JOURNAL OF INFORMATION SYSTEMS
Vol. 31, No. 3
Fall 2017
pp. 45–61

American Accounting Association
DOI: 10.2308/isys-51803

An REA Ontology-Based Model for Mapping Big Data to Accounting Information Systems Elements

Uday S. Murthy
University of South Florida

Guido L. Geerts
University of Delaware

ABSTRACT: The term "Big Data" refers to massive volumes of data that grow at an increasing rate and encompass complex data types such as audio and video. While the applications of Big Data and analytic techniques for business purposes have received considerable attention, it is less clear how external sources of Big Data relate to the transaction processing-oriented world of accounting information systems. This paper uses the Resource-Event-Agent Enterprise Ontology (REA) (McCarthy 1982; International Standards Organization [ISO] 2007) to model the implications of external Big Data sources on business transactions. The five-phase REA-based specification of a business transaction as defined in ISO (2007) is used to formally define associations between specific Big Data elements and business transactions. Using Big Data technologies such as Apache Hadoop and MapReduce, a number of information extraction patterns are specified for extracting business transaction-related information from Big Data. We also present a number of analytics patterns to demonstrate how decision making in accounting can benefit from integrating specific external Big Data sources and conventional transactional data. The model and techniques presented in this paper can be used by organizations to formalize the associations between external Big Data elements in their environment and their accounting information artifacts, to build architectures that extract information from external Big Data sources for use in an accounting context, and to leverage the power of analytics for more effective decision making.

Keywords: analytics; big data; business transaction; data sources; decision making; REA.

I. INTRODUCTION

Laney (2001) originally proposed the notion of "3D" data management, describing how rapid increases in data volume, velocity, and variety were exceeding the limits of traditional data management practices. As popularized by IBM (2012), the term "Big Data" is typically associated with "the four Vs": volume, velocity, variety, and veracity. The term "Big Data" refers to immense datasets (volume) that include a wide variety of data types (variety) that are constantly changing at a high rate (velocity). The fourth V, veracity, pertains to the "truth" or integrity of Big Data, an issue of particular interest to accountants and auditors. The term "Big Data" can refer to a wide range of data objects that vary vastly in type and structure, from video and audio files at the relatively unstructured end to website logs and social media posts at the relatively structured end.[1] According to IBM (2012), the five "game changing" use cases of Big Data are (1) Big Data exploration

We thank Kevin Moffitt, the discussant at the 2016 Accounting Information Systems (AIS) Midyear Meeting, for helpful comments. This paper has also benefited from comments by participants at the accounting research workshop at Rutgers, The State University of New Jersey, particularly Alex Kogan and Miklos Vasarhelyi. We are also grateful to the anonymous reviewers and editor for their feedback in helping shape the paper.

This paper was presented at the 2016 *Journal of Information Sytsems* Research Conference in Pleasanton, California. An earlier version of this paper was presented at the 2016 Accounting Information Systems (AIS) Midyear Meeting in Houston, Texas.

Editor's note: Accepted by A. Faye Borthick.

Submitted: May 2016
Accepted: May 2017
Published Online: May 2017

[1] Accounting transaction data within an enterprise system also constitute Big Data, in the sense that plummeting data storage costs and advances in technology have made it feasible to store and access millions, if not billions, of historical transactions (e.g., in-memory computing platforms such as SAP HANA). The problem we address in this paper is how the highly structured accounting transaction data *inside* the boundaries of the enterprise system can be connected with relevant Big Data elements *outside* the boundaries of the traditional enterprise system.

(finding, visualizing, and understanding all Big Data to improve decision making), (2) obtaining a 360° view of the customer by incorporating additional internal and external data, (3) security intelligence extension (fraud detection, real-time cybersecurity monitoring), (4) operations analysis (analyzing a variety of machine and operational data to improve operational performance), and (5) data warehouse modernization (optimizing the data warehouse by integrating Big Data capabilities).[2]

There has been considerable recent interest in the accounting academic and practitioner communities regarding how Big Data can be applied in accounting and assurance-related judgment and decision-making settings. Moffitt and Vasarhelyi (2013) discuss Big Data issues as they relate to accounting and auditing, arguing that the advent of Big Data has significant implications for research in accounting measurement and representation methods, formalization of accounting-related procedures, semantic understanding of accounting-related phenomena, assurance procedures, and issues related to social welfare and accounting education. A recent issue of *Accounting Horizons* (Griffin and Wright 2015) includes a forum on Big Data, with eight commentaries regarding the implications of Big Data for accounting and auditing. The commentaries address issues of how Big Data can transform accounting and auditing research, how Big Data affects the design and operation of management control systems, the shortcomings of current accounting and auditing standards in an era of Big Data, the opportunities and constraints in leveraging Big Data in auditing, and the behavioral implications of Big Data and advanced analytics on auditor judgment and decision making. Absent in the extant literature on Big Data in the accounting and auditing domain, however, is any model or formal specification that relates Big Data to accounting information system (AIS) elements in the context of an enterprise system.

A model or formal specification would aid architects of Big Data applications in designing systems that optimally link accounting data within enterprise systems to Big Data outside enterprise systems. We present such a model based on the REA ontology (McCarthy 1982; Geerts and McCarthy 2002) for extending business transactions with Big Data. We argue that the initial foray into linking accounting data within enterprise systems and Big Data should begin with the relatively structured components of Big Data outside the enterprise system. Accordingly, the model we propose specifies associations among relatively structured external Big Data elements and highly structured accounting data from resources, events, and agents associated with routine business transaction processing. The formal definition of these associations starts with identifying the five relatively structured external Big Data sources that are most amenable to association with data in enterprise systems: website logs, social media, sensor logs, communication repositories, and intranet/extranet logs. They are then linked to the REA primitives used to define the five business transaction phases: planning, identification, negotiation, actualization, and post-actualization (ISO 2007). Examples of such REA primitives include offering, proposal, commitment, economic event, agent, and agent type. The formal links between REA primitives and specific Big Data elements provide guidance for integrating transactional and non-transactional data in accounting and auditing-related applications. Using the most popular Big Data technologies—Hadoop and MapReduce—we specify a series of information patterns for extracting accounting-related information from external Big Data sources. Finally, based on the links in the model between accounting data and Big Data elements, we present seven analytics patterns that rely on the integration of transactional and non-transactional information. These analytics can inform both internal management accounting-related decisions (budgeting, forecasting) and auditing decisions involving the identification of risk areas and detection of anomalous transactions.

The remainder of this paper is organized as follows. The next section summarizes background material and discusses prior literature. In Section III, we present and discuss the model associating certain external Big Data sources to REA ontology primitives. In Section IV, seven MapReduce patterns are presented showing how information relevant for accounting purposes can be extracted from specific external Big Data sources. In Section V, the integrated use of transactional data and Big Data for analytics are discussed and illustrated by means of seven analytics patterns. Section VI concludes and summarizes the paper and discusses future research opportunities related to the association between Big Data and accounting.

II. BACKGROUND AND PRIOR LITERATURE

Big Data in Accounting and Auditing

In an editorial in the *Journal of Information Systems,* Moffitt and Vasarhelyi (2013) discuss the effects of Big Data on various aspects of accounting and auditing, including (1) accounting measurement and representation methods, (2) the need for formalization of standards and explicit consideration of digital information provisioning, (3) incorporation of semantic data from multiple sources, (4) traditional assurance procedures, and (5) the economics associated with the adoption of new accounting and auditing processes. Moffitt and Vasarhelyi (2013) argue that businesses are already finding value in nontraditional data (i.e., Big Data) for marketing and other purposes, despite such data not being integrated within the

[2] Source: http://www-01.ibm.com/software/data/bigdata/use-cases.html

boundaries of traditional accounting and enterprise resource planning (ERP) systems. They argue that unless accounting and auditing procedures evolve to explicitly incorporate Big Data, "current accounting and auditing methods are in danger of becoming anachronistic" (Moffitt and Vasarhelyi, 2013, 2).

Vasarhelyi, Kogan, and Tuttle (2015) argue that the term "Big Data" carries different meanings across domains. From a size standpoint, the notion of what is truly "Big" data (in terms of both volume and variety) depends on the size of the organization and its available computational capacity. Vasarhelyi et al. (2015) contend that depending on the kind and complexity of computational algorithms applied to Big Data, the processing needs would also vary dramatically from one organization to another. They suggest that the exponentially expanding enterprise data ecosystem calls for the development of a new theory of information that recognizes the nature of data capture (manual or automatic), the volume of the data capture, the extent to which new forms of data are integrated with existing data, the process of converting data to information, the granularity of the data, and the types of operational and other decisions to be made using the information. The model we present in this paper addresses the problem of integration between relatively structured external Big Data elements and transactional processing elements within the traditional enterprise information system.

