportunities," by IT trade group CompTIA, computer security, and data privacy.

Credibility. John McGlinchey is the Vice President at CompTIA for Europe and the Middle East. Mr. McGlinchey is an executive with over 25 years of experience in information management.

Summary. The author cites a study by CompTIA in which a majority of executives reported that they are not realizing the full potential of business intelligence derived from big data. The author presents a strategy for information technology leaders interested in big data system implementations in which the quality of an organization's current data is determined, suggests open source approaches to big data analytics systems and advises polices to manage, store and secure large data sets. The author advises that as a key component of the data governance framework is ensuring that information technology staff interacting with big data sets are properly trained and preferably certified by a respected third party organization. Furthermore, their access should be governed by data handing access controls directly aligned with the data governance framework. After data management policies are presented to the organization, internal data-handlers should formally acknowledge their commitment to compliance.

Ohlhorst, F. (2012). Big data analytics: Turning big data into big money. New York: Wiley. Retrieved from http://onesearch.uoregon.edu/metasearch/record?group=2013-04-04-000383&resultSet=002350&startRecord=1

Abstract. Unique insights to implement big data analytics and reap big returns to your bottom line Focusing on the business and financial value of big data analytics, respected technology journalist Frank J. Ohlhorst shares his insights on the newly emerging field of big data analytics in Big Data Analytics. This breakthrough book demonstrates the importance of analytics, defines the processes, highlights the tangible and intangible values and discusses how you can turn a business liability into actionable material that can be used to redefine markets, improve profits and identify new business opportunities.

Credibility. Frank J. Ohlhorst is an information technology journalist and consultant with over 25 years of professional experience. Mr. Ohlhurst has published in *ComputerWorld, TechTarget, CRN, Network Computing* and *PCWorld*. Mr. Ohlhurst holds a degree in computer science.

Summary. The author describes big data within four separation dimensions:

- **Volume.** This dimension regards the enormous size of big datasets and the degree to which it is amassed.
- **Variety.** Big data extends beyond text and includes various forms including video files, audio files, data regarding how users interact with digital channels and more.
- **Veracity.** The author notes that due to the large volumes of data inherent in big data analytics there is a common risk of poor data quality leading to mistakes in interpreting information.

- **Velocity.** A key value in big data analytics is making decisions as close to real-time as possible. The author advises that data analyzed and utilized as it enters the organization.

The author refers to these aspects of big data throughout the text and provides a roadmap for preparing an organization to implement a big data analytical system. The author notes that data governance is essential to a successful big data program and advises the information technology department institutionalize data governance. The author describes an optimal data governance program as including dedicated resources tasked with ensuring data is secure, normalized and properly analyzed. Within organizations data should be considered an asset and subject to risk management practices to prompt compliance and manage risk. The author notes that in large organizations with mature data governance programs include executive sponsorship and representation on governance councils.

Rosenbush, S., & Totty, M. (2013). How big data is changing the whole equation for business. *Wall Street Journal - Eastern Edition, 261* (57), R1-R2.

Abstract. The article discusses the various ways in which companies are exploiting big data to improve their operations. Corporate human-resources departments are using big data to hire more efficiently and reduce health-care costs. Examples are provided of how product-development at car-manufacturer Ford Motor Co. and Internet-game producer Zynga Inc. was aided by big data.

Credibility. Steven Rosenbush is deputy editor of the CIO Journal at *The Wall Street Journal* and is the author of *Telecommunications Business Opportunities* (Aegis

Publishing, 1998). Michael Totty is news editor at *The Wall Street Journal* San Francisco bureau and writes about technology and business trends.

Summary. The authors note that big data is commonly associated with companies such as Google, Facebook and Amazon due to their status as aggregators of large datasets; however, companies not accustomed to big data sets are engaging in big data analytics industry-wide. The authors discuss how the transition to big data analytics is made possible by new digital channels such as social-media sites and geographical location information derived from mobile devices. The authors also note that decision-making based on business intelligence from big data analytics represents a paradigm shift in most companies since decisions are most commonly based on the executive options. The application of big data analytics are discussed in multiple industries including human resources, product development, operations and marketing. In every real-world case discussed, the results of leveraging big data analytics increased efficiency and drove improvements including reducing the United Parcel Service's fuel consumption by 8.4 million gallons of fuel by correlating geo-positioning system data with fuel-efficiency sensors to determine optimal delivery routes.

