Accounting Horizons
Vol. 29, No. 2
2015
pp. 409–422

American Accounting Association
DOI: 10.2308/acch-51065

Consequences of Big Data and Formalization on Accounting and Auditing Standards

John Peter Krahel and William R. Titera

SYNOPSIS: Accounting and auditing standards help manage a compromise between users' desire for more information and the costs to prepare and transmit that information. Previously forced to manage a paucity of information, businesses now look to capitalize upon the massive volumes of data and metadata that fill petabytes of space in their servers. Likewise, data from many sources and in many forms, many of which are irrelevant, inundate investors. Accounting and auditing standards have not kept pace, maintaining an emphasis on presentation, aggregation, and sampling. This essay argues that a change in standards to focus on data, the processes that generate them, and their analysis, rather than their presentation, will add value and relevance to the accounting profession, empower end users, and improve the efficiency of the capital markets.

Keywords: Big Data; accounting standards; auditing standards; materiality; continuous audit.

INTRODUCTION

The level, breadth, and quality of externally presented financial information have always represented a compromise between the preparer's cost and the user's benefit. While preparer costs *vis-à-vis* data collection and transmission have decreased significantly, the compromises made in the paper-based era have persisted, creating a set of anachronistic accounting practices that, in our view, unfairly handicaps statement users. A similar effect can be observed in auditing practices. While data availability and standardization have increased, audit standards continue to focus on sampling and other practices indicative of a low-information environment. This paper will address the problems that result from such anachronisms, present a set of axes along which accounting and auditing standards must evolve, describe the avenues through which such changes can be accomplished, and discuss the new paradigm from academic and practical perspectives.

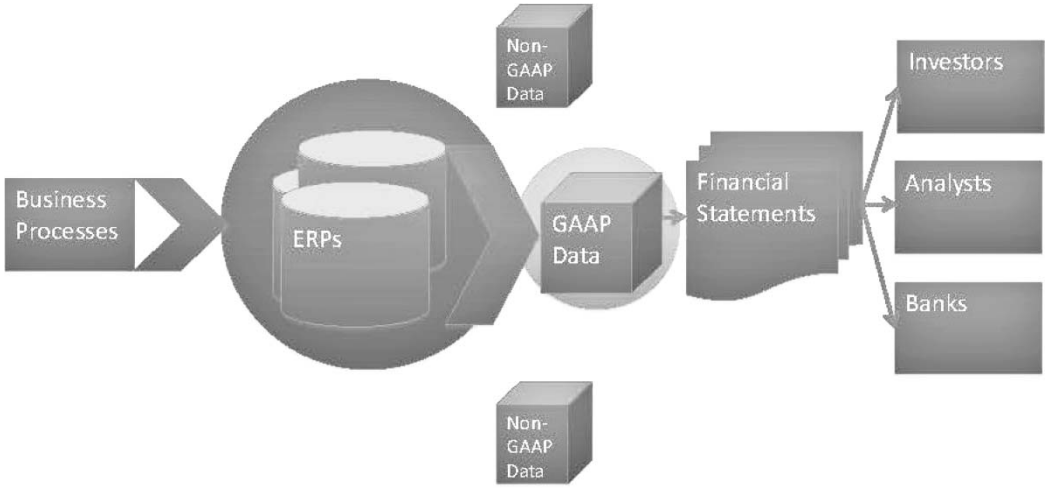
Presentation and manipulation before the Information Age was largely a slow, manual process: with compilation, tabulation, and aggregation of handwritten ledgers and accounts; physical

John Peter Krahel is an Assistant Professor at Loyola University Maryland, and William R. Titera is a retired partner, formerly of EY's national Professional Practice group.

The authors thank Trevor Stewart for his helpful comments and insight.

Submitted: January 2015
Accepted: February 2015
Published Online: February 2015
Corresponding author: John Peter Krahel
Email: jkrahel@loyola.edu

FIGURE 1
Reduction of ERP Data to One Set of GAAP-Compliant Disclosures



printing of statements and notes; and shipping paper documents to interested parties. The fundamental concepts underpinning modern accounting presentation were developed either before or during this era, and the nature of these concepts reflects these difficulties. The amount of external data presented is small and abstracted compared with the wealth of data available to insiders. In addition to management's reticence to disclose proprietary information, the costs (and potential information overload) associated with additional disclosures have long been a justification for limiting more atomized data from external users.

While the era of paper-based statements has passed, many of its artifacts remain. For example, when compared with specific unit costing, aggregation methods like FIFO or weighted average costing limit the investor's understanding of business activity by artificially assigning potentially outdated or inaccurate unit costs to cost of goods sold figures. The manipulation of bad debt expense for the purposes of earnings management (Jackson and Liu 2010) exemplifies a less innocuous use of estimates and limited choices from the paper era. Presenting an anonymized schedule of accounts receivable to accompany management's estimate would resolve this issue, but this has yet to be seen as a viable alternative.

