JOURNAL OF EMERGING TECHNOLOGIES IN ACCOUNTING
Vol. 16, No. 1
Spring 2019
pp. 000–000

American Accounting Association
DOI: 10.2308/jeta-10682

EDITORIAL

Measuring with Exogenous Data (MED), and Government Economic Monitoring (GEM)

ABSTRACT: The emergence of enormous disparate data sources, available in a multitude of alternate formats, and on very different devices, is leading to a business process revolution. This paper argues for a new approach in which these technologies are used to replace outdated blanket government intervention programs. Highly targeted interventions for social programs identified and addressed by Big Data-based methodologies would replace outdated government programs. This approach advocates for original detective methodologies, direct government action, and outcome monitoring all of which utilize Big Data-based frequent examination. The government would be in a
?1 privileged situation where it could acquire data not be made available to most other parties due to privacy laws. The data structures of the government would be protected by encryption and ~~block chain~~ privacy-protection
?2 methodologies that would only reveal targeted outcomes to the direct-action agent.

I. INTRODUCTION

During daily lives, increasing volumes of data about human behavior are generated by interacting with activities such as the use of transportation systems, shopping, online searches, and many others. More advanced technological systems now allow for the capture and processing of details of collective behavior with many interconnected data in real time at an unprecedented scale. Primarily, these large-scale datasets consist of exogenous data and analytics that are used to look for patterns in collective behavior that might recur in the future. For example, electronic media giants like Google LLC, Facebook, Inc., and Amazon.com, Inc. use "Big Data" to develop predictive behavioral models that are used to target individuals and households for revenue-generating purposes. Measurement of human activities can occur along many dimensions simultaneously. For example, the sale of a vehicle can be directly measured through records from the department of motor vehicles (open data in many locations). However, it can also be traced using data from checks issued by the buyer, or the same checks deposited by the seller. More indirectly it can be traced using various data points such as the locational data of an individual who appears at many car dealers, ads placed in newspapers, and comments made in social media. This multitude of sources presents a data paradigm that is very different from that of a corporation with mainly direct internal data. In general, as depicted in Figure 1, the characteristics of exogenous data are different from those of internal corporate data.

The use of data in this way is described by Zuboff (2019) as surveillance capitalism, a new form of capitalism that uses technology to build prediction models that anticipate what consumers will do now, soon, and later. Ultimately, surveillance capitalism is viewed as creating prediction products that are traded in what has been called the behavioral futures market (Naughton 2019). However, this disparaging definition misses the benefits and potential capabilities that ubiquitous and detailed exogenous data can bring. The use of Big Data is not only for capitalistic purposes. Predictive policing systems that aim to identify risk areas to allocate crime prevention resources more effectively (Johnson and Bowers 2004; Johnson et al. 2007) use Big Data to develop predictive models. Big Data is also being used to fight infectious diseases such as the flu. Information gathered from a variety of digital sources can be analyzed to identify patterns of behavior that may be indicative of a disease outbreak. For example, in Latin America, Twitter and Google searches predicted the spread of the Zika virus ahead of warnings by health officials (McGough, Brownstein, Hawkins, and Santillana 2017).

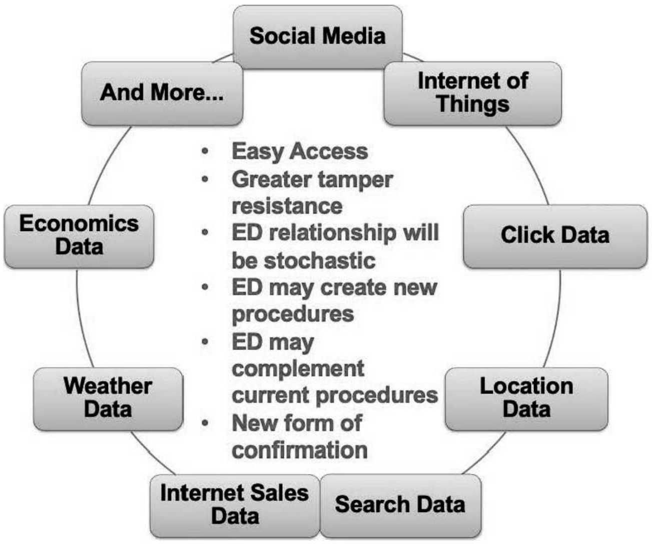
Additionally, careful examination of certain exogenous data sources may lead to increased understanding of many different issues. Stephens-Davidowitz (2017) initially focused on Google trends:

> I have spent just about every day of the past four years analyzing Google data. This included a stint as a data scientist at Google, which hired me after learning about my racism research. And I continue to explore this data as an opinion writer and data journalist for *The New York Times*. The revelations have kept coming. Mental illness; human sexuality; child abuse; abortion; advertising; religion; health. Not exactly small topics, and this dataset, which didn't exist a couple of decades ago, offered surprising new perspectives on all of them. Economists and other social scientists are

The authors are very appreciative for the many comments received in workshops at Rutgers, The State University of New Jersey, Fundacao Getulio Vargas in Rio de Janeiro, 44WCAS in Seville, Macquarie University, as well as the many colleagues of several entities including Michael Alles, Ricardo Lopes, William Glassgal, Nicola Pagani, Alex Kogan, Hussein Issa, Deniz Appelbaum, Yue Li, Kyunghee Yoon, and Tiffany Chiu.

Published Online: June 2019

?3 ?4 **FIGURE 1**
Characteristics of Exogenous Data



Source: XXX.

> always hunting for new sources of data, so let me be blunt: I am now convinced that Google searches are the most important dataset ever collected on the human psyche.

Overall, the same generic approach that electronic media giants like Google, Facebook, and Amazon are using to target individuals and households directly, can be modified. Instead of being used for capitalistic purposes, the prediction of behavior can be altered and expanded to benefit individual citizens through government action, business measurement, and assurance. Indeed, the public and private sectors, as well as academic institutions are forming partnerships to use Big Data to address complex social issues.[1] Bloomberg's Data for Good Exchange is an example of a program that brings together the public and private sectors to focus on how data science can be used to drive change in the delivery of social programs and service.[2] Thus, the purpose of this paper is to propose a data driven and preventive conceptual model, Government Economic Monitoring (GEM), which reimagines how governmental entities can provide targeted social services to the neediest in society. Over the years, governments around the world have developed and implemented a wide range of social support programs. These deal with a variety of issues, such as healthcare, education, social security, hunger, which are of great importance to the welfare of a country's population. Despite their importance, these programs are generally large, rigid, and inefficient. The emergence of Big Data in combination with exogenous data analytics has opened the door for improvement. These elements can make large social welfare programs much more efficient. Additionally, they can be used to create much narrower, time-limited programs for more specific societal problems (narrow social pathologies) such as child abandonment, opioid addiction, temporary personal crises, mental breakdowns, and marital abuse.

The current methods of economic measurement for resource allocation are very coarse and inefficient. Surveys to determine GNP or family income are grossly inaccurate (*The Economist* 2016; Costanza et al. 2014). Several of the digital giants (e.g., Google, Facebook, and Amazon) have developed methods of targeting individuals and family groups for advertising and prediction of behavior. Accordingly, researchers have utilized the information given by these digital giants
?5 (Stephens-Davidowitz 2017; Ripberger et al. 2011). Further, Mellon (2013) suggests that these types of data are more relevant and cost effective compared to traditional survey data used to infer socioeconomic measures. Thus, potential benefits to individuals and government processes may come from intervention and timely response to targeted societal problems. Additionally, there are many potential benefits to business measurement (accounting) and assurance (audit) that may accrue from the usage of exogenous data, and these may also interact with and synergistically expand GEM.

[1] See, https://hortonworks.com/article/how-companies-are-using-big-data-for-social-good/
[2] See, https://www.forbes.com/sites/ciocentral/2018/01/02/bloombergs-data-initiative-big-data-for-social-good-in-2018/#23ad08463a44

Allen Press, Inc. ■ 17 June 2019 ■ 1:14 pm Page 2
//titan/Production/j/jeta/live_jobs/jeta-16/jeta-16-01/jeta-16-01-05/layouts/jeta-16-01-05.3d

Why is the government the best entity to adopt this GEM? Most organizations do not have enough resources or lack the IT infrastructure necessary to handle the complexity or speed of the data required for a properly functioning GEM environment. Further, it is urgent that governments step up and establish data-driven approaches to policymaking. Although the transition to a more comprehensive, data driven, and preventive methodology may be expensive at the inception, this approach may allow the government to be more efficient in the delivery of its social functions. Furthermore, more precise and sophisticated criteria for intervention, compared to current "blanket" programs, can be applied. Methodologies can be developed to create greater transparency regarding government actions and provide many metrics for their evaluation.

As with most new technologies, both positive and questionable uses may arise. Exogenous data may be used for social media control, limiting the movement of citizens, helping to hack networks, and many other purposes and, as such, the ability to protect individual privacy is of utmost importance and must be weighed against any benefits that GEM provides. Privacy concerns are particularly important in situations where the government collects data given the recent concerns about the government's ability to protect their databases from cybercriminals. Given the critical importance of maintaining public trust, within the last three years, many states have enacted statues that require reasonable security measures to ensure that data are protected. In general, the recent laws require a comprehensive approach to security and security oversight, as well as specific measures to protect sensitive information. Many states also have data security laws that apply to private entities.[3] On the federal level, the Federal Trade Commission (FTC) emphasizes the need for strict regulation and safeguards to address how the data are collected and used in the system. However, while there are numerous bills proposed by Congress, there is no single law at the federal level governing data privacy and data security.

Given the extreme importance of data privacy and security, the GEM concept incorporates privacy-protection technologies. Specifically, GEM uses a privacy-preserving blockchain database (PPB) with homomorphic encryption. The use of a privacy-preserving blockchain increases the immutability of the data added to the blockchain, while application of homographic encryption lowers the possibility of potentially exposing sensitive data. The use of a private blockchain can
?6 provide a high level of confidentiality (Wang and Kogan 2018). While GEM utilizes technologies to protect privacy, it is also vital for government agencies to ensure that there are policies in place to ensure that private data are protected and implement
?33 an ethical code of conduct that encompasses the complete life cycle of data. Additionally, continuous monitoring and risk assessment will be necessary to assure that agencies comply with ethical principles. Further, because of public concerns about using data tracking/monitoring, it is vital that government agencies be transparent about how they are using the data and allow individuals the opportunity to opt out if they do not want to participate.

