International Journal of Accounting Information Systems xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Accounting Information Systems

journal homepage: www.elsevier.com/locate/accinf

Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors

Michael Alles [a,*], Glen L. Gray [b]

[a] Rutgers Business School, Department of Accounting & Information Systems, One Washington Park, Room 928, Newark, NJ 07102-3122, United States
[b] Dept. of Accounting & Information Systems, College of Business & Economics, California State University, Northridge, 18111 Nordhoff Street, Northridge, CA 91330-8372, United States

ARTICLE INFO

Article history:
Received 1 June 2015
Received in revised form 9 September 2015
Accepted 14 July 2016
Available online xxxx

Keywords:
Big Data
Auditing
Accounting information systems

ABSTRACT

With corporate investment in Big Data of $34 billion in 2013 growing to $232 billion through 2016 (Gartner 2012), the Big 4 accounting firms are aiming to be at the forefront of Big Data implementations. Notably, they see Big Data as an increasingly essential part of their assurance practice. We argue that while there is a place for Big Data in auditing, its application to auditing is less clear than it is in the other fields, such as marketing and medical research. The objectives of this paper are to: (1) provide a discussion of both the inhibitors of incorporating Big Data into financial statement audits; and (3) present a research agenda to identify approaches to ameliorate those inhibitors.

1. Introduction

1.1. Increasing interest in Big Data by auditors

A survey by Gartner (2014) found that 73% of respondents had invested or planned to invest in Big Data in the next 24 months, up from 64% in 2013. Corporate investment in Big Data is growing from $34 billion in 2013 to $232 billion in 2016 (Gartner, 2012). Likewise, accounting firms are declaring that Big Data is increasingly essential part of their assurance practice.[1] For example, EY states: "*The audit of the future will bear little resemblance to the traditional audit CFOs are accustomed to receiving today. In fact, the way organizations conduct audits will change more in the next 5-10 years given the evolution of technology and analytics. Data analytics, new technology and access to detailed industry information will all combine to help auditors better understand the business, identify risks and issues and deliver additional insights. Moreover, the ability to review and analyze entire sets of data, rather than applying sampling techniques, will help bring more confidence to the audit.*"[2] Similarly, Deloitte Chairman and CEO Joe Ucuzoglu writes: "*At Deloitte we're investing several hundred million dollars in data analytics and artificial intelligence with some cutting-edge applications that we really believe differentiate us and our audit approach. When we use these tools, we're able to get*

* Corresponding author.
E-mail addresses: alles@business.rutgers.edu (M. Alles), glen.gray@csun.edu (G.L. Gray).

[1] Although accounting firms could conduct a wide variety of engagements under a variety of professional standards; this paper focuses on financial statement audits conducted in compliance with the Statements of Auditing Standards (SASs) published by the AICPA, Auditing Standards (ASs) published by the PCAOB, and International Standards on Auditing (ISAs) published by the IAASB.

[2] http://www.ey.com/GL/en/Issues/Managing-finance/EY-cfo-need-to-know-future-of-audit?preview&HL=CON-USDD-9XAN4E&utm_source=eycom&utm_medium=homepage_PF&utm_campaign=Future%20of%20audit#introduction. Accessed 9/1/2015 1:26:52 PM.

http://dx.doi.org/10.1016/j.accinf.2016.07.004
1467-0895/© 2016 Published by Elsevier Inc.

		Data Analytic Techniques	
		Traditional (Excel, ACL, Idea)	**Extended** (Visualization, Predictive analytics)
Data Sources	**Traditional** (Accounting & Financial)	A	B
	Extended (Non-Financial Data ➡ Big Data)	C	D

Fig. 1. Paths to expand data analytics in financial statement audits.

greater coverage. We're able to more quickly identify risks. We're able to complete the audit with a higher level of quality and ultimately deliver a greater level of insight to our clients."[3] PwC is undertaking a pilot study of ten audit clients to which they are applying Big Data techniques.[4] PwC audit partner Mary Grace Davenport states that it inconceivable that auditors would not use Big Data when the marketing departments of their clients rely so heavily upon it.[5]

The year 2015 saw a sharp increase in Big Data presentations and publications by accounting academics and accounting practitioners. The June 2015 issue of *Accounting Horizons* had a special issue that included six commentary articles on Big Data. In September 2015, the American Accounting Association (AAA) organized their inaugural *Accounting IS Big Data* conference in New York City. The conference had over 200 attendees with a mix of academics, practitioners, and vendors. The 2015 AAA Annual Meeting and the subsequent 2016 Audit Midyear Meeting and AIS Midyear Meeting included several Big Data paper presentations and panels.

What is notable, however, is that most of these presentations and publications were not specifically about Big Data—at least as Big Data is commonly defined by experts outside accounting (we shall discuss these definitions in Section 2 below). The vast majority of the presentations and panels were actually discussing data analytics of essentially traditional accounting data. Data analytics is not Big Data, though, and traditional accounting data is not Big Data. When discussing Big Data in an audit context, it is important to differentiate between more of the same kind of data that auditors are already using, or more data of a different kind than what auditors have traditionally relied on to give an audit opinion. The former approach would lead, for example, to continuous auditing where the scope of data is not necessarily expanded but measurements are taken more frequently in time (Kogan et al., 2014). By contrast, Big Data pushes the domain of data far outwards from financial data to non-financial data (NFD)—from structured to unstructured data and from inside the organization to outside it—to an extent that may well be outside the comfort zone and technical capability of the current audit profession.[6]

This confusion about what the term Big Data encompasses is not uncommon even in the technical disciplines, with IBM researchers Zikopoulos et al. (2013) writing, "…*the term Big Data is a bit of a misnomer*". The real value of (the real interest in) Big Data is the value of the analytics that can be performed with that data—the ultimate insights gleaned from the data.

While Big Data and data analytics are two independent concepts, these two concepts can be interrelated. Fig. 1 illustrates how these two concepts can be related in the audit domain. For many years accounting firms have been comfortably operating in Cell A using traditional data analytic tools (e.g., Excel, ACL, and Idea) to analyze samples of accounting data. As illustrated in Fig. 1, based on practitioner presentations at 2015 and 2016 AAA meetings, accounting firms have started moving into Cell B and moving away from sampling (commonly referred as taking a 100% sample). Data visualization tools in particular, such as Tableau, are growing in popularity as an audit tool. But the focus is still on traditional accounting data and performing traditional audit procedures, such as locating duplicate invoices. Although there has been some mention by practitioners of using social media analysis as part of an audit, and thereby moving into Cell C, it appears that there is far less movement into Cell C. Cell D would be truly incorporating both Big Data and advanced data analytic tools as part of an audit. Although accounting academics discuss the broad advantages of Big Data and practitioners mentioned Big Data in a positive light in their presentations and publications, the actual use of Big Data is far beyond current audit activities.

However, as Moore (2002) illustrates (see Fig. 2), there are hurdles that must be overcome in any technology innovation diffusion within organizations. Moore's "adoption chasm" is a popular framework in the technology adoption literature. These chasms represent barriers or inhibitors to the adoption of a technology between different types of technology adopters that can help explain why a technology implementation can suddenly stop growing at accounting firms. The most challenging chasm is between Early Adopters and Early Majority, or characterized another way, between Visionaries and Pragmatists. As those latter names imply, visionaries enjoy testing new technology and speculating about what the future implementations could be like, while pragmatists do not get excited

[3] http://www2.deloitte.com/us/en/pages/audit/articles/deloitte-prepares-firm-for-audit-of-the-future-accounting-today-article.html. Accessed 9/1/2015 1:32:53 PM.
[4] Discussion by PwC partners at American Accounting Association annual meeting in Atlanta, August 2014.
[5] Discussion by PwC partners at American Accounting Association annual meeting in Atlanta, August 2014.
[6] For example, a senior technical partner at a Big-4 firm known for its statements extolling its commitment to the use of Big Data in auditing, told the authors that in reality the firm focusses almost entirely on structured data and not what he and others consider to be true, unstructured, Big Data,

Please cite this article as: Alles, M., Gray, G.L., Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors, Int J Account Inf Syst (2016), http://dx.doi.org/10.1016/j.accinf.2016.07.004

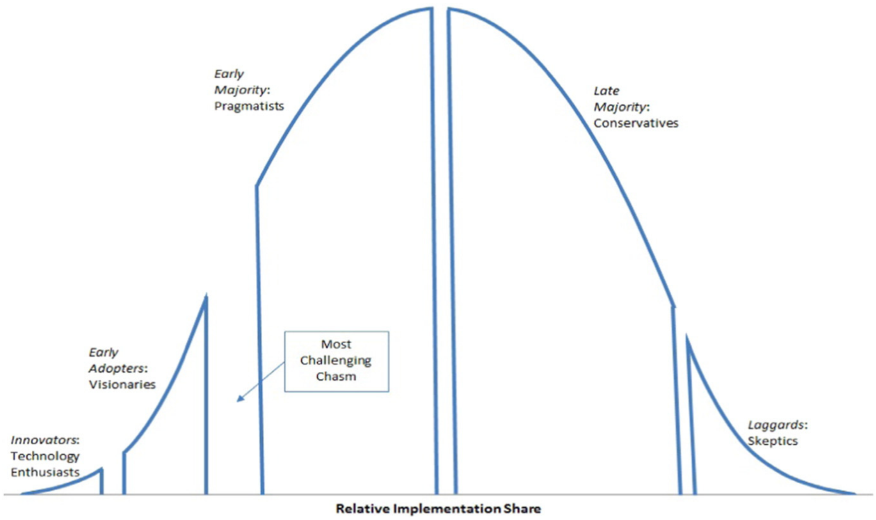

Fig. 2. Moore's (2002) Technology adoption chasms.

about new technology just because it is something new, and instead they have to be convinced that the new technology is truly superior to established technologies.