The dramatic shift in emphasis from paper to digital transmission did not leave such transmission costless. The high cost and slow speed of digital transmission remained as strong hindrances to increased disclosure. Constant and dramatic reductions in per-bit transmission costs and simultaneous increases in bandwidth and transmission speeds have changed this dynamic both internally (i.e., quickly collecting data is becoming a simpler task) and externally (i.e., transmitting those data is more readily accomplished). GAAP represents a heavily preparer-friendly compromise. Aggregation and arbitrary allocations made on static, paper-based financial statements are artifacts of a bygone era of high transmission costs and slow data-collection speeds.

As shown in Figure 1, most data are generated through transactions and stored in an ERP system. Not all of these data falls under GAAP-compliant headings, and not all data are presented externally. Those data, which are presented externally, consists of one static set of documents used for a wide range of purposes by diverse user groups.

This approach shown in Figure 1 creates knowledge asymmetry by omitting many types of information, especially external information relevant to management estimates and decisions. Several benefits associated with broader and more easily understood disclosures have been researched in recent years. Disaggregation and broader disclosure enable investors to better detect and mitigate financial reporting manipulations (Amir, Einhorn, and Kama 2014), thus empowering the individual to make better investment decisions. In addition, firms with less readable (i.e., more textually complex) quarterly disclosures suffer from greater post-earnings announcement drift than those with more easily understood disclosures (Lee 2012). Perhaps most important, firms with more frequent, corporate reporting via the Internet enjoy positive abnormal returns, higher monthly trading volume, and lower bid-ask spreads (Rahman and Debreceny 2010). While there are drawbacks to increases in disclosure,[1] the technological feasibility of such an increase is increasingly reasonable.

The following two sections develop our major contention that future accounting and auditing standards need to focus more on data and less on their presentation. The second section concentrates on Big Data and financial reporting standards; the third section details the implications of Big Data for auditing standards. The final section concludes, outlining a list of implications for practice and research.

BIG DATA AND FINANCIAL REPORTING STANDARDS

Many mechanisms enable faster and more articulate data collection and transmission, including the use of radio frequency identification technology (RFID) in inventory valuation, the availability of real-time market data for fixed asset valuation, and the development of predictive algorithms to improve accounting estimates. One possible outcome of these trends, which we describe below, is the replacement of GAAP-compliant, static financial statements with raw data to be dynamically extracted and examined by the end user.

RFID Enables Specific-Unit Costing

FIFO, LIFO, and average costing all aggregate and make simplifying assumptions about the costs of goods, with such assumptions justified by the high cost of tracking an individual item's progress through inventory. Radio frequency identification technology (RFID) reduces these costs. RFID tags are cheap, small enough to be implanted into nearly any unit of inventory, fully machine readable, and detectable over short spans of space. RFID has proven beneficial across a wide array of industries, from construction (*Concrete Products* 2009), to healthcare (Reiner and Sullivan 2005), to hospitality uniforms (*Laundry News* 2013), to logistics (Turnbull 2006). Research confirms these benefits (Hardgrave, Aloysius, and Goyal 2013). In addition to internal benefits such as spoilage prevention and error reduction, RFID tags' quicker read time implies that specific-unit costing is a reasonable reporting option for more and more businesses. Individual items can be tracked from procurement or manufacture to sale with relative ease. The benefits will increase as data storage and transmission costs shrink. French aircraft manufacturer, Airbus, has begun to experiment with RFID tags with significantly larger data capacity, promising "efficient and error-proof identification of aircraft components throughout their life" (Airbus 2012). The opportunity to provide real-time inventory costing data portends a significant decline in the cost effectiveness of LIFO, FIFO, average costing, and other aggregated valuation methods (Vasarhelyi 2012),

[1] For example, the relationship between litigation risk and increased disclosure remains a subject of academic debate (Cao and Narayanamoorthy 2011; Johnson, Kasznik, and Nelson 2001).

especially for those firms for which actual inventory acquisition and sales do not mirror these methods of aggregation.

RFID will also enable more diverse data collection and extrapolation. For example, awareness of the location of inventory, and its time in storage, will yield knowledge about obsolescence. Coupled with sufficient analytical capability, this information could be used to develop obsolescence trend models that might better inform both management and investors about the quality of inventory. Auditors will likewise benefit from RFID use. If an entire inventory can be counted and located automatically, then sampling and spot testing will become supplements to, rather than substitutes for, full population tests.

Real-Time Market Data Enable Accurate Present Value

One specific consequence of globalization and digitization is the ability to record transactions in real time. With this ability comes a need to maintain up-to-the-moment pricing data. For example, during the financial crisis of 2008, many banks did not accurately reflect the value of their loan portfolios, possibly because they were using backward-looking historical loss data that were not reflective of current conditions. Analysis of the characteristics of borrowers and broader macroeconomic trends, in addition to cost-basis accounting, might have provided a richer overall picture of banks' loan portfolios, reducing expectations and lessening the impact of the crisis. While historical cost has some value, its exclusive use omits many dimensions of information. Maintaining inventory and fixed asset values at both historical cost and current market value, allowing the user to manipulate balance sheet presentations according to personal preference, provides a richer profile of a firm's standing, which is lacking in modern disclosures (Warren, Moffitt, and Byrnes 2015). Going one step further and incorporating macroeconomic forces into a valuation scheme may run contrary to GAAP's current micro-level, historic focus, but it will also provide more relevant, in-depth information.