Krahel and Titera (2015) contend that as businesses are seeking ways to leverage the petabytes of transactional and non-transactional data within their computing domains, extant accounting and auditing standards have lagged, with a decades-old focus on presentation, aggregation, and data sampling. They argue that the availability of massive amounts of transactional-level data does not imply that all such data should be made available to all users. Instead, they suggest that the type, volume, and frequency of data should be tailored to the needs of different user types, with casual investors receiving only key ratios quarterly or annually and professional investors receiving "near-transactional-level" data on a daily or even hourly basis. The model presented in this paper provides guidelines for deciding which subsets of the massive amounts of internal transactional Big Data are most relevant for decision making in an accounting context.

Warren, Moffitt, and Byrnes (2015) discuss how various categories of Big Data, such as audio, video, image files, and non-transactional textual data (e.g., email messages) can impact practices in managerial accounting and financial accounting. They contend that in managerial accounting, such Big Data can inform the design of management control systems such as the Balanced Scorecard (Kaplan and Norton 1996). Specifically, website and email logs can be linked to learning and growth goals, email and call logs can be linked with customer-oriented goals (e.g., satisfaction), and intranet web logs and internal emails can be associated with internal business process-related goals. They also suggest that Big Data outside of traditional ERP systems, for example, census and macroeconomic data, can be used to enhance "beyond budgeting" practices that seek to use alternative sources of information for planning and performance evaluation. In the financial accounting realm, Warren et al. (2015) contend that Big Data could be used to support the existence and valuation assertions for assets. For example, they argue that web-based software agents could be programmed to automatically retrieve information from external market sources to confirm Level 1 and Level 2 fair value estimates. The model we present in this paper addresses the associations between external Big Data and firm assets that fall under the "resources" category in the REA ontology.

From an assurance perspective, Yoon, Hoogduin, and Zhang (2015) contend that Big Data can complement traditional audit evidence with sufficient, reliable, and relevant information from nontraditional sources. They contend that Big Data can be sufficiently reliable because it is often externally generated and collected directly by the auditor, rather than being provided by the client. For example, rather than relying on client-provided shipping documents, auditors can use global positioning system (GPS) data directly from a shipper to verify that shipments were delivered to customers. Social media posts can be mined to gauge customer sentiments regarding products, which could serve as useful inputs into analytical procedure models designed to estimate whether sales revenues are within expected bounds. However, as noted by Yoon et al. (2015), one of the more significant challenges in using Big Data for audit purposes is to identify and establish appropriate relationships between Big Data elements and traditional audit evidence. Our model, based on the REA ontology, provides a roadmap for relating certain relatively structured external Big Data to accounting transactional artifacts and thereby establishing the necessary links between traditional audit evidence and relevant Big Data elements.

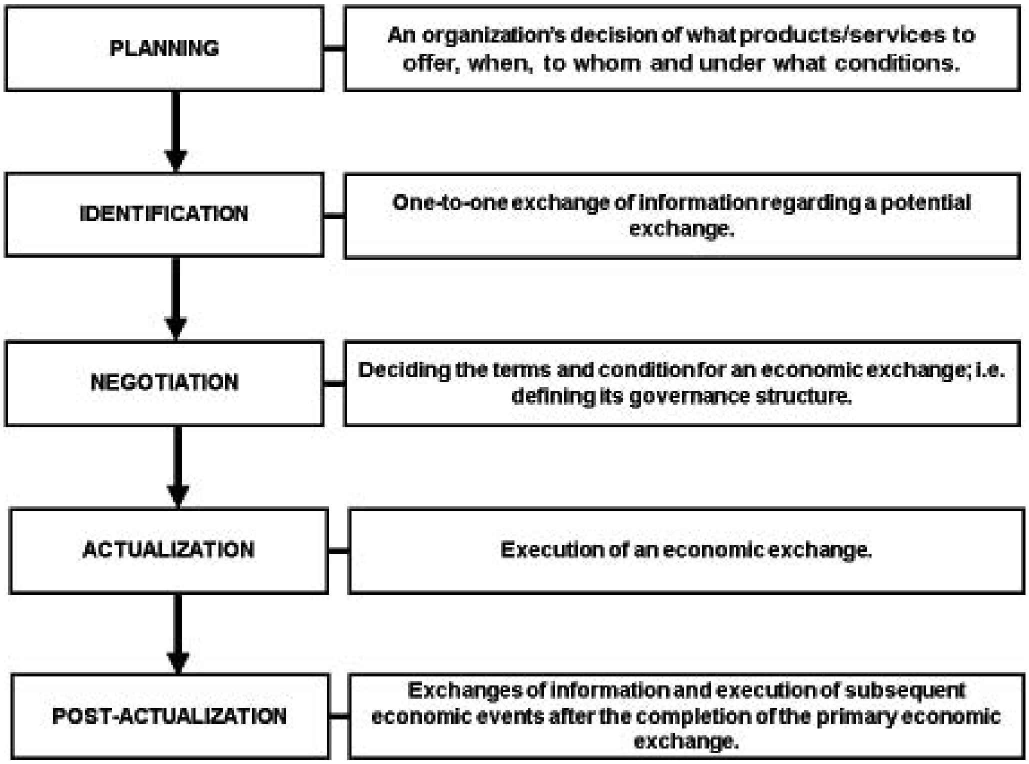
The Resource-Event-Agent Enterprise Ontology (REA)

A key issue addressed by our model is how traditional transaction systems can be extended with external Big Data. Based on empirical validation, O'Leary (2004) concludes that economic phenomena being captured as part of ERP systems, SAP in his research, strongly overlap with the economic phenomena being captured by the REA ontology. We have, therefore, chosen REA to represent the traditional transaction component of enterprise systems.

Possible non-transaction extensions to REA have been discussed by others. For example, Denna and McCarthy (1987) emphasize the importance of outside data sources, such as competitor information and regional economic indicators, for decision-making purposes. Also, Geerts and McCarthy (1999) underscore the importance of integrating non-transactional data such as commodity prices or information on product substitutes or complements with traditional transactional data. Further,

American Accounting Association

FIGURE 1
Business Transaction Phases
Adapted from ISO (2007)



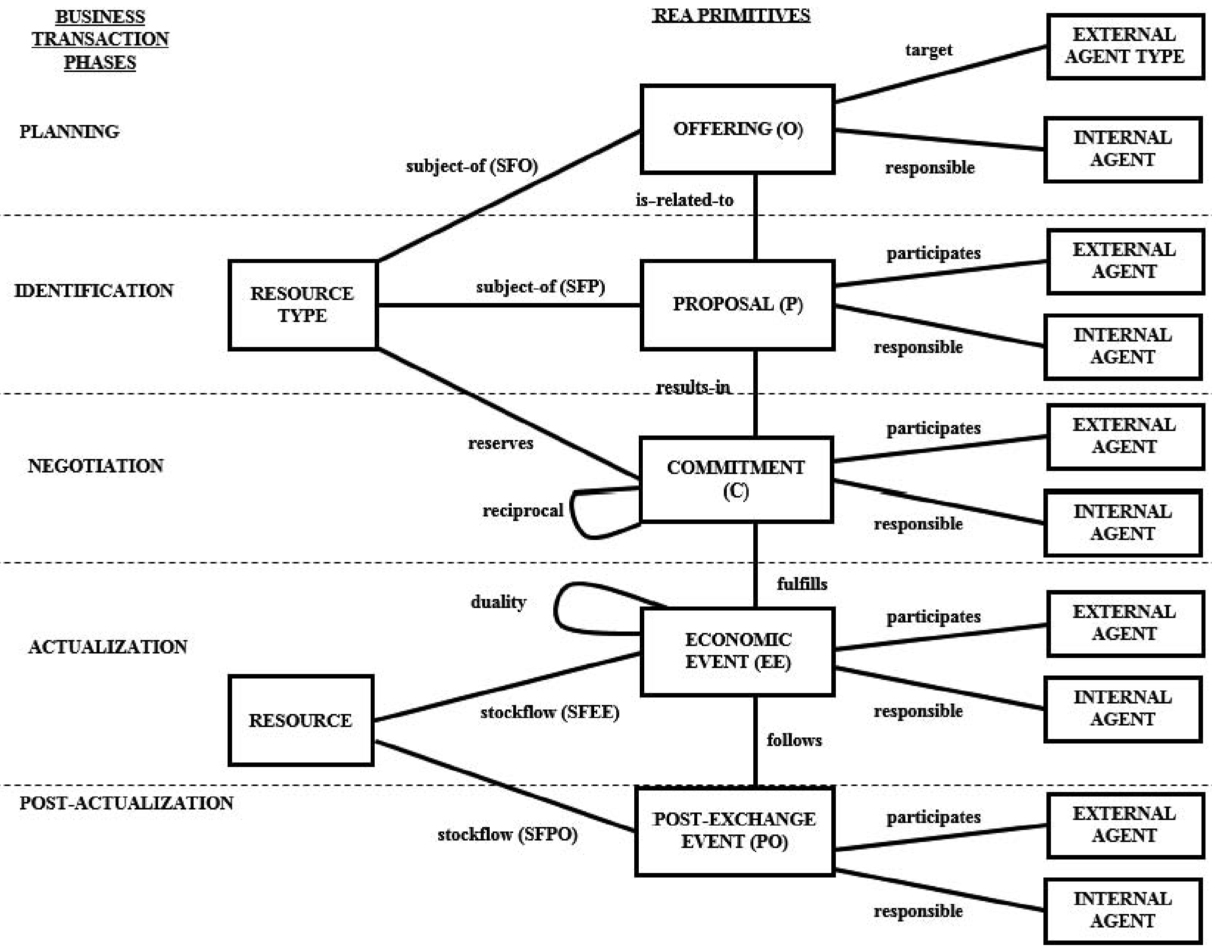
Geerts and O'Leary (2014) extend transactional data by location-based information generated as part of the Internet of Things (IoT); i.e., the network of devices seamlessly communicating over the Internet to share information.