Shute, W. (2012, November-December). Information governance takes center stage in 2013: Spotlight shines on ig pros. *Information Management Journal, 46(6),* 22+. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=85687973&site=ehost-live&scope=site

Abstract. The article focuses on the heightened information governance in 2013 that would provide opportunity for information governance professionals in the U.S. It

mentions that 1.8 trillion gigabytes of electronic information were produced only in 2011 which lead to more greater electronic information in the future, according to the study conducted by IDC. It states that information technology (IT) leader should implement comprehensive information governance programs, including inventory, organize and control, to cope with the increasing changes of volumes. It says that cloud-based tools for sharing of files and collaboration are in growing use in 2012.

Credibility. William Shute is the senior vice president of Viewpointe and is responsible for the direction of Viewpointe's information governance solutions. Mr. Shute has experience in leadership roles and information management. The article appears in *Information Management Journal*, a peer-reviewed journal and cites a study conducted by the research firm IDC. IDC is a leader in market intelligence and advisory consulting within information technology and has provided strategic services for over 48 years.

Summary. The author discusses how the function of records management within organizations will change due to big data analytics initiatives. Those responsible for managing organizational information will drive efforts to determine business intelligence from big data sets from both traditional data sources such as corporate databases to social media channels in real-time. The author contends that records managers will implement what he refers to as *defensible disposition* strategies to ensure data is relevant, timely and meaningful to the organization. The author notes that the governance of information will continue to drive efforts to reduce risk and encourage legal compliance to the entire organization. The author discusses a shift to businesses utilizing cloud-based services to more efficiently deal with the demands of big data analytical processing, storage and access. The author cites a study by IDC where banks moved many information

technology functions to cloud-based services where they realized lower costs, efficiencies in information management and greater automation of processes and systems.

Conclusions

This annotated bibliography analyses 30 references including peer-reviewed, scholarly works appearing in journals, reports and books. This presentation of selected references examines big data governance models and their requisite aspects to realize the potential of data as an asset (Khatri & Brown, 2010). The purpose is to identify the most effective aspects of a data governance framework .for consideration by information technology leaders.

Vinayak et al. (2012) note that utilizing data warehousing technology to analyze large data sets for potential gains was previously a practice limited to large organizations; however, the use of big data has expanded into organizations of all sizes. They indicate that the application of big data analytics is made possible for organizations with lower information technology budgets (as compared to large companies) due to advances in database systems and the availability of open source systems (2012). Chen, Chiang and Storey (2012) report that this proliferation of big data technology enables the application of big data analytics in domains largely underrepresented in the past including (a) government organizations and political groups, (b) scientific and technological research, (b) healthcare, and (c) public safety.

The increased utilization of big data analytics has resulted in an expectation among stakeholders that IT departments will expand services to provide big data collection, analysis, retrieval and storage capabilities (Gendron, 2012). Chen, Chiang, and Storey (2012) contend information technology leaders must determine how to manage big data sets in order to fully realize their potential as a form of business intelligence. To guide the enterprise-wide utilization of big data analytics to realize the maximum benefit to the organization, the assumption underlying this study is that information technology leaders should implement a data governance framework.

Khatri and Brown (2010) define data governance as a set of big-picture executive decisions about where the responsibility for day-to-day data management should be located, and what goals should be set in that context. They present a framework for developing a data governance strategy, titled the *Framework for Data Decision Domains*, which emphasizes five main areas of data governance elements. Given the application of big data systems across the multiple domains identified by Chen, Chiang and Storey (2012), there is no one data governance model to satisfy all potential instances. However several other potentially useful data governance frameworks are identified in the selected literature (Ashford, 2013; Breaux & Alspaugh, 2011; Coleman, Hughes & Perry, 2009; Hua, Huang & Yen, 2010; Khatri & Brown, 2010; Madsen, 2012; McGlinchey, 2013; Österle, Otto & Weber, 2009; Parssian, Sarkar & Jacob, 2004; Rifaie, Alhajj & Ridley, 2009; Rouse, 2010; Storey, Dewan & Freimer, 2012). Using the Khatri and Brown framework as a base, each of the five domains is augmented below as a way to describe a larger set of options to consider in a data governance framework for managing big data systems initiatives.