Overall, this paper aims to explore the positive uses of ubiquitous data. In particular, this paper examines linking exogenous data to continuous monitoring and assurance technology. Moreover, there are innumerous potential applications of exogenous data for business, not-for-profit, and government. The focus of this paper is on exogenous data applications for governments. The goal is to abandon the current inefficient blanket approach laden with bureaucracy and move to a more pointed, time sensitive method. To generate a friendlier, more efficient, and targeted government approach to helping citizens, GEM provides an approach that can tackle less prevalent, but critical, social problems.

Section II discusses the elements of measurement using exogenous data, followed by a definition of GEM in the context of its application to government programs in Section III. Then the GEM methodology is described in Section IV. Next, in Section V, an illustrative case study is presented applying the methodology to food stamps. This is followed by a discussion of a series of related issues in Section VI. Finally, concluding remarks and ideas for future research are discussed in Section VII.

II. MEASURING WITH EXOGENOUS DATA (MED)

"External data sources continue to expand in terms of both content and interconnectedness" (Brown-Liburd and Vasarhelyi 2015) and, as a result, there has been an explosion of data sources. For example, Google's data come from a new form of human activity (Google searches). Other examples of new forms of activity include the Internet of Things (IoT), social media, and others that did not exist two decades ago. Traditional sources contain regular natural events and human activities such as weather, speech, movement, or macroeconomic trends. Furthermore, although these traditional sources are not new, electronic capture of the data is only recently possible. The consequence of these innovations is the emergence of what is called Big Data (Vasarhelyi, Kogan, and Tuttle 2015; Moffitt and Vasarhelyi 2013; McAfee and Brynjolfsson 2012). A highly heterogeneous global data environment has emerged, enriching the analytic profile of the environment, as discussed by Stephens-Davidowitz (2017):

> This dataset, (*Google searches*) however, is not the only tool the internet has delivered for understanding our world. I soon realized there are other digital gold mines as well. I downloaded all of Wikipedia, pored through Facebook

[3] See, http://www.ncsl.org/research/telecommunications-and-information-technology/data-security-laws-state-government.aspx

> profiles, and scraped Stormfront. In addition, PornHub, one of the largest pornographic sites on the internet, gave me its complete data on the searches and video views of anonymous people around the world. (emphasis added)

The key elements[4] of measuring with exogenous data (MED) are:

- An extensive set of data sources that are used complementarily.
- A set of continuous measurement analytics.
- A set of benchmark levels for critical parameters and states-of-the-world.
- Alerts for specific conditions.
- Methods of measurement for each of the different data sources.
- Taxonomies for the integration of datasets and insights into broader data ecosystems.
- Privacy-preserving methods and algorithms (Kogan and Yin 2018).

Leveraging Big Data for analytic purposes creates the need for extensive computational facilities and new computational methods. Consequently, there is a substantive need for summarization and rational selection of large datasets:

> In fact, the smartest Big Data companies are often cutting down their data. At Google, major decisions are based on only a tiny sampling of all their data. You don't always need a ton of data to find important insights. You need the right data. A major reason that Google searches are so valuable is not that there are so many of them; it is that people are so honest in them. People lie to friends, lovers, doctors, surveys, and themselves. But on Google, they might share embarrassing information, about, among other things, their sexless marriages, their mental health issues, their insecurities, and their animosity toward black people. (Stephens-Davidowitz 2017)

In addition to the above technical methodologies, to provide private, loosely coupled event identification to society through its programs, the government would have to develop processes, approaches, and resources to apply these methods. Collaborations with the private sector (e.g., Bloomberg Data for Good Exchange) and academic institutions can be formed to leverage technologies and data science skills.

III. WHAT IS GEM?

As noted earlier, the current methods of economic measurement for resource allocation are very coarse and inefficient. Thus, a more data-driven and preventive method that takes advantage of advances in technical methodologies to gain insights and conduct analyses is timely and important.

In general terms, GEM would enable governments to harvest comprehensive sets of data. Some of these may be unavailable to the public or restricted by privacy rules. These data would be used to identify and monitor societal pathologies. Continuous audit/monitoring technology is then used to issue "alerts" when there is an indication of a problem at the individual or family level. These alerts could be relayed to a variety of intervention mechanisms. These alerts could notify government social workers. They could also result in immediate action through automated processes (e.g., send food to a family, find housing for the homeless). The alert could also be used to define patterns of socioeconomic attributes that could help the government and other selected organizations to identify potential individuals who need immediate help. The process would not stop there. GEM can be extended and utilized to understand the success or failure of the chosen intervention, deactivate the action upon resolution of a case, or take other relevant steps (see Figure 2).

Several applications of GEM have been suggested, and examples are provided in Appendix A. Additionally, Appendix B illustrates a list of potential sources of data. GEM has the potential to change, replace, or supplement the six largest U.S.
?7 governmental programs.[5] However, for illustrative purposes, it may be easier to target narrower pathologies such as poor school performance by students within a particular school district.

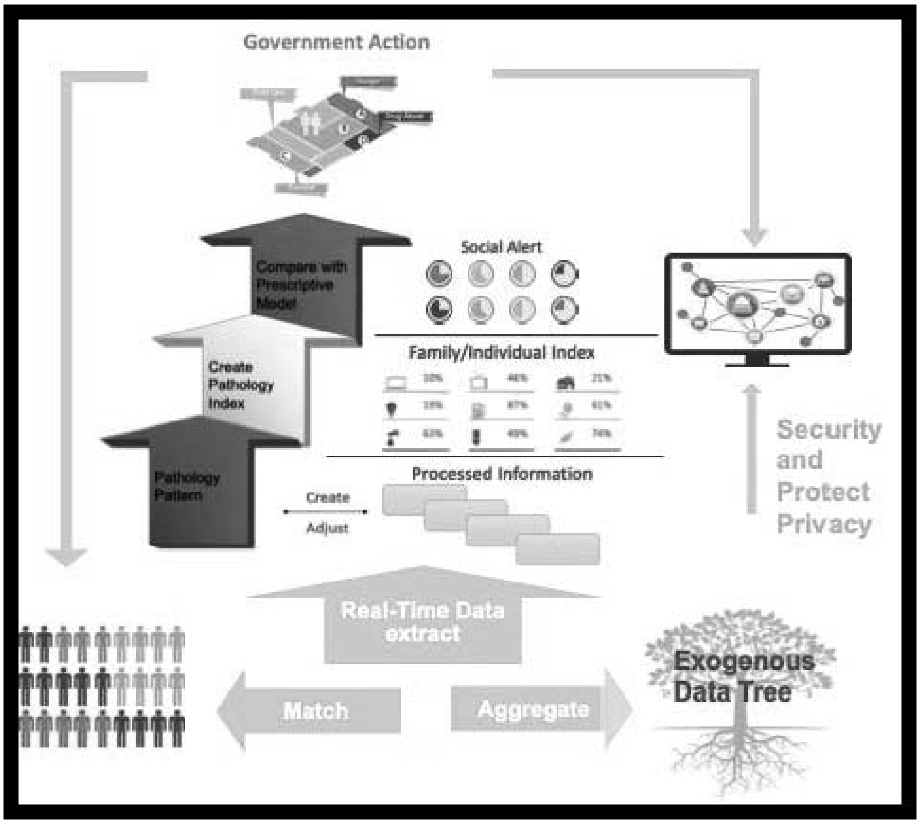
The following is a selection of applications along with potential data sources:

- Opioids addiction detection (e.g., large pharmacy bills with opioids, financial distress, health effects, hospital visits, car and residence sales, social media utterances).
- Children's family problems (e.g., lack of attention, poor appearance, lack of attendance, many family pharmacy bills, low family food spending, low electricity spending, hospital visits).

[4] The first five elements are analogous to the key elements of continuous auditing: measurements, analytics, standards, alarms, and method of data access (Vasarhelyi and Halper 1991).

[5] These programs are Temporary Assistance for Needy Families (TANF), Medicaid, Children's Health Insurance Program (CHIP), Supplemental Nutrition Assistance Program (SNAP, also called food stamps), and Supplemental Security Income (SSI).

FIGURE 2
The GEM Dynamic Processes



Source: XXX.

As illustrated in Figure 2, there are several steps in the proposed GEM process:

1. The linkage of many data sources for identification of a given social pathology. Suspicion scores (Issa 2013) and functions for this can be developed using machine learning models. Existing identified cases can be used for creating supervised machine learning approaches. Continuous monitoring of these variables is then used to identify critical points where urgent action or even police intervention may be needed. Continuous audit methodology (Vasarhelyi and Halper 1991; Alles, Kogan, and Vasarhelyi 2002; Vasarhelyi, Alles, and Williams 2010; AICPA 2015; Kogan, Alles, Vasarhelyi, and Wu 2014) is also used to generate alerts.
2. New methods of government intervention must be imagined and developed. Some may involve social workers contacting the family block (FB), others may involve third parties being asked to intervene/monitor (e.g., teachers), or perhaps direct automatic action such as payments to a pharmacy, automatic delivery of food or medications, deposits of money, dispatching an ambulance, immediate help from the police.
3. Following GEM action, evidence monitoring is performed to detect several possible issues or outcomes. It is possible that the original variables in the suspicion function have become outdated, or an error caused a false suspicion alert to occur. Alternatively, if the system worked correctly, the target goals of intervention may have been reached. It is essential to track these factors to improve the processes and models continuously.

GEM Is Not Surveillance Capitalism

GEM is designed to facilitate the efficient allocation of social resources. The goal is to assist in treating large and small-scale social problems, not to control the population. Accordingly, researchers have addressed the implication of Big Data-based approaches to increase social welfare (Cao 2012; Kim, Trimi, and Chung 2014; Williams 2015; Gillingham and Graham 2017).

Allen Press, Inc. ■ 17 June 2019 ■ 1:14 pm Page 5
//titan/Production/j/jeta/live_jobs/jeta-16/jeta-16-01/jeta-16-01-05/layouts/jeta-16-01-05.3d

TABLE 1

The Good and the Bad of Measuring with Exogenous Data

Source	The Bad	The Good
Crawling social network data (Twitter, Facebook)	• Controlling social media while hampering the freedom of speech	• Preventing cyberbullying • Preventing suicide
Capturing personal information (multiple sources)	• Marketing purposes	• Supporting children in hunger • Supporting victims of domestic violence
Dark Web usage	• Selling intellectual property obtained illegally • Trading illegal drugs	• Preventing sex trafficking • Detecting fraud and money laundering

On the other hand, Zuboff (2019) and many others who rebuke Big Data-based methodologies do not take into consideration the enormous benefits that can be gleaned from using such methodologies (O'Neil 2017; Schneier 2015; Degli Esposti 2014).