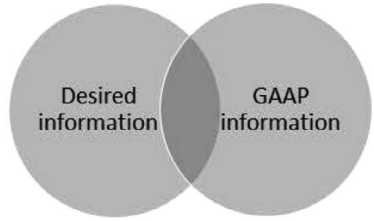
Fair value disclosures for a broader range of assets may follow a tiered framework, akin to the FASB's current three-tiered approach. Those assets more easily valued (e.g., company cars with Kelley Blue Book® values, or off-the-shelf factory equipment with a large buyer base) will fall into a higher tier than assets with less clear valuation (e.g., custom-built oil wells, or corporate headquarters). As is currently the case with the presentation of investments, fair value disclosure for a broader range of assets is not without risk. Current research analyzes the impact of fair value accounting on a variety of related phenomena, including market price volatility (Brousseau, Gendron, Belanger, and Coupland 2014); management behavior, both economic (Chen, Tan, and Ying 2013) and accounting-based (Bryan and Lilien 2013); and the banking crisis of 2008 (Badertscher, Burks, and Easton 2012; Laux and Leuz 2010; Laux and Leuz 2009).

Predictive Algorithms and Ubiquitous Computing Improve Estimates

The increasing sophistication of statistical software and the growing power of computing hardware enable more intelligent estimates of the value and likelihood of future events. The ubiquity and convenience of computing power, ranging from remotely accessible servers to portable smartphones and tablets,[2] enable advanced analytics and large amounts of data storage on demand. The same power used to anticipate customer needs and behavior can be used to estimate loan loss reserves and provisions for other bad debts, accruals for incurred but not reported claims,

[2] Dongarra and Luszczek (2012) have determined that the iPad® 2 is able to perform calculations faster than the Cray-2, the world's fastest supercomputer from 1985–1990.

FIGURE 2
Overlap between a User's Desired Information Disclosures and GAAP-Mandated Disclosures



warranty and pension expenses, and future revenue streams from current contracts. The design of the calculations used in these estimations can be transmitted to users for individual analysis. As with inventory costing, the ease of data retention and transmission can remove a level of abstraction and assumption from the presentation of estimate-heavy values on the financial statements. This expansion may cut both ways, however; sophisticated investors may be pleased with the added level of transparency, while casual investors may find themselves overloaded and unable to process the additional information.

Adjustable Granularity

The degree to which a user can drill down into deeper levels of a given value's component parts is currently limited to the depth required by GAAP. While the complexity of footnote disclosures and statements themselves may hinder investor and analyst analysis, the level of disaggregation already presented in financial statements has been shown (Fairfield and Sweeney 1996) to improve earnings prediction. Adjustable granularity will increase the pure volume of data transmitted, but this increase is not the end goal. Assuming the development of appropriate decision aids, the end result will be a level of detail specified by the user, not the preparer. Presentation of more granular data will allow for greater reporting and analytical flexibility. If a company's amount for given line items is seen as an aggregate that can be drilled down, and not as a standalone value, then deeper analysis is possible.[3] Additional context-specific data could be provided at more granular levels of reporting. To return to the subprime crisis example, information regarding the proportion of subprime debtors within a bank's portfolio would likely have proven useful for investors. Of critical importance is the idea that this information should be "pullable," not "pushed"; that is, casual investors should not immediately be presented with an overwhelmingly large and detailed dataset, but sophisticated investors should instead be given the option to receive it.

Modular Reporting

The current one-size-fits-all approach to financial reporting means that while a large volume of information is transmitted,[4] this volume does not imply relevance.[5] Figure 2 shows the overlap between a user's desired information disclosures and GAAP-mandated disclosures.

[3] The benefits of disaggregation are already observable in the form of GAAP-required segment reporting.
[4] For example, GM's most recent pension footnote occupies 16 pages and 13 tables.
[5] For more information, see Marschak's (1954) work on information economics.

TABLE 1

Data Needs of Different Investor Types

User Type	Data Needs	Ideal Dataset	Ideal Frequency
Casual investor	Low	Key ratios; aggregated balances	Quarterly or annually
Experienced investor	Moderate	Full financial statements; abbreviated notes	Daily or weekly
Professional analyst/day trader	High	Deeper, near-transaction-level data; external data on broader market trends	Daily or hourly

On-demand, modular reporting can alleviate the problem of over-reported and unnecessary data. Short-term investors may be more interested in an in-depth analysis of current and recent profitability trends, perhaps ignoring solvency concerns. Long-term creditors will be more concerned with deeper data regarding continuing liquidity and cash flow developments. A shift from the aforementioned one-size-fits-all paradigm will mean that the needs of each user will determine the frequency and volume of data received. These filtrations and requests can be easily accomplished with some combination of a subscription service and purpose-built browsing software modules. As more and more data types become quantifiable and data transmission costs continue their downward trend, transmission frequency and content become pliable variables, more attuned to the end user's needs than the provider's cost constraints. This trend implies a future in which roles are changed: the user has more responsibility for demanding available data, rather than depending on periodic output from firms, while firms are responsible for continuous provision of auditor-assured data, rather than template-based financial statements. Table 1 describes the potential needs of different investor types and the ability of a new paradigm to provide an appropriate amount of detail for those individuals.