Specifically in the auditing context there are adoption chasms between the innovators (senior-level partners) who speak at conferences and the downstream technology adopters (i.e., frontline audit partners) in the accounting firms. Auditors will face major challenges of crossing those chasms as they strive to migrate to Cell C and truly incorporate Big Data and sophisticated analytics in Cell D. In this paper we focus on the various inhibitors that will have to be overcome for auditors to move into Cells C and D. We will show that the application of Big Data in auditing is less straightforward than it is in the other fields, such as marketing and medical research. Big Data and advanced data analytics are disruptive technologies and, as such, functioning in Cell D would be a paradigm shift as to how financial statement audits are conducted: it would require significant changes to current audit standards; unprecedented access to proprietary and sensitive client data that are outside traditional data requested during an audit; a significantly higher reliance on non-financial data (NFD), which auditors have been reluctant to use in the past because it's not clear how to validate this non-GAAP data; increased technical skills on the part of the audit team; and increased business acumen to determine what Big Data elements to analyze and how to interpret the results.

With much of the current literature emphasizing the benefits of Big Data and data analytics, our paper focuses on the inhibitors. In the paper we present ways that these inhibitors can be addressed and present a research agenda to advance Big Data and data analytics in financial statement audits. We are not advocating the idea that Big Data should or should not be part of future financial statement audits—only ongoing research and practice will answer the question as to whether the benefits outweigh the costs.

1.2. The inevitability of Big Data in auditing versus the historical record

At a theoretical or normative level it seems logical that auditors will incorporate Big Data into future financial audits (Littley, 2012; Moffitt and Vasarhelyi, 2013; Setty and Bakhshi, 2013; IIA, 2013; Whitehouse, 2014; AICPA, 2014). Big Data potentially provides the most robust picture of the client's activities; far beyond that captured using only the smaller population of GAAP-compliant data. A key requirement of auditing standards is setting expectations at the beginning of the audit—an issue that PCAOB inspectors are particularly critical of in their reports. A core component of audit planning is establishing "*expectations developed from data derived from other sources*" (AU 329.16) before collecting and analyzing financial data.[7] Setting expectations based on "disaggregate data" is a recurring theme in Staff Audit Practice Alert No. 12 (PCAOB, 2014) to improve auditing of revenue, which was motivated by the PCAOB's frequent criticism of revenue auditing in its accounting firm inspection reports. The PCAOB is also pushing accounting firms to boost their fraud detection procedures in compliance with AU 316.

Zhang et al. (2015); Cao et al. (2015) and Yoon et al. (2015) argue that Big Data will enable auditors to improve their effectiveness. These articles, all in the Big Data Forum published in the June 2015 issue of *Accounting Horizons*, assume the inevitability of incorporating Big Data into financial statement auditing. The aforementioned September 2015 AAA *Accounting IS Big Data* conference also simply took for granted that Big Data will be the future of auditing, even though few examples of its specific use in accounting or auditing were presented.

[7] http://www.aicpa.org/Research/Standards/AuditAttest/DownloadableDocuments/AU-00329.pdf.
Accessed 9/2/2015 9:26:51 AM.

Please cite this article as: Alles, M., Gray, G.L., Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors, Int J Account Inf Syst (2016), http://dx.doi.org/10.1016/j.accinf.2016.07.004

In this paper we temper this optimism about the use of Big Data in auditing in light of the empirical evidence that accounting firms have had a mixed history of implementing assurance-related technologies. The 1996 *Special Committee on Assurance Services Report* (the *Elliott Report*) predicted that the annual fees for the new assurance services recommended in that report would reach $14 to $21 billion (compared to $8 billion paid in annual financial audit fees in 2013[8]). E-commerce assurance services (later called WebTrust) alone were predicted to generate $1.25 to $6.25 billion annually, far in excess of the reality today.

Regarding technology for financial statement audits specifically, articles equally optimistic as the *Accounting Horizons* Big Data articles (some by the same authors) were published in the past about artificial intelligence (AI), expert systems, data mining, and continuous auditing [CA]. Over 30 years ago, the then Big 6 were spending millions of dollars developing expert system applications. Brown (1991) listed 43 expert systems in use or under development by the Big 6; yet according to Gray et al. (2014), none of these stand-alone applications were incorporated into regular financial audit activities.[9] Researchers have been promoting continuous auditing and continuous monitoring [CM] for nearly a quarter century (Vasarhelyi and Halper, 1991), but as the AICPA (2012) stated, "*The general view is that not much is currently being done with CA/CM.*"[10] More recently, articles (e.g., Debreceny & Gray 2010 and 2011 and Gray and Debreceny, 2014) have illustrated how various forms of data mining could be valuable resources for auditors, but generally this technology is isolated in accounting firms to special forensic engagements.

Notwithstanding this discouraging record, we are not saying the accounting firms made mistakes not implementing these and other technologies as a regular part of traditional financial audits. Audit standards AS No. 15, SAS No. 122 (AU-C sec. 500) and ISA 500, all essentially say [Quoting from SAS No 122.04]: "*The objective of the auditor is to design and perform audit procedures that enable the auditor to obtain sufficient appropriate audit evidence to be able to draw reasonable conclusions on which to base the auditor's opinion.*" [Emphasis added]. Since tens of thousands of audits are completed each year, auditors currently believe they are collecting *sufficient appropriate* audit evidence to draw *reasonable conclusions* without these advanced technologies. As such, the overarching research question regarding new technologies and financial statement audits is: what are the inhibitors that seem to block the incorporation of advanced technology and how can these inhibitors be overcome? Specific to Big Data analytics: for what parts of the current audit data gathering and analysis procedures can Big Data analytics be cost-effectively substituted? We say substituted because we assume Big Data analytics will not simply be added to current audit procedures—and that accounting firms will just increase their audit fees to reflect this addition.

1.3. Costs and benefits of using Big Data in auditing

Just because the benefits of a technology outweigh the costs for a client's organization does not automatically mean that the same technology is cost-beneficial for the auditors, particularly in the heavily regulated and litigious environment in which accounting firms must function. Cost/benefit decisions, by their very nature, are very different for partnerships vs. corporations. For example, consider a situation where the client corporation and their Big 4 auditor are essentially the same size. If someone in the corporation proposes a $25 million technology modernization project, the CEO, CFO, or other executives do not calculate what that's going to personally cost them (What part of that $25 million will come out of my pocket?). At a partnership, partners can quickly determine what that project is going to personally cost them based on the prior year's profit distribution formula. In other words, conceptually, each partner does his or her own cost/benefit calculation. Although it's 25-years old, to see an example of this thinking, see the Harvard Business School case study called "KPMG Peat Marwick: The Shadow Partner," (Eccles and Gladstone 1991).[11]

1.4. Outline of the paper

The objectives of this paper are to: (1) to identify inhibitors regarding incorporating Big Data into financial statement audits; and (3) present a research agenda to identify specific aspects of Big Data that could benefit auditors. The remainder of this paper is organized as follows. Section 2 illustrates the many definitions of Big Data and places it into the context of the auditing environment. Section 3 identifies how and where Big Data *could* enhance financial audits in a normative sense. Then in Section 4 we identify the unique characteristics of what Big Data analysis requires as inputs and what it produces as outputs and compare that with the unique needs and constraints of auditing. As we show, this is where significant inhibiters arise that will require auditors to proceed very carefully and systematically if they are not to make some of the mistakes that have already arisen in other Big Data applications. That naturally leads in Section 5 to a proposed research agenda to identify how to close the gap between traditional audit procedures and the appropriate use of Big Data. Section 6 has concluding comments.

[8] http://www.auditanalytics.com/blog/audit-fees-and-non-audit-fees-a-twelve-year-trend/. Accessed 11/27/2015 2:50:30 PM.

[9] Although these *stand-alone* expert systems disappeared, some of this technology was subsequently embedded in other audit tools (Dowling and Leech, 2007).

[10] http://www.aicpa.org/interestareas/frc/assuranceadvisoryservices/downloadabledocuments/whitepaper_current-state-continuous-auditing-monitoring.pdf. Last accessed 2/22/2016 11:18:55 AM.

[11] Eccles, Robert G., Jr. "KPMG Peat Marwick: The Shadow Partner." Harvard Business School Case 492-002, July 1991. (Revised October 1995.) [Citations for this case vary. Sometimes Gladstone is first author. The latest seems to drop Gladstone.]

Please cite this article as: Alles, M., Gray, G.L., Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors, Int J Account Inf Syst (2016), http://dx.doi.org/10.1016/j.accinf.2016.07.004

2. An overview of Big Data

2.1. Defining Big Data

The first issue faced in exploring Big Data is that "Big Data" lacks a consistent definition. One website lists 32 definitions and another website had seven more.[12] Wikipedia defines Big Data as: "*Big Data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization. The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to spot business trends, determine quality of research, prevent diseases, link legal citations, combat crime, and determine real-time roadway traffic conditions.*"[13]

Although not mutually exclusive, some Big Data definitions focus on the dimensions or characteristics of Big Data and other definitions focus more on examples of the contents of Big Data. On the characteristics side, frequently, Big Data is defined in terms of volume, velocity, variety, and veracity (commonly referred as the "4 Vs").[14] Volume refers the overall amount of data included in a Big Data dataset. Velocity is how frequently the data are changing. Many Big Data installations are collecting real-time sensor data that are being continuously updated. Variety is the broad scope of data that organizations are collecting. Veracity relates to the integrity of the data. Veracity may be particularly problematic to auditors—i.e., how does the auditor develop an appropriate level of confidence in a client's Big Data with massive amounts of non-financial data?

