INTELLIGENT SYSTEMS IN ACCOUNTING, FINANCE AND MANAGEMENT
Intell. Sys. Acc. Fin. Mgmt. **20,** 53–65 (2013)
Published online in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/isaf.1336

'BIG DATA', THE 'INTERNET OF THINGS' AND THE 'INTERNET OF SIGNS'

DANIEL E. O'LEARY*
University of Southern California, Los Angeles, CA, USA

SUMMARY
This paper examines the relationship between so-called 'Big Data', the 'Internet of Things' (the 'Internet of People and Things,' and the 'Internet of Everything'), and the 'Internet of Signs.' In particular, we investigate how the 'things' in the 'Internet of Things' generate 'Big Data', and how both are used to generate semiotic 'signs'. In addition, we analyse the importance of context in and the relationships between 'Big Data', the 'Internet of Things', and the 'Internet of Signs'. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: 'Big Data'; 'Internet of Things'; 'Internet of Signs'; context; 'Internet of People and Things'; 'Internet of Everything'; Semiotics; Big Context

1. INTRODUCTION

The purpose of this paper is to investigate the integration of three emerging concepts: 'Big Data', the 'Internet of Things' and the 'Internet of Signs'. Ultimately, this paper argues that the 'Internet of Things' and people, etc. generates 'Big Data' and that 'Big Data' and the 'Internet of Things' can be used to generate an 'Internet of Signs'. In addition, each of these three interrelated concepts is examined for their effect on context and the information available about context.

Apparently, Cox and Ellsworth (1997) were among the first to use the term "big data" literally referring to using larger volumes of data for visualization of scientific data (the term "large data" also has been used). (Diebold (2012) was among the first to use the term in statistics and econometrics in roughly 2000.) At that time, the term literally referred to bigger data sets than had been the norm. However, since that time the term has evolved to include a range of characteristics, such as integrating different types of data and analyses. Accordingly, along the evolutionary trail a number of different sources have developed definitions of so-called 'Big Data'. In this paper we review some of those definitions and summarize some of the similarities among them. We will use those definitions to review the contributions of some applications as 'Big Data'.

As noted by Ashton (2009), the term the 'Internet of Things', apparently developed in 1999, initially was meant to describe the following situation:

> Today computers – and, therefore, the Internet – are almost wholly dependent on human beings for information. . . . The problem is, people have limited time, attention and accuracy – all of which means they are not very good at capturing data about things in the real world. . . . We need to empower computers with their own means of gathering information, so they can see, hear and smell the world for themselves . . .

As a result, the 'Internet of Things' provides a linked set of computer programs and sensors that do not incur the same limitations of people. However, in some contexts it appears that the 'Internet of

* Correspondence to: Daniel E. O'Leary, University of Southern California, Los Angeles, CA, USA. E-mail: oleary@usc.edu

Things' is beginning to be configured to include inputs from humans, linked to the internet, ultimately being referred to as the 'Internet of Everything' (e.g., SRA, 2009). In this paper we review some of the evolution of the original concept and use it as a tool to review context.

Further, this paper elicits the notion of the 'Internet of Signs' (O'Leary, 2012a). The 'Internet of Signs' indicates that the data generated on the internet from the broad range of sources, including devices in the 'Internet of Things', information from social media (e.g. blogs) and other internet sources (often associated with 'Big Data'), provide 'signs', such as the 'sentiment' toward some issue (e.g. O'Leary, 2011). Those 'signs' generated from information associated with the internet provide an 'Internet of Signs'. That 'Internet of Signs' can be helpful in providing potential information about events and situations.

1.1. This Paper

This paper proceeds as follows. Section 2 examines the notion of 'Big Data'. Section 3 analyses the 'Internet of Things'. Section 4 briefly reviews semiotics and drills down on the notion of the 'Internet of Signs'. Section 5 analyses notions of context, and the contribution of 'Big Data', the 'Internet of Things' and the 'Internet of Signs' to facilitating context. Section 6 summarizes some potential applications of the integration of these three concepts. Section 7 briefly summarizes the paper, examines its contributions and investigates some extensions.

2. 'BIG DATA'

'Big Data' is a term that has evolved to account for the rapidly expanding amounts of digital information that are being generated, the efforts to make that information analysable and the actual use of that data as a means to improve productivity, generate and facilitate innovation and improve decision making. There are a number of definitions that have been offered for 'Big Data'. As an example, Gartner (http://www.gartner.com/it-glossary/big-data/) defined 'Big Data' as '. . . high volume, velocity and variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making'. This section summarizes three other detailed discussions from leading vendors actually offering services in 'Big Data', including IBM, Teradata and EMC.

