Big Data information governance by accountants

Emily M. Coyne

School of Information, San Jose State University, Jose State, California, USA, and

Joshua G. Coyne and Kenton B. Walker

School of Accountancy, University of Memphis, Memphis, Tennessee, USA

Abstract

Purpose – Big Data has become increasingly important to multiple facets of the accounting profession, but accountants have little understanding of the steps necessary to convert Big Data into useful information. This limited understanding creates a gap between what accountants can do and what accountants should do to assist in Big Data information governance. The study aims to bridge this gap in two ways.

Design/methodology/approach – First, the study introduces a model of the Big Data life cycle to explain the process of converting Big Data into information. Knowledge of this life cycle is a first step toward enabling accountants to engage in Big Data information governance. Second, it highlights informational and control risks inherent to this life cycle, and identifies information governance activities and agents that can minimize these risks.

Findings – Because accountants have a strong ability to identify the informational and control needs of internal and external decision-makers, they should play a significant role in Big Data information governance.

Originality/value – This model of the Big Data life cycle and information governance provides a first attempt to formalize knowledge that accountants need in a new field of the accounting profession.

Keywords Big data, Information governance, Information life cycle

Paper type Technical paper

1. Introduction

We present the life cycle by which Big Data becomes useful information and to explain how accountants can and should become involved in the governance of this process. We follow three steps in the pursuit of this goal. First, we review the traditional information life cycle as a foundation for the process by which data becomes information. Second, we identify modifications to the information life cycle to address the idiosyncrasies of Big Data. Third, we introduce the concept of information governance and explain the role of accountants in the information governance process. A proper understanding of the Big Data life cycle and information governance will allow accountants to continue in and expand their role as information custodians in the current Information Era.

Previous studies show that information governance is important to the accounting profession. Poor data governance standards impact accounting information quality (Song, 2016) and simply adopting previously successful models does not guarantee a positive governance outcome (Ji *et al.*, 2015). Also, good internal governance reinforces external governance mechanisms to create firm value (Huang and Boateng, 2016). Furthermore, corporate governance research from Taiwan suggests that independence of boards of

directors increases market-to-book ratios (Lin and Liu, 2015). Taken together, these studies bolster the view that accountants have a vested interest in information governance structures and policies.

The information life cycle has governed the creation, use and maintenance of data since before the proliferation of digital information systems at a time when business data were more uniform and shared along predetermined channels (Leahy, 1949). However, Big Data is different. Big Data's primary differences were characterized by the 3-D model proposed by Doug Laney (2001), sometimes referred to as the three Vs of Big Data[1]. *Volume* is the reason that Big Data is referred to as "Big" as businesses receive large amounts of data from machines, transactions and social media interactions. *Velocity* references the high rate at which large amounts of data are both created and become obsolete; real-time transaction processing in e-commerce has been a big driver of the competitive acceleration in the velocity of data processing. *Variety* describes the lack of uniformity in data sources containing text, audio, video, image and other data types. We propose a life cycle model that addresses the differences between traditional business data and Big Data. To our knowledge, this is the first attempt to develop an information life cycle that is specific to Big Data.

Our primary motivation for introducing a revised life cycle is the current interest in Big Data analytics. However, more than a life cycle is necessary to convert Big Data into useful information. Information governance assigns business, legal and IT specialists the task of developing and maintaining information systems that meet consumers' informational and control needs (Smallwood, 2016). Collaboration among these parties stands juxtaposed to the traditional model of siloed IT departments, which model has come under increased criticism in recent years (Zetlin, 2014). Accountants, in their capacity as business specialists, have unique expertise in business intelligence, regulatory compliance and internal controls, which make them valuable collaborators in information governance. As a result, in addition to introducing a Big Data life cycle, we describe the activities that accountants can engage in to increase their participation in information governance. Despite this appeal to accountants, we also recognize that information governance is a shared responsibility and that IT specialists must also increase their willingness and ability to work with accountants on information system solutions that will result in improved collection and analysis of Big Data.

This paper has implications for both business and academia. First, business leaders have used popular IT and business media outlets in their appeal for added collaboration between business and IT specialists. We support this effort by encouraging accountants to become involved in information governance. Second, by proposing a new life cycle model to accommodate the idiosyncrasies of Big Data, we push the frontiers of both academic and business knowledge regarding sound treatment of Big Data in the current Information Era.

This paper proceeds as follows. In Section 2, we review the steps in the traditional information life cycle. In Section 3, we identify necessary modifications to convert the information life cycle into the Big Data life cycle. In Section 4, we discuss the role of accountants in the governance of the Big Data life cycle. In Section 5, we conclude.

2. A review of the traditional information life cycle
The information life cycle is the process by which data become usable information; it covers all aspects of the data's use from their creation to their deletion. This is the core of the information science discipline, and its principles govern the design of information systems (Hoke, 2011). The steps in the information life cycle are needs assessment, data acquisition,

classification, conversion, ingestion, storage, indexing, refreshing, interpretation, searching, analytics, reporting and disposition (Coyne *et al.*, 2016; Dederer and Dmytrenko, 2015)[2].

Although a life cycle connotes linearity, the information life cycle is first a blueprint for information systems and only then a sequence of steps. Before the execution of any step in the life cycle, the information system should be able to perform all steps in the life cycle. Furthermore, as the information system repeatedly operationalizes the information life cycle, the system will not always execute the steps in the same order (Upward, 1997). Consistent with this notion, we portray the information life cycle as containing three phases: creation, maintenance and use. Within each phase, the individual steps occur sequentially, but the life cycle can transition back and forth between phases in a nonlinear fashion[3]. These phases group similar steps together, and the three labels provide a succinct overview of the primary roles of the information life cycle. Figure 1 displays this representation of the information life cycle.

Creation is the first phase, and the activities in the creation phase are the first to occur in the information life cycle. These activities are needs assessment, acquisition, classification, conversion and ingestion. Needs assessment is the first and arguably most important step. It is the systematic analysis of an organization's information, reporting and data needs. This analysis can then be expanded to describe not only the information required, but also the data used to create that information and the forms it can and should take. Acquisition of data from internal departments or external counterparties follows. Classification involves the application of proper metadata to each data item to give the data clear meaning and context. Conversion is modification of a data item's format to become consistent with the current or expected needs of the information system and to fit into the existing storage infrastructure. The final step is the ingestion of the data into the organization's information system (i.e. manual or automated data entry). These steps, taken together, prescribe the appropriate ways to ensure that the data entering into the system is clean and usable[4].