Measuring with exogenous data (MED) may certainly serve behavioral surplus vendors by providing targeted advertising. Additionally, MED can result in the creation of new data-driven products, or the modification of traditional business practices. MED can also be leveraged for adverse actions such as population control, limitation of free expression, and criminal activities. Table 1 contrasts how certain bad activities can be repurposed for social good.

One use of MED that can lead to the most immediate social benefits would be to identify and create more direct measures of economic activity (e.g., GDP). These could be incorporated into the current schemata of government actions and would improve the efficiency of existing programs. A more dramatic and revolutionary path of action would be to adopt a philosophical approach such as the one discussed in Figure 2. In this case, applications in other areas of business such as accounting and assurance can be linked or piggybacked on/into by GEM programs.

GEM Is the 21st Century Adam Smith

The effectiveness of socioeconomic policies (e.g., food ~~stamp~~) or measurement (e.g., GDP) has been debated for a long time. Samuelson (1948), one of the most preeminent figures in the field of economics, claimed economic analysis should be based on ergodic axioms (i.e., explained in mathematical formulas), as is done in physics. He believed the rigor of economic study comes from adopting mathematics as a language. Unfortunately, such economic theories have strict predefined assumptions in order to conform to mathematical niceties. They have therefore been criticized as being irrelevant to the real world (Stiglitz 1991; North
?8 1995; Davidson 2012; Kristein 2015). Furthermore, the economic measurements and policies, resulting from these theories, do not always reflect reality accurately. For instance, Gross Domestic Product (GDP) is a measure that has been criticized as it ignores everything other than market transactions. It fails to consider vital factors such as social costs and human welfare (Costanza et al. 2014; *The Economist* 2016). Additionally, Rymph (2012) addressed public child welfare systems and concluded that these systems do not represent what children want but, instead, represent "economic wants."

As Hahn (1991) emphasized, developing a dynamic methodology is crucial to counter dynamic socioeconomic change. Additionally, static theories cannot be used to solve modern socioeconomic problems. To make better decisions, a modern Adam Smith cannot rely only on elegant mathematical formulas. Modern society cannot be depicted in such equations since the union of science and technology is complicating society (North 1995). Compared to the period when Smith (1937) scripted *The Wealth of Nations* in 1776, the society became more sophisticated where the policymaker could not rely only on the "invisible hands." Thus, any measure should be dynamically adaptable to socioeconomic change. Moreover, as Rymph (2012) asserted, public policy has to be more tailored to individuals and bettering their welfare instead of fulfilling the more substantial overall economic wants of society (i.e., solving the equations).

This paper proposes a data-driven and preventive methodology that addresses the issues mentioned above. GEM will provide dynamic on-demand measures instead of static measures such as GDP. Adam Smith's (1937) invisible hand is not sufficient for modern society. In the 21st century, GEM will transform this from an invisible hand into a smart, efficient hand.

GEM Application in the Audit and Accounting Setting

The business measurement and the assurance function (Romero, Gal, Mock, and Vasarhelyi 2012) can also benefit from MED and be used in GEM efforts. Furthermore, the more comprehensive conceptual and measurement structure allowed by Big Data not only focuses on government direct social action, but also allows for joint compliance, measurement, and assurance efforts. For example, MED can track shipments to the customer and validate this with data from invoices and subsequent payments, all the while checking for compliance with customs and trade rules.

Allen Press, Inc. ■ 17 June 2019 ■ 1:14 pm Page 6
//titan/Production/j/jeta/live_jobs/jeta-16/jeta-16-01/jeta-16-01-05/layouts/jeta-16-01-05.3d

- Instead of measuring sales, intention to buy can be captured and disclosed in more future-oriented reporting.
- Declared sales can be validated by activities such as parking lot activity, social media utterances, purchases over the internet, pedestrian traffic into stores, frequency of trucks out of the warehouse, orders to suppliers, and Internet of Things (IoT) data.
- Acquisition of inventory can be measured by similar methods as above.
- Labor can be identified and measured by face recognition.
- Organizations, suppliers, and customers can keep data in one place and share them in a blockchain. They can activate some clauses of their relationship through a smart contract (Rosario 2019).
- Continuous monitoring and auditing can be activated based on both internal and external cues.
- GPS or ISP data can be used for measuring and validating sensitive geographical activities and, for example, detecting trading with non-authorized countries.

Business measurement and assurance is one example of how the business cycle can help and be integrated into a MED/GEM approach. Exogenous data need to be progressively integrated into the assurance environment by being accepted as audit evidence (Yoon, Hoogduin, and Zhang 2015). Audit methodologies today are mostly internal, except for the external validation that is presented by confirmations. These are inefficient ways to link the external to the internal data environments. GEM and the prerogatives that come with being the government and taxing authority allow for much more effective internal and external linkage abilities. The next section discusses a methodology for this linkage.

IV. METHODOLOGY

MED methodologies can be applied in many contexts such as government environmental monitoring here proposed. The essential features of the GEM methodology are that it is data driven, preventive, focuses on privacy, and is architected to be flexible, private, and dynamic.

Data Driven

People generate an enormous quantity of data every day. Not only are more data being generated than before, but many new types of data are being generated. For example, "Google Trends" data (Stephens-Davidowitz 2017) were not available 20 years ago. These data can be categorized into internal business-related data, automatically machine-generated data such as IoT (Chui, Löffler, and Roberts 2010; Kopetz 2011) or GPS (Kaplan and Hegarty 2005), and human-generated data. In a GEM environment, society is linked by different forms of exogenous data. This linkage helps to improve government efficiency and effectiveness as well as provides more accurate macroeconomic measures. Exogenous data contain information from various sources that reflect different aspects of societal activities. Data can be collected from different entities such as the public or private sectors, social enterprises, and other publicly available sources (such as the Dark Web).[6]

In general terms, there are four social sectors (Sabeti 2011). Figure 3 divides social sectors into four categories, classified by ownership and the purpose of the sector. Government is publicly owned and serves social purposes; NPOs/NGOs are privately owned but also serve social objectives. The social enterprises are doing business for the public good, and businesses serve private needs with commercial benefits. The public sector (government and NPO/NGO) covers databases such as government agency databases, police report data, energy usage data, transportation data, tax data, housing data, and donation data. The private sector databases (social enterprise and business) encompass business data such as supermarket shopping data, stock market data, online retailer data, drug production data, telephone bill data, and clothes-shopping data. The last element is the publicly available data. This portion can be used to link databases together to create a series of virtual connections to extract meaning through suspicion functions from different databases.

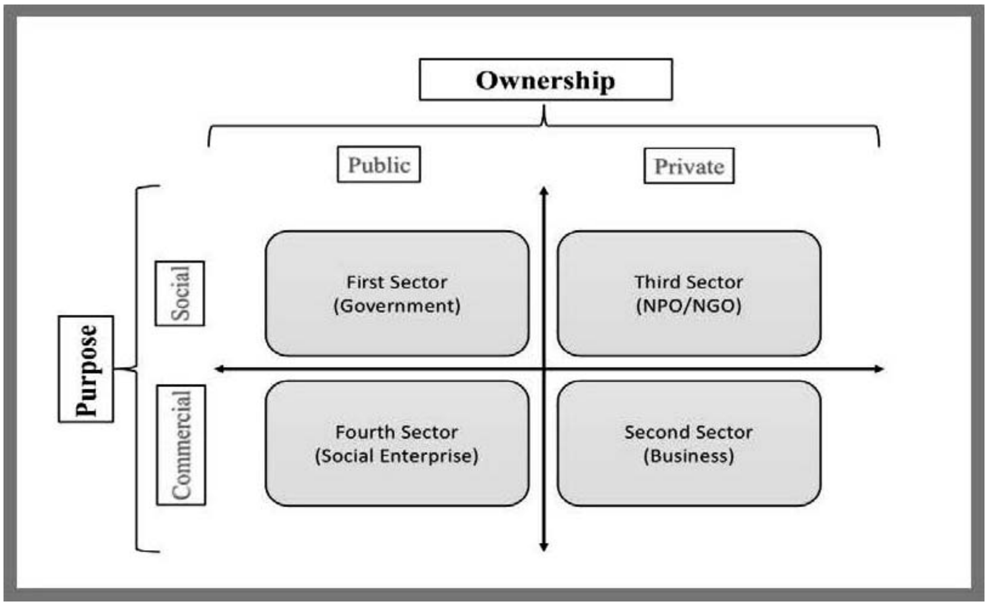
By aggregating/linking these different data sources, a knowledge tree comprised of different data sources can be created, as
?9 depicted in Figure 4.

A GEM data tree is an illustration in which each data source or information component is represented as a "branch" being brought together in a common utilization framework. In each of the branches, individuals have limited authority to write, validate, and review the data. Every activity that involves the exchange of goods, services, or money will be introduced into the big picture. The government may have the authority of collecting, validating, analyzing, and reacting, making it the trunk of the

[6] "The Dark Web is a term that refers specifically to a collection of websites that exist on an encrypted network and cannot be found by using traditional search engines or visited by using traditional browsers" (https://www.techadvisor.co.uk/how-to/internet/dark-web-3593569/). Several commercial entities, in addition to criminal groups and hackers, use these data for legitimate and illegitimate purposes Most all sites on the Dark Web hide their identity using the Tor encryption tool. More can be found at https://www.techadvisor.co.uk/how-to/internet/dark-web-3593569/

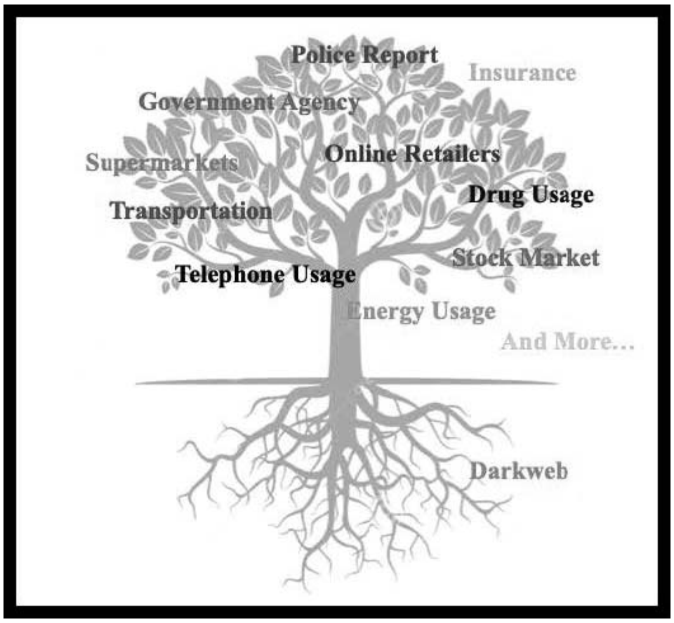
Allen Press, Inc. ■ 17 June 2019 ■ 1:14 pm Page 7
//titan/Production/j/jeta/live_jobs/jeta-16/jeta-16-01/jeta-16-01-05/layouts/jeta-16-01-05.3d

FIGURE 3
Social Sector Categories



Source: XXX.