2.1. IBM – Volume, Velocity, Variety, Veracity, Value

Zikopoulos *et al*. (2012), in an IBM publication, describe 'Big Data' as consisting of a set of three 'V-words': volume, velocity and variety. Volume indicates that there are increasing amounts of data over traditional settings. Velocity suggests that information is being generated at a rate that exceeds those of traditional systems. Variety is indicative of there being multiple emerging forms of data that are of interest to enterprises. For example, Twitter and other social media have become a source of big data. In mid 2010, Twitter tweets hit 65 million per day and there were 190 million users (Schonfeld, 2010). These three 'V-words' provide the classic definition for 'Big Data' applications.

In a subsequent IBM publication, Zikopoulos *et al*. (2013) introduced the additional concepts of veracity and value into 'Big Data'. Veracity refers to the accuracy, truthfulness and reliability of the data. This factor, although clearly desirable, is difficult to ensure, particularly with data generated from multiple sources. 'Value' refers to the potential for big data to provide a cost-beneficial addition to an enterprise's technology portfolio. Of course, ultimately, 'value' provides the key cost-beneficial criteria in terms of

DOI: 10.1002/isaf

determining whether or not 'Big Data' should be used. However, development of 'Big Data' also requires an infrastructure to support gathering, storing, processing and using the accumulated information.

2.2. Teradata: Diverse Data, Different Data Structures and Analytics

Teradata's Bawa (2011) suggests that the number of applications that are generating data is exploding and that the number of programs being written to consume that data is growing rapidly. In addition, Bawa (2011) noted that the structure of that diverse data is likely to be highly variant and a function, in part, of the program generating it. Further, with that growing amount of data there is now a need for a massive amount of analysis, and that analysis is likely to vary based on the data.

2.3. EMC: Varied and Unstructured Data Needs Rapid Analysis

EMC has a perspective grounded in 'Big Data' as a service (BDaaS). However, EMC also notes data variety, data complexity and less structured data. In addition, EMC stresses that the data must be analysed more rapidly than in the past. In particular, as noted in EMC (2012: 7):

> As data is increasingly becoming more varied, more complex and less structured, it has become imperative to process it quickly. Meeting such demanding requirements poses an enormous challenge for traditional databases and scale-up infrastructures. . . . Big Data refers to new scale-out architectures that address these needs. Big Data is fundamentally about massively distributed architectures and massively parallel processing using commodity building blocks to manage and analyze data.

2.4. Definitions of 'Big Data'

This section summarized a number of vendor-oriented definitions of 'Big Data'. Those definitions appear highly correlated, and are consistent in stressing that 'Big Data', as currently conceived, is more than just 'more' data. 'Big Data' also includes the increased speed with which data are generated and with which enterprises must respond. 'Big Data' definitions also seem to consistently note the complexity, diversity and unstructured nature of the data generated. Accordingly, the analysis of such diverse and varied data will also be diverse and varied, since it takes equivocality to remove equivocality (Ashby, 1965). Finally, the definition from EMC stresses the need for changes in information technology architectures, as there is move toward more parallel processing.

3. INTERNET OF THINGS

Chui *et al*. (2010) define the 'Internet of Things' as '. . . sensors and actuators embedded in physical objects – from roadways to pacemakers – are linked through wired and wireless networks, often using the same Internet Protocol (IP) that connects the Internet'. The 'Internet of Things' generally refers to the notion that many different 'things' are connected to the internet and thus can be connected to each other.

'Things' can be sensors, databases, other devices or software. Sensors could include pacemakers, location identifiers, such as global positioning system (GPS), and individual identification devices, such as radio-frequency identification (RFID) tags. Sensors can provide different information characteristics, typically of interest in the particular setting. For example, RFID sensors may indicate time and location,

DOI: 10.1002/isaf

pacemakers capture information about heart rate; other sensors may capture the status of the item the sensor is monitoring, the number of automobiles, the presence of an RFID tag and other information.

'Things' can be intelligent and aware of other 'Things'. As a result, in some cases 'Things' will want or need to communicate with other 'things'. One 'Thing' might find the location of a related or interesting 'Thing' and initiate a dialogue, gather information from each other and communicate implications of that information to some decision maker. For example, tagged vats of chemicals that could spontaneously combust if placed adjacent to each other, could communicate that conclusion to some decision maker to facilitate the safe storage of those chemicals.