The second phase is maintenance. Maintenance activities are vital to ensure long-term access and to discard information that no longer holds value. These activities are storage, indexing, refreshing, interpretation and disposition. Storage involves the selection of appropriate formats, media and locations for data. There are a variety of storage options, and the choice should take into account the type of data being stored and its life expectancy. Indexing maps data objects to their location in a data set to expedite searching. Refreshing is the updating of data stores to ensure data and metadata currency and to prevent

Notes: This figure portrays the steps in the traditional information life cycle in three phases: Creation, Maintenance, and Use. These phases are identified by primary chevrons in each row. Each phase includes multiple, sequential steps. The secondary chevrons in each row contain the labels for each step

Figure 1.
Information life cycle

degradation. Interpretation is the selection and preservation of hardware and software necessary to convert data sets into human- and computer-readable formats. Finally, disposition refers to any activity that marks the final resting place of information, such as returning data or intellectual property to a party from which it is licensed, archiving data into a long-term storage facility or destroying it.

Use is the third and final phase, and searching, analytics and reporting make up the activities in this phase. The use phase highlights the goals of the creation and maintenance phases and the fundamental purpose of an information system: to access and analyze information to assist decision-makers. The steps in this phase are not all preceded by the activities of the maintenance phase (e.g. storage precedes use, but disposition follows use), and in this respect, the life cycle is not linear (Upward, 1996). Searching is a simple but vital step that is supported by the activities in the other phases of the information life cycle. As an example, finding information quickly is only possible if the information has been properly classified and stored. Analytics, a prominent focus of both the academic and business world, involves a large number of activities that are used to derive useful information from stores of data. Reporting is the step in the information life cycle that is most familiar to accountants. It is the communication of information to internal and external decision-makers.

Combined, these steps serve to convert data into useful information for decision-makers. This life cycle has accurately depicted this process since before the introduction of digital information systems, and it continues to have relevance for many forms of data (Hoke, 2011).

3. Big Data life cycle

Organizational needs for managing Big Data have proven sufficiently different from the needs for managing traditional data to require a modified life cycle, the Big Data life cycle, to effectively use Big Data to create value (Dumbill, 2012). Big Data itself is fundamentally different from traditional data in a number of ways. A closer examination of these characteristics, the three Vs of Big Data, highlights the need for a distinct Big Data life cycle (Laney, 2001).

The first "V" is volume. Big Data first and foremost is big and has the potential to grow even bigger (Wall, 2014). It is tempting for some organizations to attempt to address the issue of this ever-increasing volume of data by choosing to store everything as digital storage costs have fallen significantly in the twenty-first century:

> Contrary to what many believe about storage costs getting cheaper, the exact opposite is true for organizations today, (which rely on online access to enterprise storage), due to rapidly increasing volumes (Smallwood, 2016).

The vastness of the data available to organizations, as well as the size of the data that is constantly created within their organizations, can easily overwhelm the information system in place.

The second "V" is velocity. On average, Big Data is created more quickly than traditional data, but it also varies more than traditional data in its rate of creation (SAS, 2016)[5]. For example, trending on social media drives "peaks" in information search and acquisition volume and "valleys" that result as interest declines. The increased rate has demonstrated the need for *increased* processing, and the variation has generated a need for *scalable* processing, or processing that can readily increase or decrease to accommodate demand while minimizing cost and drain on resources. Additionally, regardless of the current rate of Big Data creation, all Big Data requires immediate analysis to avoid obsolescence (Laney, 2001).

The third "V" is variety. Big Data is notorious for non-uniformity. It can arrive in any digital format, from text-based log files to audio or video, and many organizations deal with several different formats simultaneously (Arthur, 2013). However, variety extends beyond data type to include variation in cleanliness and structure (e.g. duplicate data or inconsistent metadata classification; Dumbill, 2012). In-place information systems, such as traditional relational databases, may struggle to integrate Big Data with existing business data owing to this lack of data uniformity (Coyne *et al.*, 2016).

Volume, velocity and variety limit the applicability of the information life cycle to Big Data. The inability to store all data necessitates evaluation of data prior to ingestion; the rate of data generation and obsolescence requires additional processing, as well as revisions to analysis and reporting practices; and the inconsistency in data formats demands data stores with increased flexibility (Dumbill, 2012). Addressing the issues caused by Big Data traits is the purpose of the revised Big Data life cycle. These revisions include the following:

- five steps dropped from the traditional information life cycle: acquisition, classification, conversion, indexing and searching;
- five steps added: collection, sifting, synchronization, preprocessing and monitoring; and
- three steps modified: needs assessment, storage and analytics.

We highlight these revisions as we explain the complete Big Data life cycle. Figure 2 provides a graphical representation of this life cycle.

The creation phase of the Big Data life cycle, such as the information life cycle, contains five steps, but the tasks in the Big Data creation phase differ significantly from those in the traditional information life cycle. The five steps for Big Data are needs assessment, collection, sifting, ingestion and synchronization[6]. Needs assessment is the foundation of both life cycles, and is arguably the most important step in both of them (Hoke, 2011; Praminick, 2013). A needs assessment is the process in which managers evaluate the existing information within the organization and compare it with their informational needs. The assessment identifies data sources, information technologies and employee training necessary to satisfy internal and external reporting requirements. The needs assessment

Notes: This figure portrays the steps in the Big Data life cycle in three phases: creation, maintenance and use. These phases are identified by primary chevrons in each row. Each phase includes multiple, sequential steps. The secondary chevrons in each row contain the labels for each step. The bolded labels are unique to the Big Data life cycle and not present in the traditional information life cycle. The asterisks identify steps in the Big Data life cycle with identical *labels* in the information life cycle but distinct *execution*

Figure 2.
Big Data life cycle

aids in the design of an information system, which is necessary before data collection, especially for Big Data. A system that is not equipped to handle Big Data will not provide business value. "Gather business requirements before gathering data" is first in the list of Big Data best practices (Praminick, 2013). A manufacturer would not acquire every item offered by a supplier in the hopes that some will prove useful. Similarly, data is an asset only if they satisfy specific business goals.

The next step in the Big Data life cycle is collection. In the information life cycle, an organization ideally acquires only data that are relevant and useful within the context of the information system's goals. This is not always possible with Big Data. The problematic volume and velocity of Big Data require a broader approach, because organizations cannot always know in advance which Big Data will prove useful. Instead of using informational needs to identify specific data for acquisition, organizations use informational needs to identify potential data sources. They then evaluate the various pools of data from each source and collect all of the data from the pools that are most likely to provide useful information. This information will then undergo sifting, cleaning and analysis to create said useful information (Hoch, 2014)[7].