Events, Not Values

Sorter (1969, 13) argues that the function of accounting should be "one level removed from the decision-making process." Since little is known about the manner in which accounting data are used, any arbitrary aggregation reduces the benefit of providing such information by restricting the breadth of its potential use. Sorter concludes that accounting is useful when it communicates information regarding events, not the valuation of items, subsequently arguing against the use of aggregation and for a change in perspective regarding the use of financial data. How do items co-vary, and what additional information is provided by making data more granular? The present-day phenomenon of data transmission at low cost and high speed makes Sorter's perspective more tenable, even allowing for users to alternate between lower and higher levels of aggregation with minimal effort.[6]

Queries as Compromise

Companies have an understandable desire to avoid unnecessary disclosure; while transparency increases market efficiency, it also exposes negative news to skeptical investors and, potentially,

[6] In addition to concurrent experiments demonstrating its appeal to investors of certain analytical mindsets, Sorter's (1969) work has previously provided the theoretical underpinning for McCarthy's (1982) REA ontology, a watershed event in the development of modern business database theory and practice.

gives up competitive advantage to other firms in a given industry. This concern will be addressed in greater depth later in this paper. One compromise between the desires of management and the growing data needs of the market is a creation of a query-based set of accounting principles. Under this paradigm, users would be able to request financial statements with given parameters. One user might request a completely cash-basis income statement, while another might request accruals and cash, but only for domestic transactions. An investor looking to make more equitable comparisons might request statements from the largest five firms in a given industry making common GAAP-compliant accounting choices (show all firms with FIFO inventory, the same receivables aging schedule, the same depreciation method, etc.). This would still allow the company the protection of aggregation, but would enable the user to make better decisions with better data. Investor information demands may still conflict with management's desire to maintain proprietary information, but this is an issue of information attributes, not quantity or customizability.

User Competence

One concern that can be raised regarding this paradigm is the average financial statement user's capacity for information consumption. A casual investor with limited free time may lack the ability or desire to either analyze a company's complete dataset or create a personalized set of data requests. Far from being a deterrent, this development creates a new source of value and revenue for those who possess such competencies. Financial analysis professionals can develop "branded" GAAP disclosures, varying in degrees of conservatism and aggregation methods, to be offered to casual investors as a separate information service. The fact that some users may not have the time or the inclination to take full advantage of broader data transfer is not, in itself, a reason to avoid anticipating these changes. On the contrary, it may prove an opportunity for those firms willing to leverage their expertise.

At the same time, as financial disclosures begin to incorporate more and more nonaudited, external information, responsibility for trust and verification may shift back onto the user in another tradeoff between relevance and reliability. For example, a company may indicate that its projections are based on the sustained strength of its home currency, but analysis of that presumption is and will continue to be beyond the scope of an audit. The implicit risk therefore becomes the burden of the investor.

The Shape of GAAP

As GAAP shifts focus from framing and restricted disclosure to a more open approach, its nature will undergo a fundamental change. Discussions about the type of presentation will be replaced with discussions about the quantity of data presented. Static data will be replaced with dynamic data. How deeply will an investor be able to delve into a target company's information? Far from rendering GAAP irrelevant, these changes can serve to enhance its relevance in a marketplace of dynamic information.

What Gets Presented?

Disclosure decisions revolve around a fundamental opposition: investors' desire to be informed conflicts with management's aversion to overexposure. A swing too far toward information symmetry results in a loss of privacy and competitive advantage; a swing toward asymmetry results in frustrated shareholders and a higher cost of capital. GAAP represents the current compromise, but it remains the descendant and inheritor of a paper paradigm. While data transmission (and therefore disclosure) is cheaper, a common concern is that increased disclosure may reduce a company's competitive advantage. Although evidence shows that opting for a higher quantity of

reported information reduces bid-ask spread and increases trading volume (Leuz and Verrecchia 2000), concerns regarding over disclosure and the leakage of sensitive data are not to be taken lightly.

The Effects of Expanding Formalization

Once an item becomes quantifiable, its inclusion in the realm of easily disseminated, automated, and aggregated disclosures becomes a natural next step. As more data types become quantifiable[7] (through processes such as textual analysis of emails and other communications, GPS tracking of inventory, etc.), accounting standards will need to address an ever-expanding dataset. Levels of minimum granularity can be specified across all transactions and data types, much in the same way that FASB rules currently govern intrafirm segment reporting.