In terms of defining Big Data in term of diverse content examples, the *data* in Big Data could include some mix of traditional structured financial and non-financial data (NFD), logistics data, sensor data, emails, telephone calls, social media data, blogs, as well as other internal and external data. Auditors' traditional focus on *transactional* accounting data, hence, a particularly relevant *content* definition of Big Data in the auditing context is that by Connolly (2012), which takes transactions as its starting point[15]:

He goes on to explain and illustrate this equation:

> "*ERP, SCM, CRM, and transactional Web applications are classic examples of systems processing Transactions. Highly structured data in these systems is typically stored in SQL databases. Interactions are about how people and things interact with each other or with your business. Web Logs, User Click Streams, Social Interactions & Feeds, and User-Generated Content are classic places to find Interaction data.Observational data tends to come from the 'nternet of Things'. Sensors for heat, motion, pressure and RFID and GPS chips within such things as mobile devices, ATM machines, and even aircraft engines provide just some examples of 'things' that output Observation data.*"

Connolly's (2012) framework (see Fig. 3) is useful because it puts the data currently used by auditors (in a small box in the lower-left corner) into perspective and shows how much additional data that Big Data offers to expand that input into the auditing process. Moving to Cells C and D (moving into Big Data) in Fig. 1 means the auditor will be expanding outside of that small box in the corner into a vast population of NFD.

Connolly (2012) goes on to identify what he sees as seven drivers of Big Data in business:

Business

1. Opportunity to enable innovative new business models
2. Potential for new insights that drive competitive advantage

Technical

1. Data collected and stored continues to grow exponentially
2. Data is increasingly everywhere and in many formats
3. Traditional solutions are failing under new requirements

Financial

1. Cost of data systems, as a percentage of IT spend, continues to grow
2. Cost advantages of commodity hardware & open source software

[12] http://www.opentracker.net/article/definitions-big-data; http://timoelliott.com/blog/2013/07/7-definitions-of-big-data-you-should-know-about.html respectively. Accessed 9/3/2015 9:53:42 AM.

[13] http://en.wikipedia.org/wiki/Big_data. Academic researchers will appreciate the amorphous nature of defining a methodology by reference to "typical database software tools" considering that according to such a definition, the commonplace use by empirical capital market researchers today of SAS to analyze COMPUSTAT data would be considered an example of Big Data by the standards of Ball and Brown (1968). This particular definition has the defect of encompassing almost any kind of data analysis more complex than basic regression and any data beyond the general ledger.

[14] The first three V′s are generally traced back to a META Group (now Gartner) 2001 blog (Laney 2001), which actually pre-dates the common use of the term Big Data.

[15] http://hortonworks.com/blog/7-key-drivers-for-the-big-data-market/. An alternative characterization of these three types of data are Process-Mediated Data, Human-Sourced Information, and Machine-Generated Data.

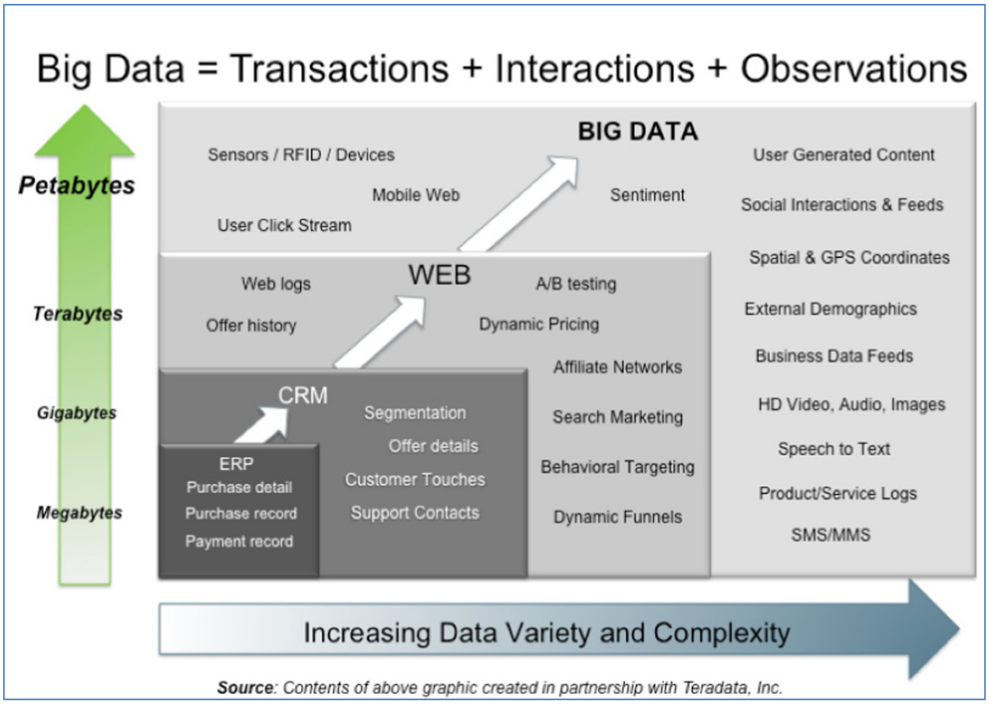

Fig. 3. Connolly's (2012) definition of Big Data.

These drivers are general and are not specific to auditing; however, if Big Data is to impact auditing more directly, then it will be through the first two drivers (new models and new insights). How precisely will Big Data do all this? Lucas (2012) has an insightful characterization of Big Data in which he argues that it: "*divides the world by intent and timing rather than the type of data. The 'old world' is about transactions, and by the time these transactions are recorded, it's too late to do anything about them: companies are constantly 'managing out of the rear-view mirror'. In the 'new world', companies can instead use new 'signal' data to anticipate what's going to happen, and intervene to improve the situation. Examples include tracking brand sentiment on social media (if your 'likes' fall off a cliff, your sales will surely follow) and predictive maintenance (complex algorithms determine when you need to replace an aircraft part, before the plane gets expensively stuck on the runway).*"[16]

The claim of "managing out of the rear-view mirror" is one often made against accounting information in general, and auditors in particular. As Lucas (2012, emphasis in original) goes on to say: "***The signals of health in a business we have been trained to look for haven't changed for a century or more, yet the amount of information available today that could indicate the relative health of a business is radically different.*** *It's sad to say, but Ebenezer Scrooge would be as adept in a board meeting today as the day that Charles Dickens conjured him up in 1843!*"

It is worth quoting at length from Lucas (2012) because he is a Big Data definer who makes his arguments in explicit accounting terms. Hence, he makes his case that "every [accounting] firm needs a real time data platform" by stating (emphasis in original):

> "*The Income Statement and Balance Sheet are at best rear-view measures of the top line and bottom line. They provide a snapshot in time of all that has* ***happened****, but very little, if any, indication of what is* ***happening*** *in the enterprise. For example, an online retailer with a subscription model experiences a massive drop in stock price because of poor Income Statement results. Further analysis indicates that there was a dramatic drop in subscribers (Churn) in one of their most profitable segments. It would stand to reason that a pre-emptive pulse-check on the churn could have helped stem the bleeding and perhaps prevented the reaction on Wall Street. This churn is an example of a new signal that could have helped this one enterprise run its business more proactively rather than look for explanations with a rear-view perspective. So how does a company gain the ability to know midstream that it has a problem or there is an opportunity going unaddressed? How will a company gain the intelligent actionable insights to make this happen? What is needed is the ability to ask complex questions going across the volume and variety of data in an interactive manner, and most importantly in real-time.*"

But as convincing as this argument may seem to be about the long criticized shortcomings of traditional accounting statements, Lucas is talking about accounting and not about auditing. Thus, while marketing managers in a business should certainly pay attention to the

[16] http://www.saphana.com/community/blogs/blog/2012/08/21/beyond-the-balance-sheet-run-your-business-on-new-signals-in-the-age-of-big-data.

Please cite this article as: Alles, M., Gray, G.L., Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors, Int J Account Inf Syst (2016), http://dx.doi.org/10.1016/j.accinf.2016.07.004

churn rate of subscribers, should auditors? Just because Big Data would enable the mangers of an "*enterprise [to] run its business more proactively rather than look for explanations with a rear-view perspective*" does not necessarily imply that auditors also have to adopt such a "proactive" perspective, and they are unlikely to want to do so in the absence of a rigorous analysis of what Big Data brings to auditing.

2.2. The context for Big Data

Defining Big Data is an incomplete exercise unless that definition is put into context. The emergence of Big Data does not mean that companies are necessarily creating new data elements (although they may be doing that as part of their Big Data activities). Big Data is assembling many different datasets together in either a data warehouse or data lake to allow analysis of dissimilar data to, in turn, discover new patterns, relationships, and correlations in the data. While it's possible to do data analysis across separate isolated databases using techniques such as open database connectivity (ODBC), this approach would be extremely inefficient across separate massive databases. Besides bringing dissimilar datasets together in one data environment, the incoming data may go through various extract, transform, and load (ETL) functions to make subsequent data analytics more efficient.[17] Hence, while Big Data certainly has its own unique characteristics it can, nonetheless, be also framed as an evolutionary phenomenon in data storage and data analysis and that evolution can take place at different rates in different companies.

The development of enterprise resource planning (ERP) systems is a major step in that evolutionary process. In the pre-ERP environments, companies were characterized as having numerous islands of technology (silos) with isolated databases specific to isolated applications that do not share data with each other. These pre-ERP databases were a mix of internally developed and proprietary acquired databases, mixed standards for naming and creating data elements (fields) and tables, and a mix of database structures. ERP software integrated the content of many (but not all) of these isolated databases into one large integrated database shared by several applications. As storage costs continued to significantly decrease, massive data warehouses beyond ERP databases became practical. To deal with these warehouses and new types of data, data mining algorithms and techniques became more sophisticated and efficient. The convergence of these and other technological advancements naturally evolved into what we now call Big Data.