The decision to collect larger amounts of data leads to the following two steps in the creation phase of the Big Data life cycle to separate useful business data from irrelevant data (Press, 2016). The first step following collection is sifting. Sifting is essentially an evaluative process. The broader perspective required by Big Data means that organizations will likely acquire more data than will be useful and whose business value may be minimal:

> We can amass all the data in the world, but if it doesn't help to save a life, allocate resources better, fund the organization, or avoid a crisis, what good is it? (Davenport, 2014).

Sifting ensures that of the data collected, an organization keeps only that which will prove useful within the context of organizational goals, and disposes the rest. This step prevents the retention of too much data, but it can also exclude data that are incomplete or full of errors, potentially minimizing a future need to clean the data, which occurs during the Maintenance phase (Horodyski, 2014). Only sifted data should enter the information system through the next step: ingestion, which is identical to the step of the same name in the information life cycle.

Synchronization is the final step in the Creation phase of the Big Data life cycle. This step takes collected, sifted and ingested Big Data, and pairs it with existing business data:

> To unleash the value of big data, it needs to be associated with enterprise application data. Enterprises should establish new capabilities and leverage prior investments in infrastructure, platform, business intelligence, and data warehouses, rather than throwing them away. Investing in integration capabilities can enable knowledge workers to correlate different types and sources of data, to make associations, and to make meaningful discoveries (Praminick, 2013).

Synchronization serves as a bridge, linking what an organization already knows to what they would like to know, creating situational insights that rely on data integration (Grimes, 2014).

The steps in the maintenance phase of the Big Data life cycle are storage, preprocessing, refreshing, interpretation and disposition. While the name and concept behind the storage step are comparable to the same step in the information life cycle, the reality has proven different from the traditional practice of storing all business data in relational databases (Dumbill, 2012).

Because of the variety of Big Data, non-relational (i.e. NoSQL) databases of various types (e.g. key-value stores, column stores, document stores, graph stores) have emerged that not only accommodate data variety, but that also promote synchronization by storing Big Data

and related traditional business data together (Patel, 2016; Zhang *et al.*, 2015). One example of data synchronization using a NoSQL solution would be to integrate online customer data, such as clicks and page views tied to geographic location, with foot traffic and sales data in physical stores. Although no individual NoSQL offering may currently address Big Data variety fully, future developments may shortly lead to a single platform being able to provide multiple processing functionalities (Patel, 2016). One of the valuable benefits of NoSQL stores are their superior scalability (Henschen, 2013). Because of the volume and velocity of Big Data, as well as the variability of velocity, data stores must grow and, in some cases be reallocated to accommodate higher, more volatile levels of data flow [Amazon Web Services (AWS), 2016][8].

The next maintenance step is preprocessing. Data cleaning and manipulation, if necessary, occur during this step. Not all data are subject to preprocessing if it is clean and ready for analysis, and adherence to the steps in the creation phase, especially sifting, will minimize preprocessing time. Preprocessing can include normalization, removing duplicates, checking for integrity violations and filtering, sorting, and grouping appropriate data (Bansal and Kagemann, 2015). Preprocessing also involves pre-calculation and reduction, in the case of Online Analytical Processing (OLAP) or MapReduce, which are useful for deep analysis of data sets that are not subject to frequent change, such as examining long-term business trends (Patel, 2016). However, with an increase in processing capabilities, the rate of Big Data obsolescence has discouraged the use of analytical tools that require pre-calculation in favor of real-time analytics (Satell, 2015). It is likely that pre-calculation will, in the near future, cease to play a role in preprocessing as modeling and analytics become increasingly *ad hoc*.

Refreshing and interpretation are similar to their corresponding steps in the information life cycle. For the former, the goal is again to update data stores with current data for analysis; for the latter, the goal is to acquire and maintain the hardware and software necessary to render compiled data human- and computer-readable. However, these steps also differ slightly from their information life cycle counterparts. First, a primary goal of refreshing and interpretation in the information life cycle, as well as the maintenance phase in general, is long-term preservation, whereas Big Data reaches the end of its useful life quickly. Second, many sources of Big Data, especially social media, require continuous refreshing because of the rate of data generation (Nemschoff, 2013).

The last step in the maintenance phase is also the last step in the life of Big Data, and that is disposition. Disposition has the same motivation in the Big Data life cycle as in the traditional information life cycle: to free storage, satisfy privacy and confidentiality requirements and reduce exposure to potential data breaches (Smallwood, 2016). However, one difference is that the vastness of Big Data, as well as its velocity, necessitates a high rate of deletion relative to traditional data that may be subject to archiving (Smallwood, 2016).

The last phase of the Big Data life cycle is use. Just as with any other data, the ultimate goal of collecting, preparing and analyzing Big Data is to discover insights that will enhance business value. In the information life cycle, searching is the first part of use. Although real-time analytics is on the rise, non-automated searching of Big Data remains rare and automated monitoring of trends and outliers is more common (Reeves, 2011). Recent studies in accounting have begun to discuss the implications of automated monitoring of client data in continuous auditing (Zhang *et al.*, 2015). One reason for monitoring, instead of searching, is that unlike traditional business and accounting data, managers will frequently not know what story Big Data tells or even what to look for (Bayer and Taillard, 2014). By using automated monitoring, employees with programming, data manipulation and analysis skills can find the message hidden within Big Data and respond accordingly.

Perhaps the most widely publicized difference between the information life cycle and the Big Data life cycle is the execution of the analytics step. Analytics is the use of software and programming tools to derive valuable decision-making information from data, and since the rise of Big Data, data analytics has become a hot topic (Menon, 2014). The reason for the current level of corporate investment in data analytics is the same as the reason for including monitoring as a step in the use phase: managers believe that Big Data has valuable information, and they will pay for the skill to find the message in the data. The form of the data, as well as the answers sought, influences the choice of analytical tools and techniques, but one part of Big Data analytics that is constant is modeling:

> In a world that's flooded with data, it becomes harder to use the data; there's too much of it to make sense of unless you come to the data with an insight or hypothesis to test (Bayer and Taillard, 2014).

Iteratively modeling and mining using methods already familiar to archivalists promotes the discovery of valuable information for decision-making. Conceptually, modeling is about developing hypotheses; mining involves identifying and extracting data used to test hypotheses.