'Things' can gather information and knowledge from their interaction with other 'things'. Things can either save that information and knowledge locally or they can relay it to some location in the 'Cloud', where the information would be broadly available for others. For example, in hospital usage of RIFD in patients, the tag typically only contains patient number, and patient information is kept on a secured server. Similarly, uses of RFID in automobiles often include tag number only (e.g. toll devices). These uses suggest multiple information technology architectures to store, analyse and process 'thing' information.

'Things' are potentially autonomous, semi-autonomous or not autonomous. However, as 'things' are networked they can become more autonomous, as they interact with other 'things'. Further, the composite of the network and 'things' can be more than the individual 'things' as 'network effects' develop among the 'things', where the network information ultimately is greater than the information associated with any one 'thing'. For example, O'Leary (2006, 2008) investigated notions of the development of autonomic supply chains that combine many different data sources and capabilities.

3.1. 'Internet of Things' Generates 'Big Data'

The 'Internet of Things' can generate 'Big Data' for a number of reasons. The volume of data attributable to the 'Internet of Things' is substantial. As sensors interact with the world, 'Things' such as RFID tags generate volumes and volumes of data. As a result, digital processing becomes a requirement of feasibility. The velocity of data associated with the 'Internet of Things', compared with traditional transaction processing, explodes as sensors can continuously capture data.

The variety of data associated with the 'Internet of Things' also is expansive as the types of sensors and the different sources of data expand. The veracity of data in the 'Internet of Things' may also be improving as the quality of sensor and other data improves over time. For example, use of RFID tags generates much more reliable information than a decade ago. Such high volumes of data, coupled with an increasing velocity of data, along with an increased variety of data, illustrate the push by the 'Internet of Things' to generate 'Big Data'.

3.2. Evolution of the 'Internet of Things'

Probably, it is unnecessarily limiting to talk about an 'Internet of Things' at the exclusion of people (or other dimensions), particularly in a world where many 'things' are automations of people and much of the work of 'things' is for or about people. Further, 'things' and the quality of the information that they produce are affected by people. As a result, it is important to have a concept that is bigger than just 'things'.

One approach would be to extend the 'Internet of Things' to the 'Internet of People and Things', providing a larger base of connections and relationships (UK Future Internet Strategy Group, 2011). People-based information could include sensors 'representing' people; for example, capturing their

DOI: 10.1002/isaf

location or other variables. In addition, people-based information could include social media, providing additional context information. In that setting, the 'Internet of People and Things' would provide access to and connection with other entities in a relevant context. Further other researchers have begun to talk about the "Internet of Everything" (SRA, 2009), where virtually everything is connected to the Internet and can communicate with everything else. However, as in SRA (2009) we will focus on the term "Internet of Things" for the rest of the paper.

Consider an example of the integration of sensor and people data. There are a number of sensors available to inform drivers about traffic that have various manifestations on the internet. For example, Sigalert.com provides sensor-based analysis of traffic density on highways. In addition, there is qualitative information about traffic. For example, 'Waze' is an application of social media that generates social data about traffic. Users can provide information about hazards, traffic density, location of police and other data. Combining Waze with sensor-based media provides users with a unique view of traffic and the context (hazards, police, etc.). In addition, by using data from both, the veracity of the data can be improved, ultimately providing improved value to the user.

4. SEMIOTICS AND THE INTERNET OF SIGNS

The purpose of this section is to briefly review semiotics and discuss how the 'Internet of Things' and 'Big Data' generate the 'Internet of Signs'.

4.1. Semiotics

Semiotics is the study of signs, with a history that includes Greek philosophers, Ferdinand de Saussure, Charles Sanders Peirce, Joseph Schumpeter and others (Dorsey, 2003). However, those signs are generally tied to social and cultural concerns. Chandler (2009) notes that semiotics is '. . . a science which studies the role of signs as part of social life'. Further, Culler (2005: 28) suggests that with semiotics we think of 'our social and cultural world as a series of sign systems . . .'.

Many different phenomena can generate signs. Eco (1978), in his *A Theory of Semiotics*, developed a range of topics in semiotics, including formalized languages, written languages, text theory, mass communication and other issues. Accordingly, signs permeate communication. Further, semiotics occurs within specific disciplines, including, for example, medical semiotics (Eco, 1978) and economics, such as Schumpeter's analysis of entrepreneurism and business cycles that apparently paralleled Peirce's semiotic research (Swedberg, 2012). However, signs also may be generated from nonlinguistic information. For example, visual communication also is included by Eco (1978).

