3-1-2018

Making Government Data Valuable for Constituents: The Case for the Advanced Data Analytics Capabilities of the ENHANCE Framework

Stephen Kozlowski
Eastern Illinois University

Hussein Issa
Rutgers - The State University of New Jersey, Newark

Deniz Appelbaum
Montclair State University, appelbaumd@mail.montclair.edu

JOURNAL OF EMERGING TECHNOLOGIES IN ACCOUNTING
Vol. 15, No. 1
Spring 2018
pp. 155–167

American Accounting Association
DOI: 10.2308/jeta-52096

Making Government Data Valuable for Constituents: The Case for the Advanced Data Analytics Capabilities of the ENHANCE Framework

Stephen Kozlowski
Eastern Illinois University

Hussein Issa
Rutgers, The State University of New Jersey, Newark

Deniz Appelbaum
Montclair State University

ABSTRACT: Constituent demands for improved transparency in governmental reporting have increased since the 2008 financial crisis in the U.S. that impacted the financial well-being of many individuals and entities, including state and local municipalities. Since then, several governmental entities have provided an open data environment. However, these early initiatives do not incorporate robust analytic capabilities that are able to satisfy numerous and varied constituent inquiries. This research proposes development of the ENHanced ANalytic Constituent Environment (ENHANCE) framework in response to the analytic nature of the inquiries from various governmental stakeholders such as citizens, analysts, bond investors, creditors, vendors, auditors, and oversight officials. The ENHANCE framework bridges the gap between constituents' current analytic capabilities and their demands for increased transparency. This research contributes to the literature by highlighting the capabilities of the proposed ENHANCE tool and the underlying series of analytic "apps" that can provide meaningful information to a governmental entity's stakeholders.

Keywords: open governmental reporting; data analytics; governmental data.

INTRODUCTION

Most average citizens that have an interest in understanding how their governments operate, also referred to as "armchair auditors" (O'Leary 2015), are not likely to have the tools or expertise required to produce sophisticated analytics on their own. While the availability and scope of governmental data has increased in the form of open data portals, it is not obvious that the recipients of this information have sufficient skills to obtain value from it. This condition is aggravated by the quasi-absence of analytics on such data portals. Accordingly, the objective of this paper is to examine this problem and to propose a framework (ENHANCE) of a user-friendly decision support system that could exploit the proliferation of these open data portals. Such a framework has the potential to empower nonsophisticated users with expert-like analytical skills to turn raw data into meaningful information and, as such, improve transparency.

Although the currently proposed ENHANCE framework will produce user-friendly analytic capabilities focused on governmental expenditure data, the concept is also intended to include the flexibility necessary to meet any number of heretofore undefined needs as new technologies emerge that may impact the accounting function. For example, there is much discussion about the application of a newly emerging technology, blockchain, that can potentially transform current accounting and assurance practices (Dai and Vasarhelyi 2017). The proposed benefits of blockchain include reduced costs, increased speed in recording transactions, and reduced fraud risk, among others (Dai and Vasarhelyi 2017). It can be envisioned that governmental agencies, in

Supplemental materials can be accessed by clicking the links in Appendix A.

Editor's note: Accepted by Miklos A. Vasarhelyi.

Submitted: May 2017
Accepted: March 2018
Published Online: March 2018

looking to reduce their operating costs and the financial burden they place on their constituency, will seriously consider blockchain technology as it develops. Blockchain incorporates a decentralized public ledger infrastructure for recording transactions among unfamiliar parties (Dai and Vasarhelyi 2017). Dai and Vasarhelyi (2017) also discuss applying the concept of a "triple-entry accounting system" as a means for documenting the accounting entries generated in a blockchain environment. ENHANCE is presently intended to simplify the search for the user-specified governmental dataset to be investigated. These search capabilities can also be leveraged to locate governmental data in a decentralized public ledger, "triple-entry" accounting data that are stored in a blockchain ledger, and data configured in a traditional double-entry system. With the potential for significant dispersal of accounting data in a blockchain environment, the search capabilities envisioned for ENHANCE can also be used to support the assurance function as it grapples with significant changes to the structure and location of the accounting data that must be audited, and thus expand the potential users that would benefit from it. Financial auditors, as well as armchair auditors, can be incorporated in such a system as a monitoring node on the blockchain.

The Pew Research Center conducted a nation-wide survey[1] to gauge public sentiment concerning the recent governmental initiatives fostering data transparency and the use of such data. The report documenting the results of the survey (Pew Research Center 2015) provides insights into the public's level of awareness of governmental initiatives to provide data, whether these initiatives are fulfilling citizen needs to measure government performance, public opinion regarding the ability of these initiatives to improve government performance and accountability, and the specific online governmental activities that citizens undertake. In summary, the survey results indicated the following (Pew Research Center 2015):

- 65 percent of Americans have accessed the internet in the past 12-month period to find government-related data or information
- Approximately 5 percent of respondents indicated that their governments are highly effective in sharing their data
- Approximately 19 percent can relate instances where their local governments either did or did not provide sufficiently useful information

The term, and concept "armchair auditing" originated in the United Kingdom, with an early definition stated as:[2] "a website that uses open spending data provided by councils to dynamically generate reports for users on council spending according to various criteria." Former British Prime Minister David Cameron, in a commentary in 2009, included transparency as an item in his proposal for political reform. "Transparency tears down the hiding places for sleaze, overspending and corruption. Soon enough all MPs' expenses are going to be published online for everyone to see: I and the rest of the shadow cabinet are already doing it. And if we win the next election, we're going to do the same for all other public servants earning over £150,000" (Cameron 2009). In his commentary Cameron also stated: "Just imagine the effect that an army of armchair auditors is going to have on those expense claims" (Cameron 2009). If this form of governmental data analysis is undertaken on a large scale by a multitude of constituents, it is possible that these users will not all possess a significant level of technical expertise to allow them to develop and execute sophisticated analytic tools without additional support.