The final step, reporting, is comparable to the reporting step in the information life cycle in that it involves using the fruits of data analytics to convey information to decision-makers. "Organizations succeed with analytics only when good data and insightful models are put to regular and productive use by business people in their decisions and their work" (Morison, 2014). One notable trend is that Big Data reporting has begun to rely heavily on visualizations. These modern visual displays of information go beyond traditional charts and graphs, relying on those pre-attentive characteristics that human vision can distinguish at lightning speeds, such as color, size, tilt and flicker (Healey and Enns, 2012). Unlike traditional accounting information, Big Data reporting is almost exclusively for internal consumption, so there are no restrictions on the visualization process and the primacy of user needs is fully recognized. However, because current regulatory reporting requirements are adapted to Big Data, this could change (Edinger, 2014). Regulatory reporting is heavily dominated by financial information but the increasing availability of other forms of information of interest to the public (e.g. social media data) may expand legal reporting requirements and subsequently regulator interest in Big Data applications for external use. One recent example of this is the implementation of IFRS 9, requiring financial institutions to identify, acquire and analyze data to predict which loans face a future risk of default and reflect this information on financial statements (Kimner, 2017).

Just as proper implementation of the information life cycle has allowed internal and external decision-makers to benefit from the intellectual product of an information system, adherence to this revised Big Data life cycle will provide decision-makers with valuable insights from the rapidly expanding amount of data available for consumption.

4. Information governance
A well-defined Big Data life cycle is not sufficient to convert Big Data into relevant and reliable information. Proper governance of that life cycle is also necessary. Smallwood (2016) defines information governance as "the activities and technologies that organizations employ to maximize the value of their information while minimizing associated risks and costs". In other words, information governance pairs the Big Data life cycle with processes, information technology and controls to create a functional information system. Without this component, the Big Data life cycle is no more than an idea. It's important to note that information governance encompasses all data activities, not just those involving Big Data.

The governing principles behind traditional data and Big Data are the same, but the volume, velocity and variety of Big Data can exponentially increase the risks as well as complicate the activities and technologies required to derive maximum value.

Without proper information governance, organizations are open to risks at various stages of the Big Data life cycle. Information governance provides the means to maximize value, minimize risk and achieve effective coordination, which is required to successfully implement the Big Data as well as the traditional information life cycles. Although the specific set of risks varies by organization, we highlight three common symptoms of a poorly governed maintenance phase. We select this phase because it is the primary focus of the Information Governance Initiative (2015), whose report details the consensus definitions of information governance and reports anecdotal and statistical evidence of what information governance professionals are currently doing. With each risk, we also identify specific governance activities to address the risk.

The first risk affects the storage step. The primary purpose of storing information within an organization is to render it accessible in the future. However, without proper governance, data can become irretrievable. This applies to traditional data as well as Big Data, but the risk is amplified by Big Data's innate characteristics of volume, velocity and variety. The primary consequence of this risk is the inability to derive financial benefit from inaccessible data (Information Governance Initiative, 2015). However, in addition to the obvious flaw that irretrievable data is useless, additional penalties arise from the inability to supply data for legal proceedings or contractual obligations. For example, Morgan Stanley was found guilty of discovery abuses owing to its inability to locate information; as a result, a jury awarded the plaintiff in the case, Coleman (Parent) Holdings, Inc., $1.4 billion in damages (LexisNexis, 2007). Although this example involved traditional data, and not Big Data, these penalties apply to digital data of any form (Fed. R. Civ. P., 2018). However, the volume of Big Data relative to its traditional counterpart, as well as the high velocity with which it arrives, dramatically increases the risk of irretrievability and the resulting importance of information governance of data storage.

To address this risk, an organization should implement consistent storage practices to ensure that instead of being siloed, disorganized or inaccessible, all business data is organized according to a company standard, rendering it catalogued and retrievable. To achieve consistency, IT leaders must develop a detailed architecture for how and where assets will be stored and organized. Then, working with the digital assets directly, IT staff can ensure compliance with established storage schemes, rendering all assets identified and retrievable (Tandberg Data, 2011).

However, this alone is insufficient to ensure that as-yet uncreated data will be findable in the future, as many members of an organization outside of the IT function create new data on a daily basis. As with all aspects of information governance, the most meaningful solution is education. Efforts to educate all employees in proper data creation and storage, underscored by the ownership of the education task by company executives across multiple business functions, will help instill and reinforce the knowledge and buy-in necessary to preserve the findability of corporate resources (Wambler, 2013).

The second risk affects the disposition step. Interestingly, this risk involves keeping too much retrievable data:

> Some information is ore that we should spend tremendous amounts of time and money refining and exploiting [. . .] while other information (some would say most) is just an industrial byproduct that represents potential cost, risk, and pain (Information Governance Initiative, 2015).

The informational risk of under-disposition and over-storage is, on the one hand, the exacerbation of the retrievability problem described earlier and, on the other hand, the increased likelihood of outdated data. Consistent storage practices require resources and time. However, over time, data becomes outdated or corrupted. This informational risk is also not limited to Big Data, but the large volume of Big Data significantly increases the amount of time and resources to manage, and in the meantime, data can become unwieldy, inaccessible, irrelevant or even misleading (Ferguson, 2012). Additionally, from a cost perspective, retaining all data is both unwise and infeasible (Rosenthal *et al.*, 2012). Counter to this, some believe that because the unit cost of storage has fallen and continues to fall, complete data retention is a cost-effective solution; however, just as the volume and velocity of Big Data reduce the retrievability of stored data, they also negate even the possibility of complete retention of stored data.

Over-storing also represents a severe control risk. In 2014, Sony Pictures Entertainment faced a prolonged attack by hackers to obtain a large cache of information including e-mails, digital films and employees' identifying information (Krebs, 2014). Sensitive but no longer useful, and in Sony's case damning, information such as outdated e-mail exchanges between executives should not have existed anywhere in the system at the time of the hack. This is not a risk unique to Big Data, but the risk is significantly amplified by the volume of Big Data and the difficulty in managing its velocity and variety. Furthermore, because much of Big Data is consumer data, compliance with expectations of privacy is critical. While scrubbing data to remove personally identifying information is one solution, it isn't always enough to avoid the appearance of a privacy breach as individuals become aware that their personal habits or choices have been recorded. The use of lengthy privacy notices may be legally sufficient to acquire the right to collect and analyze user data, but in the eyes of consumers, it also may not be enough. Regardless of the legality of collecting user data, consumer outrage can have negative effects on a company's bottom line. In 2016, Uber, Pokemon Go and WhatsApp, three well-known smartphone applications, came under fire for a perceived over-collection of private data (Hautala and Kerr, 2016). The Pew Research Center found that 88 per cent of American adults say that "not having someone watch you or listen to you without your permission" is important to them (Madden and Rainie, 2015). This is consistent with the President's Council of Advisors on Science and Technology, which "believes that the responsibility for using personal data in accordance with the user's preferences should rest with the provider rather than with the user" (Executive Office of the President, President's Council of Advisors on Science and Technology, 2014).