Semiotics has been used to capture multiple types of signs (e.g. 'symptom signs'). For example, semiotics can be used to investigate apparent 'symptoms' of behaviour and events. Culler (1981: 30) noted as part of semiotics that 'Someone . . . might investigate symptoms as signs of prior causes and seek to reconstruct a history of . . . events'. 'Symptoms' provide 'signs' of potential causes of behaviours and events. Further, changes in those 'symptoms' also may be 'signs'. As an example, symptoms signs may include blogs that suggest that there is something wrong with some product. Accordingly, there is interest in determining how data can be used to generate insights into those symptom and signs. For example, based on available information, there can be 'signs' of bankruptcy, fraud or reputation (Spangler *et al*., 2009).

Although there is an interest in symptoms, Culler (2005: 43) noted 'semiotics attempts to make explicit the knowledge which enables signs to have meaning . . .'. That explicit knowledge has been generated using multiple categories of information. For example, Eco (1978) has offered a number of

DOI: 10.1002/isaf

categories of signs, including written language, natural language, cultural codes, aesthetic codes, codes of tastes and a number of others. Making the knowledge explicit can require multiple types of analysis. To facilitate use and integration with 'Big Data' and the 'Internet of Things' generally requires development of explicit knowledge. For example, explicit knowledge might be teased from blogs or other social media (e.g. O'Leary, 2011).

4.2. Things Generate Signs

Increasingly there is information generated from computer-based devices (e.g. sensors) for people to monitor devices, often automating tasks previously done by people. Accordingly, one perspective in this paper is that signs can be generated by those 'things'. In particular, 'Things' generate a wide range of measurements that can be used to provide insight into signs. For example, sensors indicating no available parking places provide a sign that a facility is busy or that some event is going on.

Although there has been a limited analysis of 'things' in semiotics, the concern has not been with the 'things' but instead with the signs, symbols and concepts that the 'things' present and represent. For example, as noted by Langer (1942):

> Symbols are not proxy for their objects but are vehicles for the conception of objects . . .. In talking about things we have conceptions of them, not the things themselves; and it is the conceptions, not the things, that symbols directly mean.

We interact with the conceptions and representations of the 'things'.

In addition, 'things' ultimately generate the information on which signs about those 'things' can be based. As an example, in economics, semiotics has used metrics based on issues such as price trends, sentiment and even slogans as a means to investigate phenomena such as behaviour and events in the stock market (Dorsey, 2003). Ultimately, either directly or indirectly, 'things generate signs'.

Further, in semiotics, in analysis of Saussure, Chandler (2009) notes 'primacy is given to relationships rather than to things – the meaning of signs is seen as lying in their systematic relation to each other, rather than deriving from any inherent reference to material things'. From the perspective of semiotics, 'the relationships between things generate signs'. Similarly, in the 'Internet of Things', relationships between 'Things' play a key role in defining signs. The 'Internet of Things' would provide substantial data about those relationships. Thus, there are data about the 'Things' and data about relationships between 'Things' generating 'Big Data', as discussed in Section 4.3.

As a result, from the perspective of semiotics, rather than concern for an 'Internet of Things' there is concern or interest in what I would call the 'Internet of Signs'. In particular, how does the 'Internet of Things' manifest itself as 'signs' or the 'Internet of Signs' and what are the relationships between the 'things' and signs of 'things'? Ultimately, the relationships between 'things', conceptions of 'things' and symptoms of behaviours can provide a basis to better understand events, situations, behaviours and other issues.

4.3. Use Big Data to Generate Signs

Historically, semiotics has focused on human-generated information. Signs based on that human-generated information are embedded throughout the internet, generated from a wide range of other data besides sensors. For example, signs are in different internet media, such as blogs, wikis, comments, Twitter messages, Youtube and so on. Whatthetrend.com provides an analysis of Twitter messages to

DOI: 10.1002/isaf

see which issues are of most frequent occurrence. Yahoo.com provides a summary of what is 'trending now'. Such summaries of social media activity provide signs as to what is occurring or has occurred, including what events or situations are seen as important or interesting.

There are a number of contemporary examples of 'signs' from information on the internet as symptoms or problems. For example, Crowdsourcing (2012) examines 'New signs that Wikipedia began a long decline'. Similarly, before the American Football College Championship game between Notre Dame and Alabama in January 2013, it was reported on Twitter that the Alabama team had a player's only meeting, which was a sign that the players from Alabama were not 'focused'.

Researchers have taken the implicit information and knowledge available in blogs and so on and begun to determine the nature of 'signs' in those media, in order to make signs and knowledge about them more explicit. As an example, research has been done on finding signs of 'sentiment' in blogs or 'reputation' from information on the web (e.g. Spangler *et al*., 2009; O'Leary, 2011). Accordingly, these data sources capture indicators and potential signs of activity. This research can be interpreted as resulting in the generation of signs from what is sometimes seen as a data source in 'Big Data'.