Big Data represents unprecedented access to a wide variety of data resources, which we formally link to the REA-based five-phase business transaction specification presented in ISO (2007). The left side of Figure 1 shows the ISO business transaction phases, while the right side provides a brief description of each phase. Similar business phases have been recognized by others, including Maes, Guttman, and Moukas (1999) and Hümmer, Lehner, and Wedekind (2002). We have chosen the ISO (2007) specification because of its strong ties to the REA model. As pointed out by Maes et al. (1999), phases are an approximation and simplification of complex behaviors, often overlap, and migration from one to another can be nonlinear and iterative.

Figure 2 defines each of the business transaction phases in terms of REA primitives. Some of the primitives, such as "offering," were developed for the purpose of this paper. With the exception of the "participates" and "responsible" primitives, of which the semantics are similar across phases, we have defined unique names for the primitives for reference purposes; e.g., subject-of-offering (SFO) versus subject-of-proposal (SFP). During the planning phase, a business "formulates an abstract vision of an exchange" (ISO 2007, 27): what products and/or services the business plans to sell, to whom, when, and under what conditions. Communication of planned exchanges, i.e., offerings, can take many forms, including catalogs and web advertising campaigns. The REA primitives used for the definition of offerings are shown in the Planning segment of Figure 2. Instances of Resource Type define "what is being offered." Offerings typically refer to product types, not individual products.[3] For example, a car manufacturer plans offerings of models (resource type): when a new model will be launched; how many will be produced; what is the price for the model; etc. An "Offering" represents the conditions under which a resource type is offered, including the time frame within which the offer is valid and the price or price range. For simplicity, we assume that an offering is being defined for a specific resource type. External Agent Type represents the target of the offering, such as a

[3] Exceptions exist. For example, a catalog of paintings for sale or the specific houses listed on a real estate website represent offerings of individual items for sale. See Geerts and McCarthy (2006) for an in-depth discussion of the "Resource Type" REA primitive.

FIGURE 2
REA Business Transaction Definition



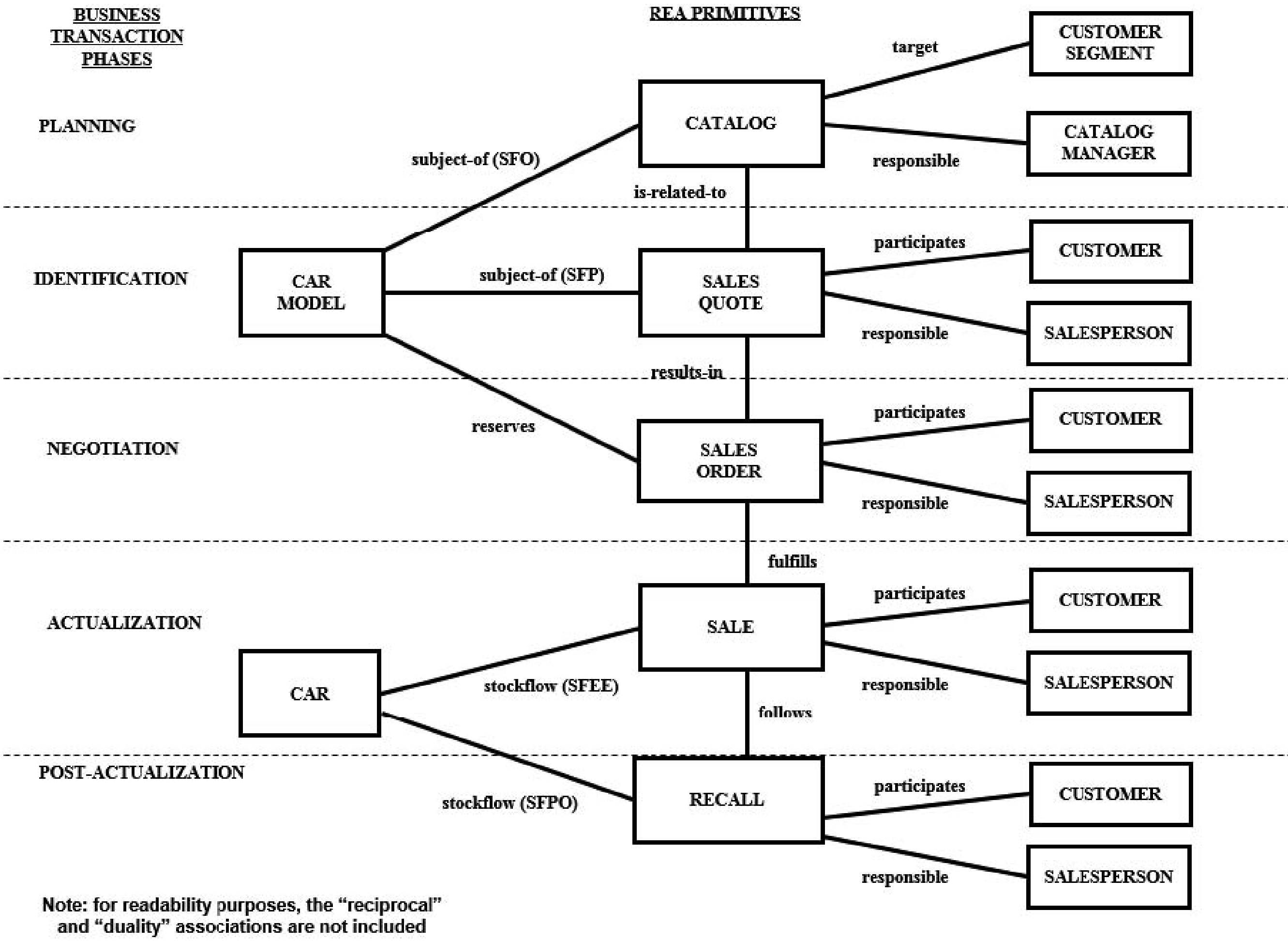
specific customer segment. An important characteristic of the planning phase and offerings is that no specific agents are identifiable (yet). Finally, as for all transactions, Internal Agent defines who within the organization is responsible for an offering.

An important characteristic of the identification phase is the establishment of a one-to-one linkage with an external agent, such as a specific customer or vendor. We have labeled transactions resulting from the identification phase as "proposals." Similar to offerings, a proposal is typically for a resource type (subject-of-proposal) and an internal agent is responsible for it. To the extent possible, proposals are linked to offerings, as represented by the "is-related-to" association in Figure 2. Tracking such linkages helps determine the effectiveness of offerings. Proposals can take many different forms, including a web-enabled information request, an email, a sales quote, or unilateral information sharing.

The negotiation phase is characterized by information sharing in order to achieve an explicit agreement regarding the terms and conditions for an exchange: the commitment. Key elements of a commitment specification are the resource types being committed (reserves), to whom (external agent), and who is responsible for the commitment (internal agent). The reciprocal association links "give" and "take" commitments. Further, to determine effectiveness, it is useful to link commitments with proposals. ISO (2007) further bundles commitments into contracts. We ignore contracts here for simplification purposes.