Whatever the characteristics of Big Data are, that is not what defines its value to auditors or any other users. The data → information → knowledge → wisdom hierarchy literature (e.g., see Rowley, 2007 and Zins, 2007; or, in an accounting context, Elliott 1998) can be drawn on to make the point that Big Data is indeed just data. In other words, Big Data—whether considered as an evolutionary or revolutionary development in technology—remains a means towards an end and not, as the hype sometimes expresses it, as an end in itself. If auditors are to find value in it, Big Data and related analytics must lead to more efficient, effective and/or higher-quality audits.

2.3. Extraordinary claims require extraordinary evidence

In this regard, it is worth recalling the old adage that "extraordinary claims require extraordinary evidence".[18] The context for Big Data to be of value to auditors is that the corollary also holds: **extraordinary use of data requires extraordinary questions that need answering**. What has yet to be articulated is whether auditing contains such extraordinary questions that cannot be answered by fuller use of already available transactional data as well as traditional audit procedures to analyze that data. In other words, are there strong benefits that will either push or pull auditors out of Cell A and into Cells C and D in Fig. 1? Only when the limits of traditional data have been reached are auditors likely to turn to other types of Big Data. Moreover, as Zikopoulos et al. (2013) point out above, what is special about Big Data is the relatively easier ability today to analyze large data sets, not the availability of those data sets. For example, in Fig. 3, one component of Connolly's conception of Big Data is RFID chips, but these have been in widespread use for at least a decade. Hence, auditors need not have waited for the emergence of Big Data as a concept to have started to use RFID in their audit procedures, but there is little evidence that they have chosen to do so.

3. Potential advantages of incorporating Big Data into audit practice

Table 1 summaries the advantages associated with using Big Data as a part of audits. A significant way Big Data analytics adds value is by discovering patterns (e.g., unexpected correlations) that are not discernible in limited smaller data sets (such as typical audit samples). As Cukier and Mayer-Schoenberger (2013) state: "*It is tempting to understand Big Data solely in terms of size. But that would be misleading. Big Data is also characterized by the ability to render into data many aspects of the world that have never been quantified before; call it "datafication." For example, location has been datafied, first with the invention of longitude and latitude, and more recently with GPS satellite systems. Words are treated as data when computers mine centuries' worth of books. Even friendships and "likes" are datafied, via Facebook.*"

[17] Depending on the infrastructure, alternatively, the order could be extract, load, and transform (ELT).

[18] Marcello Truzzi called "the skeptic's skeptic" and investigator of protosciences and pseudosciences is credited with originating the phrase "Extraordinary claims require extraordinary proof" and, later, Carl Sagan popularized "Extraordinary claims require extraordinary evidence". (https://en.wikipedia.org/wiki/Marcello_Truzzi)

Please cite this article as: Alles, M., Gray, G.L., Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors, Int J Account Inf Syst (2016), http://dx.doi.org/10.1016/j.accinf.2016.07.004

Table 1
Potential advantages of incorporating Big Data into audits.

Potential advantages	Comments
Strong predictive power, which is a powerful tool for setting expectations for financial statement auditor.	Events/transactions included in Big Data can predate accounting transactions by days, weeks, months, and even years. For example, knowing someone is pregnant (by analyzing sales of pregnancy test kits) enables prediction of a permanent change in purchase patterns.
Rich data sources to identify potential fraudulent activities.	Difficult for fraudster to change all upstream non-financial transactions to cover up financial statement fraud. For example, trade based money laundering can be detected by comparing invoices with the actual weight of shipping containers. Some potentially valuable data may be unavailable due to privacy considerations.
Analyzing all data increases probability of discovering red flags, "smoking guns," and suspicious outliers.	Fraud represents a very small percentage of transactions and could be easily not included in the small samples auditors traditionally select.
Developing more predictive models of going concern, using leading indicators of sales and costs.	Going concern issues are not a major part of the audit of most well established businesses in normal circumstances and there is a limit to the resources that auditors will devote to its estimation.

Zikopoulos et al. (2013) provide the following example of structured Facebook data in the JavaScript Object Notation (JSON) format showing the large amount of business related information that is obtainable from even a small example of Big Data:

{"data":[
{"id":"53423432999_23423423_19898799",
"from":{"name":"PaulZikopoulos","id":"Z12"},
"message": "Thinking of surprising my wife with a quality time gift that lets her know
she's special, any ideas? I thought about taking her to the driving range, perhaps play a
round and caddie my game.", "created_time":"2012-08-02T21:27:44+0000","likes:5,"

"comments":{"data":[{11 id11 : 2847923942_723423423", "from":
{"name":"MaryAnneInfanti","id":"948574763"}.
"message": "Paul! Purses and gold! Costco's got a great Kate Spade purse
on sale this week that says I love you without having to lift a pen.
If you go with your idea, the only thing driving will be you: alone!"• "created_time
2012-00-02Tll:27:44+0000", "likes:64 } }

It is this kind of detailed information, not just about preferences for Kate Spade purses, but also contextual information that the husband plays golf and that the family has a Costco membership that leads people to say that there is "gold in that data".[19]

Most of the data in Big Data can be viewed as leading indicators because they precede accounting transactions from a few days to one or more years. Some of this other data is tightly coupled to future accounting transactions. For example, companies like Boeing and Airbus have backorders for aircraft that will be delivered many years in the future, and hence, that order data is highly correlated with future revenues and costs. Other data may be more loosely coupled to accounting transactions, but still informative. For example, a trend of negative comments in Twitter could signal a future decline of revenue. Big Data thus has the potential to be a powerful means of setting and adjusting auditor expectations at the beginning (planning phase) and throughout the audit.

Because much of Big Data contents can be separated both physically and conceptually from accounting data, fraudsters are not likely to be able to manipulate all applicable Big Data elements to disguise their fraud. Because of the substantial contents of Big Data and the concept of "100% sampling" (analyzing the whole population), the discovery of "red flags" and other suspicious outliers is greatly increased. There are many other factors involved, but Big Data analytics—particularly with its emphasis on NFD—clearly had the *potential* to detect historical examples of massive frauds.

Indeed, EY, 2014 argues that an obvious use of Big Data is fraud detection. Along those lines, in Brazil it is commonplace for employers to require that employees use a given bank in order that the employer can then check that account for evidence of corrupt practices, such as expenditures in excess of salary. Such practices would never be allowed in most Western nations due to privacy concerns, but this example is indicative of the value added that non-traditional data can provide in a well-defined prediction problem.

Given the criticism auditors faced for not issuing going concern opinions for businesses that failed in the recent financial crisis such as Bear Stearns, improving auditor assessment of client risk may be another promising application of Big Data. On the other hand, it has to be kept in mind that this represents only a small part of the audit engagement and that the number of going concern opinions issued has declined lately to a small minority of all audit clients examined.[20] Nonetheless, this may be the most immediately promising area for research into incorporating Big Data techniques into audit practice.

[19] http://data-informed.com/understand-your-business-with-behavioral-data/. Accessed 9/2/2015 5:20:44 PM.
[20] http://www.complianceweek.com/going-concern-warnings-drop-to-12-year-low/article/273186/.

Table 2
Inhibitors associated with incorporating Big Data into audits.

Inhibitors	Comments	Amelioration
Auditors will need direct access to client's Big Data	Accounting firms' legal department discourage direct access because of fears of corrupting data and/or negatively impacting performance of the client's system by writing poorly-designed queries. Client concerned about auditors having unlimited access to propriety data.	Start small. The incorporation of Big Data in financial statement can be an evolutionary (not revolutionary) process. Part of the pre-audit will be to identify NFD that could be valuable based on the pre-audit risk analysis and brainstorming session.
Could "miss" smoking gun in 100% population.	Accounting firms' legal department discourage 100% "samples" because the firm has no defense as to why they missed the smoking gun. Current defense: smoking gun was not in the small sample analyzed during the audit.	Not increasing the population is going to be increasingly harder to defend as Big Data, user-friendly analytical software, and inexpensive computer resources become more available. Firms are going to be driven toward improving their analytical skills.
Requesting specific data send signals to client regarding areas of concern to the auditors.		Pre-requesting broad varieties of data at the start of the audit will reduce signaling.
NFD is messy	Could require a significant amount of time to validate the data to determine level of messiness (particularly in terms of accuracy and timing) and usability/applicability to the audit.	The Big 4, in particular, have numerous clients to which they provide internal audit services. Because internal audit performs broader activities and has freer access to internal Big Data, accounting firms can use these opportunities to hone their skills to deal with messy NFD.
Results can be messy (e.g., numerous false positives and spurious relations/patterns)		Improving queries (in terms of variables included/excluded in the queries) can greatly reduce false positives. False positives are related to moving along the learning curve.
Skills needed to select and use appropriate Big Data analytics.	Requires skills beyond the usual skill set of Excel, Idea or ACL.	Working with Big Data vendors, consultants, and academic, the accounting firms can develop training materials to improve needed Big Data skills.
Skills needed on audit team to interpret correlations, patterns, and outliers.	Requires a higher level of business acumen and insights.	

4. Inhibitors of incorporating Big Data into audit practice

Section 3 provides an overview of the potential advantages of Big Data that builds on the multitude of positive articles regarding Big Data in accounting and auditing. Section 4 provides an overview of the many inhibitors or challenges that auditors will face as they expand their use of Big Data and advanced analytics. As the next subsection discusses, at a very pragmatic level, using Big Data as an integral part of audits is disruptive technology that would require a paradigm shift in terms of major operational changes on several dimensions. The remaining subsections take a closer look at the broader issues associated with integrating Big Data into external financial statement audits.

4.1. A paradigm shift

Even with significant potential benefits of Big Data in financial statement audits, the audit practice community (firms, regulators, and other stakeholders) will have to go through a paradigm shift to exploit those benefits. Table 2 presents the inhibitors that will have to be addressed before the benefits of Big Data can be fully realized. In broad terms, auditors will need unprecedented access to client data (which raises concerns for both the audit firms and their clients); an increase in the audit team's data analytical skills (which can have both short and long-term cost consequences); and an increase in the understanding of the relationships between the vast population of NFD elements and financial data (which means enhancing the business acumen of audit team members).