Motivated by some concerns raised by O'Leary (2015) with respect to the armchair auditor concept, this paper proposes the ENHANCE framework to answer the following research questions:

RQ1: Can technology be leveraged by governments to engage potential armchair auditors?

RQ2: What kinds of tools and analytics can be developed to facilitate armchair auditing?

RQ3: Can advanced analytical tools increase the quality of the armchair auditor contributions?

The contribution of this paper to the accounting literature is as follows: First, this research describes how the capabilities of the proposed ENHANCE tool can support advanced data analytic capabilities and provide meaningful information to a governmental entity's stakeholders that will permit them to make educated assessments regarding the financial health of their governmental entity. These analytic capabilities are not generally available to constituents who do not possess significant technical expertise. ENHANCE provides the means for engaged constituents to generate meaningful expenditure information about their governmental entity. Second, this paper discusses typical outputs and formats of the analytic results that would be presented to ENHANCE users, as well as a practical example illustrating the value provided by ENHANCE's data analytic capabilities. Third, a literature review of transparency and current practices in government financial reporting is provided. In addition to the practical implications (for governments as well as other stakeholders), this research contributes to the academic literature by proposing different venues for future research.

The remainder of this paper is structured as follows: The "Background" section presents a review of relevant literature, examples of U.S. and U.K. open data initiatives, and a discussion of how advanced analytic tools can enhance governmental

[1] This survey can be found at http://www.pewinternet.org/2015/04/21/open-government-data/
[2] The full text of Prime Minister Cameron's speech is available for download, see Appendix A.

transparency efforts. The third section presents an illustration of the ENHANCE framework. The fourth section is a discussion of ENHANCE as a decision support system. The next section illustrates several examples of the ENHANCE framework for an enabled armchair auditor, and the following section presents an example of ENHANCE analytics in practice. The "Conclusion" section completes this paper.

BACKGROUND

A lack of analytic tools to undertake sophisticated analytics, as well as other issues with several of the early open data initiatives, limits the ability of average citizens, and especially armchair auditors, to effectively monitor governmental activities. Some of the early data published do not include relevant date information for either the recording of the transaction or the occurrence of the underlying activity (O'Leary 2015). Additionally, many of the early data were published in PDF or flat file formats, or even CSV format, that limit the ability to analyze or link the data (O'Leary 2015). Descriptions of the exact items purchased were also not necessarily provided, nor was there access to comparative cost information to support cost-benefit analyses (O'Leary 2015).

When studying one of the first authorities to adopt the U.K. Local Government Transparency Code,[3] Frank and Oztoprak (2015) conducted interviews with officers, politicians, and potential user groups. At that early point in the open data initiative, the respondents indicated that the available datasets were being used to a very limited extent, which could be expected as there had been little effort to promote the availability of the datasets. Some local authorities had created "profiles" that included data relevant to the locality from several sources and had provided tools to present and interpret the data (Frank and Oztoprak 2015). However, although the users had to interpret the results they were provided, these early initiatives did provide the users a sense of what could be accomplished.

Worthy (2015) conducted research to understand the impact of the U.K. Local Government Transparency Code by means of surveys, freedom of information (FOI) requests, interviews, and media analysis. It was determined that business users comprised about 39 percent of all users, followed by the media at 31 percent, the public (the targeted user group) at 21 percent, and non-governmental organizations at 7 percent. It was envisioned that an army of armchair auditors would force the municipalities to maintain accountability for the data, but non-governmental organizations (NGOs), which have access to analytic capabilities not available to the general public, were already fulfilling that role (Worthy 2015). Greater participation in governmental monitoring is linked to greater transparency, but the primary participants appear to be NGOs, driven by controversial situations such as scandals or the cessation of local services or functions, and not the envisioned army of armchair auditors (Worthy 2015). As for information transmission, that activity had been undertaken primarily by the media, but only to a limited extent (Worthy 2015). A possible explanation is that the raw data themselves did not supply valuable information without further explanation, as could be provided by advanced analytic tools (Worthy 2015).

One obstacle at this point for successful open data initiatives, was a lack of integration of the raw data with analytics to provide meaningful information (Worthy 2015). An obstacle to robust armchair auditor participation was an inherent weakness in relying on crowdsourcing techniques, which is an unstable activity dependent on typically very small groups (Worthy 2015). In order to have effective and efficient armchair auditing, a certain level of expertise and technical skills is required. If the necessary skills are available, coupled with the existence of open data portals, armchair auditing can prove informative and powerful. For instance, Dai and Li (2016) propose a set of audit analytic "apps" to identify irregularities in governmental procurement contracts in Brazil.

The State Budget Crisis Task Force (http://www.statebudgetcrisis.org/wpcms/) was formed to develop an understanding of the extent of the fiscal problems faced by U.S. states subsequent to the financial crisis. Some of the more serious financial issues identified in these states included the use of cash-based budgeting, the absence of relevant midyear financial planning, and a lack of clarity regarding future financial obligations.