During the actualization phase, exchanges are executed. Exchanges consist of a network of economic events that are connected by means of duality associations. Unlike the three earlier business transaction phases—planning, identification, and negotiation—which are typically specified in terms of resource types, economic events, and post-exchange events typically relate to actual resources, for example, the sale (economic event) of a specific car (economic resource). Assuming that the tracking of economic resources is cost-effective, economic events are connected to specific economic resources by means of

FIGURE 3
REA Business Transaction Instantiation: "Sales"



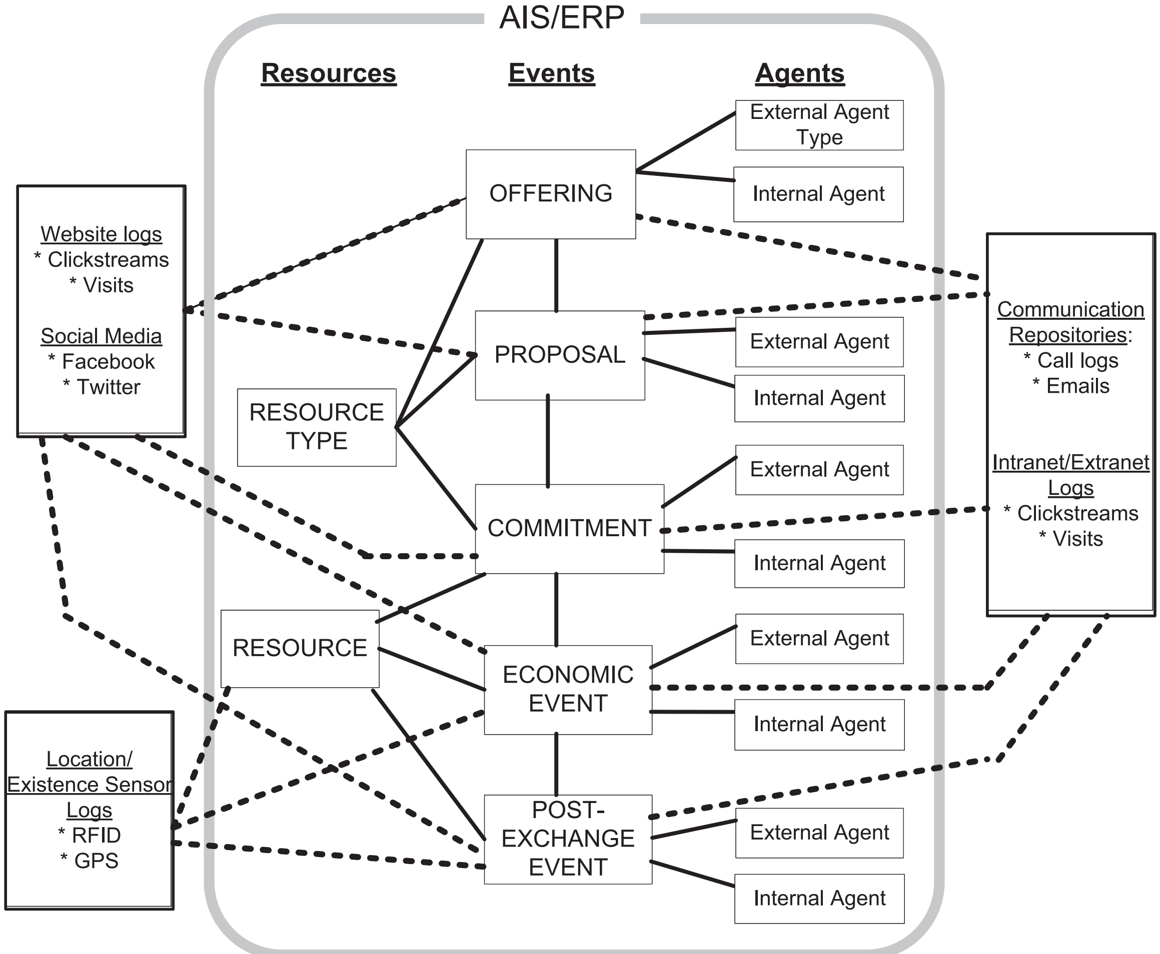
stockflow associations. In REA vocabulary, specific outside agents participate in an event, while specific inside agents are responsible for the event (McCarthy 1982). Commitments can be seen as governance structures for exchanges, and adherence to them is monitored by means of the "fulfills" association.

Post-actualization refers to transactions that occur after an economic exchange has been completed but that should be considered as part of the business transaction life cycle. A product recall is a good example of such a transaction. In addition to the agents being involved, the model in Figure 2 defines the resources involved in a business transaction as a result of such events. To understand the full cycle, post-exchange events are linked to their specific exchanges by means of the "follows" association.

Figure 3 shows an instantiation of the REA business transaction definition in Figure 2: "Sales Business Transaction."[4] A catalog is an example of an offering that can be used as part of a marketing campaign that targets a specific customer segment. Sales quotes are proposals that are related to the catalog offerings and are made to specific customers by specific salespeople with the goal of resulting in "sales order" commitments. Catalogs, sales quotes, and sales orders are all defined in terms of car models (resource type). Actual sales transfer the ownership of a specific car to a customer. Recalls are an extra cost and their tracing to specific exchanges will help determine the profit margin of the latter.

[4] Figure 3 does not show the "reciprocal" and "duality" associations. This would require the modeling of additional business transactions such as cash receipt (duality), which would make the diagram substantially more complex

FIGURE 4
Associations between Big Data Elements and REA Ontology Primitives



Solid lines = associations between REA primitives; dashed lines = primary associations between REA primitives and Big Data elements.

III. BIG DATA SOURCES AND THEIR ASSOCIATIONS TO REA BUSINESS TRANSACTIONS

Moffitt and Vasarhelyi (2013, Figure 1) describe a variety of emerging sources of Big Data; specifically, scanner data, web data, mobility data, videos, and recordings. Although many of these Big Data sources contain highly unstructured data, as Moffitt and Vasarhelyi (2013) contend, it is possible to identify certain key attributes in Big Data that might serve as the basis for structuring the data; attributes such as the source, date/time, medium, and location of the data. Identification of such attributes will, in turn, facilitate analytics comprising various statistical transformations of these attributes in combination with their association with related data within the organization's transaction system. Moffitt and Vasarhelyi (2013) also suggest that "text understanding" and "vague text understanding" could be applied to link textual elements to ERP data. They do not, however, specify exactly how these links could be identified within textual Big Data. From the multitude of external Big Data sources identified by Moffitt and Vasarhelyi (2013), we argue that five are particularly well suited for integration with data in traditional transaction-oriented enterprise systems—website logs, social media, location/existence sensor logs, communication repositories, intranet/extranet logs. A model depicting associations between these five relatively structured external Big Data sources and REA primitives within the AIS is shown in Figure 4. In the next section, we provide details of the Big Data technologies that facilitate the formal linkages between the AIS elements and the external[5] data sources.

[5] Note that the term "external" implies outside the boundaries of the traditional enterprise system (e.g., SAP). It does not imply external to the firm. Most of the data sources described in this section are within the boundaries of the firm and within the control of the firm, except that they lie outside the transaction processing enterprise system.

Website Logs

Log files from web servers contain a wealth of data that can be mined to better understand, for example, customer browsing and shopping habits, and web advertising effectiveness. Each web click creates about 100 bytes of data in a typical website log. Consequently, large websites handling millions of simultaneous visitors can generate hundreds of gigabytes or even terabytes of log data per day. The web log data typically contain the remote host, the date/timestamp, the web "request" line (i.e., which file is being requested), the status/size/referrer, and the user "agent" (browser type, e.g., Mozilla Firefox). Specifically, the "request" information that includes the specific file/folder accessed could be linked to offerings, proposals, or commitments. The "host" information could be used to identify the internal or external agent (employee or customer/supplier) responsible for the click event. Weblog parsing can, thus, reveal the *what* (request section), the *when* (timestamp), the *from where* (referrer), and the *who* (IP number) for every website visit.

Social Media

Usable social media data can be classified into two broad categories: (1) commentary data, and (2) action data. Commentary data includes product mentions, conversations between customers about specific products, posts about products, and tweets about products. Action data refers to specific actions that social media users take that can convey meaningful information about products or services. Examples of such data include Facebook "likes" and "shares" and Twitter "retweets," which are data reflecting customer sentiment (positive or negative) about products or services. For both commentary and action data, however, the difficulty lies in associating that data with ERP system artifacts within the boundary of the traditional accounting information system. The challenge with social media is that it is unlikely that explicit agent identifiers (internal—employee, external—customer) are directly identifiable in social media logs. However, it may be possible to "parse" these logs to attach such agent identifiers, perhaps by applying fuzzy match algorithms. The fields of social media data mining and social network analysis present techniques for automatically "scraping" social media sites to extract and structure content for the purpose of performing analyses (Barbier and Liu 2011; Murthy, Gross, Takata, and Bond 2013).