FIGURE 4
Data Tree



Source: XXX.

Allen Press, Inc. ■ 17 June 2019 ■ 1:14 pm Page 8
//titan/Production/j/jeta/live_jobs/jeta-16/jeta-16-01/jeta-16-01-05/layouts/jeta-16-01-05.3d

data tree. Hidden information in the Dark Web should also be incorporated for further investigation and used to prevent harmful purposes.

Preventive Approach

GEM utilizes a variety of technologies to extract informative results from Big Data sources. Not only is it used to detect socioeconomic issues efficiently, but to better prepare government agencies before adverse socioeconomic issues emerge. The GEM process is composed of three stages: (1) continuous data monitoring, (2) socioeconomic measure formation, and (3) future societal need prediction. Cutting-edge tools would be introduced to gather and preprocess data from multiple sources. These data could directly or indirectly capture the socioeconomic status. The data start at the family block (FB) level and get aggregated up to a national level. In particular, *family block* refers to a broad form of the family such as the traditional family, housemates, or just single individuals if that is the case. Notably, it is expected that most of the proposed data sources are publicly available. Despite this fact, privacy protection remains both a legal and social concern. As a result, GEM would not reveal individual identities during its analytic process, but would instead aggregate FBs into similar groups for analytic purposes. Aggregation of the data will make individual identification very difficult. The model captures events on a real-time basis making the process, to a certain degree, continuous. Accordingly, GEM could introduce dynamic examination with different levels of granularity (i.e., individual level, family block level, group level, regional level, up to national level) and with time stamps (i.e., continuous data collection and preprocess). Although most aggregations in traditional economic analysis are orthogonal extrapolations, collection of data and relationships at the atomic FB level allow for oblique linkages taking into consideration geography, family status, narrow social problems, or several of these together (Cuzzocrea 2011).

Considering other dimensions, based on preprocessed data, GEM can formulate numerous key socioeconomic measures. The measures can also take two different forms: static measures and on-demand measures. Static measures represent the well-known set of socioeconomic parameters and states-of-the-world such as happiness (Ura, Alkire, Zangmo, and Wangdi 2012) or poverty level. On the other hand, static measures do not fully depict reality. Government authorities or auditors will need parameters to reflect specific circumstances. In particular, these on-demand measures would be analogous to the information provided by decision support systems in an enterprise (Bonczek, Holsapple, and Whinston 2014). Users will describe a measure needed to the system by entering keywords and selecting categories. The system will identify the characteristics of the measure that the user needs. The system will then provide a measure that matches the user's request as well as the measure's basis of calculation. Based on the information provided, the user could further verify the measure as being correct and decide on the desirability of its usage. The purpose of these on-demand measures is to allow GEM to satisfy a greater variety of user needs. It is anticipated that accumulated static measures would be used to develop the suspicion function, but on-demand requests would also be used in action monitoring.

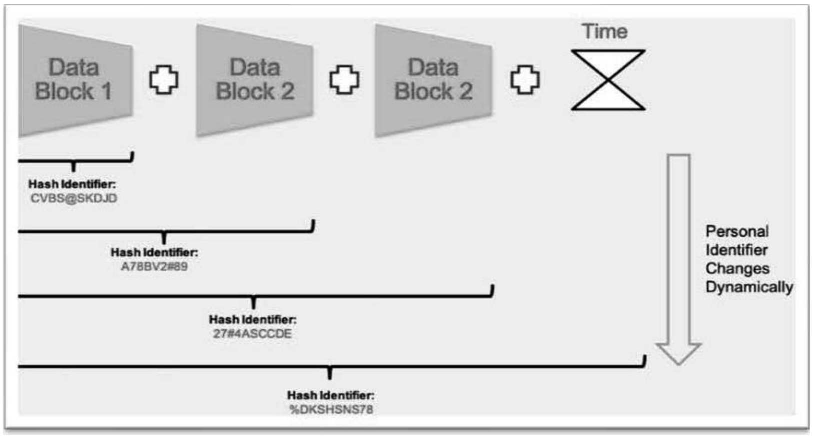
The most effective way to address adverse socioeconomic issues would be through prevention or providing anticipatory intervention. Despite this, most government support programs are only provided after the recipient provides proof of need. For example, individuals usually visit the doctor after they feel symptoms. Accordingly, the doctor only provides an *ex post* diagnosis for a patient who is already exposed to a certain disease. A better approach would be to create an alert for the patient before the pathology is acquired. This alerting system could gather exogenous data such as endemics, genomic aberrations, diet, habits, exercise schedule, air pollution level, neighborhoods, and other information such as income level. Predictive analytics are progressively being used in many domains of knowledge (Kuenkaikaew 2013). Similarly, GEM would gather the information from available sources to determine pre-conditions and high suspicion scores. Then, the system would continuously learn from multiple on-demand sources of data and predict the needs of society by aggregating and modulating individual FB predictions. Governments could take full advantage of this for policymaking. Since the government would have a better picture of future social needs, more effective and accurate budget plans could be generated. Furthermore, governments would not be anchored to obsolete rigid budgets but could operate on variable optimized spending plans, which would result in a better allocation of social goods and resources.

Multiple tools could be used for GEM. Data mining could be used to gather data and extract useful knowledge from a given data source. Machine learning algorithms could be utilized to glean unobservable knowledge from the data and most likely to create suspicion functions. Also, once a sufficient amount of knowledge is obtained (i.e., knowledge base), adaptive artificial intelligence algorithm-based systems (Issa, Sun, and Vasarhelyi 2016) could be built. Of all the issues raised by GEM, the most concerning is privacy. Encryption, multilevel segregation of data, and privacy-preserving blockchain (PPB) (Kogan and Yin 2018) could be used to address the issues of privacy and mitigate concerns. These are discussed in the next section.

Privacy

GEM has to strictly comply with privacy regulations and concerns. Under the Privacy Act of 1974, GEM should be openly announced to the public since the information is gathered at the individual level. However, GEM differs from the government

FIGURE 5
Hashed Personal Identifier



Source: XXX.

surveillance entailed in the Electronic Communications Privacy Act (ECPA) of 1986 and the Patriot Act in 2001. Further, GEM has to comply with the Stored Communications Act (SCA) while it stores non-content information (i.e., metadata). Further, the EU data protection rules[7] (GDPR) state that data processing is allowed without individual consent when the data processing is of *vital public interest or to complete public tasks*. GEM goes beyond to ensure the protection of data by mitigating the possibility of tracing information back to a specific individual.

Privacy-Preserving Blockchain Database (PPB)

To protect privacy, PPB is operationalized in GEM. The use of private blockchain technology can provide a high level of confidentiality (Wang and Kogan 2018). The system stores the data for each socioeconomic attribute selected to be extracted and transmitted in a blockchain to secure immutability. The application of homomorphic encryption will further allow
?10 socioeconomic measurement to be performed on encrypted data (Gentry 2009). This section discusses (1) disguising personal identifiers in a dynamic way to comply with the privacy act, and (2) application of homomorphic encryption on a private blockchain network.

As shown in Figure 5, initially certain data of Person A are gathered. Subsequently, A's personal identifier dynamically changes each time new data are added to A's database. Since all the data are stored in a blockchain, whenever a new data block is added to the chain, the system generates a new hash index for the blockchain. This index becomes a hashed identifier for A. An additional time block is inserted periodically to alter the hashed identifier, even though no new data have been added to the blockchain. Such an identifier will look meaningless and is not possible to map to any individual since the hash changes periodically. Only an agency with limited access to the private blockchain could determine the owner of the identifier. Accordingly, identification of whom the record belongs to is limited to the agencies providing the service.

Further, homomorphic encryption is applied alongside the usage of private blockchains. First, hashed data that fit into certain social attributes will be stored into a separate private blockchain. Primarily, Big Data is not suitable to be stored in the form of blockchain. However, data gathered by GEM could be parsed into multiple social attributes to decrease the size of each input (transactions). Second, homomorphic encryption reduces potential sensitive data exposure. The application of homomorphic encryption allows the end user to acquire the answer for the query even if the inputs are hashed in the form of ~~cypher text~~. For example, if the end user asks about the poverty level of a specific region, the system provides the answer without revealing any sensitive information stored in other attributes (blockchain).

As depicted in Figure 6, the usage of blockchain increases the immutability of each datum added to the blockchain while the application of homographic encryption lowers the possibility of potentially exposing sensitive data. This privacy-preserving blockchain, which is designed as Case 4 in Figure 6, is not the most efficient system to query and maintain. The most efficient system would be Case 1, where the immutability and security of the data are sacrificed. Notably, Case 4 is chosen since

[7] See, https://en.wikipedia.org/wiki/General_Data_Protection_Regulation

Allen Press, Inc. ■ 17 June 2019 ■ 1:14 pm Page 10
//titan/Production/j/jeta/live_jobs/jeta-16/jeta-16-01/jeta-16-01-05/layouts/jeta-16-01-05.3d

FIGURE 6
Effect of Utilizing Blockchain and Homographic Encryption

	Not Utilizing Blockchain	**Utilizing Blockchain**
Not Utilizing Homomorphic Encryption	**Case 1:** Immutability ↓ Sensitive Data Exposure ↑	**Case 2:** Immutability ↑ Sensitive Data Exposure ↑
Utilizing Homomorphic Encryption	**Case 3:** Immutability ↓ Sensitive Data Exposure ↓	**Case 4:** Immutability ↑ Sensitive Data Exposure ↓

Source: XXX.