The State Budget Crisis Task Force has not been the only entity involved in a review of state financial reporting. The Governmental Accounting Standards Board (GASB) (http://www.gasb.org/home) issued a research brief in 2011 concerning the timeliness of state and local government financial reporting.[4] A summary of this brief was published in the *Journal of Government Financial Management*.[5] The GASB research both examined the timeliness of financial reporting by state and local governments during the period 2006–2008 and conducted a survey of financial information users with respect to their perceptions of the impact of timeliness on the usefulness of financial reporting. The findings reflect that 73 percent of the larger

[3] See, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/360711/Local_Government_Transparency_Code_2014.pdf/. The U.K. Department for Communities and Local Government issued a transparency code that requires 400 of the larger U.K. authorities to publish a minimum set of open data for transparency reasons.

[4] The research brief can be found at: http://www.gasb.org/cs/ContentServer?c=Document_C&cid=1176158316214&d=&pagename=GASB%2FDocument_C%2FGASBDocumentPage

[5] The *Journal of Government Financial Management* is published quarterly by the Association of Government Accountants, and information on the journal can be found at https://www.agacgfm.org/Resources/Journal-of-Government-Financial-Management.aspx

governments studied, and 46 percent of the smaller governments, issued their annual financial reports within six months following the close of the fiscal year.

Several governmental entities in the U.S. now provide increased openness and transparency with respect to their operations, and the adoption of state of the art technology, which provides the cost-savings and convenience, has supported much of this change. Although several of these governmental entities provide online access to their data, the format of these data does not always accommodate processing and understanding for additional insights.[6] For instance, certain governments provide reports in PDF format, rendering data analysis prohibitively costly due to the required laborious and time-consuming manual work. On the other hand, numerous governments, especially those pioneering open data portal initiatives, provide data in multiple (more accessible) formats, including machine-readable formats. Such governments are motivated by their desire to offer their constituents an improved level of transparency.[7] Examples of governments that provide data portals are available in the Online Appendix (see Appendix A for the link to the downloadable file). The advances in technology that enable the re-engineering of governmental budgetary and transactional reports also allow governments to integrate electronic reporting capabilities into many of their projects.

A few data portals offer some rudimentary analytics; however, those are not user-friendly and are difficult for the nonsophisticated user to understand. They primarily provide for the extraction of data but offer very limited analytical capabilities. ENHANCE proposes to provide user-friendly and expert-like analytical skills to the average user.

In addition to providing appropriate governmental data upon which advanced analytic tools can be employed, any effort to standardize the type and format of the data provided will significantly improve the ability for an analytic tool to complete the desired analytics. In an effort to leverage the implementation of standardized data, auditing, and reporting capabilities in corporate organizations, a partnership between the Oregon State Controller's Division, the Association of Government Accountants (AGA), the Governmental Accounting Standards Board (GASB), and PricewaterhouseCoopers (PWC) was formed to demonstrate the feasibility of interactive data for public sector reporting (Mueller 2009). The goal of the Oregon project was to develop a state and local taxonomy that was compliant with the GASB, with a specific focus on taxonomy development for two of the primary Comprehensive Annual Financial Reports (CAFR): the Statement of Activities and Statement of Net Assets.[8] The project started by converting the spreadsheet data of the Oregon CAFR into tagged data elements that would be machine readable. The spreadsheet data from the Oregon controller's office were loaded into taxonomy-building software that processed the data and converted them into the elements of an XBRL schema from which an XBRL taxonomy was developed. The project was able to create a taxonomy that included around 156 GASB-compliant tagged data elements (96 elements from the Statement of Net Assets and 60 elements from the Statement of Activities). The Oregon project was viewed as a successful undertaking as well as an opportunity to showcase the implementation of XBRL in the area of governmental reporting and highlight its potential benefits for the future.

Another example that demonstrates how meaningful information is not clearly presented can be found in the comparison of costs for purchased items and purchase transaction details (O'Leary 2015). The ENHANCE framework provides the ability for spending comparisons across departments within a governmental entity (intragovernmental), and if comparable data taxonomies exist, can provide comparisons between different governmental entities (intergovernmental). Armchair auditors may be faced with information overload if they choose to undertake analyses over very large amounts, or even the complete dataset, of the expenditure data available (O'Leary 2015). ENHANCE provides the capability for not only appropriate data summarizations that enable users to comprehend the results of the analytics, but also may include the tools to undertake analyses. One such analysis could result in the output of only "exceptional exceptions" (Issa 2013). In an "exceptional exceptions" type of application, anomalies or outliers would be filtered and sorted according to their priorities. The identification and prioritization of exceptions and anomalies that O'Leary (2015) presents may need to be addressed by future research.

ILLUSTRATION OF THE ENHANCE FRAMEWORK: RESPONSE TO RQ1

The purpose of this research is to propose an ENHanced ANalytic Constituent Environment (ENHANCE) framework, facilitated by technology and the availability of open governmental data. Such a framework has the capability to fulfill the

[6] Many of the available data are provided in PDF format, which is difficult to manipulate for analytical purposes.

[7] While the ENHANCE framework proposed in this study would provide optimal outcomes if the database were in machine-readable form, it can still be applied in less optimal situations.

[8] The report can be accessed on the AGA website: https://www.agacgfm.org/getattachment/Resources/Online-Library/Research-Reports/XLBRPublicSectorSept2008.pdf.aspx

reporting and analytic requirements of the various governmental stakeholders such as citizens, analysts, bond investors, creditors, vendors, auditors, and oversight officials.[9] The steps to fulfill these requirements are as follows:

1. The availability of open governmental data, presented in a standardized and usable format.
2. The design of a series of imbedded analytic apps that can provide meaningful information to the entity's stakeholders.
3. The development of the ENHANCE framework where the analytic apps can function, acting upon the standardized governmental data to support constituent reporting requirements.