Proper information governance requires scheduled review and disposition of data that poses risks and holds little or no current or future value (McDonald and Léveillé, 2014). Disposition schedules create a formula that is used to ascertain whether and when information should be purged from a system. As regards penalties for data irretrievability, while some organizational leaders favor keeping everything just in case of lawsuit or government investigation, there is no legal penalty for documents or data that are undiscoverable because they were disposed according to an appropriate retention schedule (LexisNexis, 2007). On the other hand, all data that is retained, including Big Data, is discoverable in litigation, sometimes to an organization's detriment (Salvarezza, 2015). Putting in place policies to assess which data no longer adds value and following prescribed processes to purge it protects an organization from the need to store useless data that poses unnecessary risk.

The third risk affects the preprocessing step. The use phase represents the ultimate purpose of Big Data: to create useful knowledge from data. The biggest challenge that data analysts face in this endeavor is the fact that 80 per cent of their time is spent not on analysis

or the creation of new knowledge, but rather on collecting data and cleaning it prior to use (Lemieux *et al.*, 2014). This "data-wrangling" may be necessary to ensure accurate, useful and complete data exists to analyze, but these activities are tedious and time-consuming, especially in an uncontrolled information environment. Furthermore, dirty data can result in incomplete or erroneous reporting, consistent with the widely referenced "garbage in, garbage out" principle. Any data source can suffer from unclean data, but the variety of Big Data formats (e.g. log files, emails, audio, video, etc.) exacerbates the frequency of unclean data, especially when attempting analysis that relies on data from multiple sources.

There are two solutions to the collection of dirty data. The first is to address internal data collection (i.e. the data that are generated from within the organization). Many organizations generate their own Big Data on a daily basis; examining how this data are generated, while keeping in mind how data analysts will use it, can reduce dirty data. When data analysts collaborate with those responsible for creating or collecting internal data, it is possible for them to improve the current system, allowing for the collection of cleaner data that are organized in a manner that minimizes the need for analysts to preprocess (Hadidi, 2015). The second is to address external data collection. Unstructured data from Web sources such as webpages or social media platforms is an example of a popular source of Big Data. The use of appropriate data collection tools allows IT staff to convert it from unstructured data to structured data before storing it. This can reduce the amount of labor required to create usable data for analysis, although it does not eliminate the need for cleaning because issues such as duplicate data and incomplete data will still exist (Williams, 2013). Additionally, when new data come into the organization, IT staff can integrate it with existing data, improving analysts' ability to extend company knowledge (Ashutosh, 2012).

Information governance addresses not only proper life cycle governance activities, but also proper governing agents. Figure 3 displays the parties involved in the governance of the Big Data life cycle.

The rounded shape surrounding the Big Data life cycle implies that information governance manages (governs) the implementation of the life cycle principles. The parties listed on the left of Figure 3 beside the heading "Maintainers" are those who collaborate to develop and maintain a controlled system that converts Big Data into information. Because
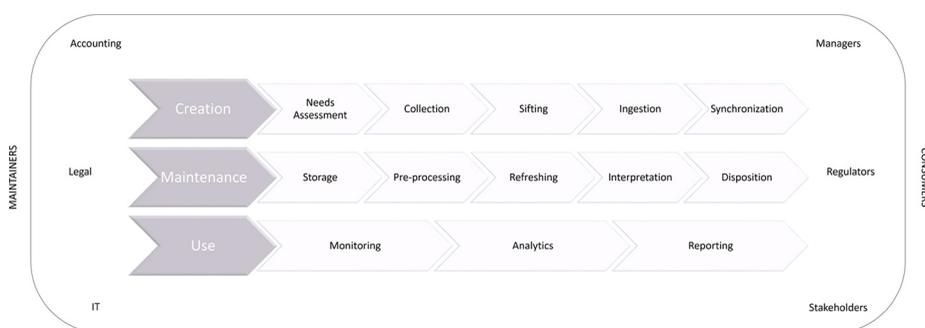
Notes: This figure displays an application of the information governance to the Big Data life cycle. The rows of chevrons display the phases and steps of the Big Data life cycle. The labels on the left of the figure list the maintainers of the information system: accounting, legal and IT. The labels on the right of the figure list the consumers of the information: managers, regulators and stakeholders (i.e. external stakeholders)

of the complexity of information technologies, businesses have historically relied on the focused expertise of IT specialists to design and maintain information systems to support business measurement and reporting needs. However, the understanding of measurement and reporting needs lies with business specialists, including accounting, finance, operations, HR and legal. A longstanding lack of comprehension of IT by business specialists, and vice versa, as well as a lack of understanding of the business value of well-governed information systems and processes, has prevented the communication and collaboration necessary to develop information systems that consistently capture relevant data in an efficient manner, convert that data into accurate, useful information for decision-making purposes and secure it against compromise (Congdon, 2015; Zetlin, 2015). "The top barrier to information governance progress is not a lack of money, but rather a set of factors including a lack of institutional education, communication, and leadership" (Information Governance Initiative, 2015). The siloing of business information and processes has long been a hindrance to the management of an organization's data, and it continues to be a primary barrier to progress in the implementation of information governance policies that successfully maximize the value of data and minimize risk (Information Governance Initiative, 2015). Historically, this siloing has not affected organizations sufficiently to evoke change in the level of collaboration, but with the speed of innovation inherent to the current economic environment, businesses no longer have the luxury of being able to miss out on information about market forces and customer behavior (Whitehurst, 2015).

Improved information governance is the answer to the need for better information and decision-making, and enhanced collaboration between business specialists and IT specialists is the answer to improved information governance. "The more coordination that occurs, the more effective the overall IG can be". (Information Governance Initiative, 2015).

Accountants have particular expertise as business specialists. As a result, accountants are valuable collaborators with IT specialists in information systems design and maintenance, not only as end-users of business information, but also as advocates for other internal and external decision-makers. Additionally, accountants are very familiar with governance issues relating to control objectives and regulatory compliance, which along with data and information management, is an important aspect of information governance. The focus of the accounting profession on controls reinforces the value of increased collaboration between accountants as business specialists and IT specialists in designing and maintaining systems that satisfy not only informational needs, but also control objectives as dictated by internal and external decision-makers[9].