Domain 1: Data Principles

Khatri and Brown (2010) propose that organizations should establish data principles "to which data is an enterprise wide asset, and thus what specific policies, standards and guidelines are appropriate" (p. 150). They advise that success at managing big data sets is contingent on first realizing that data is an important asset to the organization. By qualifying organizational data as an enterprise-wide asset, the data becomes subject to the organization's policies, control standards, and control procedures. This realization that data is asset to be used to inform business intelligence requires many of the same compliance practices applied to other types of assets, such as human resources, physical assets, and financial assets.

In order to manage data as an asset, McGlinchey (2013) recommends organizations create policies that (a) determine what the intended goals of big data analytics are to the organization, (b) mandate that the integrity and use of data is governed throughout the entire organization, and (c) define data management as the responsibility of all stakeholders – both on the business side and the information technology side. They recommend the creation of a *responsibility matrix* for ensuring that data governance policies are regularly updated and enforced.

Data governance principles should clearly describe how decisions regarding data management policies and procedures are determined. According to research conducted by Rifaie, Alhajj and Ridley (2009), these principles prompt the formation of organizational structures to:

- Determine data governance policies and procedures and authorize enforcement
- Remediate risks incurred by the failures of the data governance program
- Communicate data governance decisions and provide transparency in the formation and monitoring of data management policies within the organization
- Authorize the financial spend for big data investments including technology and staffing

Domain 2: Data Quality

Khatri and Brown (2010) propose that organizations should establish standards to ensure data is up-to-date, complete and credible. Parssian, Sarkar and Jacob (2004) state that the consideration of data as an asset indicates a move into what they refer to as an *information economy*. Within this emerging economy, the success of the organization is dependent in large part to the quality of its data. Findings note that data quality issues can contribute to unplanned expenses, employee dissatisfaction and adversely impact growth and that quality standards

reduce financial impacts of poor data quality. These costs are estimated as 8% to 12% of revenue. IT leadership should therefore determine standards concerning data quality.

Österle, Otto, and Weber, (2009) contend that while there is no single solution for all organizations to implement a uniform data quality program, there are several information technology governance frameworks to consider including centralized, decentralized, and a federated model. They describe how data governance aligns with the overall information technology governance model of the organization and discuss the benefits of data governance in aligning information technology goals to the goals of the business as a principle aspect of the governance framework.

Data quality strategy. Alhajj, Ridley, and Rifaie (2009) advise that the data governance program should provide a clear understanding of how the effective management of data maps to strategic drivers within the organization and should communicate data quality goals. The strategy component of data quality management should provide the following:

- The objectives for data quality within the organization
- A repository of all data quality policies and procedures
- A business case indicating how data quality reduces financial risk
- The process for assessing data quality operations and reviewing processes.

Data quality governance. Alhajj, Ridley, and Rifaie (2009) also note that decision-making authority for data quality and the overall data governance program resides in a role-based, hierarchical model. While multiple models are presented in the research, a common theme is a structure that includes executive sponsorship and representation from throughout the organization. IT leadership should work with the organization to gain sponsorship and formulate a Data Governance Council comprised of stakeholders from information technology and the

business units. This governing body is primarily responsible for determining data governance policies and acting upon the data quality strategy.

Storey, Dewan and Freimer (2012) suggest that it is essential to communicate concise data quality policies throughout the organization and model what data quality aspects should be readily apparent in data sets. Additionally, they recommend that individuals should be properly incentivized to practice sound data quality procedures and rewarded for their performance.

Domain 3: Metadata Management

Khatri and Brown (2010) propose that metadata "provides a mechanism for a concise and consistent description of the representation of data, thereby helping interpret the meaning" (p. 150). In other words, metadata provides a consistent means for organizations to realize efficient searching and analysis of their data. Coleman, Hughes, and Perry (2009) indicate that metadata is a key aspect in effective data governance and suggest organizations leverage shared coding schemas for capturing metadata and enabling reporting. Data governance councils should mandate a consistent approach to tagging information assets to realize the following enable the following:

- Increased efficiencies in determining the origin of data assets
- Transparency in where the data resides in its lifecycle
- Immediate reporting of what data assets have been distributed outside of the organization and within the enterprise
- Improved search capabilities and cataloguing of data assets
- Automated dissemination of information based on user behavior or context.