In the third column of Table 2, we present suggestions of how the inhibitors could be addressed to ameliorate the impact of each inhibitor. The suggested approaches to address the inhibitors fall into two broad themes. First, using Big Data is not an all or nothing choice. Auditors can start small (cherry picking) in terms of both breadth and depth. Auditors can develop their analytical skill by first using data similar to familiar accounting variables such as that included in order-entry and purchase order applications and then expand outward (illustrated by the arrow in Fig. 3) to incorporating data that is further and further removed from accounting data. The value of these suggestions can be explored in future research.

The second theme in the ameliorations is the rather pragmatic theme that Big Data will eventually be a normal part of audits (because of the changing client IT infrastructure environments) and it is better to be proactive than to be reactive in terms of incorporating Big Data into audits, as Alles (2015) also argues. The inhibitors in Table 2 are only the usual operational stumbling blocks that have to be overcome when implementing any new technology. The unique issues that Big Data will pose for its incorporation into auditing practice, its true "inhibitors", are what we discuss in the next subsections.

4.2. Big Data analysis and the nature of audit decision making

If Big Data is to establish a place in auditing, it also has to be a superior solution to current audit procedures. A key aspect of the audit that may limit the value of Big Data is that the ultimate decision variable is discrete rather than continuous: an unqualified audit opinion or not, with only a small minority of non-public clients receiving either a qualified or adverse opinion.

Please cite this article as: Alles, M., Gray, G.L., Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors, Int J Account Inf Syst (2016), http://dx.doi.org/10.1016/j.accinf.2016.07.004

Generally, with a continuous outcome variable, with each iteration of the data analysis, the analyst keeps trying to fine tune the analysis (different data and different models) to obtain better results than the prior iteration. For example, think of continuous quality techniques such as Six Sigma. Or think about medical or marketing research—constantly trying to get better and better results or better predictions. On the other hand, at the start of an audit the auditors set the targeted or acceptable audit risk (e.g., 5%). The whole audit is planned around achieving that level of audit risk. Once the audit partner is confident that the audit procedures have been completed and he/she is ready to give an unqualified opinion, the field work is declared done. Presumably, the audit achieved the acceptable level of audit risk. With each subsequent audit, the auditors are *not* trying to continuously lower the audit risk. Instead, they *may* be trying to improve the efficiency and effectiveness of (and confidence in) achieving the acceptable audit risk level.

If the auditor's conventional audit procedures yield an audit risk of 5% for a particular client, then why should the auditor now turn to Big Data unless the Big Data analysis is likely to more effectively and/or efficiently achieve that given audit risk threshold? As Keltanen (2013) states[21]: "*Diminishing returns still apply. Statements like* '*data is the new oil*' *make it sound like data is currency, when it's actually an investment. In all statistical measurements, once enough data points have been collected to establish a result, adding more data points begins to create less accuracy. This should be a pressing concern when you're investing increasing amounts of money, time and resources into capturing and analysing data*."

Of course not all decisions within an audit are dichotomous. There clearly is opportunity for Big Data to play a role in continuously refining fraud detection, for instance, or developing better analytical procedures. But these are means towards the end of an audit opinion and not ends in themselves, as auditors point out when they say that if fraud detection is the main objective then they should be hired to do a forensic audit rather than a regular one. As Norris (2011) put it: "*To the auditing industry, the fact that investors tend to blame auditors when frauds go undetected reflects unrealistic expectations, not bad work by the auditors. The rules say auditors are supposed to have a* '*healthy degree of skepticism,*' *but not to detect all frauds.* '*There is a significant expectations gap between what various stakeholders believe auditors do or should do in detecting fraud, and what audit networks are actually capable of doing, at the prices that companies or investors are willing to pay for audits,*' *stated a position paper issued in 2006 by the chief executives of the six largest audit networks.*

> *Note that last part. They suggested that if investors were really worried about fraud, they should consider paying more for a* '*forensic audit*' *that would have a better — but not guaranteed — chance of spotting fraud. Don't like our work? Pay us more.*"[22]

4.3. Implications of "datafication": Correlation over causation

As summarized by Gray and Debreceny (2014), traditional data mining approaches can be classified into two broad categories: directed (also top-down or hypotheses testing) and undirected (also bottom-up or hypotheses generating). With directed data mining, specific relationships with preselected variables are being tested (e.g., how has the relationship between revenue and cost of goods changed from year to year?). With undirected data mining, specific relations are not being tested; instead the analyst wants the software to discover relationships between populations of variables—and, particularly, to identify outliers to those newly discovered relationships. It is up to the data analyst (i.e., the auditor) to develop a hypothesis as to why there is a relationship between two variables and what do the outliers mean in that context. The phenomenon of "datafication" subsequently drives the way in which Big Data is analyzed, with pattern recognition or (undirected) hypothesis generating taking precedence over (directed) hypothesis testing. As Cukier and Mayer-Schoenberger (2013) go on to say:

> "*Using great volumes of information in this way requires three profound changes in how we approach data. The first is to collect and use a lot of data rather than settle for small amounts or samples, as statisticians have done for well over a century. The second is to shed our preference for highly curated and pristine data and instead accept messiness: in an increasing number of situations, a bit of inaccuracy can be tolerated, because the benefits of using vastly more data of variable quality outweigh the costs of using smaller amounts of very exact data. Third, in many instances, we will need to give up our quest to discover the cause of things, in return for accepting correlations. With Big Data, instead of trying to understand precisely why an engine breaks down or why a drug's side effect disappears, researchers can instead collect and analyze massive quantities of information about such events and everything that is associated with them, looking for patterns that might help predict future occurrences. Big Data helps answer what, not why, and often that's good enough*."

These three profound changes listed above—collecting more data, messy data, and correlation over causation—can be inhibitors for acceptance by auditors and we examine each in turn.

4.3.1. Collecting more data

While the cost of storing large amounts of data has steeply decreased, that only takes into account the direct cost of data storage. Another factor that would particularly concern auditors would be maintaining security of ever larger amounts of client-related data and ensuring that there are no breaches of confidentiality and privacy, most especially concerning information drawn from social networks and other interactions that come from outside the audit client—for example, customer complaints, proprietary information shared in a supply chain, customer or employee personal identification data and health care records.

[21] http://www.theguardian.com/media-network/media-network-blog/2013/apr/16/big-data-lean-strategy-business.

[22] http://www.nytimes.com/2011/06/10/business/10norris.html.

Please cite this article as: Alles, M., Gray, G.L., Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors, Int J Account Inf Syst (2016), http://dx.doi.org/10.1016/j.accinf.2016.07.004

4.3.2. Messy data

While the above data security concern may be addressed by rigorous data protection controls, the next change that Cukier and Mayer-Schoenberger (2013) postulate is necessary to succeed at Big Data poses a more challenging hurdle for auditors: "*In the past, when people collected only a little data, they often had to decide at the outset what to collect and how it would be used. Today, when we gather all the data, we do not need to know beforehand what we plan to use it for*." While this approach may make sense in the case of scientists or internet companies, it is hard to imagine that auditors would feel inclined to request/collect data without a clear reason to do so. Doing so also goes counter to the desire of auditors to be conservative and to be able to clearly justify any decisions that they make. More to the point, if data are to come from the client, then obtaining information at all, let alone with no explanation provided is difficult to attain. Indeed, the development of the Audit Data Standard (AICPA, 2013) indicates that auditors have difficulty obtaining even transactional data: "*One reason data standards are needed is to address the ongoing challenge that management and internal and external auditors face in the efficient exchange of a company's data. This process is complicated by the fact that accounting and IT personnel approach requests for such information from different perspectives. For example, in some cases, audit-related data requests are forwarded directly to a company's IT department, with limited further involvement from the accounting or finance department. In many cases, the burden is on the auditors to acquire the data.*"

The problems that auditors would have with this approach towards non-curated data gathering can be gauged by what Cukier and Mayer-Schoenberger (2013, emphasis added) say next: "*When we increase the scale by orders of magnitude, **we might have to give up on clean, carefully curated data and tolerate some messiness**. This idea runs counter to how people have tried to work with data for centuries. Yet the obsession with accuracy and precision is in some ways an artifact of an information constrained environment. When there was not that much data around, researchers had to make sure that the figures they bothered to collect were as exact as possible. Tapping vastly more data means that **we can now allow some inaccuracies to slip in** (provided the data set is not completely incorrect), in return for benefiting from the insights that a massive body of data provides.*"

While researchers or managers in search of new marketing strategies may tolerate such ambiguity, it is probable that auditors fall into that category of people for whom it runs counter to how they have worked with data for many years. Allowing "some inaccuracies to slip in" is difficult to reconcile with the focus in auditing on data integrity.[23] AICPA (2014) recognizes these inherent difficulties that Big Data (or "predictive analytics" as they call it) will pose for CPAs:

> "*This all sounds easy enough. But, in practice, it is difficult to build and validate a model. Some reasons for this difficulty include:*

1. ***It often is difficult to understand causality.*** *Does a fast car cause a driver to exceed the speed limit?*
2. ***Data quality issues affect model results.*** *One source rounded dollars up to millions but the other used standard rounding.*
3. ***It often is difficult to obtain the data that you want.*** *You did not collect historical data from one of the locations in your model.*
4. ***The real world is the lab — you can't always be flexible with experiment design.*** *Would there have been fewer audit findings if you had used a different sampling system?*
5. ***Independent variables affect each other.*** *The type of customer who is complaining behaves differently during price increases.*
6. ***The relationships between variables often are not linear.*** *How much would a client care about a price change of 1%, 5% or 500%?*
7. ***Relationships are not necessarily stable over time.*** *Who is watching the same advertisement at 2pm and 6pm?*"