5. CONTEXT

Context has been defined (Dictionary.com) as '. . . the set of circumstances or facts that surround a particular event, situation, etc.' First, consider the notions of circumstances or facts. In the case of 'Big Data' or the 'Internet of Things' there would be substantial data that could be used to characterize circumstances or facts. Such data would pre-occur, co-occur or post-occur with the particular event or situation. This would provide substantial data as a basis for characterization of context. Second, in the definition, context is defined around an event or situation. As a result, one approach would be to define a model of the world surrounding the particular events or situation that would help to define the context. For example, Schilit and Theimer (1994) defined context as consisting of location, identities of nearby objects and people, and changes to those objects. As another example, Schilit *et al*. (1994) suggest that the important part of context/events includes what resources are near (resources), who are you with (adjacent agents) and where are you located (location).

However, Dey (2001) suggests that those definitions are too specific and that it is difficult to enumerate the entire set of interesting variables, a priori. Accordingly, Dey (2001) suggests that

> Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.

As a result, Dey's (2001) definition is one that would be consistent with employing the substantial data available in 'Big Data' and the 'Internet of Things' to capture and analyse data about a particular event or situation.

5.1. 'Internet of Things' and Context

Context-based devices have been a concern of the 'Internet of Things' (Heil *et al*., 2007). As a result, the 'Internet of Things' provides potential for capturing and generating context about relevant 'things'. The context would be defined by the sensor readings for some 'thing', the sensor readings for related 'things' and the sensor readings for the interaction between related 'things'. 'Things' interact with other 'things',

DOI: 10.1002/isaf

and that interaction creates context. In addition, measures associated with 'things', such as time, date, location and other issues, create context data. However, it is still important to understand the 'things' being studied to ensure that the appropriate context information is captured and that the appropriately parsimonious model is built. For example, if the 'things' are RFID tags representing inventory, then we would have some understanding and expectations regarding behaviour because the tags are representing inventory.

5.2. 'Big Data' and 'Big Context'

At one level, context is defined by the set of 'pre-occurring', 'co-occurring' and 'post-occurring' data. As events occur, data are generated from a variety of sources. Capturing more and different data results in capturing more 'context'. If context is captured using all the available data, then 'Big Data' also should be able to provide 'Big Context'. In this setting, 'Big Context' would refer to access to a substantial amount of data, in different formats, from different sources, about a situation or event, but integrated and available for use.

Accordingly, 'Big Data' offers the opportunity to provide 'more' context than traditional settings. As a result, a recent development in 'Big Data' has been to try to embed data in context. As an example, as noted by Hernandez (2012), in the case of business transactions, the new perspective is to store 'each transaction in the context of the business activity such as make payment, search, or purchase, how well it performed, who initiated it, where the user was located, and more'.

In the case of business settings, there may be theoretic structures or schemas that can facilitate identification of the appropriate variables and the expected relationships between them (e.g. O'Leary, 1999). Thus, context identification variables are likely to require some consideration of the events, situations or settings of interest.

5.3. 'Internet of Signs' and Context

Some discussions of context in semiotics provide insight on context. For example, Eco (1981: 37) stresses the importance of contexts, noting that a sign only becomes '. . . fully meaningful when it is inserted within a larger context'. As another example, as also noted by Eco (1981: 45), ' . . . what I have to do is look for possible contexts capable of making the . . . expression intelligible and reasonable. The very nature of signs postulates an active role on the part of their interpreter'.

Semiotics suggests terms that make the information a part of the context: meaningful, intelligible and reasonable. 'Meaningful' suggests that there is a model of how the world functions that allows understanding of the data, both in local and more global contexts. 'Intelligible' indicates that the relationship between the data and the model can be understood. 'Reasonable' indicates that data behaviour in the model is appropriate.

5.4. 'Local' Context and 'Larger' Context

Constructs such as the 'Internet of Things' and the 'Internet of Signs' can facilitate the definition of both local and larger contexts. For example, the context could be defined as the set of other 'things' within some epsilon of the 'thing' of concern. Since the 'Internet of Things' forms a network, classic network approaches can be used to facilitate the analysis. For example, epsilon might be a function of paths from the specific 'thing' of some particular length or other graph theoretic measure. Alternatively, 'things' may be grouped accordingly to some model. For example, inventory 'things' might be grouped by truck (which could also be represented as a 'thing'), and trucks could be grouped by fleets and so on. Such models could have multiple relationships, such as cascading, item to truck to fleet. Such cascade groupings could be used to define local and larger contexts.