"effective" privacy protection is a crucial goal. The next section brings together these privacy-protection methodologies in an architecture to provide secure model development as well as the application of these suspicion functions to a live stream of secure extracted data.

An Architecture for Privacy Protection of the GEM Schema

As indicated above, multiple sources of data may be used in the GEM schemata. These may be open government data (e.g., Ohio checkbook, U.S. government data),[8] may be sourced by commercial enterprises (e.g., Google, Facebook) and partially public, or may be data provided by preferential agreements with the government (e.g., supermarket sales, telephone call metadata, electric billing, physical fitness measures). Each of these sources must be segmented in meaningful sectors such as zip code, nature of the FB, time stamps, and retained for *a limited time* for usage. For additional security, these sectors must have different access arrangements, and different supervisory/support personnel. Figure 7, "Data Capture, Analysis, and Distribution in GEM," describes a proposed privacy-protected data/action methodology.

Each of these data sources is tapped for each FB key and relevant data are extracted. These records contain encrypted FB IDs and homomorphically encrypted data for each record. These records are merged with other records from the alternate sources, and a suspicion score is calculated. The threshold of the suspicion score for that particular social pathology is compared with the outcome, and exceptions are generated (Vasarhelyi and Halper 1991). These are placed on a privacy-protected blockchain and distributed to the action agents that selectively can decrypt only the relevant data for the needed actions. These agents also act as miners for the blockchain to ensure the integrity of the data stream. This architecture is composed of separate layers for each type of GEM action desired, but some may be linked or overlapped due to similarity and resource considerations. The next section illustrates GEM applied to food stamp distribution.

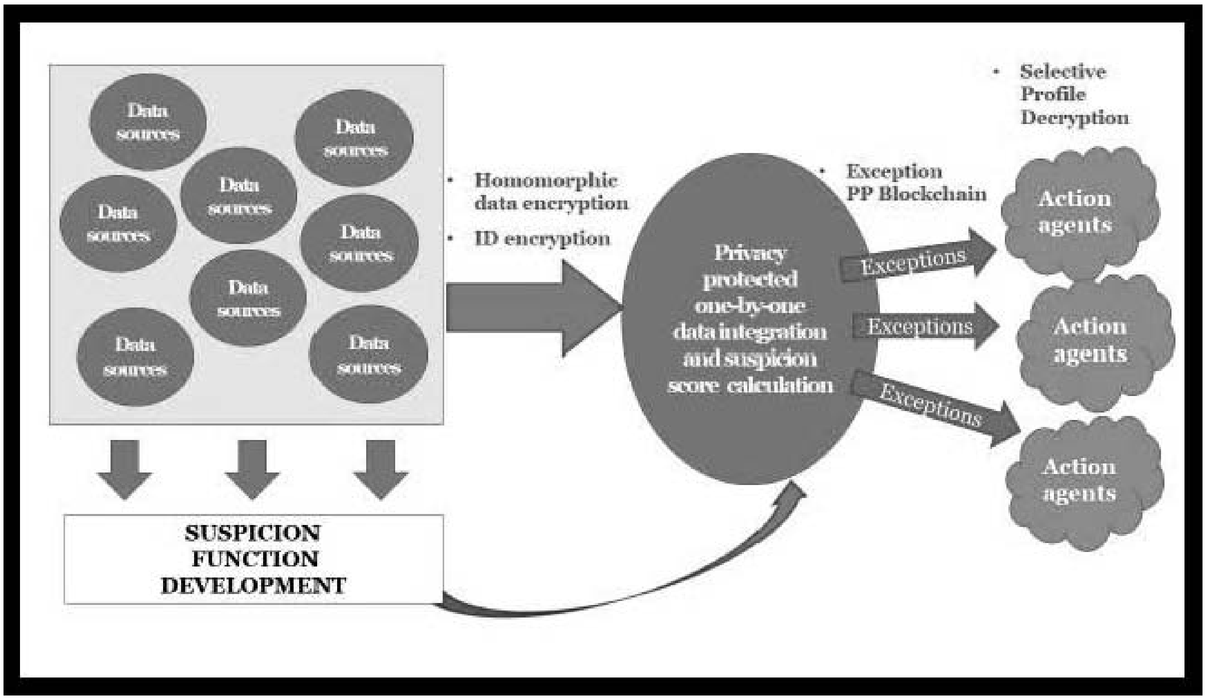
V. IMAGINARY CASE STUDY: FOOD STAMP DISTRIBUTION

New Jersey is the fourth wealthiest state in the United States. However, 298,000 children are living in food-insecure environments, and one in six children in Passaic County have faced hunger (Center of United Methodist Aid [CUMAC] 2016). Furthermore, in 2017 there were 69,765 children under 18 living in Newark: 34,967 of them are not SNAP recipients, and 25,360 live in poverty (U.S. Census Bureau 2017). To address this, GEM could help food stamp distribution to provide direct support to those in need.

As described in Figure 8, GEM first would collect the data that pertain to the neighborhood under examination. For example, to find out the poverty level in the FB, GEM would need access to available credit history, property reports, forfeiture penalty reports, supermarket consumption, locational data, classroom attendance/performance, and other necessary data. Based on this information, a series of analyses need to be conducted to determine the FB economic status while preserving personal privacy. At the same time, regional Amazon purchase history, nearby grocery store sales data, and previous food stamp distribution history would be used to find the grocery purchase patterns of the FBs. During data analysis, data are secured using
?11 ?12 the PPB (Kogan and Yin 2018; Agrawal and Srikant 2000; Aggarwal and Phillip 2008). Each FB would be evaluated using a suspicion function. Results with a suspicion score[9] of, say, above 0.9 would be deemed actionable. All input information would

[8] See, https://data.gov
[9] Suspicion score thresholds can be established analytically with cost benefit functions and parameters.

FIGURE 7
Data Capture, Analysis, and Distribution in GEM



Source: XXX.

be transformed into a hashed form, stored in a blockchain only visible to authorized agencies, and all personal information transformed into hashed indexes that change every millisecond. Even if this information is breached, there would be no way of tracing the hashed index back to the relevant individuals.

As well as considering neighborhood activities, GEM could inform school districts about cases needing direct actions. The
?13 system could suggest an area or FB where food stamps should be distributed (see Figure 9). The system would track food distribution history while continuously monitoring to identify neighborhood children in hunger. The GEM algorithm could also calculate the best timing for subsequent distribution. Therefore, the state could provide food stamps or food directly to the children, eliminating substantive distribution gaps.

When a state receives a request from a resident for food stamps, the system could provide evidence to assist the review of the request. Furthermore, after the state distributes Electronic Benefit Transfer cards to the target area, it could gather the information from the retailers that participate in the Supplemental Nutrition Assistance Program credits. Records such as the food purchased by the household, transaction amounts, and other purchase-related data could be gathered. Based on the data, validation for assistance should be implemented. The use of blockchain would permit traceability as well as validity verification. Furthermore, GEM would lead to more efficient food/food stamp allocation since it would continuously get feedback from the data being collected from a variety of sources. Ultimately the physical "food stamps" may completely disappear with electronic tokens or direct targeted food distribution to the needy FBs.

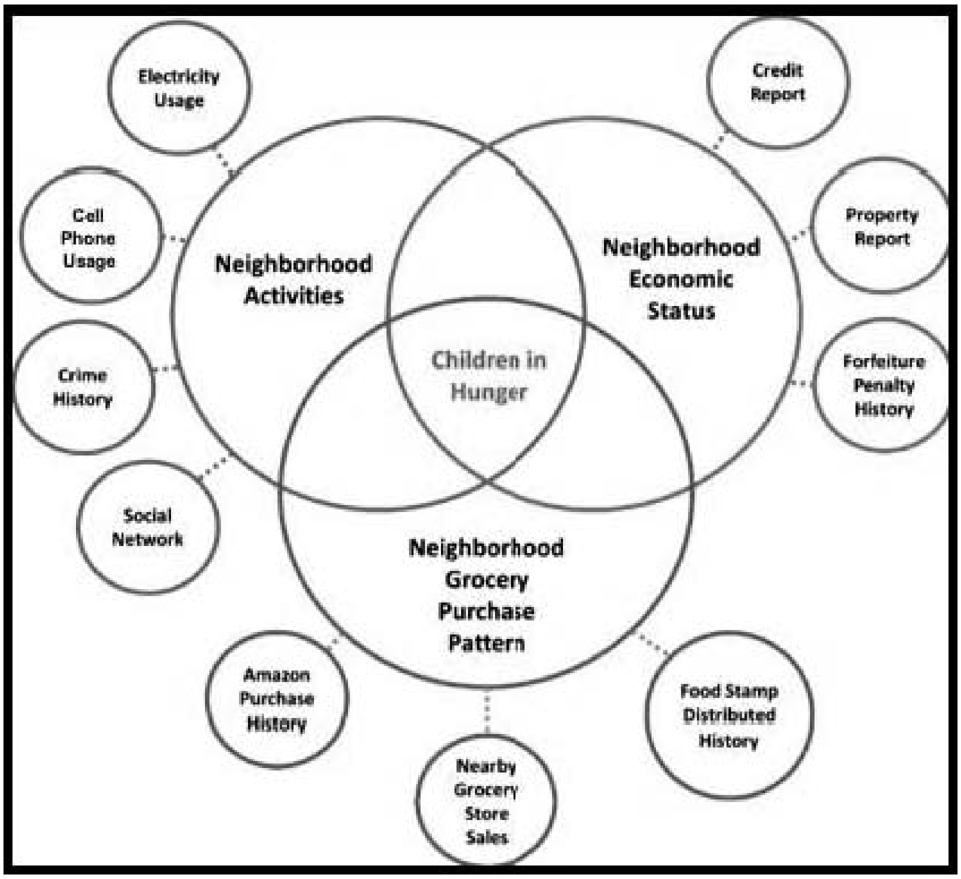
VI. RELATED ISSUES

Smart Government and Social Welfare

GEM serves as a way for the government to better support citizens who need help. Specifically, government agencies could utilize GEM to implement efficient and effective social policies. Further, policies could be more preventive of potential social problems since the government would have better insights about society. Better insights can lead to refinement and the development of social policies to yield a higher level of social welfare. Proper policy application and budget allocation will increase public trust. For instance, the Brazilian Federal Tax Authority has already implemented SPED (Sistema Publico de Escrituracao Digital or Public System of Digital Accounting) to track transactions made by firms. With the tremendous

Allen Press, Inc. ■ 17 June 2019 ■ 1:14 pm Page 12
//titan/Production/j/jeta/live_jobs/jeta-16/jeta-16-01/jeta-16-01-05/layouts/jeta-16-01-05.3d

FIGURE 8
Data Collection



Source: XXX.

transaction data stream gathered by firms, the tax authority could implement better taxation policies as well as reduce future tax evasion.