4.3.3. From causation to correlation

The emphasis on quantity rather than quality in Big Data leads to the third major change identified by Cukier and Mayer-Schoenberger (2013, emphasis added): "*These two shifts in how we think about data—from some to all and from clean to messy—give rise to a third change: **from causation to correlation**. This represents a move away from always trying to understand the deeper reasons behind how the world works to **simply learning about an association among phenomena and using that to get things done**.*"

This idea that data is sufficient in itself is a persistent theme in the Big Data literature. For example, Anderson (2008) in an article with the telling title "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete" argues that "*Correlation is enough. We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot*." The approach of using correlation from Big Data to draw inferences has had considerable success. For example, Ginsberg et al. (2009) reported on the use of Google searches for flu related terms to plot incidences of influenza. As the authors state: "*Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms, we can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day*." The methods of this study have now been made publicly available through the Google Flu Trends site.[24]

Whatever the successes of this approach, however, it is certainly startling to researchers who had the adage "correlation is not causation" drummed into them as doctoral students. And auditors too would be skeptical about basing conclusions on correlations that cannot be proven to be based on any kind of causality—for example, if the decisions based on those Big Data results were ever to be subject of subsequent litigation.[25]

[23] For example, the Audit Data Standard includes a questionnaire designed to "help auditors better understand the data and assess its completeness and integrity." http://www.aicpa.org/interestareas/frc/assuranceadvisoryservices/pages/auditdatastandardworkinggroup.aspx
[24] http://www.google.org/flutrends/us/#US.
[25] As one practitioner referee of an earlier version of this paper stated in this regard: "*Definitely true, how can you justify a conclusion based on a correlation that may or may not exist in the data. The judges and lawyers could have a difficult time with this. Also if all correlations had to be proven 100%, the efficiencies could easily be lost.*"

Please cite this article as: Alles, M., Gray, G.L., Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors, Int J Account Inf Syst (2016), http://dx.doi.org/10.1016/j.accinf.2016.07.004

The difference between managers using Big Data to create a new marketing strategies or doctors tracking disease is that while they are using correlation to drive future action, auditors use information to create controls. Thus, if Amazon finds a correlation between college-educated women under 30 who do yoga, purchase Lululemon clothing and eat organic food, then they will tailor a marketing strategy accordingly. Of course, it is only a hypothesis and the strategy is really an experiment, but if the correlation proves valid then more sales are achieved, while if it turns out to be spurious, then Amazon would move on to the next experiment.

By contrast, if auditors using Big Data find a correlation between newly divorced employees and overbilling of expenses, what precisely are they supposed to do with that information? Recommend that divorcing employees be subject to greater scrutiny or be fired? It is one thing for auditors to use correlation information to suggest new areas to investigate further, but quite another to institute formal controls based on that information. Verifiability and credibility are vital in auditing if action against particular employees is to be initiated, or even when broad controls are instituted because, unlike with a marketing strategy, actions by auditors may have to be defended in litigation. Of course, this is not true for all audit findings and it is not to say that Big Data has no role to play in auditing. But the "correlation not causation" mantra of Big Data poses a bigger hurdle in auditing than it would in the management side of the business.

4.4. The role of theory in interpreting Big Data results

It is worth noting that there is a backlash against the correlation over causation approach in the Big Data literature itself. Keltanen (2013) writes: "*…the lightest, simplest way to achieve your data analysis goals is the best one… The dirty secret of Big Data is that no algorithm can tell you what's significant, or what it means. Data then becomes another problem for you to solve. A lean data approach suggests starting with questions relevant to your business and finding ways to answer them through data, rather than sifting through countless data sets. Furthermore, purely algorithmic extraction of rules from data is prone to creating spurious connections, such as false correlations… today's Big Data hype seems more concerned with indiscriminate hoarding than helping businesses make the right decisions*."

David Brooks (2013) writing in the *New York Times* first describes the correlation focused case for Big Data and then discusses what Big Data can and cannot do[26]: "*The theory of Big Data is to have no theory, at least about human nature. You just gather huge amounts of information, observe the patterns and estimate probabilities about how people will act in the future… As Viktor Mayer-Schönberger and Kenneth Cukier write in their book, 'Big Data,'[27] this movement asks us to move from causation to correlation. People using Big Data are not like novelists, ministers, psychologists, memoirists or gossips, coming up with intuitive narratives to explain the causal chains of why things are happening. 'Contrary to conventional wisdom, such human intuiting of causality does not deepen our understanding of the world,' they write. Instead, they aim to stand back nonjudgmentally and observe linkages: 'Correlations are powerful not only because they offer insights, but also because the insights they offer are relatively clear. These insights often get obscured when we bring causality back into the picture…' One limit [of Big Data] is that correlations are actually not all that clear. A zillion things can correlate with each other, depending on how you structure the data and what you compare. To discern meaningful correlations from meaningless ones, you often have to rely on some causal hypothesis about what is leading to what. You wind up back in the land of human theorizing*."

A debate that auditors need to pay attention to in this regard is that taking place in the Big Data community about the role of theory in interpreting Big Data results. Crawford (2013) writes[28]: "*Can numbers actually speak for themselves? Sadly, they can't. Data and data sets are not objective; they are creations of human design. We give numbers their voice, draw inferences from them, and define their meaning through our interpretations. Hidden biases in both the collection and analysis stages present considerable risks, and are as important to the big-data equation as the numbers themselves. For example, consider the Twitter data generated by Hurricane Sandy, more than 20 million tweets between October 27 and November 1. A fascinating study combining Sandy-related Twitter and Foursquare data produced some expected findings (grocery shopping peaks the night before the storm) and some surprising ones (nightlife picked up the day after — presumably when cabin fever strikes). But these data don't represent the whole picture. The greatest number of tweets about Sandy came from Manhattan. This makes sense given the city's high level of smartphone ownership and Twitter use, but it creates the illusion that Manhattan was the hub of the disaster*."

The economist Susan Athey in an interview with *The Region* journal said the following (Clement, 2013)[29]: "*Region: What are your thoughts about Big Data? Does it portend, as some have suggested, an 'end to theory'*?

> *Athey: Absolutely not an end to theory. In fact, the need for theory is in some ways magnified by having large amounts of data. When you have a small amount of data, you can just look at the data and build your intuition from it. When you have very large amounts of data, just taking an average can cost thousands of dollars of computer time. So you'd better have an idea of what you're doing and why before you go out to take those averages. The importance of theory to create conceptual frameworks to know what to look for has never been larger, I think.*

> *Region: And yet some have argued that because data exist in increasingly large quantities, all you really need is to 'see what the data say.'*

[26] http://www.nytimes.com/2013/04/16/opinion/brooks-what-youll-do-next.html?hp.
[27] Mayer-Schoenberger and Cukier (2013).
[28] http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/.
[29] http://www.minneapolisfed.org/publications_papers/pub_display.cfm?id=5112.

Please cite this article as: Alles, M., Gray, G.L., Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors, Int J Account Inf Syst (2016), http://dx.doi.org/10.1016/j.accinf.2016.07.004

> *Athey: I think what is true is that when you have large amounts of data, if you ask it the right questions, you have a greater ability to let the data speak, and so you can be much less reliant on assumptions. But you still need a strong conceptual framework to understand what's coming out*."

While one may question whether taking an average can cost thousands of dollars of computer time even with very large data sets, the key takeaway from both these citations is that the correlation not causation school of Big Data is not the only perspective and that making use of Big Data results may require a proper understanding of the context from which those results are obtained. As Crawford (2013) argues: "*We get a much richer sense of the world when we ask people the why and the how not just the 'how many'. This goes beyond merely conducting focus groups to confirm what you already want to see in a Big Data set. It means complementing data sources with rigorous qualitative research. Social science methodologies may make the challenge of understanding Big Data more complex, but they also bring context-awareness to our research to address serious signal problems. Then we can move from the focus on merely 'big' data towards something more three-dimensional: data with depth*."

A theory-driven approach may prove to be a better fit in the application of Big Data in auditing. For one thing, developing a theory to explain Big Data results found in one client engagement would facilitate the transfer of that learning to other clients. Moreover, auditors unlike other Big Data users may feel more comfortable specifying to at least some extent what they are looking for (a directed, hypothesis testing approach) rather than leaving it all to the data to speak (an undirected, hypothesis generating approach). Keltanen (2013) argued that it is more efficient to start with relevant questions and find ways to answer them through data, rather than sifting through countless data sets in the hope of finding something useful. But it may also fit better with the conservative culture of auditing, with "conservative" not meant as a pejorative but in due recognition of the greater societal and economic consequences of audit decisions.

Indeed, Crawford (2013) offers a corrective to the previously cited Big Data success story of Google Flu Trends: "*But as we increasingly rely on Big Data's numbers to speak for themselves, we risk misunderstanding the results and in turn misallocating important public resources. This could well have been the case had public health officials relied exclusively on Google Flu Trends, which mistakenly estimated that peak flu levels reached 11% of the US public this flu season, almost double the CDC's estimate of about 6%. While Google will not comment on the reason for the overestimation, it seems likely that it was caused by the extensive media coverage of the flu season, creating a spike in search queries*."

A recent *New York Times* (2014) article provides auditors with insights both into the potential and the pitfalls of using Big Data.[30] The article points out that Big Data analytics enables the search of millions of patient records for diagnoses and treatments, something which is particularly helpful in the case of rare conditions for which it is difficult to make the economic case for a targeted research program: "*…large, costly and time-consuming clinical trials are rarely carried out for uncommon complications of this sort. In the absence of such focused research, doctors and scientists are increasingly dipping into enormous troves of data that already exist — namely the aggregated medical records of thousands or even millions of patients to uncover patterns that might help steer care*." An important question that auditors will have to ask is whether the situations they face is more akin to treatment of a rare condition or whether audit engagements share broadly similar issues that can be better investigated more systematically than through Big Data techniques. Certainly the article makes the point that Big Data is not the ultimate analytic method that supersedes all others (such as formal clinical trials) but one that adds value in some circumstances and not in others.