DOI: 10.1002/isaf

6. APPLICATIONS

There are a number of potential applications of the nexus of these concepts. This section briefly reviews some of those applications.

6.1. Continuous Monitoring of the 'Internet of Things' as a Source of 'Big Data'

Continuous monitoring is done in a broad base of disciplines: heart patients have a history of heart problems, inventory goods are monitored using RFID tags, hospital patients are managed using RFID tags, parking place availability is managed using a range of sensors, and so on. Such classic use of sensors is consistent with the historical concepts of the 'Internet of Things'.

Continuous monitoring of such data apparently meets the definitional requirements of 'Big Data'. The volume of information deriving from the tags is substantial generating 'Big Data'. Because there are multiple types of information, there are a variety of different types of information available to monitor. Finally, sensor information operates in real time, speeding the velocity, resulting in a continuous monitoring environment that is classic "Big Data".

6.2. Continuous Monitoring of Social Media as a Source of 'Big Data'

There also is continuous monitoring of human-based social media information for different purposes. For example, blogs and other information have been continuously monitored as a means of capturing information about a firm's reputation and brand (e.g. Spangler *et al*., 2009). In these types of applications, monitoring of social-media-generated information is done continually. The analysis is based on a wide range of sources and media, and is characterized by large volumes of data, high-velocity data (e.g. millions of Twitter messages every day) and a high variety of different types of data, qualifying this type of potential application as 'Big Data'.

6.3. Continuous Monitoring of Accounting Information as 'Internet of Things' and 'Big Data'

Continuous monitoring of information provides the ability to investigate the quality and validity of that information. Accordingly, there is interest by enterprises in analysing and auditing their accounting information for issues ranging from errors to potential fraud.

Transaction information for accounting systems can come from three basic sources. First, accounting information can be captured as part of the 'Internet of Things', using sensor-based information. For example, RFID tags can be used to track the sale and movement of inventory. Second, as part of enterprise resource planning systems, and other contemporary systems, accounting transaction entries may be automated. The software is programmed to develop transactions, based on timing or certain events. Third, human-based transactions can be manually initiated and entered.

The notion of the 'Internet of Things' suggests that monitoring and auditing the first two types of transactions would use different models than the third, because of the digitalization of automated transactions, compared with the human-entered transactions.

Further, recently there has been a suggestion to include a broader base of data while monitoring and auditing accounting transaction information, in the spirit, of 'Big Data'. For example, O'Leary (2012b) suggested integrating blogs, message boards and other types of information in the analysis of accounting data as a part of continuously monitoring financial information. This approach would require monitoring, integrating and analysing increased volumes of information, different structures of information

DOI: 10.1002/isaf

and more rapidly generated information. In particular, this would call for continuous financial assurance using 'Big Data' that consists of not only accounting transaction data, but also data from a larger context. That context could be defined by 'social media discussions by employees in finance and accounting', 'social media discussions by employees that mention financial information' or even 'social media discussions by anyone that mentions financial information of the particular company'. Those contexts would include multiple sources of data.

6.4. Generating and Monitoring Location-Based 'Big Data' and the 'Internet of Things'

Enterprises have a number of sources of information about location: RFID (on products, assets and badges), GPS, mobile phones and so on. Much of that information is continuously monitored. For example, badge locations are monitored so that the location of people is known and trucks are monitored as they journey to their destinations. Often, those are stand-alone applications and do not use additional context information from other related settings. Notions of 'Big Data' suggest integrating multiple data sets with the original data sets to broaden the context. In addition, the information in those applications could be used in other settings for alternative uses. For example, in the case of an asset management system, context could consider collocation of tagged assets and employee badges or mobile phones as an approach to effectively controlling the assets and minimizing asset losses.

6.5. Flash Mobs and Monitoring Location in the 'Big Data' and the 'Internet of Things'

Flash mobs are groups that spontaneously decide to meet at some location for some agreed-upon purpose. Tucker and Watkins (2011) examined how flash mobs are increasingly being used for criminal purposes, rather than 'dance parties'. Enterprises which could be damaged by flash mobs or law enforcement authorities could be interested in continuously monitoring for such activities.

Tucker and Watkins (2011) indicated that such mobs were often put together using social media. Since location is embedded in both social media devices and flash-mob meeting locations, a potentially important tool is continuously monitoring social media, such as Twitter, for evidence of criminal flash-mob activity and the corresponding locations of devices. Analysis of the potential development of a flash mob truly would require 'Big Data' and integration of social media with classic sensor information available from the 'Internet of Things'.