Data analytics as presented in this paper include techniques developed by other researchers, such as "exceptional exceptions" (Issa 2013) and advanced data clustering techniques (Byrnes 2015). This research also includes data analytic techniques as described by Provost and Fawcett (2013):

- Exploratory data analytics/descriptive statistics
- Anomaly/outlier detection
- Time-series analysis
- Threshold levels
- Cluster analysis

The capabilities of the ENHANCE framework can be extended beyond that described in this paper to include:

- Exceptional exceptions
- Cross-entity analysis
- Development of KPIs
- Pattern recognition

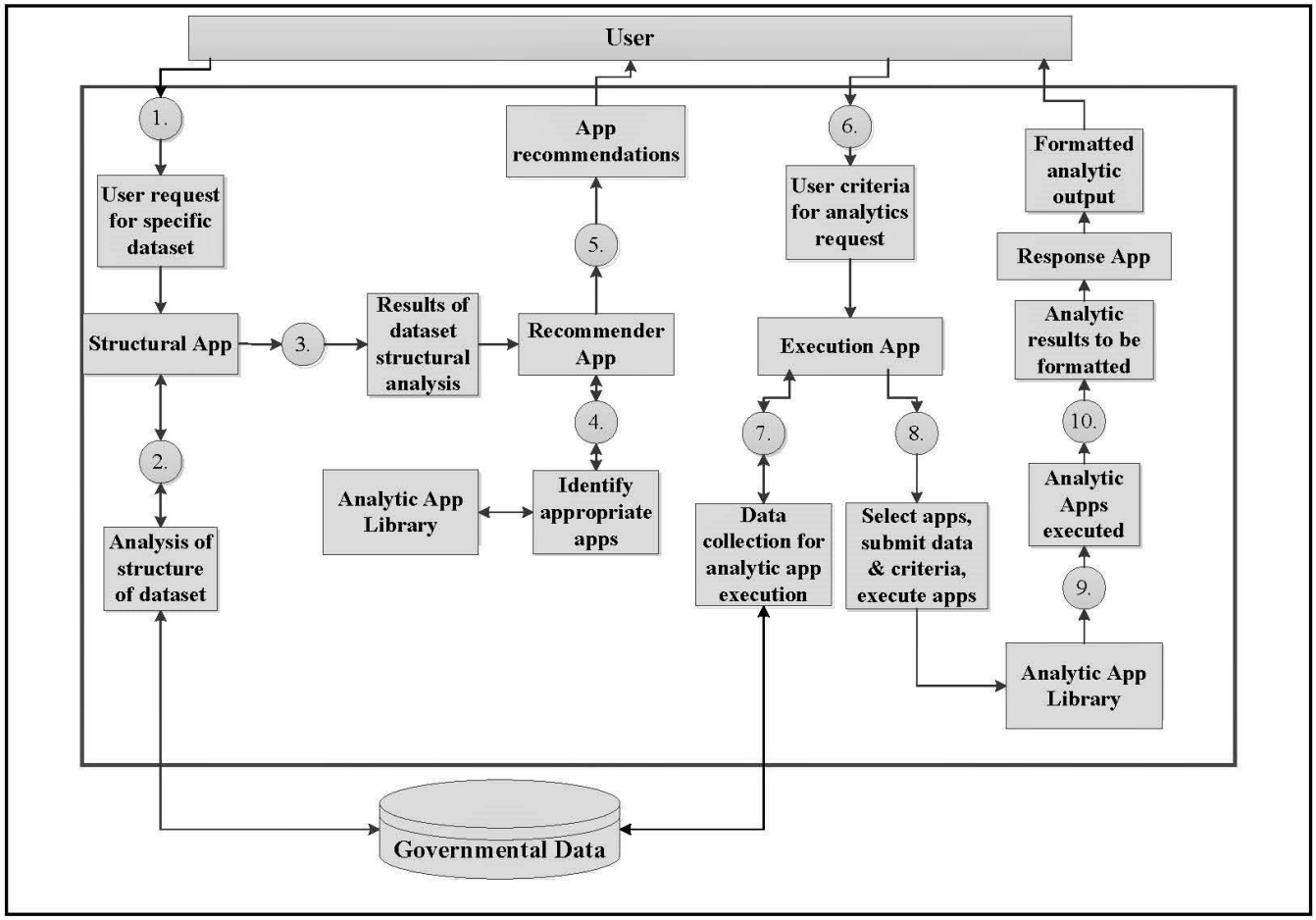
The ENHANCE framework acts upon inputs from external agents to undertake robust analyses. The primary external agents include the user that is requesting the analytics and the governmental dataset of interest. The steps involved in the ENHANCE-user interaction may be outlined as follows:

1. The user accesses ENHANCE to initiate a data analytic request. The user provides ENHANCE with the name of the governmental entity whose data they wish to analyze.
2. The information from Step 1 is passed to the Structural App within the ENHANCE framework. The Structural App includes a search capability that identifies the specific website where the requested data reside, based on a description of the governmental entity, such as city or state name. The Structural App performs an analysis of the structure and attributes of the user-requested dataset.
3. The Structural App passes the attribute information to the Recommender App to identify potential analytics to execute.
4. The Recommender App takes the attribute information and matches this against the collection of Analytic Apps available in the Analytic App Library to identify which Analytic Apps are appropriate to execute, given the data attributes available.
5. The Recommender App provides the user a list of possible tests and analytics that can be launched with the dataset, what typical data selection criteria might include, and also a recommendation as to what an appropriate output format for the results of the analytics might be.
6. The user responds with their selection of Analytic Apps, requested output format, and applicable selection criteria for the analytics; for example, fiscal years and periods to be investigated. This information is provided to the Execution App within the ENHANCE framework.
7. The Execution App accesses the target governmental dataset for the analytics and extracts the appropriate data.
8. The Execution App accesses the App Library and selects the specific Analytic Apps to be executed, passes relevant criteria to each Analytic App, and provides the appropriate data to the Analytic Apps.
9. The Analytic Apps will each complete their analyses over the governmental data and generate the results.
10. The Analytic Apps pass their results to the Response App, which formats the information into the user-selected formats and provides the results to the user.