IT leaders should also assume that the approach to managing metadata will change in the long term. Khatri and Brown (2010) note that governance framework is also intended to take a proactive approach in navigating the inevitable changes to how an organization manages its metadata due to shifting business operations and objectives. Regardless of the particular approach used to capture metadata, research by Hua, Huang and Yen (2010) advise that data warehouse systems should be internet-accessible and coding schemas should utilize extensible mark-up language (XML).

Domain 4: Data Access

Khatri and Brown (2010) advise that organizations should effectively control access to data assets to manage risk and compliance. They note that access standards establish rules for organizations to control what data shall be accessed and by whom. Accordingly, data governance frameworks should quantify the value of data for business intelligence and mandate security controls for big data sets. While big data enables the ability of organization to realize insights regarding their processes and client behaviors, the inherent value of these datasets could be misused for nefarious means without applicable, transparent access control standards and procedures.

Ashford (2013) notes that information technology leaders and data governance councils should amend security programs to compensate for the large volumes of data incoming to organizations. The goal is to ensure procedures for collecting, analyzing and reporting meaning from data are tightly coupled with enterprise governance risk and compliance programs. Ashford also notes that one industry leader in big data storage, security and analytics advised that information technology leaders should transform their approach to security to meet the real-time requirements of big data systems. Data governance frameworks should ensure the various digital

channels from which data is collected are secure and highly reliable. Madsen (2012) notes that big data governance also requires clearly defined roles and a responsibility matrix to both authorize and inform data stewards on the proper data handling procedures from the point of collection to distribution inside and outside of the organization.

Domain 5: Data Lifecycle Management

The final domain identified by Khatri and Brown (2010) concerns the flow of data throughout the organization. The data lifecycle provides organizations with the means to understand how data is sourced from various contributors – both inside and outside of the organization – and its progression from active use to its eventual archival is a critical piece of an effective data governance program.

Rouse (2010) advises that all data residing in big datasets should be immediately accessible based on its age. Rouse notes that since data newer to the organization is commonly accessed in greater frequency, it should reside on storage devices optimized for improved search capabilities. Governance programs should therefore authorize information technology departments to modify their storage and retrieval architecture to facilitate these search requirements.

Breaux and Alspaugh (2011) describe an optimal data lifecycle that controls the flow of data by implementing *rule sets*. These rule sets may include mappings to authoritative sources mandating data handling standards based on the context of the data. Leveraging the approach suggested by Breaux and Alspaugh (2011), digital licenses containing rule sets are imbedded within data sets to provide controls against noncompliant usage from the moment the data is categorized to its eventual archival.

References

Aiken, P., Gillenson, M., Rafner, D., & Zhang, X. (2011). Data management and data administration: Assessing 25 years of practice. *Journal of Database Management. 22, 3* (July 2011), 24-45. doi:10.4018/jdm.2011070102. Retrieved from http://go.galegroup.com/ps/retrieve.do?sgHitCountType=None&sort=DA-SORT&inPS=true&prodId=AONE&userGroupName=s8492775&tabID=T002&searchId=R1&resultListType=RESULT_LIST&contentSegment=&searchType=AdvancedSearchForm&currentPosition=2&contentSet=GALE%7CA267609846&&docId=GALE|A267609846&docType=GALE&role=

Alhajj, R., Ridley, M., & Rifaie, M. (2009). Data governance strategy: A key issue in building enterprise data warehouse. *Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services* (iiWAS '09). ACM, New York, NY, USA, 587-591. doi:10.1145/1806338.1806449. Retrieved from http://doi.acm.org/10.1145/1806338.1806449

Allen M., & Cervo, D. (2011). Master data management in practice: Achieving true customer MDM. Retrieved from http://orbis.eblib.com.libproxy.uoregon.edu/patron/FullRecord.aspx?p=693520

Alter, A. (2006). Information governance's big payback. *CIO Insight,* (68), 69-76. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=cph&AN=21057431&site=ehost-live&scope=site

Ashford, W. (2013). Embracing big data can lead to greater security. *Computer Weekly, 6-7.* Retrieved from

http://web.ebscohost.com.libproxy.uoregon.edu/ehost/pdfviewer/pdfviewer?sid=7a48c664-6d0b-4153-a2d0-3b6c44bf0425%40sessionmgr113&vid=2&hid=125

Bell, C., & Frantz P. (2012). Critical evaluation of information sources. Retrieved May 19, 2013 from http://library.uoregon.edu/guides/findarticles/credibility.html.