7. SUMMARY, CONTRIBUTIONS AND EXTENSIONS

This paper has investigated some of the relationships between three different sets of concepts: 'Big Data', the 'Internet of Things' and the 'Internet of Signs'. This paper also has investigated some reasons for how those concepts are interrelated and how those concepts relate to notions of context. Finally, this paper applied those concepts to a set of applications based in continuous monitoring.

7.1. Contributions

This paper argues that not only is there an 'Internet of Things', but there also is an 'Internet of Signs'. This paper also suggested that the 'Internet of Things' and 'Big Data' provide the data basis for the 'Internet of Signs'.

DOI: 10.1002/isaf

This paper analysed the evolution of the three concepts. In addition, this paper suggested that there are a number of reasons that the 'Internet of Things' will evolve to an 'Internet of People and Things', and 'Internet of Everything' social media and other types of information being integrated into the 'Internet of Things'. For example, we argued that since tasks are increasingly automated, the line is blurring between human-generated data and sensor-generated data. Further, we noted that people are responsible for and use sensor-based information. In addition, we also noted that people are generating huge amounts of data that can be related to sensor-based data. Through discussion and examples, we illustrated that it is often helpful to integrate sensor information and human-based information to ultimately create more valuable applications.

In addition, this paper analysed some notions of context, and how context related to 'Big Data', the 'Internet of Things' and the 'Internet of Signs'. This paper suggested that 'Big Data' drove 'Big Context' because of the potential availability of context data. This paper also suggested that 'Big Data' and the 'Internet of Things' be used for analysis and generation of contexts.

Finally, this paper discussed a number of applications that illustrated the concepts. Those applications included continuous monitoring of accounting data, continuous assurance and auditing of and using 'Big Data,' monitoring location information and other data such as that for 'flash mobs'.

7.2. Extensions

This paper can be extended in a number of directions. First, this paper has not investigated the roles of business intelligence or more classic data warehousing. That data capture and analysis is typically more concerned with traditional transaction data. However, data from the 'Internet of Things' and 'Big Data' that includes social media and other forms of unstructured data could be integrated into business intelligence and data warehousing, including the 'Internet of Signs'.

Second, in this paper we noted that sensors might capture information and keep that information locally or put that information in some cloud-based system. In some cases there might be substantial sensor information that would need to be communicated. As a result, an extension of this research might be to analyse when communication or self-storage is more appropriate. Further analysis could focus on other related architecture-based issues.

Third, there is increasing research in understanding and inferring context. This paper suggests that additional sign-based information might be used as part of that analysis of context. This paper also suggested additional models for understanding context; for example, local to larger context and cascading contexts.

Fourth, although the 'Internet of Things' has received increasing attention, additional efforts need to be directed to the 'Internet of Signs' and the 'Internet of People and Things'. Additional research might focus on their use in issues such as context and better understanding links between different kinds of signs.

Fifth, new approaches to using 'Big Data' to facilitate auditing or continuous financial assurance may be required as new and emerging data sources become available. Further, as increasing amounts of data become available new analytic approaches are likely to be needed for auditing or continuous financial assurance of 'Big Data.'

ACKNOWLEDGEMENTS

An earlier version of part of this paper was given as part of a presentation at the IC3K 2012 meeting in Barcelona, in October 2012. I would like to thank participants at that meeting for their comments on that presentation.

DOI: 10.1002/isaf

REFERENCES

Ashby R. 1965. *Introduction to Cybernetics*. John Wiley & Sons.

Ashton K. 2009. That 'Internet of Things' thing: in the real world, things matter more than ideas, http://www.rfid-journal.com/article/view/4986, 22 June (accessed January 2013).

Bawa M. 2011. Multi-structured data: platform capabilities required for big data analytics. Teradata Aster – The Data Blog, http://www.asterdata.com/blog/2011/06/multi-structured-data-platform-capabilities-required-for-big-data-analytics/#Citation, 13 June (accessed January 2013).

Chandler D. 2009. Semiotics for beginners, http://www.aber.ac.uk/media/Documents/S4B/sem02.html (accessed January 2013).

Chui M, Loffler M, Roberts R. 2010. The Internet of Things. *McKinsey Quarterly* (**2**): 1–9.

Cox M, Ellsworth D. 1997. Managing big data for scientific visualization. ACM Siggraph.

Crowdsourcing. 2012. New signs that Wikipedia began a long decline in 2012, http://www.crowdsourcing.org/editorial/new-signs-wikipedia-began-a-long-decline-in-2012/22895 (accessed January 2013).