Location/Existence Sensor Logs

There has been a dramatic explosion in the number of sensors producing streams of data all around us. Once exclusively in the domain of industrial control systems or major transportation systems, technologies such as radio frequency identification (RFID), infrared, wireless, and GPS now permeate a variety of products and services at the consumer level. Sensors are either embedded into mechanical systems to track system activity, or they can be programmed to monitor (human) consumer behavior. Mobile phones provide location information, while in-store sensors and video equipment can monitor consumer shopping behavior. Some have estimated that the volume of data available from sensors will soon grow so rapidly that we will be engulfed in a massive "data deluge" (Baraniuk 2011). When embedded into product inventory or pallets carrying inventory, RFID tags can be read during product movement to automatically update work-in-process and finished goods inventory records in systems (Borthick 2012).

As noted by O'Leary (2013), location data of "things" provide valuable context data that can be mined to generate useful information when linked with other data. Leveraging location information, in addition to other primitives, Geerts and O'Leary (2014) developed the EAGLET (event, agent, location, equipment, and thing) ontology, which facilitates the visibility and interoperability of "things" along the supply chain. Thus, as has been recognized in prior research, sensor data, especially RFID and GPS, could be associated with "resources" (e.g., inventory) in the REA ontology, assuming that resource identifiers are embedded or could be extracted from the sensor logs. Borthick, Bowen, and Gerard (2008) demonstrate the application of REA modeling to RFID sensor data to make it usable with structured transaction data. As with social media data, the challenge with data from sensors is to structure the inherently unstructured data such that at least some subset of it becomes usable from an accounting and auditing perspective.

Communication Repositories

Beyond the primary transaction system, most companies maintain a plethora of systems designed to capture many different kinds of digital communication, including, but not limited to, email and instant messages. From an external communication perspective, as Moffitt and Vasarhelyi (2013) contend, call logs between salespersons and customers (both existing and potential) can be a "treasure trove" of information to help understand who calls, when they call, what they call about, how long calls last, and whether the calls are positive, negative, or neutral in "tone." For data in communication repositories to be usable from an accounting and auditing standpoint, it will be necessary to embed or somehow extract "agent identifiers" (internal agent—employee and/or external agent—customer/supplier).

Intranet/Extranet Logs

Most companies have intranets for the dissemination and consumption of a range of internal electronic documents and knowledge bases. Every employee access to an electronic document and knowledge base is logged. Assuming the log includes an employee identifier, at a minimum, the following data should be accessible in the intranet log: timestamp, employee ID, and e-document type. Additional related data can be provisioned by linking the log with related data in the corporate database. For example, the employee type, rank, tenure with the firm, and age are some types of additional data that can make the basic intranet log data more informative. Intranets also typically support some form of internal employee communication (e.g., message boards, discussion forums, chat rooms). Many companies have "extranets," as well, which are controlled access sites that their business partners interact with using a web browser (Vlosky, Fontenot, and Blalock 2000). For example, Ford has an extranet that provides its worldwide network of 15,000+ dealers access to promotions, inventory, and financial information through a portal (Vlosky et al. 2000). Since intranets and extranets use the same Internet technologies as regular open-access websites, intranet and extranet log files can be mined in much the same way as webserver log files. Unlike webserver logs, however, intranet/extranet logs are easier to mine since the agent identifiers (internal/external agents) are automatically included in these logs. Thus, intranet (extranet) logs of employee (partner) access to various e-documents and knowledge repositories can be mined for useful information that can be associated with components of the transaction system.

External Big Data sources provide additional information outside of the traditional enterprise information boundary that can be used in decision processes related to the different business transaction phases. Economic sensing based on external data sources can be used to evaluate and reconfigure offerings. Location-based data can be used to generate customized proposals; e.g., an offer for a product based on a customer's position in a grocery store. Price elasticity determined based on external sources can be employed in the negotiation process. Sensor data can be used to monitor the effectiveness of exchange execution. Social media might provide an indication of issues that cause the occurrence of post-exchange events.

Table 1 establishes possible associations between REA primitives and external Big Data sources. The columns in Table 1 represent the five Big Data sources discussed above; the rows represent the REA primitives presented in Figure 2. The entries in Table 1 indicate whether the association is a *primary* association (P) or a *secondary* association (*S*). Since transactions are the triggers of all interactions with resource types, resources, internal, and external agents, associations between external Big Data sources and each of the five different transaction types—Offering, Proposal, Commitment, Economic Event, and Post-Exchange Event—are considered *primary* associations. The association between location/ existence sensor logs and resources is also deemed to be a *primary* association, since the primary focus of sensor logs is to track resources. Each primary association, in turn, has secondary associations with related resources, internal, and external agents, which are shown in Table 1 as *secondary* (or derived) associations. In effect, Table 1 contains the operational details of the visual links shown in the model linking the AIS to external Big Data presented in Figure 4. Intersections delineate what external Big Data sources are likely useful for decision making related to specific REA primitives. Regarding the various external Big Data sources, website logs primarily add value by observing the behavior of external agents with regard to their business transaction expectations. The same is true for intranet logs, but here, the focus is on the behavior of internal agents and, in particular, their effectiveness. Social media and communication repositories apply to most REA primitives. Social media platforms such as Facebook and Twitter are often used by customers to convey sentiments (both positive and negative) about specific transactions. Sensors add value in the execution phases by being able to physically observe the movements of resources and agents.

We briefly discuss a number of the specific intersections in Table 1 for illustrative purposes. Website logs might provide useful information for decision making in the planning phase, such as the configuration, valuation, and reconfiguration of offerings. One example is monitoring of website traffic attributable to a specific offering, such as an advertising campaign or a promotional event. For a website that is e-commerce enabled, logs also provide useful information regarding commitments. Social media enable the economic sensing of products or offerings (Facebook "likes," Twitter "tweets" and "retweets"). Also, links to e-commerce functionality included on these social media sites generate data regarding the is-related-to and results-in REA primitives. Sensors such as RFID can observe the movement of economic resources (such as inventory) and, thus, collect additional real-time data regarding the REA stockflow primitive when triggered by an economic event. Communication repositories such as call logs and email are useful in all business process phases, and the following are a few specific examples: there can be call logs and emails regarding requests (proposals) and orders (commitments); they can provide confirmatory information regarding the completion of economic exchanges; they can provide information regarding a specific economic event and the resources involved in it (stockflow: SFEE). Intranet logs also cover all aspects of a business transaction and are especially useful to monitor the effectiveness of employees (internal agents). Extranet logs observe the involvement of business partners in an organization's business transactions, but the interaction might be limited to specific phases.

TABLE 1

Associations between Big Data Sources and REA Primitives

	Website Logs	Social Media	Location/ Existence Sensor Logs	Communication Repositories	Intranet/ Extranet Logs
Resource Type	*S*	*S*		*S*	*S*
Resource		*S*	P	*S*	*S*
Offering	P	P		P	P
subject-of (SFO)	*S*	*S*		*S*	*S*
Proposal	P	P		P	P
subject-of (SFP)	*S*	*S*		*S*	*S*
is-related-to	*S*	*S*		*S*	*S*
Commitment	P	P		P	P
reserves	*S*	*S*		*S*	*S*
results-in	*S*	*S*		*S*	*S*
reciprocal	*S*	*S*		*S*	*S*
Economic Event	P	P	P	P	P
stockflow (SFEE)		*S*	*S*	*S*	*S*
fulfills		*S*		*S*	*S*
duality		*S*		*S*	*S*
Post-Exchange Event	P	P	P	P	P
stockflow (SFPO)		*S*	*S*	*S*	*S*
follows		*S*	*S*	*S*	*S*
External Agent Type	*S*	*S*		*S*	*S*
target	*S*	*S*		*S*	*S*
External Agent	*S*	*S*		*S*	*S*
participates	*S*	*S*		*S*	*S*
Internal Agent		*S*		*S*	*S*
responsible		*S*		*S*	*S*

P = primary association; *S* = secondary association.

IV. EXTRACTING BUSINESS TRANSACTION INFORMATION FROM BIG DATA SOURCES: TECHNOLOGIES AND INFORMATION PATTERNS

The generally accepted standard technology for implementing Big Data solutions is Hadoop: an open-source framework for developing and executing distributed applications that must process extremely large datasets (Dean and Ghemawat 2008; Shvachko, Kuang, Radia, and Chansler 2010). The components that comprise Hadoop are: the Hadoop distributed file system (HDFS), MapReduce, partitioner, combiner, InputFormat, and OutputFormat. HDFS is the Hadoop file system—a distributed file system for storing massive amounts of data across multiple nodes in a cluster. Included in the Hadoop framework are an application program interface (API) and a command-line interface for interacting with HDFS.