Borkar, V., Carey, M., & Li, C. (2012a). Big data platforms: What's next?. *XRDS* 19, 1 (September 2012), 44-49. doi:10.1145/2331042.2331057. Retrieved from http://doi.acm.org/10.1145/2331042.2331057

Borkar, V., Carey, M., & Li, C. (2012b). Inside "big data management": Ogres, onions or parfaits?. EDBT/ICDT 2012 Joint Conference Berlin, Germany. Retrieved from http://www.edbt.org/Proceedings/2012-Berlin/papers/keynotes/a3-carey.pdf

Breaux, T.D., & Alspaugh, T.A. (2011). Governance and accountability in the new data ecology. *2011 Fourth International Workshop of Requirements Engineering and Law (RELAW)*, 2011. 14, 30-30 Aug. 2011. doi: 10.1109/RELAW.2011.6050267. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6050267&isnumber=6050260

Busch, C., De Maret, P., Flynn, T., Kellum, R., Le, S., Meyers, B., Saunders, M., White, R., & Palmquist, M. (2013). Content analysis. Writing@CSU. Colorado State University Department of English. Retrieved from http://writing.colostate.edu/guides/research/content/.

Chan, S.; Rhodes, W.; Atencio, C.; Kuo, C.; Ranalli, B.; Miao, A.; Sala, S.; Serene, S.; Helbling, R.; Rumbley, S.; Clement, M.; Sokol, L.; Gary, L. (2012). Robust Decision Engineering: Collaborative big data and its application to international development/aid, CollaborativeComputing: Networking, Applications and Worksharing (CollaborateCom),

2012. 8th International Conference, pp.597,604, 14-17 Oct. 2012. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6450957&isnumber=6450881

Chen, H., Chiang, R. H. L., & Storey, V. C. (January 01, 2012). Business intelligence and analytics: From big data to big impact. *Management Information Systems Quarterly, 36*, 4, 1165-1188. Retrieved from http://uolibraries.worldcat.org/oclc/820907587

CNSS. (2010). National information assurance (IA) glossary. Committee on National Security Systems. CNSS Instruction No. 4009. Retrieved from http://www.cnss.gov/Assets/pdf/cnssi_4009.pdf

Coleman, D., Hughes, A., & Perry, W. (2009). The role of data governance to relieve information sharing impairments in the federal government, *Computer Science and Information Engineering, 2009 WRI World Congress on Computer Science and Information Engineering*, vol.6. pp.267 - 271, March 31 2009-April 2 2009. doi: 10.1109/CSIE.2009.630. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5171000&isnumber=5170943

Cukier, K. , & Mayer-Schoenberger, V. (2013). The rise of big data. *Foreign Affairs*, *92*(3), 27-40. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=87000329&site=ehost-live&scope=site

Data Governance Institute. (2012). Definitions of data governance. Retrieved from http://www.datagovernance.com/adg_data_governance_definition.html

Dewan, R., Freimer, M., & Story, V. (2012). Data quality: Setting organizational policies. *Decision Support Systems. 54, 1* (December 2012), 434-442.

doi:10.1016/j.dss.2012.06.004. Retrieved from http://dx.doi.org/10.1016/j.dss.2012.06.004

Du Mars. R. (2012). Mission impossible? Data governance process takes on 'big data'. SearchDataManagement. Retrieved from http://searchdatamanagement.techtarget.com/feature/Mission-impossible-Data-governance-process-takes-on-big-data

Dysart, J. (2013). The dawn of big data. *ABA Journal, 99*(5), 42-47. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=87445429&site=ehost-live&scope=site

English, L. (2010). To a high iq! Defining information quality: More than meets the eye. *IDQ Newsletter,* 1(3). Retrieved from http://iaidq.org/publications/doc2/english-2005-04.shtml#english2

Gartner. (2012). Big data. IT Glossary. Retrieved from http://www.gartner.com/it-glossary/big-data/