An effort toward standards formalization may lead to concerns about removing judgment or quantifying the unquantifiable. However, instead of transforming the nature of the underlying information contained in such previously unformalized content as an MD&A disclosure, proper standards formalization will attempt to determine those *preexisting* attributes of such data that can, given sufficient processing power, be treated as formal elements. Syllable count, synonyms, sentence length, tone, and paragraph count are examples of items that approach quantifiability without any effect on the underlying data. Textual analysis has been used to characterize the culture of accounting firms (Holmes and Marsden 1996), shareholder expectations management (Huang, Siew, and Zhang 2014), and classification of auditor disclosures of internal control weaknesses (Boritz, Hayes, and Lim 2013). Formalization of standards to deal with such elements is an acknowledgment of the reality of a digitized future.

IMPLICATIONS FOR AUDITING STANDARDS

Along with anticipated changes in financial reporting standards and the data they govern, changes in the level, nature, and frequency of auditing and assurance can be expected. A change toward more granular disclosures will likely bring about changes to the audit profession tasked with assuring these data. The advent of continuous auditing and the shift from sampling to population testing will eventually make auditing large datasets common. While such assurance may be difficult to conceptualize and deploy initially, a much more pressing issue is that the current set of auditing standards are geared toward a paper paradigm in terms of data collection, error response, and auditor competencies.

The Nature of Data Collection and Error Response

Current PCAOB auditing standards are geared toward intermittent and incomplete analyses of available data. AU Section 350 (PCAOB 2011) deals with statistical and nonstatistical sampling and the inherent uncertainties and personal judgment inherent in such methodologies. A selection from Paragraph 7 of this standard illustrates the compromising nature of sampling use and the corresponding standards that govern it.

> Some degree of uncertainty is implicit in the concept of "a reasonable basis for an opinion" referred to in the third standard of fieldwork. The justification for accepting some uncertainty arises from the relationship between such factors as the *cost and time required to examine all of the data* and the adverse consequences of possible erroneous decisions

[7] This expansion is demonstrated by the growing scope of XBRL (Sheridan and Drew 2012), formalized textual analysis (Loughran and McDonald 2011; Humpherys, Moffitt, Burns, Burgoon, and Felix 2011), and advanced Big Data analytics (Chen, Chiang, and Storey 2012).

> based on the conclusions resulting from examining only a sample of the data. If these factors do not justify the acceptance of some uncertainty, the only alternative is to examine all of the data. Since this is seldom the case, the *basic concept of sampling is well established in auditing practice*. (PCAOB 2011; emphasis added)

The availability of continuous auditing techniques, combined with the progressive digitization of transactional data, diminishes the costs incurred by auditors during data collection and analysis. No longer is population testing a costly method of last resort. Costs associated with false positives, trivial exceptions, and information overload may become more relevant issues under the new paradigm.

If audit standards are modified to advocate the availability of more continuous, automated, and population-level techniques, rather than treating them as a hopefully avoided contingency, then we believe the audit will become more standardized and effective.[8] With an increase in the percentage of data tested, and a decrease in the amount of primary detection work that is prone to human error, the likelihood of error detection will naturally increase. The increased volume of flagged transactions and other errors will have implications for error response.

The availability of Big Data sets and audit data analytics will necessitate changes in standards regarding error and misstatement discovery and evaluation. AS No. 14, *Evaluating Audit Results* (PCAOB 2010a), takes pains to define the nature of "clearly trivial" misstatements, noting that the auditor must estimate the total volume of error, whether he[9] has specifically detected individual misstatements. In an environment of data-level assurance, standards must address not only more atomistic errors and their appropriate response, but describe the evaluation of potentially error-prone internal processes. When broader data (and data types) can be more readily and frequently analyzed, determining the impact of individual misstatements may prove secondary in importance to identifying broader trends that, taken as a whole, may provide a material risk to a company's performance.

Quantifying Materiality

The concept of quantification of judgment and anticipation is already in practice in the case of receivables aging. A company may apply a predetermined uncollectibility percentage to a given account or group of accounts based on age. Similarly, an auditor may apply various analysis techniques that are based on historical trends in the aggregate. It is simply impractical to manually analyze each account, especially if a company has thousands or millions of credit customers. However, in a Big Data environment, such types of analyses become not only possible, but both preferable and disclosable. A full suite of materiality and judgment quantifications—likelihood percentages, levels of management analysis, etc.—can be included in a set of financial disclosures at relatively low cost. Auditing firms can likewise include their own quantifications for materiality determination as a supplement to the audit opinion.

Quantification of judgment within auditing standards, when used in conjunction with formalization and a proper measurement scheme, will yield several benefits. Market participants will be able to judge the adequacy of a firm's materiality determinations, assessing management's ability to accurately quantify and predict future events and controlling for any over- or under-conservatism that presents itself. These quantifications can be disclosed as a matter of general

[8] The auditing of estimates will persist as a source of uncertainty, even when incorporating broader data sources and techniques. However, it is in such estimate-heavy areas where continuous audit and data analytic techniques may prove even more beneficial by helping to reduce uncertainty and stronger, evidence-based foundations to the judgment of decision makers.