In addition to accountants and IT, Figure 3 also includes an additional agent as a maintainer: legal. Accountants are experts on regulatory compliance, but legal specialists are experts on contracts. Because many control objectives, and some informational needs, serve to satisfy contractual obligations, legal specialists are valuable information governance partners (Smallwood, 2016)[10]. One straightforward example of how legal specialists can assist in the IT governance process is to facilitate discovery, along with the required supervision of any informational holds during proceedings, in the case of legal action against an organization. Another example is the need to ensure adherence to government mandated privacy laws such as HIPAA and FERPA, as well as an organization's own privacy policy.

Consumers of information, including internal and external users, also play an important role in information governance. In 2001, software developers created the Agile Manifesto to improve flexibility, speed and value in software development by creating cross-functional teams that include business specialists, developers and customers (Beck et al., 2001).

Additional movements in the development community, especially DevOps, have increased the number of players on these cross-functional teams to include not only those who develop the information system, but also those who deploy and maintain it (van Orman, 2014). All the while, the emphasis on needed collaboration with business partners and internal and external stakeholders has increased (Zetlin, 2015).

On the right-hand side of Figure 3, beside the heading "Consumers", are the internal and external decision-makers who rely on the information system. First and foremost are managers in their capacity as internal decision-makers. These are the primary consumers of business information, and they rely the most on a correct application of the Big Data life cycle. Although this observation would apply to the Big Data life cycle and the traditional information life cycle alike, it especially applies to Big Data because unlike other forms of business data, much of the collection, maintenance, and use of Big Data are not governed by regulation (yet), and much of the resulting information is not shared with external agents.

These internal decision-makers represent many business functions, such as operations and marketing, as well as accounting itself because accountants interface with IT to develop systems that meet the informational needs of other stakeholders, as well as their own informational needs. The importance of accountants' own informational needs are reinforced by a recent observation by a representative of a Big-4 accounting firm that accountants are the largest internal consumer of business information[11].

Despite the significance of internal decision-makers with respect to the consumption of Big Data, regulators and other stakeholders, such as customers, vendors, auditors and equity and debt holders also matter. In the first place, the controls that the system maintainers select satisfy requirements stipulated by consumers. From that perspective, consumers consume not only information, but also the benefits of internal controls. These parties may also consume the informational outputs of the system, but the degree to which this will occur depends on influence – debt holders can place considerable pressure on borrowers to share private information – and trends in the regulatory environment; regulators can begin to require additional disclosure of information derived from Big Data. Already research has begun to investigate the implications of increasing the pool of data used by external auditors, as well as the amount of data to make available in financial disclosures. The response by regulators to the rapid growth in Big Data will affect what external stakeholders will be able to consume in the future (Cao et al., 2015; Krahel and Titera, 2015).

With proper collaboration and processes, information governance can preserve the quality of Big Data throughout its life cycle. Accountants are already familiar with many internal control practices, especially those that govern financial reporting, but additional practices become necessary to handle the risks unique to the volume, velocity and variety of Big Data.

5. Conclusion
In this paper, we build on the traditional information life cycle to model a life cycle that accommodates the idiosyncrasies of Big Data. We then apply the concepts of information governance to the principles of a Big Data life cycle. The resulting model circumscribes an information system that converts Big Data into information in a manner that satisfies the informational and control needs of both internal and external decision-makers. Also, we make a case for accountants to play an important role in Big Data governance.

The recent upsurge in the creation of structured and unstructured data has encouraged businesses to increase their ability to extract business intelligence from previously

underutilized data sources. This endeavor requires modifications to existing information systems or the creation of new information systems to address the characteristics of Big Data that render both the data and the process of converting it into information distinct from existing business data. In addition, the availability of new and non-traditional information increases demands on internal and external reporting and exposes organizations to new risks and attention from regulators that must be addressed. Reporting and control responsibilities are typically the purview of accountants.

The design of efficient and effective systems requires more than a sequence of life cycle steps, and businesses cannot assume that IT expertise endows IT specialists with all necessary knowledge to design a useful information system. The volume, velocity and variety of Big Data give rise to both informational and control risks, and IT specialists cannot address these risks alone. Instead, individuals and teams with expertise in the informational needs of internal and external decision-makers and regulatory and contract compliance with respect to reporting and internal controls should collaborate with IT specialists to develop and maintain information systems. Much of the necessary expertise resides with accounting professionals; however their education and training will need to be expanded to adequately fulfill these new roles.

Finally, we highlight some of the risks connected with the big data environment, but our list is far from comprehensive. Much future research is necessary to explore the risks that Big Data creates, investigate the potential for new kinds of reports, new roles for regulatory bodies, document information governance policies from practice and study factors that lead to effective governance structures and policies.

Notes

1. Some have proposed additional Vs, such as veracity, value, variability, viability and victory. However, unlike the original three Vs that are inherent characteristics of Big Data, these other attempts to rely on a convenient mnemonic (aka "wanna-vs") "mistake interpretive, derived qualities for essential attributes" (Grimes, 2013). As a result, we restrict our perspective to the original three Vs.

2. Some information life cycle models include fewer steps or combine multiple steps into one. We choose this expanded list to provide a thorough review of the activities necessary to convert data into information. An expanded list also allows us to highlight the differences between the information and Big Data life cycles with increased granularity.

3. It is infeasible to portray a perfectly linear information life cycle because no conceptual prescription exists dictating the amount of time between data acquisition and information reporting. Additionally, the same data could be used and reused on a repeated basis within a sub-loop of the steps on the information life cycle (Upward, 1996).

4. Data analysts report that 60 per cent of their time is spent cleaning data before analysis (Press, 2016). One reason for this need is a lack of correct application of the steps in the creation phase of the information life cycle. "Lack of data quality can be thought of as a polluted lake. The lake represents the databases, water is the data, and streams are the sources of new data. Data pollution in this metaphor is caused by breaks in business processes" (Redman, 2014). However, in the presence of Big Data, even a well-defined system may require data cleaning. We discuss this issue below in the context of the Big Data life cycle.

5. "Variability" in creation rate is a commonly cited "wanna-v." Although we exclude variability from our list of Big Data characteristics, we reference it as a "derived quality" of velocity: the increase in average creation rate has opened the door for increased variability in creation rates (Grimes, 2013).