Gendron, M. (2012). Business intelligence applied: Implementing an effective information and communications technology infrastructure. New York: Wiley. Retrieved from http://onesearch.uoregon.edu/metasearch/record?group=2013-04-04-000383&resultSet=002350&startRecord=12

Goldston, D. (January 01, 2008). Big data: Data wrangling. *Nature, 455*, 7209. doi: 10.1038/455015a. Retrieved from http://www.nature.com/news/2008/080903/full/455015a.html

Gopalkrishnan, V., Guszcza, J., Lewis, H., & Steier, D. (2012). Big data, big business: Bridging the gap. In *Proceedings of the 1st International Workshop on Big Data, Streams and*

Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications (BigMine '12). ACM, New York, NY, USA, 7-11. doi:10.1145/2351316.2351318. Retrieved from http://doi.acm.org/10.1145/2351316.2351318

Gorton, I. & Gracio, D. (2012). *Data-Intensive Computing : Architectures, Algorithms, and Applications*. Cambridge University Press. Retrieved from http://orbis.eblib.com.libproxy.uoregon.edu/patron/FullRecord.aspx?p=1042395&userid=8r16fsFJMWM%3d&tstamp=1369014206&id=78740BE31F39B1357B660549FB3F1EC170A5BCB2&conl=uoregon

Hua, J. , Huang, S. , Yen, D. (2012). Architectural support for business intelligence: A push-pull mechanism. *Online Information Review*, *36*(1), 52-71. Retrieved from http://www.emeraldinsight.com/journals.htm?issn=1468-4527&volume=36&issue=1&articleid=17019377&show=html

Hsu, V., & O'Neal, K. (2009). How technology enables data governance. Oracle corporation. Retrieved from http://www.oracle.com/us/products/applications/master-data-management/042941.pdf

Jacob, V., Parssian, A., & Sarkar, S. (2004). Assessing data quality for information products: Impact of selection, projection, and cartesian product. *Management Science*, (2004) 50:967-982; doi:10.1287/mnsc.1040.0237. Retrieved from http://mansci.journal.informs.org/content/50/7/967.full.pdf+html

Janssen, C. (2013). Data analytics. Definition – what does data analytics mean? Retrieved from http://www.techopedia.com/definition/26418/data-analytics

Khatri, V., & Brown, C. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148-152.Retrieved from http://cacm.acm.org/magazines/2010/1/55771-designing-data-governance/fulltext

Kooper, M., Maes, R., & Lindgreen, E. (2011). On the governance of information: Introducing a new concept of governance to support the management of information. *International Journal of Information Management 31*, 3 (June 2011), 195-200. doi:10.1016/j.ijinfomgt.2010.05.009 http://dx.doi.org/10.1016/j.ijinfomgt.2010.05.009

Literature Reviews. (n.d.) The University of North Carolina. Retrieved from https://writingcenter.unc.edu/handouts/literature-reviews/

Madsen, L. (2012). *Healthcare business intelligence: A guide to empowering successful data reporting and analytics + website*. New York: John Wiley & Sons.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. Retrieved from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

McEvoy, B., & Machi, L. (2009). The literature review: Six steps to success [Kindle version]. Thousand Oaks, CA: Corwin Press.

McGlinchey, J. (2013). Putting the right processes in place to benefit from big data. *Computer Weekly*, 14. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=cph&AN=86374841&site=ehost-live&scope=site

Mone, G. (2013). Beyond hadoop. *Communications of the ACM, 56*(1), 22-24. doi:10.1145/2398356.2398364. Retrieved from http://doi.acm.org/10.1145/2398356.2398364

Myers, P. (2012). Information governance-control content chaos. *KM World, 21* (8), 9. Retrieved from http://web.ebscohost.com.libproxy.uoregon.edu/ehost/pdfviewer/pdfviewer?sid=3c767cef-7502-4138-a91c-f2b0db52c85b%40sessionmgr115&vid=2&hid=125

NASCIO. (2009). Data governance part ii: Maturity models – a path to progress. Retrieved from http://www.nascio.org/publications/documents/NASCIO-DataGovernancePTII.pdf

NISO. (2004). Understanding Metadata. NISO Press. Retrieved from http://www.niso.org/publications/press/UnderstandingMetadata.pdf