This process is presented as a flowchart in Figure 1.

[9] See, http://www.gasb.org/jsp/GASB/Page/GASBSectionPage&cid=1176156741809

FIGURE 1
ENHANCE Framework



ENHANCE AS A DECISION SUPPORT SYSTEM: RESPONSE TO RQ1

Decision Support Systems (DSS) are computer-driven solutions to support complex decision making and problem solving (Shim et al. 2002). DSS encompass an area within the information systems (IS) discipline that is engaged in supporting and improving managerial decision making by developing and deploying IT-based systems (Arnott and Pervan 2008). DSS provide decision makers with analytical capabilities and timely information to improve decision making (Power 2009). Specifically, as defined by Power (2009), DSS:

- Provide structured information to decision makers
- Help decision makers analyze specific situations by using various types of models
- Store knowledge and make the knowledge available to decision makers
- Support decision making by individuals, small groups, and large groups

Power (2009) provides the following definition of DSS, as posted on http://dssresources.com/: "An interactive computer-based system or subsystem intended to help decision makers use communication technologies, data, documents, knowledge and/or models to identify and solve problems, complete decision process tasks, and make decisions."

The ENHANCE framework provides the user with robust analytics over the selected governmental entity's expenditure data, enabling the user to make an informed decision as to the propriety of the expenditures under investigation, in other words a DSS. The decision depends on the actions of the user receiving the analytic output; if the analytics indicate a possible issue with the governmental entity, then the user may undertake additional analytics and/or contact governmental representatives to discuss the issue further, for example. The discussion below illustrates how ENHANCE fits the description of a DSS.

Characteristics and abilities of a typical DSS, as provided by Power (2009), and examples of how these can be employed by the user, an armchair auditor, for example, include:

- Facilitate and support decision-making activities: Do the analytic results obtained prompt the user to contact proper governmental authorities to request further explanation?
- Interact with decision makers, or the users who control the interactions; that is, ENHANCE provides the requested analytic results directly to the user.
- Provide ancillary support to decision makers but do not replace decision makers; the user will determine what actions to take, if any, based on the analytic results. Also, the analytic results may provide new insights to governmental authorities by reviewing their data in a new manner.
- Intended for repeated use either on a routine basis or *ad hoc;* the ENHANCE analytics can be run or re-run as required by the user.
- Task-oriented in supporting the following: data analysis, identification and/or design of alternatives, a choice among alternatives, and implementation of the decision; by design, ENHANCE will primarily support the data analysis activity.
- An identifiable system itself, or a specified subsystem of a larger, integrated IS; the ENHANCE framework is intended to be a self-contained entity.
- Impacts the decision by improving the accuracy, timeliness, quality, and effectiveness of decision (or decisions); the analytic results, as generated by ENHANCE, will be as accurate as the underlying governmental data.

Power (2009) also describes the types of DSS based on the dominant architecture component. The ENHANCE framework represents a primarily data-driven type of DSS, with a strong secondary classification as model driven due to its ability to provide numerous analytic techniques for the user to select from. The features of a data-driven DSS (Power 2009) that are incorporated into the ENHANCE framework include:

- *Ad hoc* data filtering and retrieval, where the filtering may offer drop-down menus, predefined queries, and drill-downs from summaries to detailed information; the ENHANCE framework is intended to incorporate all of these capabilities to support nontechnical users in their analytic requests and analyses.
- Creation of data displays, which may include scatter diagrams, bar graphs, and pie charts; the ENHANCE framework is intended to incorporate a variety of presentation techniques to support a variety of analytic requests.
- Data management and summarization that provides users the ability to view and/or create pivot tables, request custom aggregations, and extract/download data; the resulting summarized data that represent the analytic output from ENHANCE can be downloaded and further analyses using other tools can be undertaken.
- View predefined data displays such as dashboards and scorecards; as noted above, if the user requires very tailored presentation formats, the data can be downloaded into other, sophisticated tools.

ILLUSTRATIONS OF ENHANCE FOR AN ENABLED ARMCHAIR AUDITOR: RESPONSE TO RQ2

Data availability is generally not an issue, but the availability of robust analytic techniques that can provide meaningful results over that data has yet to be addressed. Without advanced analytic applications to facilitate and enhance understanding, the value of the available data is limited. ENHANCE intends to place that capability into the hands of the typical user of governmental data. The proposed framework undertakes the role of an armchair auditor enabler.

Academic research presents analytic techniques that are appropriate to incorporate into the apps library supporting the ENHANCE framework. Issa's (2013) "exceptional exceptions" and Byrnes's (2015) data clustering techniques are examples of such analytics. Data analytics included in this present research have also been drawn from research by Provost and Fawcett (2013), as well as Dai and Li (2016), and include:

- Exploratory data analytics/descriptive statistics
- Anomaly/outlier detection
- Time-series analysis
- Stratification
- Cluster analysis

It is noteworthy that the majority of existing analytical tools developed and used by audit firms are proprietary and require a certain level of analytical skills. ENHANCE, on the other hand, is primarily intended for nontechnically sophisticated users and the general public, rather than auditors, who are considered experts and sophisticated users. Moreover, additional research (for example a survey) is necessary to determine the analytics that are most important to various stakeholders to implement them in ENHANCE. For practicality, the paper describes the set of analytics presented below, which should be viewed as a

sample of the analytics applications library that would constitute ENHANCE, to provide some illustrative examples of the framework.