This is particularly the case when one considers the drawbacks of Big Data, for the *New York Times* the article goes on to make some of the caveats that we have discussed above, such as the problems of "less than pristine data": "*In the lab, ensuring that the data-mining conclusions hold water can also be tricky. By definition, a medical-records database contains information only on sick people who sought help, so it is inherently incomplete. Also, they lack the controls of a clinical study and are full of other confounding factors that might trip up unwary researchers. Daniel Rubin, a professor of bioinformatics at Stanford, also warns that there have been no studies of data-driven medicine to determine whether it leads to positive outcomes more often than not. Because historical evidence is of 'inferior quality,' he says, it has the potential to lead care astray*."

As with the prior section on pros of Big Data, this discussion is not an exhaustive examination of its cons relative to auditing practice, and nor is it meant to suggest that there is no possibility of reconciling the two. Rather, this discussion is meant as a corrective to the presumption that because Big Data is being embraced by businesses implies that it will be facile to apply it to auditing too. As before, more research is needed on this topic, as we discuss next.

5. Research opportunities in the application of Big Data to auditing

The inhibitors of the implementation of Big Data in auditing arise from the contrast between the unique characteristics of auditing versus those of Big Data, which suggests the need for further research on better understanding those inhibitors and how they can be best overcome. To begin that process, we pose four broad and interrelated questions about the degree and speed at which Big Data will be incorporated into financial statement audits:

1. What are the unique characteristics of Big Data (vs. traditional accounting data) that have the greatest potential to improve the efficiency, effectiveness, deterrent ability, confidence and/or quality of financial statement audits?
2. What are the costs (software, training, and staffing) of incorporating Big Data into financial statement audit practice?
3. What are the technical, business acumen, and cognitive skill sets auditors will need in a Big Data environment?

[30] http://www.nytimes.com/2014/10/05/magazine/can-big-data-tell-us-what-clinical-trials-dont.html?ref=magazine.

Please cite this article as: Alles, M., Gray, G.L., Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors, Int J Account Inf Syst (2016), http://dx.doi.org/10.1016/j.accinf.2016.07.004

4. What would be the optimum approach to begin incorporating Big Data into financial statement audits, particularly, placing specific Big Data elements and data analytics into specific phases of the audit process?

Based on our review of the literature and the comments made by practitioners at the 2014 and 2015 AAA Annual Meetings and 2015 and 2016 Audit and AIS Midyear Meetings, accounting firms are in the very early and limited stages of actually incorporating Big Data activities into their current financial statement audits. Because of the newness of these activities at the accounting firms there is likely to be a lack of quantitative data in the immediate future to conduct empirical research. Hence, most of the Big Data research that can be conducted in the near term will be some mix of design science, interviews, case studies, surveys, and other qualitative research, such as focus groups. A variety of research approaches will have to be employed in order to obtain insights and perspectives to answer the questions posed above. The discussion below divides potential subjects into two broad categories: subjects outside of external auditors and subjects who are external auditors. Companies, in general, as well as internal auditors are ahead of external auditors in their use of Big Data and, as such, can potentially provide rich research environments to help predict the answers to the above four overarching research questions in the external financial audit domain. On the other hand, external auditors can better identify the inhibitors that will have a negative impact on the adoption of Big Data and advanced analytics.

5.1. Research outside of the external audit domain

There is a vast and growing body of literature, both academic and nonacademic, that discusses how various industries are using Big Data as well as benefits, costs, issues, and concerns of using Big Data. Much of the Big Data activities discussed in the existing business literature are in the marketing domain, but those activities should not be discounted in the audit domain. Instead, researchers should identify parallels to auditing practice for insights on how that knowledge can be transferred from one domain to the other and then analyze how these activities could be modified for use in auditing practice. For example, Big Data marketing research is searching for patterns in human behavior that are correlated with purchase practices and to subsequently predict future purchasing decisions. For example, LinkedIn Corp. (LinkedIn.com) uses Big Data to determine the "People You May Know" lists on their Web site (Dwoskin, 2014). Auditing might take a related, but opposite perspective in terms of using Big Data to identify those who do *not* seem to follow predicted behavior in order to isolate outliers. This is already being done in the data mining of emails where various natural grouping and cliques are recognized in order to subsequently identify those who seem to be in a clique that they don't logically belong to (Debreceny and Gray, 2011). For example, why is the CFO sending emails to a warehouse worker? Interviews and case studies would be particularly valuable here.

Although general corporate Big Data activities may appear to differ from financial statement auditing, on closer examination there are parallels, particularly to the auditor's assessment of the effectiveness of the client's internal controls over financial reporting (ICFR) process. Besides marketing activities, corporate Big Data activities include analyzing compliance with corporate policies and procedures (identifying employees who are not following corporate procedures and policies) and improving the efficiency of logistics by systematically analyzing routing and warehousing activities—for example, by using GPS coordinates to determine which employees are not following prescribed routes (or visiting prohibited locations, such as bars). These and other related activities are essentially control monitoring/evaluation activities that could inspire ICFR-related audit procedures.

5.1.1. Research internal auditor Big Data activities

Compared to the Big Data uses described above in the scientific and marketing domains, the characteristics of the internal auditor domain are much closer to the external auditor domain. However, there are two characteristics different about the internal auditor domain that will be very valuable to researchers. First, internal auditors can potentially access their organization's Big Data more easily and more fully than external auditors. Both the Institute of Internal Auditors (IIA) and ISACA have published articles and other publications regarding Big Data and internal auditing. Many parallels can be drawn from these publications and be applied to external auditors. These articles and publications might also help identify specific internal auditors that could be interviewed and organizations that could provide a basis for case studies. Both the IIA and ISACA have been very helpful in working with academics and could be approached about helping organize focus groups and inviting members to complete online surveys.

Second, the internal auditors' use of Big Data would generally be broader than that of external auditors. For example, internal auditors might use a broad selection of NFD as part of an operational audit (as opposed to a financial audit). The experiences of internal auditors using NFD could provide valuable guidance for the external auditors' use of NFD.

5.1.2. Accounting firms experiences with Big Data outside of financial statement auditing

Accounting firms provide Big Data consulting services and their forensic groups are expanding their use of Big Data for their analytical activities. Through interviews, case studies, focus groups, and other qualitative research methods, valuable information could be gleaned from these non-audit Big Data activities that could answer the four broad questions posed above. Of particular interest might be the cost-benefit parameters that could be derived from those non-audit activities. It's clear that as more Big Data analytics are incorporated into financial statement auditing there will have to be a change in the mix of technical skills on the audit team. As we mentioned earlier, there is a growing general shortage of Big Data analysts and, as such, these analysts are paid premium salaries. So, even if Big Data analytics could improve the efficiency and effectiveness of the audit, a critical research question is going to be whether this improvement in efficiency and effectiveness can justify the cost of the people needed to conduct the Big Data analytics. The kinds of skills these new Big Data auditors will need to have, both technical and cognitive, and how the accounting curriculum will have to change to teach them are all matters that warrant further research.

Please cite this article as: Alles, M., Gray, G.L., Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors, Int J Account Inf Syst (2016), http://dx.doi.org/10.1016/j.accinf.2016.07.004

5.2. Gaining insights directly from external auditors

The incorporation of Big Data into financial statement audits is probably more likely to succeed if it is seen as an evolution in audit practice driven by easier access to a wider set of data rather than as a technology-driven revolution in analytics. Nearly every auditor has experience using client data and, maybe, external data to perform analytical procedures required for every audit. A growing body of auditors is expanding the data and the data analysis and data mining tools they use as part of an audit. As such, the experiences and opinions regarding drivers and inhibitors of current auditors can provide insights into the Big Data evolutionary process. These auditors can particularly address the paradigm-shifting aspects of Big Data, such as, moving away from sampling, the changing mix of skills needed on audit teams, and the increasing requirement of direct access to client data (particular NFD). Although the other groups of the above subjects could have direct or indirect experience or opinions, the following questions could be particularly addressed by experienced external auditors:

a) Given the hierarchy of data in the Connolly (2012) framework (Fig. 2), what is the marginal value to auditing of going beyond transactional data? Are higher levels of data substitutes or complements for traditional data inputs into the audit process?
b) What unique aspects of Big Data are likely to have the biggest impact (positive or negative) on audit practice?
c) To what extent is data integrity central to auditing and how can data integrity be addressed when dealing with "messier" non-financial and non-transactional data?
d) Is unguided information collection appropriate or feasible in the auditing context?
e) Is the shift from causation to correlation inescapable in auditing and can audit practice evolve to deal with that shift? Can auditors justify basing decisions on patterns observed in data for which they cannot offer an underlying theory?
f) Is it the case that Big Data requires stability in the environment and does that accurately describe a typical audit setting? When there is dynamic change in the environment, how should both Big Data and auditing respond?

6. Conclusion

Big Data is becoming an indispensable resource to many organizations and has the potential to be an extremely valuable resource to financial statement auditors. But that presumption must not be taken as a given without question, skepticism, and further research. Otherwise, there is the danger that Big Data will succumb to the same forces that have stalled the adoption of prior technologies equally extolled in their time as having the potential to transform auditing practice.

We conclude with an example from our own experience that illustrates the importance of gaining a full understanding of what Big Data is and what its pros and cons are before applying it to auditing. Earlier in the paper we cited Crawford (2013) who discusses a flawed Big Data analysis of Twitter data to assess the impact of super storm Sandy. In that case the prevalence of Twitter users in Manhattan gave the false impression that it was there that the most damage took place when in reality it was New Jersey that took the brunt of the storm. The point he and we were making was that in the absence of theory as to what Big Data analytics should be performed and on what data, it is all too easy to fall into simplistic and erroneous interpretations of spurious or misleading correlations.