Recalling the food stamp case, GEM could further be applied to simplify the Supplemental Nutrition Assistance Program (SNAP) application process. Figure 10 is the current SNAP application process. The application process takes up to 30 days to receive food stamps, since the administrator must validate multiple pre-requirements. Further, the lack of information about the applicant could result in potential misuse or misallocation of food stamp benefits.

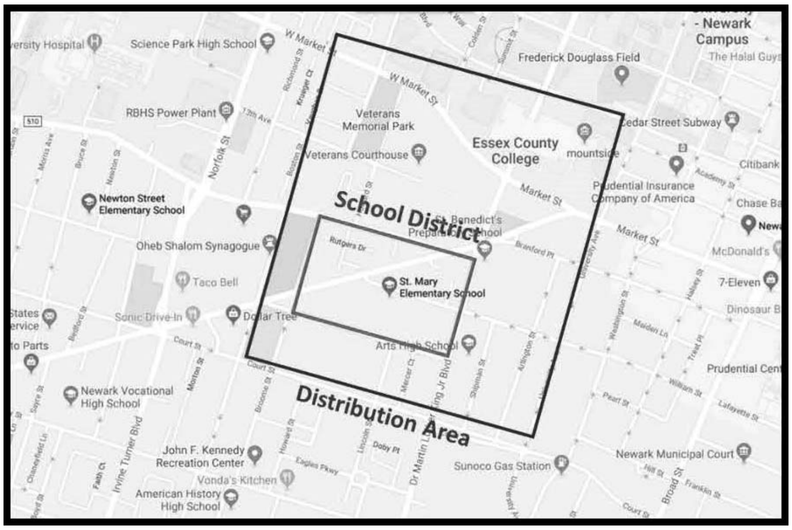
Figure 11 shows the proposed SNAP application process by utilizing a smart contract framework enabled by GEM. GEM analyzes the applicant's information to check whether they meet SNAP criteria. If the applicant's information meets these criteria, the applicant would easily enroll, or be automatically enrolled, in the program through a simple electronic interface supported by a smart contract. Simultaneously, the government could identify food providers near the applicant's residence and obtain validating evidence. Using such a process, the government could save time and effort in validating the application, and the beneficiary receives faster benefits. Benefits to both parties will lead to a higher level of social welfare by reducing the deadweight loss generated by an inefficient application and distribution process.

The GEM architecture would exist in an environment with an increasingly large set of data sources, most with unstructured data, that need to be linked/categorized in the emerging cyber ecosystem.

The Cyber Ecosystem

The three forms of data are creating exponentially expanding data sources and a disjointed/haphazard storage environment. Extensive, but still ineffective, methods are being developed to control and manage this environment. In general, data must be generated, captured, stored, analyzed, loosely linked, used, and selectively retained (Eugen and Marius 2013) for future use.

FIGURE 9
Food Stamp Distribution Area Suggestion

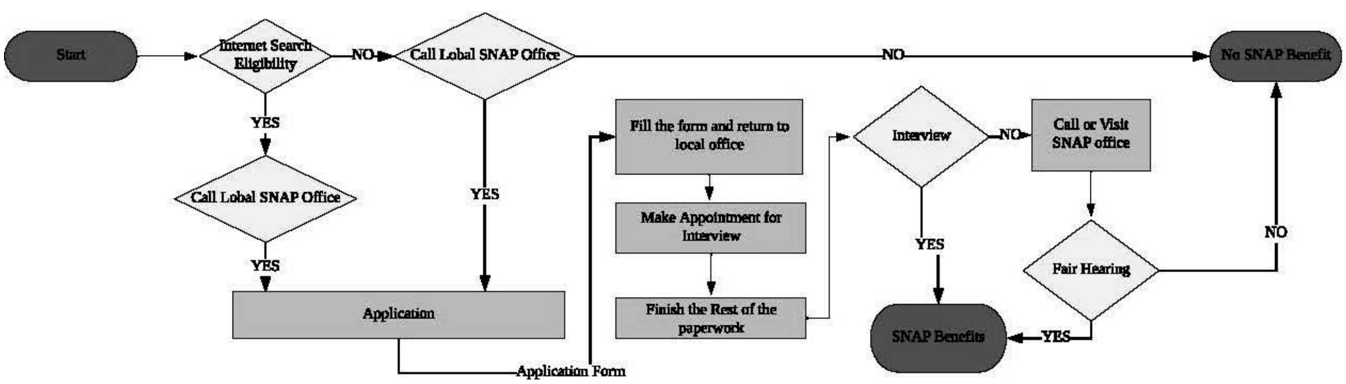


Source: XXX.

This data ecosystem that contains usable applications is in general permeated with unstructured data, typically textual but increasingly with voice and image, that are progressively somewhat usable. Social Welfare Information Networks (SWINs) are created in GEM creating logical links between interrelated information sources and fields. The business and government data ecosystem (Kozlowski and Vasarhelyi 2014) must be integrated for analysis either by links such as SWINs or by classificatory taxonomies (AICPA 2013).

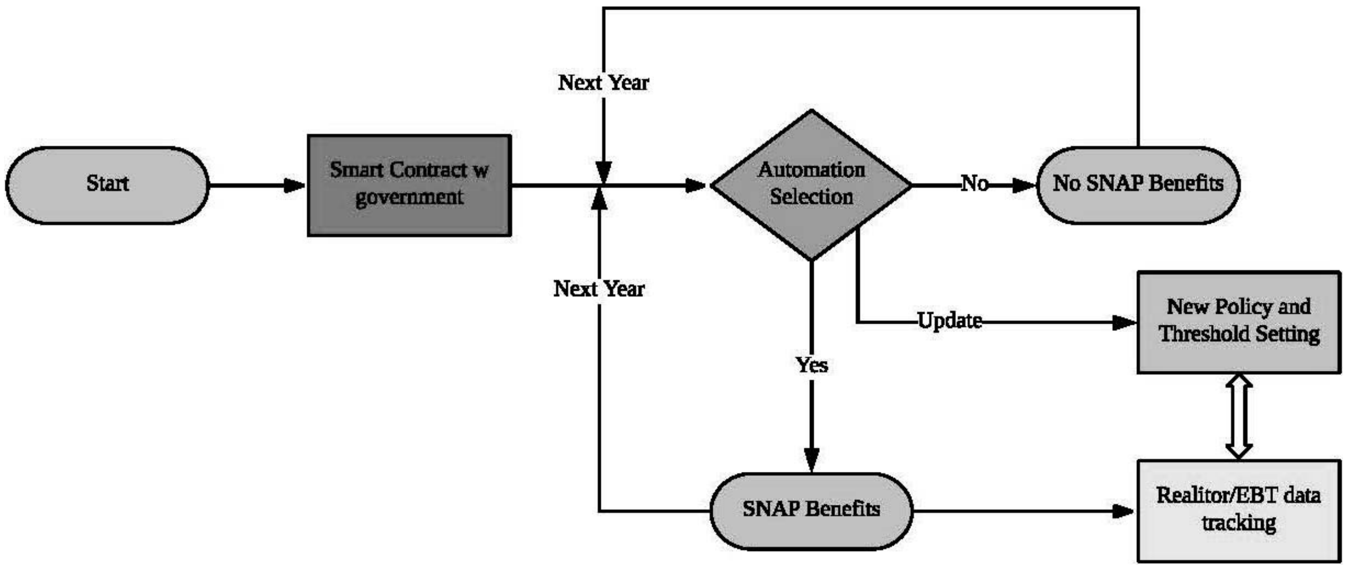
The generic approach here, named MED, is being applied to many fields. In general, what Zuboff (2019) calls corporate surveillance may be its main current business application. However, GEM is a set of ideas for active government intervention. Other forms of exogenous variable linkage and utilization are emerging in many fields and will also rely on software bots to be able to connect data sources that entail a highly heterogeneous set of data. The emerging area of bio-measurements, where devices such as health monitoring watches can be integrated through SWINs, adds a new dynamic and living environment to MED. These emerging and dynamic data sources and linkages are difficult to understand but offer highly valuable products and problems for which research is highly needed.

FIGURE 10
Current SNAP Application Process



Source: XXX.

FIGURE 11
?14 **GEM SNAP Process**



Source: XXX.

Privacy Assurance and Transparency

Using a series of privacy-preserving technologies, GEM methods aim to be as transparent as possible and assured of such by third-party assurance. The Federal Trade Commission (FTC) has emphasized that the customers' potential benefit of ?15 information sharing far exceeds customers' cost of information sharing (Muris 2013). On the other hand, the FTC emphasizes the need for strict regulation and safeguards to address how the data are collected and used in the system. Compared to "surveillance capitalism" or any type of law enforcement, GEM is engineered to support society but not to control and monitor the public. Privacy assurance by an independent third party[10] will reduce privacy concerns through examining and disclosing ?16 the control and structure (Hui 2007). Also, imposing strict guidance and ethical standards would mitigate privacy concerns and ?17 yield trust toward the system (Bansal 2008).

VII. CONCLUSION

The emergence of enormous disparate data sources, available in a multitude of alternate formats and on very different devices, is leading to a business process revolution. This revolution has been termed surveillance capitalism. Furthermore, similar technologies have been used in several countries for social control, or in some instances by individuals for criminal activities. This paper argues for an alternative approach in which these technologies are used to progressively replace outdated blanket government intervention programs. This approach advocates for original detective methodologies, direct government action, and outcome monitoring, all of which utilize Big Data-based frequent examination. The data structures termed SWINs would create links among large receptacles of data. Some of these receptacles are only available to the government. The government would be in a privileged situation where it could require data to be furnished that would not be available to most other parties due to privacy laws. Upon collection, these datasets would be encrypted/hashed in different manners aimed at achieving substantive data protection. However, once an ~~FG~~ is identified as being highly suspect of a serious pathology, either the system (if an automatic action is to be performed) or a government agent (e.g., a social worker) must be able to decode, identify, and act. The same privacy protection should exist in monitoring functions.