[9] The authors' use of gender-specific pronouns here mirrors that of the PCAOB's standards.

policy on behalf of all parties involved in financial data disclosure. Audit firms may likewise establish baseline levels of materiality in open disclosures. Broadly speaking, the auditor's role will move from statement-level assurance to data-level assurance. This implies disaggregation, more frequent and automated assurance, and a sharpened focus on processes. Since new data are being generated at a constant rate, continuous disclosure will necessitate assurance on the processes that capture and transmit information, detecting and addressing potentially material deviations before they propagate into larger levels of error and misstatement.

More Rapid/Continuous Assurance

The current conceptualization of the audit is directed toward a single piece of periodic assurance on static documents. In an era of high-cost, high-latency, and low-throughput paper financial statement transmission, such infrequent assurance was all that was necessary. If we are to move to more frequent disclosure, driven by effective and low-cost data transmission, then assurance must likewise occur with greater frequency, increasing to the point where individual statements may cease to exist, replaced by individual amounts updated on demand by the user. Should this occur, audit by exception, catching errors as they occur and not on an infrequent basis, will likely follow.

Auditor Competencies

Audit standards do not address the increasingly technical nature of auditor competence. Section 210 (AICPA 1975), which deals with auditor training and proficiency, has not been substantively updated since 1975. It makes no reference to specific skills or skill sets needed beyond a mention of "adequate . . . technical scope."

More Quantitative Auditor Skill Requirements

The existence of Big Data and the increasing ease of its transmission create a new set of problems for auditors. Audit evidence decisions in the face of data paucity are replaced by similar decisions in the face of data overload. Audit standards will likewise need to address a different set of requirements. When even the most conscientiously internally controlled organization reports voluminous errors of differing depth of impact, standards dealing with sample size decisions must be complemented with standards addressing full population analysis. To what degree is the auditor responsible for detection of misstatement when fundamentally all data are made available for review? The limitations of sampling are thoroughly addressed in audit standards (AU 350, PCAOB 2011) and well documented in academic literature (Elder, Akresh, Glover, Higgs, and Liljegren 2013), but surprisingly little attention has been paid to the issues associated with its opposite. While full population analysis may imply a greater deal of audit thoroughness, information overload will likely prevent an audit team from following up on every error uncovered over the course of an engagement. This difference will need to be articulated, lest the public come to expect a "perfect" audit from complete information.

Even in the absence of a complete dataset, or in cases where all attributes of data objects are not known, the availability of cheap, ubiquitous computing power enables more advanced audit data analytics. Clustering (Lenard, Alam, and Booth 2000), random forests (Yeh, Chi, and Lin 2014), linguistic analysis (Loughran and McDonald 2011; Humpherys et al. 2011), and rough sets (Yeh et al. 2014) are just some of the data-classification techniques that require both greater processing power and increased competence and understanding on the part of the individual auditor. Interpretation of results from large datasets will become a more pressing challenge. While it is unnecessary for audit standards to require the use of any specific tool, standards must be updated to

address the availability of higher volumes of data and processing power to improve the effectiveness of the properly trained auditor.

Analysis of Unstructured Data

The end result of the intersection of Big Data and powerful processing is the automated formalization of objects and attributes as they enter the data store of an organization. The time required to manually analyze and develop new elements and attributes in response to changing business phenomena is wasteful and largely unnecessary. Contextual technologies such as XML and natural language processing are harbingers of greater levels of standardization in erstwhile-unclassified data types. These techniques are already being used on an *ad hoc* basis by researchers. For example, Humpherys et al. (2011) are able to use Bayesian text analysis to improve fraud prediction based on word choices in MD&A sections of financial reports. While there will always be a role for human judgment and a place for unstructured discussion and data gathering, the automated standardization of unstructured formats portends a wealth of additional data to be analyzed.

Standards must adapt to allow for increased structuring of such otherwise unstructured data, but this is rarely seen in practice. For example, AU Section 329, *Substantive Analytical Procedures* (PCAOB 2010b), would seem a likely candidate standard for the discussion of appropriate analytical techniques. However, it only addresses such procedures in conjunction with the use of already quantified information. AU Section 350, *Audit Sampling* (PCAOB 2011), likewise implies the use of purely quantitative instances. These standards may be broadened in scope to allow for the use of unstructured data. The requisite processing power and analytical methodologies already exist, but auditors and audit firms will be likely be averse to taking on additional responsibility in the absence of guidance from audit standards to do so. A standard discussing the inclusion of unstructured data in audit planning and testing might serve to broaden the horizon of the auditor's competence and awareness.