At the heart of the Hadoop framework is MapReduce—a functional programming paradigm for processing records in HDFS and assembling the results into a usable solution. In essence, MapReduce involves "mapping" relevant elements from large datasets to produce a set of intermediate key/value pairs and then "reducing" those intermediate values by merging values with the same intermediate key (Dean and Ghemawat 2008). The typical example used to explain the functioning of MapReduce is the problem of counting the number of occurrences of every word in a collection of documents. Two separate functions need to be written: a "map" function, which outputs each word with the count of the number of its occurrences in each document, and a "reduce" function, which sums all counts for each word across all documents.

The next two components of Hadoop are the partitioner and the combiner. The partitioner divides a problem into workable chunks of data for use by the various Mappers. The usual method of operation is to divide work up by rows of data in the HDFS. The combiner must be invoked if it is necessary to perform a "local" reduce operation within a node before returning the data back. The final two Hadoop components are "InputFormat" and "OutputFormat." InputFormat is used for any setting where the data are not formatted in a standard way, such as "key, value" or "key [tab] value." It is likely that InputFormat will be required to define custom formats for parsing web logs, social media logs, and logs of other Big Data sources identified in this paper, particularly for the purpose of identifying and extracting the event and agent identifiers. OutputFormat is needed if the data are to be written in any format other than the standard "key [tab] value."

Table 2 represents seven possible information extraction patterns inferred from an analysis of the associations between external Big Data sources and REA primitives in each of the five business transaction phases. The first column in Table 2 defines a unique ID for each of the seven patterns. The second and third columns specify the transaction phase and the Big Data source, respectively. The fourth column specifies the REA primitive. Together, the third and fourth columns refer to intersections in Table 1. The key-value "mappings" in column 5 illustrate supplemental data regarding business transactions that can be provided by Big Data. The sixth column provides examples of aggregate functions that can be applied to "reduce" the key-values for the purpose of subsequent analyses and, thus, decision making.[6] Most of the patterns in Table 2 are self-explanatory, but we will discuss a number of them next for illustrative purposes.

Consider pattern #1 in Table 2, representing the association between "website logs" and "offering." For example, there could be a web advertising campaign (offering) that directs current and potential customers to the company's website, for which "click-through" events would be uniquely identifiable. The "map" function in MapReduce (column 5 in Table 2) could parse the website logs to extract the clicks associated with the specific offering. From each log entry, the map function could also extract the timestamp and the HTTP "referrer," which indicates the source of the click. The "reduce" function (column 6 in Table 2) would aggregate all outputs of the map function to generate the total click count for the offering, the click count within a user-specified time window, and the count by referrer group (e.g., Google search, Yahoo search, Bing search, etc.). From a decision-making perspective, such information would be helpful for the economic evaluation of the effectiveness of offerings.

One measure of the effectiveness of offerings, such as advertising campaigns, is the number of proposals generated. Website logs can be used to trace proposals—e.g., a request for a quote—to offerings. Linking website click-throughs that resulted in proposals would provide an indication of the "yield" for offerings. As pattern #4 in Table 2 shows, aggregated information that can be determined includes: total proposals generated by an offering, i.e., the yield; proposals generated in a specific time window; total number of proposals per referrer.

Information patterns are driven by the data source type and its specific implementation. For example, as illustrated by pattern #2 in Table 2, for Facebook, the map function extracts clicks, the associated timestamps, "likes," and "shares." Alternatively, for Twitter, the map function extracts "followers" and "retweets." The reduce function, in turn, applies a number of aggregate functions to generate total counts and counts within user-specified time windows for each type of social media activity.

Location and existence logs generated by sensors (i.e., RFID tag data and GPS data) can observe the physical movement of resources. As shown in Table 2 by pattern #6, the following information can be collected per item (from sensors): the item's type and where (location) it is at a specific moment. The reduce function can then apply aggregate functions to these data to generate information such as the number of items available per category (e.g., per car model), the number of items available per location (e.g., dealership), and to trace the distance traveled per item within a user-specified time window.

Text data sources such as social media and communication repositories are often used for collecting opinions. For example, as shown by Table 2, pattern #3, such sources help to understand whether, for example, a product is being discussed (on social media), frequency of discussions (per communication type), whether the discussions are extensive, and the tone of the discussions. The reduce function can then generate aggregate statistics regarding the count of communication log entries, the average communication entry length, and the count of positive, negative, and neutral comments. Such data can be useful for understanding where further exploration is needed. A similar analysis can also be done for agents, both internal and external. For example, parsing call information and emails might help understand whether the overall tone of conversations with an agent is positive, negative, or neutral.

V. TRANSACTIONAL AND BIG DATA ANALYTICS FOR DECISION MAKING IN ACCOUNTING

Having discussed the various sources of external Big Data most amenable to integration with the AIS, and the mechanism for linking such data sources with accounting transactional data using technologies such as Hadoop and MapReduce, we now consider how such information can be leveraged for decision-making purposes. Specifically, we consider the kinds of analytics that can be routinely executed to extract insights from the integration of relatively structured external Big Data with highly structured internal transactional data for decision making in accounting.

Table 3 presents seven different analytics patterns. The first column groups the patterns in terms of analytics relating to transaction, agent, and resource. The second and the third columns define, for each pattern, a unique ID and a name that

[6] The "reduce" functions in Table 2, column 6 are similar in nature to the REA-based queries discussed in Dunn (2013). However, there are two major differences. First, they are defined as MapReduce functions, as opposed to SQL queries. Second, their focus is on non-transactional data such as sentiment analysis (e.g., Count(Like) and Count(Follow)).

American Accounting Association

TABLE 2

MapReduce-Based Information Extraction Patterns for External Big Data Source-REA Primitive Associations

#	Business Transaction Phase	Big Data Source	REA Primitive	MapReduce → Map {Key → Value}	MapReduce → Reduce {Key → AggregateFunction}
1	Planning	Website Logs	Offering	{OfferingID → Click} {OfferingID → TimeStamp} {OfferingID → Referrer}	{OfferingID → Count(Click)} {OfferingID → Count(starttime,endtime,Click)} {OfferingID → Count(Referrer)}
2	Planning	Social Media	Offering	{OfferingID → Click} {OfferingID → TimeStamp} {OfferingID → Like} {OfferingID → Share} {OfferingID → Follow} {OfferingID → Retweet}	{OfferingID → Count(Click)} {OfferingID → Count(starttime,endtime,Click)} {OfferingID → Count(Like)} {OfferingID → Count(starttime,endtime,Like)} {OfferingID → Count(Share)} {OfferingID → Count(starttime,endtime,Share)} {OfferingID → Count(Follow)} {OfferingID → Count(starttime,endtime,Follow)} {OfferingID → Count(Retweet)} {OfferingID → Count(starttime,endtime,Retweet)}
3	Planning	Communication Repositories	Resource Type	{ResourceTypeID → CommunicationType} {ResourceTypeID → CommunicationLength} {ResourceTypeID → Communication Tone}	{ResourceTypeID → Count(CommunicationType)} {ResourceTypeID → Avg(CommunicationLength)} {ResourceTypeID → Count(CommunicationTone="Positive")} {ResourceTypeID → Count(CommunicationTone="Negative")} {ResourceTypeID → Count(CommunicationTone="Neutral")}
4	Identification	Website Logs	is-related-to	{OfferingID → ProposalID} {ProposalID →TimeStamp} {ProposalID → Referrer}	{OfferingID → Count(ProposalID)} {ProposalID → Count(starttime,endtime,ProposalID)} {ProposalID → Count(Referrer)}
5	Negotiation	Website Logs	Commitment	{CommitmentID → Click} {CommitmentID → TimeStamp}	{CommitmentID → Count(Click)} {CommitmentID → Count(starttime,endtime,Click)}
6	Actualization	Sensor Logs	Resource	{ResourceID → ResourceTypeID} {ResourceID → TimeStamp} {ResourceID → LocationID}	{ResourceTypeID → Count(ResourceID)} {LocationID → Count(ResourceID)} {ResourceID → Distance(starttime,endtime)}
7	Post-Actualization	Social Media	Resource Type	{ResourceTypeID → Comment} {ResourceTypeID → CommentLength} {ResourceTypeID → CommentTone}	{ResourceTypeID → Count(Comment)} {ResourceTypeID → Avg(CommentLength)} {ResourceTypeID → Count(CommentTone="Positive")} {ResourceTypeID → Count(CommentTone="Negative")} {ResourceTypeID → Count(CommentTone="Neutral")}