The initial technical conceptualization of ENHANCE begins with the identification of appropriate analytic techniques that can provide the analytic capability. To illustrate the functionality incorporated into ENHANCE, this paper presents some analytics capabilities that are not currently available in any of the governmental data portals. Examples of output from the following analytics are available in the Online Appendix. It is noteworthy that while these analytics were undertaken using commercially available data analytics software, it is possible to replicate them using free and open source software programs.[10]

Anomaly Detection

Anomalous activities or behaviors are defined as those that do not coincide with established normal profiles (Lee and Xiang 2001). In an analysis of governmental expenditure data, duplicate payment records can be considered as such a behavior. The attributes that define duplicate payments are entity specific, with the exact criteria for duplicate payment investigations being user defined.

The identification of duplicate payments is of importance to citizens who expect government entities to manage and spend their tax dollars efficiently. Unfortunately, many citizens would likely lack the technical skills necessary to identify such inappropriate expenditures. It would be possible for these citizens to act as armchair auditors, once enabled with the tools provided by ENHANCE.

Time-Series Analyses

Time-series analyses are a form of data summarization that can provide insights into spending trends and/or patterns over time; for example, with the data presented in a matrix format. Data points are sequenced at equal intervals of time. Time-series analytics may be useful when pattern analysis is not sufficient, for example when the data exhibit characteristics such as "systematic nonrandom patterns" (Alwan and Roberts 1988).

While sophisticated users may execute a time-series analysis to examine trends regularly, normal citizens in general lack the technical knowledge to conduct similar analyses. However, if empowered by ENHANCE, they will be able to evaluate the trends and patterns in government expenditures.

Stratification

Stratification has been defined as "arranging something, or something that has been arranged, into categories" (https://www.vocabulary.com/dictionary/stratification). Stratification chunks or layers the data according to defined ranges. In the context of social research, stratification may also be used to divide a sample into subgroups from which random samples are selected, and then these focused subsamples are combined to form an overall sample. This process may be observed, for example, when dealing with complex survey data (Sturgis 2004). Social researchers also analyze the structure of a stratification system, that is, the composition of the population under investigation with respect to the attributes of interest such as education, income, and occupation (Treiman 1970). As applied to financial data, expenditure data for example, stratification can indicate where anomalous behavior may be occurring in order to circumvent required approvals for purchases above a specified cut-off level.

Clustering

Clustering represents a gathering, or cause to gather into a grouping, a number of similar things (https://www.vocabulary.com/dictionary/cluster). Cluster analysis is undertaken to identify subgroups within the data being analyzed (Fraley and Raftery 1998) and can be considered a subset of pattern recognition. Clustering is also considered as a procedure within a data summarization activity (Nassar, Sander, and Cheng 2004). Unlike discriminant analysis that assigns items to predefined groups, cluster analysis identifies the appropriate groupings based on the data (Fraley and Raftery 1998). Cluster analysis allows the data to tell the story in an exploratory fashion.

From a research perspective, clustering represents an unsupervised classification of patterns or observations that is regarded as a type of exploratory data analysis (EDA) (Jain, Murty, and Flynn 1999). Patterns can be represented, for example, as a vector of measurement or a point in multidimensional space (Jain et al. 1999). Clustering techniques have gained in prominence in recent years as data mining activities have increased. Specific examples include the identification of customer

[10] For example, CaseWare IDEA, Version 9.2.0.630 (x86).

and product groupings in retail databases, analyses of web usage data, and image analyses for segmentation and quantization (Fraley and Raftery 2002).

A cluster analysis uses a variety of algorithms as there is no one definition of what a cluster is because it can take many different shapes, and it also depends on the type of underlying data, such as continuous or discrete, and whether the clustering involves identifying similarities or dissimilarities. In the example using governmental expenditure data, clustering could indicate the relative magnitude of purchases by a vendor, which could be indicative of the use of favored vendors by the purchasing function.

Most open data portals do not provide any similar analytics on their websites, which leaves it up to the user to extract the data and run cluster analyses. Clustering is a powerful methodology widely used by marketing, management, computer science, and auditing. However, it requires a significant level of technical skills not commonly found in nonsophisticated users. Such users can leverage the proposed ENHANCE framework to bridge the gap and transform into skillful armchair auditors.

AN EXAMPLE OF ENHANCE IN PRACTICE: RESPONSE TO RQ3

The following is an example of employing the ENHANCE framework to provide a user with a tailored information dashboard from which they can select a specific action or decision. This illustration answers RQ3 by showing how the use of tools can increase the quality of armchair auditor contributions, following the steps highlighted in Figure 1. RQ3 seeks to understand the potential increase in the quality of the armchair auditor contributions. To illustrate, Arthur Armchair is a concerned citizen who would like to ensure that the officials in the City of Austin, TX are not involved in unethical activities, such as favoritism. For example, Arthur is interested in examining the possibility for a new purveyor of computer products and services to enter the market, in addition to ensuring that the city treats various suppliers fairly and equally. To accomplish these tasks, Arthur needs to examine what other firms are providing similar services to the City of Austin. Armed with this information, Arthur can identify if there are opportunities for a new firm to provide additional services to the City of Austin. Arthur is aware of the open data portal that the City of Austin provides, and is aware of the availability of the ENHANCE framework and the capabilities it encompasses.[11] Below is a step-by-step illustration of Arthur's experience with ENHANCE.