GEM is a research program for the future. This paper explores the ideas around using technology for a government monitoring and intervention process. Each social problem that emerges, either by GEM prediction or by political processes, requires serious quantitative and qualitative research to:

[10] See, for example, https://www.aicpa.org/press/pressreleases/2017/aicpa-unveils-cybersecurity-risk-management-reporting-framework.html

Allen Press, Inc. ■ 17 June 2019 ■ 1:14 pm Page 15
//titan/Production/j/jeta/live_jobs/jeta-16/jeta-16-01/jeta-16-01-05/layouts/jeta-16-01-05.3d

1. Identify relevant data sources;
2. Identify real cases of the pathology;
3. Link the diagnostic variables to the outcomes (diagnosed pathologies) using machine learning to create the suspicion functions for both diagnostic and monitoring purposes;
4. Parameterize costs and benefits of each process to optimize the threshold of intervention in a suspicion function;
5. Create the schemata of data protection;
6. Create the databases to support such an effort;
7. Develop process flows of the remediation processes;
8. Enable remediation processes;
9. Study cases of remediation to further define the suspicion functions used for monitoring;
10. Create quantitative protocols for suspending GEM action when cure is achieved, or the action was ineffective;
11. Create quantitative schemata to continually evaluate process effectiveness; and
12. Develop or identify an ethical framework for the governance of data privacy and monitoring.

A new and expanding research program into the usage and linkage of exogenous variables from internal business measurement process to a wide range of variables being progressively collected and ubiquitous in location is extremely needed.

—Helen Brown-Liburd
—Arion Cheong
—Miklos A. Vasarhelyi
—Xinxin Wang
Rutgers, The State University of New Jersey

REFERENCES

Adams, J. 1854. John Adams to John Jebb: September 10, 1785. The works of John Adams, second president of the United States: With a
?18 ?19 life of the author, notes, and illustrations, 9.

Alles, M. G., A. Kogan, and M. A. Vasarhelyi. 2002. Feasibility and economics of continuous assurance. *Auditing: A Journal of Practice & Theory* 21 (1): 125–138. https://doi.org/10.2308/aud.2002.21.1.125

American Institute of Certified Public Accountants (AICPA). 2013. *Audit Data Standards*. Available at: https://www.aicpa.org/content/dam/aicpa/interestareas/frc/assuranceadvisoryservices/downloadabledocuments/auditdatastandards-gl-august2013.pdf

American Institute of Certified Public Accountants (AICPA). 2015. *Continuous Auditing and Audit Analytics. Monograph*. New York, NY: AICPA.

Appelbaum, D. 2016. Securing Big Data provenance for auditors: The Big Data provenance black box as reliable evidence. *Journal of Emerging Technologies in Accounting* 13 (1): 17–36. https://doi.org/10.2308/jeta-51473

Appelbaum, D., A. Kogan, and M. A. Vasarhelyi. 2017. Big Data and analytics in the modern audit engagement: Research needs.
?20 *Auditing: A Journal of Practice & Theory* 36 (4): 1–27. https://doi.org/10.2308/ajpt-51684

Bansal, G., and F. Zahedi. 2008. *The Moderating Influence of Privacy Concern on the Efficacy of Privacy Assurance Mechanisms for Building Trust: A Multiple-Context Investigation*. Proceedings of 2008 International Conference on Information Systems (ICIS),
?21 Paris, France, December 14–17.

Bonczek, R. H., C. W. Holsapple, and A. B. Whinston. 2014. *Foundations of Decision Support Systems*. Waltham, MA: Academic Press.

Bregman, R. 2017. *Poverty Isn't a Lack of Character; It's a Lack of Cash*. Available at: https://www.ted.com/talks/rutger_bregman_
?22 poverty_isn_t_a_lack_of_character_it_s_a_lack_of_cash?language=en

Brown-Liburd, H., and M. A. Vasarhelyi. 2015. Big Data and audit evidence. *Journal of Emerging Technologies in Accounting* 12 (1): 1–16. https://doi.org/10.2308/jeta-10468

Cao, L. 2012. Social security and social welfare data mining: An overview. *IEEE Transactions on Systems, Man and Cybernetics. Part C: Applications and Reviews* 42 (6): 837–853.

Center of United Methodist Aid (CUMAC). 2016. *Hunger in NJ*. Available at: https://www.cumac.org/hunger-in-nj

Chui, M., M. Löffler, and M. Roberts. 2016. *The Internet of Things*. Available at: https://www.mckinsey.com/industries/high-tech/our-insights/the-internet-of-things

Costanza, R., I. Kubiszewski, E. Giovannini, H. Lovins, J. Mcglade, K. Pickett, K. Ragnarsdottir, D. Roberts, R. De Vogli, R. Wilkinson. 2014. Development: Time to leave GDP behind. *Nature* 505 (7483): 283–285.

Cuzzocrea, A. 2011. Retrieving accurate estimates to OLAP queries over uncertain and imprecise multidimensional data streams. In *International Conference on Scientific and Statistical Database Management*. (July): 575–576. Heidelberg, Germany: Springer.

Dai, J., and Q. Li. 2016. Designing audit apps for armchair auditors to analyze government procurement contracts. *Journal of Emerging*
?23 *Technologies in Accounting* 13 (2): 71–88. https://doi.org/10.2308/jeta-51598

Davidson, P. 2012. Is economics a science? Should economics be rigorous? *Real-World Economics Review* 59: 58–66.

Degli Esposti, S. 2014. When Big Data meets data surveillance: The hidden side of analytics. *Surveillance & Society* 12 (2): 209–225. https://doi.org/10.24908/ss.v12i2.5113

Eugen, C., and C. Marius. 2013. Right to be forgotten. *Anales Universitatis Apulensis Series Jurisprudentia*. 16.

?24 Gentry, C., and D. Boneh. 2009. *A Fully Homomorphic Encryption Scheme*. Stanford, CA: Stanford University.

Gillingham, P., and T. Graham. 2017. Big Data in social welfare: The development of a critical perspective on social work's latest "electronic turn." *Australian Social Work* 70 (2): 135–147. https://doi.org/10.1080/0312407X.2015.1134606

Hahn, F. 1991. The next hundred years. *Economic Journal (London)* 101 (404): 47–50. https://doi.org/10.2307/2233836

Hui, K. L., H. H. Teo, and S. Y. T. Lee. 2007. The value of privacy assurance: An exploratory field experiment. *Management Information*
?25 *Systems Quarterly* 31 (1): 19–33. https://doi.org/10.2307/25148779

Issa, H. 2013. *Exceptional Exceptions*. Doctoral dissertation, Rutgers, The State University of New Jersey.

Issa, H., T. Sun, and M. A. Vasarhelyi. 2016. Research ideas for artificial intelligence in auditing: The formalization of audit and workforce supplementation. *Journal of Emerging Technologies in Accounting* 13 (2): 1–20. https://doi.org/10.2308/jeta-10511

Johnson, S. D., and K. J. Bowers. 2004. The burglary as clue to the future: The beginnings of prospective hot-spotting. *European Journal of Criminology* 1 (2): 237–255. https://doi.org/10.1177/1477370804041252

Johnson, S. D., W. Bernasco, K. J. Bowers, H. Elffers, J. Ratcliffe, G. Rengert, and M. Townsley. 2007. Space-time patterns of risk: A cross national assessment of residential burglary victimization. *Journal of Quantitative Criminology* 23 (3): 201–219. https://doi.org/10.1007/s10940-007-9025-3

Kaplan, E., and C. Hegarty. eds. 2005. *Understanding GPS: Principles and Applications*. Norwood, MA: Artech House.

Kim, G. H., S. Trimi, and J. H. Chung. 2014. Big-data applications in the government sector. *Communications of the ACM* 57 (3): 78–85. https://doi.org/10.1145/2500873

Kogan, A., and C. Yin. 2018. *Privacy Preserving Information Sharing within an Audit Firm*. Working paper, Rutgers, The State University of New Jersey, CarLab.

Kogan, A., M. Alles, M. 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. https://doi.org/10.2308/ajpt-50844

Kopetz, H. 2011. Internet of things. In *Real-Time Systems*, 307–323. New York, NY: Springer.

Kozlowski, S., and M. A. Vasarhelyi. 2014. *An Audit Ecosystem: A Starting Point with Definitions, Attributes, and Agents*. Working paper, Rutgers, The State University of New Jersey, CarLab.

Kuenkaikaew, S. 2013. *Predictive Audit Analytics: Evolving to a New Era*. Ph.D. dissertation, Rutgers, The State University of New Jersey.

McAfee, A., and E. Brynjolfsson. 2012. Big Data: The management revolution. *Harvard Business Review* (October): 60–66.

McGough, S., J. Brownstein, J. Hawkins, and M. Santillana. 2017. Forecasting Zika incidence in the 2016 Latin America outbreak: Combining traditional disease surveillance with search, social media, and news report data. *PLoS Neglected Tropical Diseases* 11 (1): . https://doi.org/10.1371/journal.pntd.0005295

Mellon, J. 2013. Where and when can we use Google Trends to measure issue salience? *Political Science & Politics* 46 (2): 280–290. https://doi.org/10.1017/S1049096513000279

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

Muris, T. J. 2004. *The Federal Trade Commission and the Future Development of U.S. Consumer Protection Policy*. Remarks before the
?26 Aspen Summit: Cyberspace and the American Dream. The Progress & Freedom Foundation, Aspen, CO, August 19.

Naughton, J. 2019. *"The Goal Is to Automate Us": Welcome to the Age of Surveillance Capitalism*. Available at: https://www.theguardian.com/technology/2019/jan/20/shoshana-zuboff-age-of-surveillance-capitalism-google-facebook

North, D. C. 1995. Economic theory in a dynamic economic world. *Business Economics* 30 (1): 7–12.

O'Leary, D. 2015. Armchair auditors: Crowdsourcing analysis of government expenditures. *Journal of Emerging Technologies in*
?27 *Accounting* 12 (1): 71–91. https://doi.org/10.2308/jeta-51225

O'Neil, C. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Portland, OR: Crown-Broadway Books.