TABLE 3

Transactional and Big Data Analytics for Decision Making in Accounting

#	Pattern	Data Source	REA Primitive	Analytics	Decisions
	Transaction				
1	Transaction Analytics	Transaction System	Offering, Proposal, Commitment, Economic Event, Post-Exchange Event	• Transaction frequency by time window. • Transaction monetary amount comparison to prior transactions.	• Monitoring transaction volumes and size ($ amount), including comparison across periods
2	Transaction Sequence Analytics	Transaction System	is-related-to, results-in, fulfills, follows	• Number of inter-phase associations (frequency); e.g., number of proposals generated per offering (is-related-to) • Inter-phase latency; e.g., latency between offering and proposal. • Intra-phase latency; e.g., time to complete an exchange—example: invoice to cash.	• Monitoring success rate (yield) • Monitoring latency issues
3	Transaction Interaction and Reaction Analytics	Website Logs Social Media	Offering, Proposal, Commitment Offering, Proposal, Commitment, Economic Event, Post-Exchange Event	• Number of (frequency) website clicks. • Latency between transaction-related social media clicks. • Comparison of like counts, share counts, follower counts, and retweet counts for current event time window to previous event time windows.	• Predicting number and size of transactions • Economic sensing
		Communication Repositories	Offering, Proposal, Commitment, Economic Event, Post-Exchange Event	• Comparison of communication type counts. • Comparison of average communication length. • Comparison of positive, negative, and neutral comment counts.	
	Agent				
4	Agent Analytics	Transaction System	target, responsible, participates	• Latency between successive transactions by internal agent. • Transaction frequency grouped by internal agent. • Transaction monetary amount comparison by internal agent. • Latency between successive transactions by external agent. • Transaction frequency grouped by external agent. • Transaction monetary amount comparison by external agent.	• Monitoring internal agent (employee) effectiveness • Monitoring growth (frequency and volume) for external agents and market segments • Monitoring external agent loyalty • Agent prioritization
		Intranet/Extranet Logs	target, responsible, participates	• Latency between successive transactions by internal agent.	

(continued on next page)

TABLE 3 (continued)

#	Pattern	Data Source	REA Primitive	Analytics	Decisions
5	Agent Interaction and Reaction Analytics	Website Logs	External Agent	• Agent-specific number (frequency) of website clicks. • Comparison of click counts between time windows.	• Monitoring internal agent (employee) effectiveness • Agent profiling • Predicting market growth • Relationship building • Social network management
		Intranet/Extranet Logs	Internal Agent	• Agent-specific number (frequency) of website clicks. • Comparison of click counts between time windows.	
		Social Media	Internal Agent, External Agent	• Number of organization-related communications (frequency). • Comparison of positive, negative, and neutral comment counts.	
		Communication Repositories	Internal Agent, External Agent	• Identification of social relationships.	
	Resource				
6	Resource Analytics	Transaction System	subject-of, reserves, stockflow	• Proposal frequency by resource type. • Proposal monetary amounts by resource type. • Resource inflow/outflow latency. • Resource inflow/outflow quantity by time-window.	• Inventory Management
		Location/Existence Sensor Log	stockflow	• Resource inflow/outflow latency. • Resource inflow/outflow quantity by time-window.	
7	Resource Interaction and Reaction Analytics	Website Logs	subject-of, reserves	• Resource-Type-specific number (frequency) of website clicks. • Comparison of click counts between time windows.	• Predicting market growth • Economic sensing
		Social Media	subject-of, reserves, stockflow	• Comparison of like counts, share counts, follower counts, and retweet counts for current event time window to previous event time windows.	
		Communication Repositories	subject-of, reserves, stockflow	• Comparison of communication type counts. • Comparison of average communication length. • Comparison of positive, negative, and neutral comment counts.	

describes the nature of the analytics, respectively. The fourth column indicates the data sources for each pattern. The fifth column indicates the REA primitives associated with each pattern, and the sixth column shows representative analytics for each pattern. The last column shows types of decisions that could benefit from specific patterns. Through the "REA Primitive" column, Table 3 can be related to the ISO (2007) business transaction phases shown in Figure 1. The main objective of Table 3 is illustrating how various kinds of analytics can extend decision making in accounting.

Table 3 presents two different types of analytics patterns. The first group of analytics patterns (1, 2, 4, and 6) have transaction systems as their main data source. They can be categorized as *descriptive* and the consumers for such analytics will be primarily accountants and both internal and external auditors. The information integrity threshold for such analytics is often high, given the cost of a decision error, implying that a relatively high degree of precision would be required for decisions to be made based on them. The analytics related to those patterns are often used to spot anomalies in accounts and transactions. The external auditor can use such analytics to identify areas that warrant additional audit attention. Table 3 further illustrates how external sources of relatively structured data might provide additional descriptive analytics based on more granular data. For example, for agent analytics (#4), intranet logs are able to provide a much more detailed picture of an employee's activities and, thus, effectiveness. Similarly, for resource analytics (#6), sensing technologies such as RFID are able to precisely track a resource's location, which results in advanced inventory management systems that are able to indicate, among others, the need for future movement of inventory to warehouses that are experiencing higher inventory depletion rates.

The second group of analytics patterns in Table 3 (3, 5, and 7) are more akin to *predictive* and *prescriptive* analytics. For example, positive sentiments in social media activity that can be associated with a particular product would suggest increased sales for that product. For such predictive and prescriptive analytics, the information integrity expectations are often lower, meaning that there is an understanding of the lack of precision when they are considered as part of business decisions.

It should further be noted that the same analytics pattern can be applied to different primitives. For example, in Table 3, transaction analytics (#1) applies to five different REA primitives: offering, proposal, commitment, economic event, and post-exchange event. Furthermore, analytics often span multiple REA primitives; e.g., one would like to know the number of (frequency) post-exchange events for a specific offering. Next, we discuss some of the specifics of the transaction, agent, and resource analytics patterns.

Transaction

Pattern #1 in Table 3, transaction analytics, pertains to the frequency and size (monetary value) of each transaction type, from offerings to post-exchange events. With the exception of post-exchange events, increasing trends signal good news. The main data source for pattern #1 is the transaction system. Pattern #2, transaction sequence analytics, monitors how business transactions evolve over time and whether initial investments result in actual exchanges. Specific analytics involve calculation of the latency (time lag) between successive transactions and yield; i.e., success in moving a business transaction to the next phases, such as turning a proposal into commitments and its execution by means of economic events. The monitoring of the execution of economic exchanges in compliance with their commitments is at the heart of accounting information systems. Again, post-exchange events are the exception here. The main data source for pattern #2 is the transaction system. Pattern #3 goes beyond conventional transaction analytics. Website clicks are especially useful to measure interest, while social media and communication repositories allow more in-depth analysis of sentiments regarding, for example, offerings. Such economic sensing helps with predicting the number and size of transactions.

Agent

Pattern #4 in Table 3, agent analytics, aims at determining the effectiveness of internal agents and understanding the development of business relationships with external agents over time. Common analytics supported by transaction systems for both internal and external agents are latency between successive events and transaction frequency. For transactions that have economic impacts, the size—$ amount—is important to analyze, as well. For example, the monetary amount of a transaction can be compared to the average amount for that transaction type by the internal or external agent involved. These amount comparisons can involve the generation of Z-scores to identify transactions that are outliers (e.g., more than two or three standard deviations from the mean). As indicated above, such analytics can be refined by means of emerging data sources such as intranet/extranet logs.

Pattern #5 in Table 3 relies on external data sources and goes beyond traditional agent analytics. Website logs can be used to determine interest by external agents and, therefore, for building agent profiles and to predict market growth. Social media and communication repositories can be used for relationship-building purposes. A unique application of analytics involving external data sources is social network analysis. For example, Tyler, Wilkinson, and Huberman (2003) demonstrate how internal social networks can be uncovered from an organization's email corpus.

Resource

Resource analytics, pattern #6 in Table 3, involves categorizations by resource type: number of proposals and commitments per resource type, as well as the size of such transactions. In addition, it involves metrics dealing with the physical flow of the actual resources and how flow rates change over time, which are key to inventory management. As indicated above, sensors such as RFID technology enable real-time detailed descriptions of physical flows. Pattern #7 shows how external data sources extend inventory management by market analysis based on interest and sentiment information. Such analytics are often aggregated by resource type: fluctuations in interest and demand, and capturing of sentiments.