1. Arthur accesses ENHANCE and requests access to the Austin checkbook dataset.
2. This request is sent to the Structural App, which in turn analyzes the requested data and their structure.
3. The Structural App forwards the results of the analysis to the Recommender App.
4. This in turn identifies the possible and appropriate Analytic Apps that can be used with the requested data.
5. Next, the Recommender App provides the list of recommended Analytic Apps to Arthur.
6. Arthur wishes to understand city spending for the prior fiscal year, the latest available full year, the department where computer-related purchases primarily occur, the charge-to accounts within the department where goods and services most similar to what his firm provides are recorded, and the dollar amount of the charges. Arthur requests listings of the valid values for the attributes he selects and ENHANCE provides that information to him. Upon reviewing the valid value listings, Arthur selects the desired criteria (e.g., fiscal year, department name).
7. This information is sent to the Execution App, which requests and collected the data.
8. Arthur's selection of Analytic Apps is passed to the Execution App.
9. The Execution App next runs the analysis based on Arthur's selection of Analytic Apps and criteria.
10. The results are next forwarded to the Response App in order to format it and finally deliver it to Arthur in Excel, as he had requested.

Arthur wishes to obtain results that are summarized by vendor, instead of detailed line item data as are provided in the dataset, so he selects the Data Summarization analytic, by vendor, which ENHANCE provides in the list of potential analytics. Arthur selects to receive the output from the analytic in Excel format, as he may wish to further massage the output. In reviewing the output, Arthur determines that he is only interested in vendors that supplied over $100,000 of goods and services during 2015, so he first sorts the data in descending amount order and then truncates the output. The results are presented in Table 1.

In another example, Arthur is conducting cross-city comparative analysis. He would like to compare vendors' statistics in Austin, TX to those in Las Vegas, NV. Arthur first needs to identify the vendors in Las Vegas that are providing goods and services to the city. He mimics the actions he conducted using the ENHANCE framework with Austin data, but now based on Las Vegas checkbook data. After submitting his request for Las Vegas data to ENHANCE, he receives a response listing the attributes that the Las Vegas dataset provides. Arthur is interested in the more significant IT vendors to the city and first sorts

[11] This example, along with the analytics included in this section, are for illustration purposes only. The objective of such a specific example is to provide the user with a potential practical application of the ENHANCE framework in a real-life scenario.

TABLE 1

Results from ENHANCE Summarization Analytic for Austin, TX, as Adjusted by User to Reflect > $100,000 Vendors

Vendor	Amount
Freeit Data Solutions Inc. Total	1,488,552.11
Insight Public Sector Inc. Total	1,316,291.56
Future Com Ltd. Total	679,606.60
M&S Technologies Inc. Total	334,699.80
Mark III Systems Inc. Total	330,895.23
Solutions-II, Inc. Total	281,354.70
Imtech Corporation Total	224,754.74
Dell Marketing LP Total	211,131.25
Sirius Computer Solutions Inc. Total	138,240.00
Total Amount	5,005,525.99

the output by descending amount and then truncates the data to display only those vendors providing over $100,000 of goods and services, as shown in Table 2.

In reviewing these results, Arthur notes the top two vendors are not of interest to him, as Oracle Corporation most likely is providing ERP or similar software to the city, and Verizon Wireless is most likely providing communication services to the city, and he removes these from the listing, resulting in Table 3.

Arthur can now research these vendors by reviewing vendor websites and/or social media sites to determine what goods and services each provides. Arthur can also contrast the data for each city in a qualitative fashion by comparing the total amount each city expends to its top vendors, which is similar in amount based on these analyses: $5,005,000 for Austin versus $4,997,000 for Las Vegas. Arthur has identified the populations for both cities[12] and has found that Austin's population, at approximately 790,000, is about 35 percent greater than Las Vegas's, at approximately 584,000. Given the population differences, it appears from Arthur's initial IT expenditure analyses that Las Vegas may spend a disproportionately greater amount on IT goods and services than Austin does, or at least relatively larger amounts to its top vendors based on population. It might be necessary for Arthur to undertake additional analyses, but in a fairly brief period of time, and with relatively minimal effort on his part, Arthur has obtained preliminary information to support his business venture.

[12] See, https://suburbanstats.org for population and demographics information. The population information for both Austin, TX and Las Vegas, NV was accessed on June 5, 2018.

TABLE 2

Results from ENHANCE Summarization Analytic for Las Vegas, NV, as Adjusted by User to Reflect > $100,000 Vendors

Vendor	Amount
Oracle America Inc. Total	1,258,674.28
Verizon Wireless Total	759,286.00
Dell Marketing LP Total	633,491.57
Dell Total	495,882.02
DynTek Services Inc. Total	406,217.64
Infor Public Sector Inc. Total	365,783.86
Zunesis Inc. Total	340,156.85
Gartner Inc. Total	168,200.00
Clark County Nevada Total	131,596.64
Black Box Network Services Total	115,914.76
ISIS Solution Corporation Total	109,319.00
CDW Government Inc. Total	107,048.68
Environmental Systems Research Institute Inc. Total	105,934.93
Total Amount	4,997,506.23

TABLE 3

User-Adjusted Results from ENHANCE Summarization Analytic for Las Vegas, NV

Vendor	Amount
Dell Marketing LP Total	633,491.57
Dell Total	495,882.02
DynTek Services Inc. Total	406,217.64
Infor Public Sector Inc. Total	365,783.86
Zunesis Inc. Total	340,156.85
Gartner Inc. Total	168,200.00
Clark County Nevada Total	131,596.64
Black Box Network Services Total	115,914.76
ISIS Solution Corporation Total	109,319.00
CDW Government Inc. Total	107,048.68
Environmental Systems Research Institute Inc. Total	105,934.93
Total Amount	2,979,545.95