Culler J. 1981. *The Pursuit of Signs*. Routledge: London and New York.

Culler J. 2005. *The Pursuit of Signs*. Routledge: London.

Dey AK. 2001. Understanding and using context. *Personal and Ubiquitous Computing* **5**: 4–7.

Diebold FX. 2012. A personal perspective on the origin(s) and development of "Big Data": the phenomena, the term and the discipline, http://www.ssc.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf (accessed January 2013).

Dorsey W. 2003. Market semiotics: behavioral market diagnosis, http://www.marketsemiotics.com/docs/pdf/MSBrochure.pdf (accessed January 2013).

Eco U. 1978. *A Theory of Semiotics*. Indiana University Press: Bloomington, IN; 9–13.

Eco U. 1981. The theory of signs and the role of the reader. *The Bulletin of the Midwest Modern Language Association* **14**(1): 35–45.

EMC. 2012. Big data-as-a-service: a market and technology perspective, http://www.emc.com/collateral/software/white-papers/h10839-big-data-as-a-service-perspt.pdf, July (accessed January 2013).

Heil A, Knoll M, Weis T. 2007. The Internet of Things – context-based device federations. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences* (*HICSS 2007*).

Hernandez P. 2012. App employs context for Big Data analytics efficiency, Enterprise Apps Today, 18 September, http://www.enterpriseappstoday.com/business-intelligence/app-employs-context-for-big-data-analytics-efficiency.html (accessed January 2013).

Langer S. 1942. *Philosophy in a New Key: A Study in the Symbolism of Reason, Rite, and Art*. Harvard University Press: Cambridge, MA.

O'Leary DE. 1999. REAL-D: a schema for data warehouses. *Journal of Information Systems* **13**(1): 49–62.

O'Leary DE. 2006. Using mirror worlds to support supply chain management. *Applications of Management Science* **12**: 203–217.

O'Leary DE. 2008. Supporting decisions in real-time enterprises: autonomic supply chains. In *Handbook on Decision Support Sytems 2: Variations*, Burstein F, Holsapple CW (eds). Springer: Berlin; 19–37.

O'Leary DE. 2011. Blog mining: from each according to his opinion. *Decision Support Systems* **51**(4): 821–830.

O'Leary DE. 2012a. Knowledge discovery for continuous financial assurance using multiple types of digital information. In *Contemporary Perspectives in Data Mining*, vol. **1**, Lawrence KD, Klimberg R (eds). Information Age Publishing: Charlotte, NC; 103–122.

O'Leary DE. 2012b. Semiotics and the Internet of Signs. Presented at the IC3K Conference, Barcelona, Spain, October.

Schilit B, Theimer M. 1994. Disseminating active map information to mobile hosts. *IEEE Network* **8**: 22–32.

Schilit B, Adams N, Want R. 1994. Context-aware computing applications. In *First International Workshop on Mobile Computing Systems and Applications*; 85–90.

Schonfeld E. 2010. Costolo: Twitter now has 190 million users Tweeting 65 million times a day, http://techcrunch.com/2010/06/08/twitter-190-million-users/, 8 June (accessed January 2013).

Spangler S, Chen Y, Proctor L, Lelecu A, Behal A, He B, Davis T. 2009. COBRA – mining web for corporate brand and reputation analysis. *Web Intelligence and Agent Systems* **7**(3): 243–254.

SRA. 2009. The Internet of Things Strategic Road Map, http://sintef.biz/upload/IKT/9022/CERP-IoT%20SRA_IoT_v11_pdf.pdf.

DOI: 10.1002/isaf

Swedberg R. 2012. Schumpeter's theories of organizational entrepreneurship. In *Handbook on Organisational Entrepreneurship*, Hjorth D (ed.). Edward Elgar Publishing: Cheltenham; 31–48.
Tucker E, Watkins T. 2011. Flash mobs no longer just dance parties and pillow fights, pose growing criminal threat: cops, *Huffington Post*, 9 August, http://www.huffingtonpost.com/2011/08/09/flash-mobs-no-longer-just_n_921973.html (accessed January 2013).
UK Future Internet Strategy Group. 2011. Future Internet Report, May 2011, https://connect.innovateuk.org/c/document_library/get_file?folderId=861750&name=DLFE–34705.pdf (accessed January 2013).
Zikopoulos P, Eaton C, DeRoos D, Deutsch T, Lapis G. 2012. *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw-Hill.
Zikopoulos P, DeRoos D, Parasuraman K, Deutsch T, Corrigan D, Giles J. 2013. *Harness the Power of Big Data*. McGraw-Hill.

DOI: 10.1002/isaf