6. One glaring absence from the Big Data life cycle is the classification step and its assignment of metadata. Big Data is an accumulation of data objects individually governed by the traditional information life cycle. These data objects received metadata assignment previously during their respective classification steps, and when they are used, essentially reused, by the Big Data life cycle, they require no assignment of new metadata (Horodyski, 2014). For example, a form on a customer-facing website assigns metadata to each field when a customer creates a new record. However, Big Data, which is the accumulation of customer records, receives no additional metadata. Machine-generated data, the fastest growing Big Data source, provides another example (EMA, 2014). Each communication between digital devices has metadata attributes and values. Big Data, as a collection of these communications, needs no additional classification.

7. Value is another one of the derived traits of the inherent characteristic of volume (a "wanna-V"), which helps to mediate the amount of data collected and retained during the creation phase of the Big Data life cycle (Grimes, 2013).

8. This need for storage that readily expands, as needed, is a primary motivator for current cloud computing initiatives and the associated transition to on-demand, cloud-based storage solutions (AWS, 2016).

9. Despite this preexisting familiarity with informational needs and responsibility for control issues, accountants require additional understanding of information technology and training in technical skills. As the introduction of SOX, accountants have occupied themselves with the audit of IT controls, the responsibility for the implementation of these controls has rested squarely on the shoulders of IT specialists (Desmond, 2016). An increased knowledge of IT would increase accountants' ability to engage in not only information governance, but also in the more traditional IT audit function, which despite the importance of IT controls, has not been a core competency of many accounting professionals (Coyne *et al.*, 2016).

10. It would be possible to add additional business specialists to the list of maintainers. We limit the number of maintainers to these three, not in an attempt to exclude other business specialists, but rather to highlight a core minimum of collaborators. Attempts to expand the network of collaborators will likely have additional positive consequences for information governance.

11. This statement was part of a presentation at the 2016 American Taxation Association Midyear meeting.

References

Amazon Web Services (AWS) (2016), "Big data analytics options on AWS", available at: https://d0.awsstatic.com/whitepapers/Big_Data_Analytics_Options_on_AWS.pdf

Arthur, L. (2013), "What is big data?", available at: www.forbes.com/sites/lisaarthur/2013/08/15/what-is-big-data/#621a5a813487

Ashutosh, A. (2012), "Best practices for managing big data", available at: www.forbes.com/sites/ciocentral/2012/07/05/best-practices-for-managing-big-data/#2091c52ef028

Bansal, S.K. and Kagemann, S. (2015), "Integrating big data: a semantic extract-transform-load framework", *Computer*, Vol. 48 No. 3, pp. 42-50.

Bayer, J. and Taillard, M. (2014), "Story-driven data analysis", From Data to Action, available at: www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/hbr-from-data-to-action-107218.pdf

Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R.C., Mellor, S., Schwaber, K., Sutherland, J. and Thomas, D. (2001), "Manifesto for Agile software development", available at: www.agilemanifesto.org/

Cao, M., Chychyla, R. and Stewart, T. (2015), "Big data analytics in financial statement audits", *Accounting Horizons*, Vol. 29 No. 2, pp. 423-429.

Congdon, L. (2015), "Digital education presents new challenges and opportunities for IT", available at: https://enterprisersproject.com/article/2015/6/gaining-perspective-its-digital-education-challenge

Coyne, J., Coyne, E. and Walker, K. (2016), "A model to update accounting curricula for emerging technologies", *Journal of Emerging Technologies in Accounting*, Vol. 13 No. 1, pp. 161-169.

Davenport, T. (2014), "What to ask your numbers people", From Data to Action, available at: www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/hbr-from-data-to-action-107218.pdf

Dederer, M.G. and Dmytrenko, A. (2015), "8 Steps to effective information lifecycle management", *Information Management Journal*, Vol. 49 No. 1, pp. 32-35.

Desmond, P. (2016), "Holding all employees accountable: ensuring security across the enterprise", The Enterprisers Project, available at: https://enterprisersproject.com/article/2016/4/holding-all-employees-accountable-ensuring-security-across-enterprise

Dumbill, E. (2012), "What is big data? O'Reilly Media", available at: www.oreilly.com/ideas/what-is-big-data

Edinger, S. (2014), "The metrics sales leaders should be tracking", *From Data to Action*, available at: www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/hbr-from-data-to-action-107218.pdf

EMA (2014), "Big data: operationalizing the buzz", available at: www.sas.com/en_us/offers/sem/ema-operationalizing-the-buzz-big-data-107198.html

Executive Office of the President, President's Council of Advisors on Science and Technology (2014), "Big data and privacy: a technological perspective", available at: www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf

Federal Rules of Civil Procedure (2018), "Title 5, Rule 26", available at: www.federalrulesofcivilprocedure.org/frcp/title-v-disclosures-and-discovery/rule-26-duty-to-disclose-general-provisions-governing-discovery/

Ferguson, R. (2012), "The storage and transfer challenges of big data", available at: http://sloanreview.mit.edu/article/the-storage-and-transfer-challenges-of-big-data/

Grimes, S. (2013), "Big data: Avoid 'wanna V' confusion", Information Week, available at: www.informationweek.com/big-data/big-data-analytics/big-data-avoid-wanna-v-confusion/d/d-id/1111077?

Grimes, S. (2014), "Metadata, connection and the big data story", available at: www.huffingtonpost.com/seth-grimes/metadata-connection-and-t_b_5225861.html

Hadidi, S. (2015), "Data is ugly: tales of data cleaning", available at: www.kdnuggets.com/2015/08/data-ugly-tales-data-cleaning.html

Hautala, L. and Kerr, D. (2016), "When apps collect more data, outrage is powerful – sometimes", available at: www.cnet.com/news/when-apps-collect-more-data-outrage-is-powerful-sometimes/

Healey, C.G. and Enns, J.T. (2012), "Attention and visual memory in visualization and computer graphics", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 18 No. 7, pp. 1170-1188.

Henschen, D. (2013), "When NOSQL makes sense", *Informationweek*, Vol. 1376, pp. 8-15.

Hoch, M. (2014), "Google on launching an analytics MOOC and taking data-driven actions", From Data to Action, available at: www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/hbr-from-data-to-action-107218.pdf

Hoke, G.E.J. (2011), "Records life cycle: a cradle-to-grave metaphor", *Information Management*, Vol. 45 No. 5, pp. 28-32.

Horodyski, J. (2014), "Breaking down big data: the value in metadata", available at: www.cmswire.com/cms/information-management/breaking-down-big-data-the-value-in-metadata-026985.php

Huang, W. and Boateng, A. (2016), "On the value relevance of analyst opinions and institutional shareholdings in China", *International Journal of Accounting and Information Management*, Vol. 24 No. 3, pp. 206-225.