As the role of the auditor shifts from fact checking to analysis, the behavior governed by auditing standards must do likewise. For instance, if data-collection procedures were to transition from manual processes to automatic ones, then the auditor's time would be freed for more cognitively complex tasks such as in-depth regression analysis or other kinds of predictive statistical work. Superior processing power must be involved at the preliminary stages of analysis. When automatic data collection and preliminary, rule-based analysis yields orders of magnitude more errors, the auditor's responsibility must shift from exception detection to exception *prioritization,* determining which of many errors are most worthy of further investigation (Issa and Kogan 2014). While analysis has long been a value added service offered by audit firms, the advent of Big Data will tilt the balance further toward such higher-level thinking.

CONCLUSION

In light of cost reductions in data generation, storage, retrieval, and transmission, the inherent compromises within the paper paradigm are of little benefit. Users are entitled to more in-depth, granular values that they can manipulate, drilling down and up for more or less detail where needed. Financial reporting standards that govern presentation and arbitrary aggregation must likewise give way to rules regarding the limits and frequency of data transmission, as well as the quality of those data.

Audit standards must change as well. Error detection and risk quantification are no longer sufficient targets, but must be seen as small components of an audit of broader scope. The deep analysis of tremendous volumes of data and potentially thousands of exception reports necessitates a different paradigm of reporting and assurance. The role of auditing standards, far from being

diminished in the face of increasing automation, must shift from governing sampling procedures to embracing the broader, deeper data availability and analysis of the modern era in an effort to create a better, more thorough audit.

The following non-exhaustive list contains implications for the future from the perspectives of practice and academic research.

Practice Implications

- Accounting and reporting standards must adapt to deal with the frequent (possibly even continuous) transmission of granular data, not only their presentation in the aggregate.
- Such standards need to consider addressing company-specific data, as well as macro-level data that may be important to the analysis of a company's financial condition. They also need to consider enhancing historical reporting to include other data elements that may enable predictive analysis by users.
- Auditing standards must address situations where data are abundant, not only where data are sparse. The concept of materiality in relation to a company's financial statements, taken as a whole, needs to be reevaluated.
- Auditing standards must also do more to address the concept of process auditing. When data are available on a continuous basis, the processes generating those data must be continuously assured. Internal and external auditor competencies must be broadened to include more advanced types of data analytics.
- All parties along the financial reporting value chain must recognize the latent value in unstructured and semi-structured data.
- Care must be taken to minimize the expectations gap between users and auditors in the face of increasing data and analytical capacity. A user's role (and responsibility), which could change, must also be considered.

Research Implications

- Will incorporating and disclosing external information significantly reduce a firm's cost of capital?
- Will firms that pioneer broader disclosures lose significant competitive advantage?
- What impact will a more data-heavy audit have on investor confidence?
- Will users ascribe more credibility to self-customized financial reports?
- How can firms minimize the risk to casual investors of information overload?
- What is the ideal disclosure frequency?

REFERENCES

Airbus. 2012. *Airbus Extends RFID Part-Marking across All its Aircraft Families*. Available at: http://www.airbus.com/presscentre/pressreleases/press-release-detail/detail/airbus-extends-rfid-part-marking-across-all-its-aircraft-families/

American Institute of Certified Public Accountants (AICPA). 1975. *Training and Proficiency of the Independent Auditor*. SAS 1, Section 210. Available at: http://www.aicpa.org/Research?Standards?AuditAttest/DownloadableDocuments/AU-00210.pdf

Amir, E., E. Einhorn, and I. Kama. 2014. The role of accounting disaggregation in detecting and mitigating earnings management. *Review of Accounting Studies* 19 (1): 43–68.

Badertscher, B., J. Burks, and P. Easton. 2012. A convenient scapegoat: Fair value accounting by commercial banks during the financial crisis. *The Accounting Review* 87 (1): 59–90.

Boritz, J. E., L. Hayes, and J. Lim. 2013. A content analysis of auditors' reports on IT internal control weaknesses: The comparative advantages of an automated approach to control weakness identification. *International Journal of Accounting Information Systems* 14 (2): 138–163.

Brousseau, C., M. Gendron, P. Belanger, and J. Coupland. 2014. Does fair value accounting contribute to market price volatility? An experimental approach. *Accounting & Finance* 54 (4): 1033–1061.

Bryan, S., and S. Lilien. 2013. How fair values and accounting structures allow triple-counting income: Implications for standard setters, market participants, and academics. *Journal of Accounting, Auditing & Finance* 28 (1): 79–98.

Cao, Z., and G. Narayanamoorthy. 2011. The effect of litigation risk on management earnings forecasts. *Contemporary Accounting Research* 28 (1): 125–173.

Chen, H., R. Chiang, and V. Storey. 2012. Business intelligence and analytics: From Big Data to big impact. *MIS Quarterly* 36 (4): 1165–1188.

Chen, W., H. Tan, and E. Ying. 2013. Fair value accounting and managers' hedging decisions. *Journal of Accounting Research* 51 (1): 67–103.

Concrete Products. 2009. *Evolving RFID System Key to Quality, Inventory Control*. Available at: http://www.concreteproducts.com/news/1176-blank-96724202.html

Dongarra, J., and P. Luszczek. 2012. Anatomy of a globally recursive embedded LINPACK benchmark. *2012 IEEE Conference on High Performance Extreme Computing*, Waltham, MA, September 10–12.