Ohlhorst, F. (2012). Big data analytics: Turning big data into big money. New York: Wiley. Retrieved from http://onesearch.uoregon.edu/metasearch/record?group=2013-04-04-000383&resultSet=002350&startRecord=1

Olavsrud. T. (2013). 4 barriers stand between you and big data insight. Retrieved from http://www.cio.com/article/731503/4_Barriers_Stand_Between_You_and_Big_Data_Insight?page=3&taxonomyId=600010

O'Reilly (2012). *Big data now: 2012 edition*. O'Reilly Media. Retrieved from http://oreilly.com/data/radarreports/big-data-now-2012.csp?intcmp=il-strata-ebooks-big-data-now-2012-strata-right-rail

Österle, H., Otto, B., & Weber, K. 2009. One size does not fit all---a contingency approach to data governance. *Journal of Data and Information Quality* 1, 1, Article 4 (June 2009), 27

pages. doi:10.1145/1515693.1515696. Retrieved from http://doi.acm.org/10.1145/1515693.1515696

Parssian, A., Sarkar, S., & Jacob, V. (2004). Assessing data quality for information products: Impact of selection, projection, and cartesian product. *Management Science*. 2004 50:967-982; doi:10.1287/mnsc.1040.0237. Retrieved from http://mansci.journal.informs.org.libproxy.uoregon.edu/content/50/7/967

Peyret, H. (2013). Why maturity models for data governance are irrelevant in the data economy? Retrieved from http://blogs.forrester.com/henry_peyret/13-05-14-why_maturity_models_for_data_governance_are_irrelevant_in_the_data_economy

Rajan, S., McMillan, L., & O'Connor, G. (2011). IBM transforms data at work, accelerates big data analytics. Retrieved from http://dotnet.sys-con.com/node/2033349

Rosenbush, S., & Totty, M. (2013). How big data is changing the whole equation for business. *Wall Street Journal - Eastern Edition, 261* (57), R1-R2. Retrieved from http://online.wsj.com/article/SB10001424127887324178904578340071261396666.html

Rouse, M. (2010). Data lifecycle management (DLM). Retrieved from http://searchstorage.techtarget.com/definition/data-life-cycle-management

Sarsfield, S. (2009). *The data governance imperative*. IT Governance Publishing.

Shute, W. (2012). Information governance takes center stage in 2013: Spotlight shines on ig pros. *Information Management Journal, 46(6),* 22+. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=85687973&site=ehost-live&scope=site

Soares, S. (2012). Big data: A boon for governance professionals. Retrieved from http://ibmdatamag.com/2012/04/big-data-a-boon-for-governance-professionals/

Soares, S. (2013). *Big data governance: An emerging imperative*. Mc Press.

Soares, S., Deutsh, T., Hanna, S., & Malik, P. (2012). Big data governance: A framework to access maturity. Retrieved from http://ibmdatamag.com/2012/04/big-data-governance-a-framework-to-assess-maturity/

Solis. B. (2012). The human algorithm: Redefining the value of data. Retrieved from http://www.briansolis.com/2012/12/the-human-algorithm-redefining-the-value-of-data/

Stanford University. (2013). What is data governance? DG @ Stanford. Retrieved from http://www.stanford.edu/dept/pres-provost/cgi-bin/dg/wordpress/

Tamm-Daniels, R. (2013). The key to smart big data: Know thy technology. *Information Today*, *30*(4), 19. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=86640870&site=ehost-live&scope=site

Thomas, G. (2004). Data Governance Definitions. Data Governance Institute. Retrieved from http://www.datagovernance.com/adg_data_governance_definition.html.

Thomas, G. (2012). The DGI data governance framework. The Data Governance Institute. Retrieved from http://datagovernance.com/dgi_framework.pdf

Violino, B. (2013). 5 strategic tips for avoiding a big data bust. *Infoworld*. http://www.infoworld.com/d/business-intelligence/5-strategic-tips-avoiding-big-data-bust-215296

Wisker, G. (2005). The postgraduate handbook. Palgrave Macmillan.

Yap, J. (2011). APAC enterprises unprepared for big data. Zdnet. Retrieved from http://www.zdnet.com/apac-enterprises-unprepared-for-big-data-2062302632/