Information Governance Initiative (2015), *Information Governance Initiative Annual Report 2015-2016*, Information Governance Initiative LLC, September, available at: http://iginitiative.com/wp-content/uploads/2015_IGI-Annual-Report_Final-digital-use.pdf

Ji, D., Ahmed, K. and Lu, W. (2015), "The impact of corporate governance and ownership structure reforms on earnings quality in China", *International Journal of Accounting & Information Management*, Vol. 23 No. 2, pp. 169-198.

Kimner, T. (2017), "CECL and IFRS 9: the challenges of new financial standards", available at: www.sas.com/en_us/insights/articles/risk-fraud/challenge-of-new-financial-standards.html

Krahel, J.P. and Titera, W.R. (2015), "Consequences of big data and formalization on accounting and auditing standards", *Accounting Horizons*, Vol. 29 No. 2, pp. 409-422.

Krebs, B. (2014), "Sony breach may have exposed employee healthcare, salary data", available at: https://krebsonsecurity.com/2014/12/sony-breach-may-have-exposed-employee-healthcare-salary-data/

Laney, D. (2001), "3D data management: controlling data volume, velocity, and variety", Application Delivery Strategies, available at: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Leahy, E.J. (1949), "Modern records management", *American Archivist*, Vol. 12 No. 3, pp. 231-242.

Lemieux, V.L., Gormly, B. and Rowledge, L. (2014), "Meeting big data challenges with visual analytics", *Records Management Journal*, Vol. 24 No. 2, pp. 122-141.

LexisNexis (2007), "Elements of a good document retention policy", available at: www.lexisnexis.com/applieddiscovery/lawlibrary/whitePapers/ADI_WP_ElementsOfAGoodDocRetentionPolicy.pdf

Lin, J. and Liu, C. (2015), "R&D, corporate governance, firm size and firm valuation evidence from taiwanese companies", *International Journal of Corporate Governance*, Vol. 6 No. 2, pp. 87-97.

Madden, M. and Rainie, L. (2015), "Americans' views about data collection and security", available at: www.pewinternet.org/2015/05/20/americans-views-about-data-collection-and-security/

McDonald, J. and Léveillé, V. (2014), "Whither the retention schedule in the era of big data and open data?", *Records Management Journal*, Vol. 24 No. 2, pp. 99-121.

Menon, S. (2014), "Stop assuming your data will bring you riches", From Data to Action, www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/hbr-from-data-to-action-107218.pdf

Morison, R. (2014), "How to get more value out of your data analysts", available at: https://hbr.org/2013/12/how-to-get-more-value-out-of-your-data-analysts

Nemschoff, M. (2013), "Social media marketing: how big data is changing everything", available at: www.cmswire.com/cms/customer-experience/social-media-marketing-how-big-data-is-changing-everything-022488.php

Patel, J.M. (2016), "Operational NoSQL systems: what's new and what's next?", *Computer*, Vol. 49 No. 4, pp. 23-30.

Praminick, S. (2013), "Ten big data implementation best practices", available at: www.ibmbigdatahub.com/blog/10-big-data-implementation-best-practices

Press, G. (2016), "Cleaning big data: most time-consuming, least enjoyable data science task, survey says", available at: www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#85238517f758

Redman, T.C. (2014), "Getting in front of data quality", From Data to Action, www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/hbr-from-data-to-action-107218.pdf

Reeves, A. (2011), "What is different about big data governance?", available at: https://infocus.emc.com/april_reeve/what-is-different-about-big-data-governance/

Rosenthal, D.S., Rosenthal, D.C., Miller, E.L., Adams, I.F., Storer, M.W. and Zadok, E. (2012), "The economics of long-term digital storage", *Memory of the World in the Digital Age*, Vancouver, BC.

Salvarezza, M. (2015), "Records & information management: 2015 risk perspective", available at: http://corporatecomplianceinsights.com/records-information-management-2015-risk-perspective-2/

SAS (2016), "What is big data?", available at: www.sas.com/en_us/insights/big-data/what-is-big-data.html#dmhistory

Satell, G. (2015), "IBM's latest move signals a new era for data", available at: www.forbes.com/sites/gregsatell/2015/06/22/ibms-latest-move-signals-a-new-era-for-data/#4f2d24c44710

Smallwood, R. (2016), *Information Governance for Executives: Fundamentals and Strategies*, Bacchus Business Books, San Diego, CA.

Song, L. (2016), "Accounting quality and financing arrangements in emerging economies", *International Journal of Accounting & Information Management*, Vol. 24 No. 1, pp. 2-19.

Tandberg Data (2011), "Guide to data protection best practices", available at: http://tandbergdata.com/default/assets/File/white_papers/WP_BackupGuide.pdf

Upward, F. (1996), "Structuring the records continuum, part one: postcustodial principles and properties", *Archives and Manuscripts*, Vol. 24 No. 2.

Upward, F. (1997), "Structuring the records continuum, part two: structuration theory and recordkeeping", *Archives and Manuscripts*, Vol. 25 No. 1.

van Orman, C. (2014), "DevOps is not a synonym for application development", available at: https://enterprisersproject.com/article/2014/7/devops-not-synonym-application-development

Wall, M. (2014), "Big data: are you ready for blast-off? BBC News", available at: www.bbc.com/news/business-26383058

Wambler, S. (2013), "Agile/lean data governance best practices", available at: www.agiledata.org/essays/dataGovernance.html

Whitehurst, J. (2015), *The Open Organization: Igniting Passion and Performance*, Harvard Business Review Press, Boston, MA.

Williams, A. (2013), "Import.io turns web pages into spreadsheets for getting out the data that matters most", available at: https://techcrunch.com/2013/09/12/import-io-turns-web-pages-into-spreadsheets-for-getting-out-the-data-that-matters-most/

Zetlin, M. (2014), "Advice on how to handle shadow IT", The Enterprisers Project, available at: https://enterprisersproject.com/article/2014/9/break-down-silos-between-it-and-business

etlin, M. (2015), "CIOs should make sure their teams are as strong as their networks", The Enterprisers Project, available at: https://enterprisersproject.com/article/2015/5/cios-should-make-sure-their-teams-are-strong-their-networks

Zhang, J., Yang, X. and Appelbaum, D. (2015), "Toward effective big data analysis in continuous auditing", *Accounting Horizons*, Vol. 29 No. 2, pp. 469-476.

Corresponding author

Joshua G. Coyne can be contacted at: jgcoyne@memphis.edu