Elder, R., A. Akresh, S. Glover, J. Higgs, and J. Liljegren. 2013. Audit sampling research: A synthesis and implications for future research. *Auditing: A Journal of Practice & Theory* 32 (Supplement): 99–129.

Fairfield, P., and R. Sweeney. 1996. Accounting classification and the predictive content of earnings. *The Accounting Review* 71 (3): (07): 337–355.

Hardgrave, B., J. Aloysius, and S. Goyal. 2013. RFID-enabled visibility and retail inventory record inaccuracy: Experiments in the field. *Production and Operations Management* 22 (4): 843–856.

Holmes, S., and S. Marsden. 1996. An exploration of the espoused organizational cultures of public accounting firms. *Accounting Horizons* 10 (3): 26–53.

Huang, X., H. Siew, and Y. Zhang. 2014. Tone management. *The Accounting Review* 89 (3): 1083–1113.

Humpherys, S., K. Moffitt, M. Burns, J. Burgoon, and W. Felix. 2011. Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems* 50 (3): 585–594.

Issa, H., and A. Kogan. 2014. A predictive ordered logistic regression model as a tool for quality review of control risk assessments. *Journal of Information Systems* 28 (2): 209–229.

Jackson, S., and X. Liu. 2010. The allowance for uncollectible accounts, conservatism, and earnings management. *Journal of Accounting Research* 48 (3): 565–601.

Johnson, M. F., R. Kasznik, and K. K. Nelson. 2001. The impact of securities litigation reform on the disclosure of forward-looking information by high technology firms. *Journal of Accounting Research* 39 (2).

Laundry News. 2013. *On the Trail of Textiles*. Available at: http://laundrynews.com/node/521

Laux, C., and C. Leuz. 2009. The crisis of fair-value accounting: Making sense of the recent debate. *Accounting, Organizations and Society* 34 (6/7): 826–834.

Laux, C., and C. Leuz. 2010. Did fair-value accounting contribute to the financial crisis? *Journal of Economic Perspectives* 24 (1): 93–118.

Lee, Y. 2012. The effect of quarterly report readability on information efficiency of stock prices: The effect of quarterly report readability on information efficiency of stock prices. *Contemporary Accounting Research* 29 (4): 1137–1170.

Lenard, M., P. Alam, and D. Booth. 2000. An analysis of fuzzy clustering and a hybrid model for the auditor's going concern assessment. *Decision Sciences* 31 (4): 861–884.

Leuz, C., and R. Verrecchia. 2000. The economic consequences of increased disclosure. *Journal of Accounting Research* 38 (3): 91–124.

Loughran, T., and B. McDonald. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance* 66 (1): 35–65.

Marschak, J. 1954. Toward an economic theory of organization and information. *Decision Processes* 3: 187–220.

McCarthy, W. 1982. The REA accounting model: A generalized framework for accounting systems in a shared data environment? *The Accounting Review* 57 (3): 554–578.

Public Company Accounting Oversight Board (PCAOB). 2010a. *Evaluating Audit Results*. Auditing Standard No. 14. Available at: http://pcaobus.org/Standards/Auditing/Pages/Auditing_Standard_14.aspx

Public Company Accounting Oversight Board (PCAOB). 2010b. *Substantive Analytical Procedures*. AU Section 329. Available at: http://pcaobus.org/Standards/Auditing/Pages/AU329.aspx

Public Company Accounting Oversight Board (PCAOB). 2011. *Audit Sampling*. AU Section 350. Available at: http://pcaobus.org/Standards/Auditing/Pages/AU350.aspx

Rahman, A., and R. Debreceny. 2010. Frequency of corporate announcements via stock exchange web sites and market efficiency. *Journal of Accounting, Auditing & Finance* 25 (3): 457–490.

Reiner, J., and M. Sullivan. 2005. RFID in healthcare. *Healthcare Purchasing News* 29 (6): 74–76.

Sheridan, B., and J. Drew. 2012. The future is now: XBRL emerges as career niche. *Journal of Accountancy* 213 (6): 122–127.

Sorter, G. 1969. An "events" approach to basic accounting theory. *The Accounting Review* 44 (1): 12–19.

Turnbull, L. 2006. *Why I Can't Wait for the RFID Dream to Become a Full-Blown Reality*. Available at: http://www.canadianshipper.com/features/why-i-can-t-wait-for-the-rfid-dream-to-become-a-full-blown-reality/

Vasarhelyi, M. 2012. AIS in a more rapidly evolving era. *Journal of Information Systems* 26 (1): 1–5.

Warren, D., K. Moffitt, and P. Byrnes. 2015. How accounting will be changed by Big Data. *Accounting Horizons* 29 (2).

Yeh, C., D. Chi, and Y. Lin. 2014. Going-concern prediction using hybrid random forests and rough set approach. *Information Sciences* 254 (01): 98–110.