Ripberger, J. T. 2011. Capturing curiosity: Using internet search trends to measure public attentiveness. *Policy Studies Journal: The*
?28 *Journal of the Policy Studies Organization* 39 (2): 239–259. https://doi.org/10.1111/j.1541-0072.2011.00406.x

Romero, S., G. Gal, T. J. Mock, and M. A. Vasarhelyi. 2012. A measurement theory perspective on business measurement. *Journal of Emerging Technologies in Accounting* 9 (1): 1–24. https://doi.org/10.2308/jeta-50396

Rosario, A. 2019 *Three Essays in Audit Innovation*. Ph.D. dissertation, Rutgers, The State University of New Jersey.

Rymph, C. E. 2012. From "economic want" to "family pathology": Foster family care, the new deal, and the emergence of a public child welfare system. *Journal of Policy History* 24 (1): 7–25. https://doi.org/10.1017/S0898030611000352

Sabeti, H. 2011. The for-benefit enterprise. *Harvard Business Review* 89 (11): 98–104.

Samuelson, P. A. 1948. *Foundations of Economic Analysis*. Boston, MA: Harvard University Press.

Schneier, B. 2015. *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. New York, NY: WW Norton and Company.

Smith, A. 1937. *The Wealth of Nations*. New York, NY: Modern Library.

Allen Press, Inc. ■ 17 June 2019 ■ 1:14 pm Page 17
//titan/Production/j/jeta/live_jobs/jeta-16/jeta-16-01/jeta-16-01-05/layouts/jeta-16-01-05.3d

Stephens-Davidowitz, S. 2017. *Everybody Lies*. New York, NY: Dey Street Books. (Kindle Edition).

Stiglitz, J. E. 1991. *The Invisible Hand and Modern Welfare Economics*. *NBER Working Paper No. 3641*. Cambridge, MA: National Bureau of Economic Research.

The Economist. 2016. *The Trouble with GDP—Measuring Economies*. Available at: https://www.economist.com/briefing/2016/04/30/the-trouble-with-gdp

Ura, K., S. Alkire, T. Zangmo, and K. Wangdi. 2012. *A Short Guide to Gross National Happiness Index*. Thimphu, Bhutan: The Centre for Bhutan Studies.

U.S. Census Bureau. 2017. *2013–2017 American Community Survey 5-Year Estimates*. Available at: https://www.census.gov/data/
?29 developers/data-sets/acs-5year.2016.html

Vasarhelyi, M. A., and F. B. Halper. 1991. The continuous audit of online systems. *Auditing: A Journal of Practice & Theory* 10 (1): 110–125.

Vasarhelyi M. A., M. G. Alles, and K. T. Williams. 2010. *Continuous Assurance for the Now Economy*. A Thought Leadership Paper for
?30 the Institute of Chartered Accountants in Australia.

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

Williams, R. 2015. Using data analytics for oversight and efficiency. *Journal of Government Financial Management* 64 (2): 18.

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

Zuboff, S. 2019. *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. London, U.K.: Profile Books.

APPENDIX A

GEM Ideas Proposed in Workshops

Water Quality
Violence against Women
Unhealthy Consumption (drugs)
AIDS Detection
Air Pollution
Carbon Emissions
?31 Monitoring "NEM-NEM" s
Low Weight Babies
Early Detection of Loan Default
Public Sector Fraud and Public Sector Waste
Education Efficiency
Child Molestation
Universal Basic Income
?32 Water Quality
Suicide among the Elderly
Voting Patterns
Patterns of STD
Brazilian Bolsa Familia
Children's Vaccination
Decreasing Birth Rates

APPENDIX B

Some Miscellaneous Data Sources

Facebook Utterances
Bank Transactions[11]

[11] Data may be obtained from many sources, from the bank, from the client, from the intervening merchant, from the credit card company, from a camera in the store, from utterances in social media. Multiple sourcing may be a valuable provenance validation mechanism (Appelbaum 2016).

Supermarket Sales Records
Taxicab Routings in New York City
Amazon Sales
Supermarket Scanner Data
Weather Data
Microeconomic Data
Labor Data
Twitter Data
Highway Transit Data
Electric Consumption Data
IP Location Data
Bulletin Board Utterance Data
News Media Content
Website Content
Telephone Call ~~Meta Data~~
Email Address Lists
Web Advertising Links
Store Sales Records
Credit Card Sales Records
Pharmacy Records
Prescription Records
Doctor's Diagnostic Records
Government Paid Checks
Government Contracts
Government Employees' Salaries and Other Info
Traffic and Parking Tickets
Car Registrations
House Sales
Airline Routes
Airline Traveler Lists

Queries for jeta-16-01-05

1. Author: For the sentence beginning "The government would be in a privileged . . ." the phrase "where it could acquire data not be made available" is grammatically incorrect. Should "acquire" be revised to "require"? Or perhaps ". . .where it could acquire data not made available to most other parties"? Please review and revise as warranted. Jan
2. Author: Your copyright release form was not submitted at the time that the article page proofs were generated. If you have not already done so, then please send the copyright release form to chelsea.matthews@aaahq.org. If you need the form, then please contact Chelsea Matthews. Jan
3. Author: AAA cannot accommodate color figures in the print journal. They have been rendered as PDF greyscale and may no longer be interpretable. If this is so, please resupply graphic(s) in B&W or grayscale as hi-def PDF (preferred), JPG, PNG, or PPT. We cannot guarantee the visual quality of any graphic that is resupplied in a format other than a B&W PDF. If you prefer, we can offer the original color versions as a downloadable supplemental file. A figure note will be added to alert the reader. Please advise. Copyeditor
4. Author: Please provide a citation (and reference entry if needed) for the graphics used in Figures 1–11 (or within text), or advise that the graphic(s) is original to this article. Copyeditor
5. Author: The in-text citation "Ripberger et al. 2011" is not in the reference list. Please correct the citation, add the reference to the list, or delete the citation. Copyeditor
6. Author: The in-text citation "Wang and Kogan 2018" is not in the reference list. Please correct the citation, add the reference to the list, or delete the citation. Copyeditor
7. Author: re footnote 5: The program names have been revised per web search. Please approve as is, or mark to revise as preferred. Copyeditor
8. Author: The in-text citation "Kristein 2015" is not in the reference list. Please correct the citation, add the reference to the list, or delete the citation. Copyeditor
9. Author: AAA formatting requires a mention of numbered tables and figures. Figure 4 was not mentioned in the text. A mention has been inserted. Please approve, or mark to revise and/or move as appropriate. Copyeditor
10. Author: The in-text citation "Gentry 2009" is not in the reference list. Please correct the citation, add the reference to the list, or delete the citation. Copyeditor
11. Author: The in-text citation "Agrawal and Srikant 2000" is not in the reference list. Please correct the citation, add the reference to the list, or delete the citation. Copyeditor
12. Author: The in-text citation "Aggarwal and Phillip 2008" is not in the reference list. Please correct the citation, add the reference to the list, or delete the citation. Copyeditor
13. Author: There are two graphics labeled Figure 8 in the manuscript. The second Figure 8 has been re-labeled Figure 9 and a mention of Figure 9 has been inserted in the test. The rest of the figures in the article have been re-numbered accordingly, along with their in-text mentions. Please review carefully. Copyeditor
14. Author: re Figure 11: Should "Realitor/EBT data tracking" be 'Realtor/EBT data tracking' instead? Please resupply a corrected graphic in B&W or grayscale as hi-def PDF (preferred), JPG, PNG, or PPT. We cannot guarantee the visual quality of any graphic that is resupplied in a format other than a B&W PDF. Please note that we cannot correct the existing graphic. Jan
15. Author: The in-text citation "Muris 2013" is not in the reference list. Please correct the citation, add the reference to the list, or delete the citation. Copyeditor
16. Author: The in-text citation "Hui 2007" is not in the reference list. Please correct the citation, add the reference to the list, or delete the citation. Copyeditor
17. Author: The in-text citation "Bansal 2008" is not in the reference list. Please correct the citation, add the reference to the list, or delete the citation. Copyeditor
18. Author: Reference "Adams, 1854" is not cited in the text. Please add an in-text citation or delete the reference. If a citation is added to the text, then please provide publishing information for this work. Copyeditor
19. Author: Some reference entries and corresponding in-text citations have been revised or updated based on the information provided by CrossRef and appropriate web searches. Review the "References" section carefully, and provide approval (with an "OK" to "Confirmed" response to this query). Please note that there may be other citation- or reference-specific queries that need your attention. Thank you. Copyeditor
20. Author: Reference "Appelbaum, Kogan, and Vasarhelyi 2017" is not cited in the text. Please add an in-text citation or delete the reference. Copyeditor
21. Author: Reference "Bansal and Zahedi 2008" is not cited in the text. Please add an in-text citation or delete the reference. Copyeditor
22. Author: Reference "Bregman 2017" is not cited in the text. Please add an in-text citation or delete the reference. Copyeditor
23. Author: Reference "Dai and Li 2016" is not cited in the text. Please add an in-text citation or delete the reference. Copyeditor
24. Author: Reference "Gentry and Boneh 2009" is not cited in the text. Please add an in-text citation or delete the reference. Copyeditor
25. Author: Reference "Hui, Teo, and Lee 2007" is not cited in the text. Please add an in-text citation or delete the reference. Copyeditor
26. Author: Reference "Muris, 2004" is not cited in the text. Please add an in-text citation or delete the reference. Copyeditor
27. Author: Reference "O'Leary 2015" is not cited in the text. Please add an in-text citation or delete the reference. Copyeditor
28. Author: Reference "Ripberger 2011" is not cited in the text. Please add an in-text citation or delete the reference. Copyeditor
29. Author: re U.S. Census Bureau. 2017: The original URL did not link to a working webpage. A substitute was found. Please check this link to confirm that it lands on an appropriate webpage or provide a preferred link. Copyeditor
30. Author: Vasarhelyi et al. 2010: Please provide a sponsoring organization for the paper or publishing information if published. Copyeditor
31. Author: re Appendix A: "Monitoring "NEM-NEM" s" is grammatically incorrect. Please review and revise as warranted. Copyeditor
32. Author: re Appendix A: "Water Quality" is listed twice. OK? Please review and revise as warranted. Copyeditor
33. Author: For the sentence beginning "Additionally, continuous monitoring," copyeditor requests confirmation that "assure that agencies comply" is the correct phrasing, as opposed to "ensure that the agencies comply." Please review the sentence, and revise if warranted. Copyeditor