VI. SUMMARY AND FUTURE RESEARCH

In introducing the commentaries on Big Data in *Accounting Horizons,* Griffin and Wright (2015, 377) note "Big Data and business analytics permeate almost all aspects of major companies' decision making and business strategies." They further conclude by calling for research that demonstrates the contribution of Big Data for operations and decision making and how managers in firms could benefit from the insights provided by Big Data. This paper addresses this call for research by presenting a model for integrating relatively structured external Big Data sources and the structured transactional data within Accounting Information Systems. In explicating the model, we discuss three steps involved in the process of integrating the external Big Data sources with accounting system elements. The first step relates five distinct external sources of relatively structured data to the different phases of business transactions as defined by the REA ontology (ISO 2007) and, thus, provides a map of the extended datasets that are available for decision making in accounting. The second step focuses on information extraction, relying on Big Data application technologies such as Hadoop and MapReduce. Seven information patterns were presented that illustrate the process of generating useful information by integrating unstructured data from external data sources with structured internal (accounting) transactional data. The third step focuses on the use of information generated from external sources for decision making in accounting, and this was illustrated by means of a number of analytics patterns.

The primary contribution of this paper is in demonstrating how the formal application of the REA ontology can relate structured accounting transactional elements with external, but relatively structured, Big Data elements. The model and associated steps presented in this paper can form the basis for much future work on Big Data in accounting. First, future research could engage in a more in-depth analysis of alternative external data sources that are more unstructured (e.g., video and audio files), and the extent to which they can be integrated with the structured transactional elements in a traditional accounting system. Specifically, future research could explore the degree of processing necessary to sufficiently structure the unstructured Big Data elements to make the data amenable to connection with internal transactional data. Second, in a design science vein, researchers could construct actual implementations using technologies such as Hadoop and MapReduce to show how Big Data analytics can be materialized. A wide range of implementation issues exist that need to be addressed. For example, future research could examine the computational demands of applying Big Data analytics in near-real time. The simulation-based approach to evaluate the systems performance impact of continuous monitoring controls presented in Murthy (2004) is a good example of such research. Third, the analytics patterns presented in this paper are illustrative in nature and a more extended set should be developed. Finally, leveraging the work of Kogan, Alles, Vasarhelyi, and Wu (2014), it could be fruitful to explore the development of a system of continuity equations relating Big Data activity to each of the five business transaction phases in succession, i.e., from planning to identification to negotiation to actualization to post-actualization.

REFERENCES

Baraniuk, R. G. 2011. More is less: Signal processing and the data deluge. *Science* 331 (6018): 717–719. doi:10.1126/science.1197448

Barbier, G., and H. Liu. 2011. Data mining in social media. In *Social Network Data Analytics,* 327–352. New York, NY: Springer.

Borthick, A. F. 2012. Designing continuous auditing for a highly automated procure-to-pay process. *Journal of Information Systems* 26 (2): 153–166. doi:10.2308/isys-50233

Borthick, A. F., P. L. Bowen, and G. J. Gerard. 2008. Modeling a business process and querying the resulting database: Analyzing RFID data to develop business intelligence. *Journal of Information Systems* 22 (2): 331–350. doi:10.2308/jis.2008.22.2.331

Dean, J., and S. Ghemawat. 2008. MapReduce: Simplified data processing on large clusters. *Communications of the ACM* 51 (1): 107–113. doi:10.1145/1327452.1327492

Denna, E. L., and W. E. McCarthy. 1987. An events accounting foundation for DSS implementation. In *Decision Support Systems: Theory and Application,* 239–263. Berlin, Germany: Springer.

Dunn, C. 2013. REA Accounting Systems: Resources-Events-Agents: An Ontology for Designing, Controlling, and Using Integrated Enterprise Systems. New York, NY: McGraw-Hill Education.

Geerts, G. L., and W. E. McCarthy. 1999. An accounting object infrastructure for knowledge-based enterprise models. *IEEE Intelligent Systems & their Applications* 14 (4): 89–94. doi:10.1109/5254.784089

Geerts, G. L., and W. E. McCarthy. 2002. An ontological analysis of the economic primitives of the extended-REA enterprise information architecture. *International Journal of Accounting Information Systems* 3 (1): 1–16. doi:10.1016/S1467-0895(01)00020-3

Geerts, G. L., and W. E. McCarthy. 2006. Policy-level specification in REA enterprise information systems. *Journal of Information Systems* 20 (2): 37–63. doi:10.2308/jis.2006.20.2.37

Geerts, G. L., and D. E. O'Leary. 2014. A supply chain of things: The EAGLET ontology for highly visible supply chains. *Decision Support Systems* 63: 3–22. doi:10.1016/j.dss.2013.09.007

Griffin, P., and A. Wright. 2015. Commentaries on Big Data's importance for accounting and auditing. *Accounting Horizons* 29 (2): 377–379. doi:10.2308/acch-51066

Hümmer, W., W. Lehner, and H. Wedekind. 2002. Contracting in the days of eBusiness. *SIGMOD Record* 31 (1): 31–36. doi:10.1145/507338.507345

IBM. 2012. *The Four V's of Big Data*. Available at: http://www.ibmbigdatahub.com/infographic/four-vs-big-data

International Standards Organization. 2007. ISO 15944-4—Information Technology—Business Operational View—Part 4: Business Transaction Scenarios—Accounting and Economic Ontology. Geneva, CH: ISO.

Kaplan, R., and D. Norton. 1996. Using the balanced scorecard as a strategic management system. *Harvard Business Review* 74 (1): 75–85.

Kogan, A., M. G. Alles, M. A. Vasarhelyi, and J. Wu. 2014. Design and evaluation of a continuous data level auditing system. *AUDITING: A Journal of Practice & Theory* 33 (4): 221–245.

Krahel, J. P., and W. R. Titera. 2015. Consequences of Big Data and formalization on accounting and auditing standards. *Accounting Horizons* 29 (2): 409–422. doi:10.2308/acch-51065

Laney, D. 2001. *3D Data Management: Controlling Data Volume, Velocity, and Variety*. META Group, Gartner. Available at: http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Maes, P., R. H. Guttman, and A. G. Moukas. 1999. Agents that buy and sell. *Communications of the ACM* 42 (3): 81–83. doi:10.1145/295685.295716

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

Moffitt, K. C., and M. A. Vasarhelyi. 2013. AIS in an age of Big Data. *Journal of Information Systems* 27 (2): 1–19. doi:10.2308/isys-10372

Murthy, D., A. Gross, A. Takata, and S. Bond. 2013. Evaluation and development of data mining tools for social network analysis. In *Mining Social Networks and Security Informatics,* Lecture Notes in Social Networks, edited by Ozyer, T., Z. Erdem, J. Rokne, and S. Khoury. New York, NY: Springer. doi:10.1007/978-94-007-6359-3_10

Murthy, U. S. 2004. An analysis of the effects of continuous monitoring controls on e-commerce system performance. *Journal of Information Systems* 18 (2): 29–47. doi:10.2308/jis.2004.18.2.29

O'Leary, D. E. 2004. On the relationship between REA and SAP. *International Journal of Accounting Information Systems* 5 (1): 65–81. doi:10.1016/j.accinf.2004.02.004

O'Leary, D. E. 2013. "Big Data," the "Internet of Things" and the "Internet of Signs." *Intelligent Systems in Accounting, Finance and Management* 20 (1): 53–65. doi:10.1002/isaf.1336

Shvachko, K., H. Kuang, S. Radia, and R. Chansler. 2010. The Hadoop distributed file system. In *Proceedings of the IEEE Symposium on Massive Storage Systems and Technologies*. Los Alamitos, CA: IEEE.

Tyler, J. R., D. M. Wilkinson, and B. A. Huberman. 2003. Email as spectroscopy: Automated discovery of community structure within organizations. In *Communities and Technologies*, 81–96. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Vasarhelyi, M. A., A. Kogan, and B. Tuttle. 2015. Big Data in accounting: An overview. *Accounting Horizons* 29 (2): 381–396. doi:10.2308/acch-51071

Vlosky, R. P., R. Fontenot, and L. Blalock. 2000. Extranets: Impacts on business practices and relationships. *Journal of Business and Industrial Marketing* 15 (6): 438–457. doi:10.1108/08858620010349510

Warren, J. D., Jr., K. C. Moffitt, and P. Byrnes. 2015. How Big Data will change accounting. *Accounting Horizons* 29 (2): 397–407. doi:10.2308/acch-51069

Yoon, K., L. Hoogduin, and L. Zhang. 2015. Big Data as complementary audit evidence. *Accounting Horizons* 29 (2): 431–438. doi:10.2308/acch-51076

American Accounting Association