The above is an example of a false positive. The researcher made the mistake of not seeking separate confirming/disconfirming data such as weather data that is available in real time on the internet. Auditors will have to be particularly careful not to be misdirected by what later turns out to be false positives. This is a particular concern because of the correlation/causation discussion early. It is easy to fall into the trap of giving too much weight to a result that already supports your beliefs. In other words if the auditor can quickly come up with an explanation for an outlier, they are not likely to search for disconfirming data.

There is little recognition that Big Data is only a means towards an end—not a magic bullet—and how well it works depends on the choices made by the analyst (auditor) about what data to include and how that data is analyzed. And the failure to grasp that fact can make even highly capable statisticians make fundamental mistakes, which should raise doubts about what will happen when auditors in the field are urged to develop Big Data queries continuously.[31]

References

AICPA, 2012. The current state of continuous auditing and continuous monitoring. White Paper October http://www.aicpa.org/interestareas/frc/assuranceadvisoryservices/downloadabledocuments/whitepaper_current-state-continuous-auditing-monitoring.pdf. Last accessed 2/22/2016 11:17:34 AM.

AICPA, 2013. Audit Data Standard. Audit Data Standard Working Group http://www.aicpa.org/InterestAreas/FRC/AssuranceAdvisoryServices/Pages/AuditDataStandardWorkingGroup.aspx.

AICPA, 2014. Why predictive analytics should be "a CPA thing". http://www.aicpa.org/InterestAreas/InformationTechnology/Resources/BusinessIntelligence/DownloadableDocuments/Predictive_Analytics_Final.pdf.

Alles, M., 2015. Drivers of the use and facilitators of the evolution of Big Data by the audit profession. Account. Horiz. 29 (2), 439–449.

Anderson, C., 2008. The end of theory: the data deluge makes the scientific method obsolete. Wired Mag. July 16. http://www.wired.com/science/discoveries/magazine/16-07/pb_theory.

Ball, R., Brown, P., 1968. An empirical evaluation of accounting income numbers. J. Account. Res. 6 (2), 159–178.

Brooks, D., 2013. "What you'll do next,". New York Times A27 April, 16, 2013.

Brown, C.E., 1991. Expert systems in public accounting: current practice and future directions. Expert Syst. Appl. 3 (1), 3–18.

Cao, M., Chychyla, R., Stewart, T., 2015. Big Data analytics in financial statement audits. Account. Horiz. 29 (2), 423–429.

[31] This is not idle speculation. Speaking at the EY reception at the 2015 American Accounting Association annual meeting the head of the firm's Tax practice stated that he expected each of this associates to come up with a question that required Big Data analysis every single day.

Please cite this article as: Alles, M., Gray, G.L., Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors, Int J Account Inf Syst (2016), http://dx.doi.org/10.1016/j.accinf.2016.07.004

Clement, D., 2013. Interview with Susan Athey. The Region: The Federal Reserve Bank of Minneapolis June 10. http://www.minneapolisfed.org/publications_papers/pub_display.cfm?id=5112.

Connolly, S., 2012. 7 key drivers for the Big Data market. Blog Posting May 14th. http://hortonworks.com/blog/7-key-drivers-for-the-big-data-market/.

Crawford, K., 2013. The hidden biases in Big Data. Harvard Business Review Blog Network April 1. http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/.

Cukier, K., Mayer-Schoenberger, V., 2013. The Rise of Big Data: How It's Changing the Way We Think About the World. Foreign Affairs May/June.

Debreceny, R.S., Gray, G.L., 2010. Data mining journal entries for fraud detection: An exploratory study. Int. J. Account. Inf. Syst. 11 (3), 157–181.

Debreceny, R.S., Gray, G.L., 2011. Data mining of electronic mail and auditing. Res. Agenda J. Inf. Syst. 25 (2), 195–226.

Dowling, C., Leech, S., 2007. Audit support systems and decision aids: current practice and opportunities for future research. Int. J. Account. Inf. Syst. 8 (2), 92–116.

Dwoskin, E., 2014. Big Data's High-Priests of Algorithms. Wall Street Journal August 8. http://www.wsj.com/articles/academic-researchers-find-lucrative-work-as-big-data-scientists-1407543088.

Eccles, R.G., Gladstone, J., 1991. KPMG Peat Marwick: The Shadow Partner. Harvard Business School, case 492-002 July 1991 (Revised October 1995).

Elliott, R.K., 1998. Accounting in the 21st Century, Newman Library Digital Collection, Baruch College, CUNY, New York, NY. http://www.baruch.cuny.edu/library/alumni/online_exhibits/digital/saxe/saxe_1998/elliott_98.htm.

EY, 2014. Global forensics data analytics survey. http://www.ey.com/GL/en/Services/Assurance/Fraud-Investigation—Dispute-Services/EY-Global-Forensic-Data-Analytics-Survey-2014.

Gartner, 2012. Gartner says Big Data will drive $28 billion of IT spending in 2012. Press Release, October 17th http://www.gartner.com/newsroom/id/2200815.

Gartner, 2014. Gartner survey reveals that 73 percent of organizations have invested or plan to invest in Big Data in the next two years. Press Release, September 17th http://www.gartner.com/newsroom/id/2848718.

Ginsberg, J., Mohebbil, M., Patel, R., Brammer, L., Smolinski, M., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. Nature 457, 1012–1014.

Gray, G.L., Debreceny, R.S., 2014. A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. Int. J. Account. Inf. Syst. 15 (4), 357–380.

Gray, G.L., Chiu, V., Liu, Q., Li, P., 2014. The expert systems life cycle in AIS research: what does it mean for future AIS research? Int. J. Account. Inf. Syst. 15 (4), 423–451.

IIA, 2013. Big Data: collect it, respect it. Institute of internal auditors: tone at the top. Issue 63. August https://global.theiia.org/knowledge/Public%20Documents/TaT_August_2013.pdf.

Keltanen, M., 2013. Why 'lean data' beats Big Data. Guardian Newspaper http://www.theguardian.com/media-network/media-network-blog/2013/apr/16/big-data-lean-strategy-business.

Kogan, A., Alles, M., Vasarhelyi, M., Wu, J., 2014. Design and evaluation of a continuous data level auditing system. Audit. J. Theory Pract. 33 (4), 221–245.

Laney, D., 2001. 3D Data Management: Controlling Data Volume, Velocity, and Variety, Application Delivery Strategies, META Delta, META Group, Inc., File: 949, Stamford, CT.

Littley, J., 2012. Leveraging data analytics and continuous auditing processes for improved audit planning, effectiveness and efficiency. KPMG White Paper http://www.kpmg.com/US/en/IssuesAndInsights/ArticlesPublications/Documents/data-analytics-continuous-auditing.pdf.

Lucas, S., 2012. Beyond the balance sheet: run your business on new signals in the age of Big Data. Blog posting August 21st. http://www.saphana.com/community/blogs/blog/2012/08/21/beyond-the-balance-sheet-run-your-business-on-new-signals-in-the-age-of-big-data.

Mayer-Schoenberger, V., Cukier, K., 2013. Big Data: A Revolution That Will Transform How We Live, Work, and Think. Eamon Dolan/Houghton Mifflin Harcourt ISBN: 9780544002692.

Moffitt, K., Vasarhelyi, M., 2013. AIS in an age of Big Data. J. Inf. Syst. 27 (2), 1–19 Fall.

Moore, G.A., 2002. Crossing the Chasm. Harper Business, New York.

New York Times, 2014. Can Big Data tell us what clinical trials don't? October 5th http://www.nytimes.com/2014/10/05/magazine/can-big-data-tell-us-what-clinical-trials-dont.html?ref=magazine

Norris, F., 2011. Troubled Audit Opinions. N.Y. Times June 10. http://www.nytimes.com/2011/06/10/business/10norris.html.

PCAOB, 2014. Matters Related to Auditing Revenue in an Audit of Financial Statements, Staff Audit Practice Alert No. 12 September 9.

Rowley, J., 2007. The wisdom hierarchy: representations of the DIKW hierarchy. J. Inf. Sci. 33 (2), 163–180.

Setty, K., Bakhshi, R., 2013. What is Big Data and what does it have to do with it audit? ISACA J. 3, 1–3.

Vasarhelyi, M.A., Halper, F.B., 1991. The continuous audit of online systems. Audit. J. Pract. Theory 10 (Spring), 110–125.

Whitehouse, T., 2014. Auditing in the era of Big Data. Compliance Week. April 29 http://www.complianceweek.com/auditing-in-the-era-of-big-data/article/7XljSe2GNuI1/t/.

Yoon, K., Hoogduin, L., Zhang, L., 2015. Integrating different forms of data for audit evidence: markets research becoming relevant to assurance. Account. Horiz. 29 (2), 431–438.

Zhang, J., Yang, X., Appelbaum, D., 2015. Towards effective Big Data analysis in continuous auditing. Account. Horiz. 29 (2), 469–476.

Zikopoulos, P.C., deRoos, D., Parasuraman, K., Deutsch, T., Corrigan, D., Giles, J., 2013. Harness the Power of Big Data: The IBM Big Data Platform. McGraw-Hill, New York, NY.

Zins, C., 2007. Conceptual Approaches for Defining Data, Information, and Knowledge. Journal of the American Society for Information Science and Technology 58 (4). Wiley Periodicals, Inc., pp. 479–493.

Please cite this article as: Alles, M., Gray, G.L., Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors, Int J Account Inf Syst (2016), http://dx.doi.org/10.1016/j.accinf.2016.07.004