CONCLUSION

Many businesses these days are collecting and analyzing Big Data, and governmental entities are no exception. For instance, municipalities at many levels are now collecting and publishing data of all sorts, from purchase orders, salaries, and 911 calls to health inspections results, grant expenditures, and daily transactions. Additionally, the government sector is facing the same challenge that concerns these businesses: how can stakeholders and interested parties obtain useful information and knowledge from these massive data sets? Often the data are inconsistent, unorderly, and massive, forcing the consumer of such data to seek assistance elsewhere to obtain understanding. Often, this data consumer has certain queries and objectives when looking at the data, but has little knowledge of how to acquire the results.

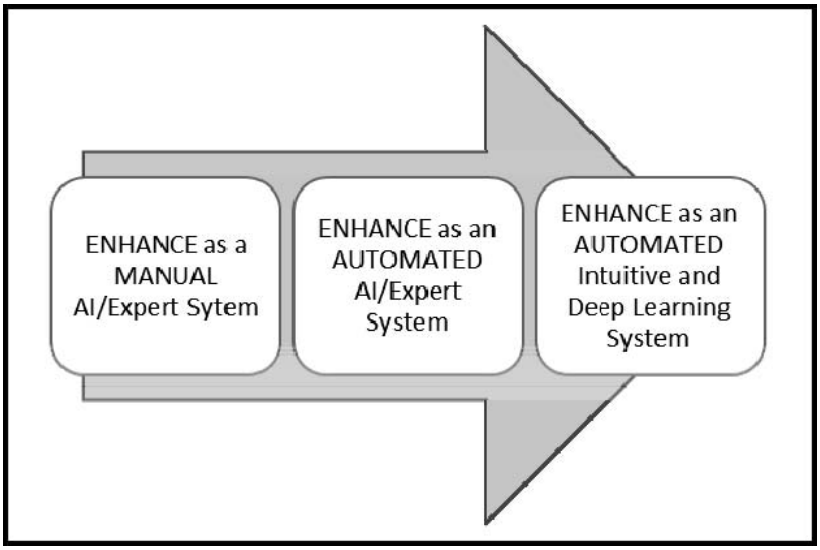
Contribution

Dai and Li (2016) propose a series of audit apps to support the requirements of armchair auditors in analyzing governmental expenditure data to identify potential fraud, waste, and abuse of taxpayer dollars. This research extends the research by Dai and Li (2016) by proposing an advanced analytic framework incorporating a series of analytic (audit) apps that, facilitated by open governmental data, can fulfill the analytic requirements of the various governmental stakeholders such as citizens, analysts, bond investors, creditors, vendors, auditors, and oversight officials. The contribution of this research to academic literature includes presenting both the capabilities of the proposed ENHANCE tool and the scenario in which the "apps" that underlie the proposed framework can provide meaningful information to a governmental entity's stakeholders. The capabilities provided by the ENHANCE framework are not presently available to governmental stakeholders. The constituents would need to possess significant database and information technology expertise to replicate the advanced analytics provided by ENHANCE. ENHANCE provides that level of technical sophistication to constituents that would be required to generate meaningful expenditure information from their governmental entity. With ENHANCE, the gap between open data and actionable intelligence would appear to be narrower.

Additionally, the ENHANCE analytic tool should be regarded as an introductory artificial intelligence (AI) application. That is, once the user provides the dataset(s) and the general questions to be answered, ENHANCE recommends the best options and analytical tests and provides the results. AI is basically intelligence demonstrated by machines or statistical/machine learning applications,[13] where machines mimic the expected cognitive behaviors of humans. ENHANCE is not dissimilar to an expert system, a version of AI popular in the 1980s, which mimics human decision-making expertise based on predefined, if-then rules and logic (Issa, Sun, and Vasarhelyi 2016). Behind every decision within ENHANCE lie many lines of code, crafted and based on the expertise of accounting analytics researchers. ENHANCE exhibits the cognitive expert qualities of reasoning, knowledge discovery, problem solving, and decision making with its activities. This is accomplished behind the scenes, as ENHANCE does not require the user to be knowledgeable of the coding that is involved in the system's analytical methods. The framework would eventually act as a platform enabling the addition of new emerging technologies, in the same

[13] See, https://en.wikipedia.org/wiki/Artificial_intelligence

FIGURE 2
The Proposed Implementation Phases of ENHANCE



manner that smartphone apps are added to app stores on a daily basis. Smartphone users do not necessarily possess the technical skills to develop (or even understand the coding that goes behind the development of) such apps.

Furthermore, regarding ENHANCE as an AI application expands its future possibilities. As more and varied data are made available (Big Data) on open governmental portals, a more automated version of ENHANCE could exist whereby tests and results are generated when new data become available. In this case, all of the tests discussed earlier in the examples would be generated automatically when any new information is available on the web portal, demonstrating process automation of routine tasks. The final iteration of ENHANCE would be enhanced with qualitative data such as sensor readings, picture/video files, social media, and audio files, so that ENHANCE could exhibit more intuitive deep learning, reasoning, and decision making. The possible proposed implementation phases of ENHANCE are illustrated in Figure 2.

It is hoped that ENHANCE will stimulate debate and analysis in this rapidly evolving domain of open governmental data. ENHANCE presents major potential implications for both the quality and process of open governmental data efforts. Practitioners and researchers should collaborate going forward to allow ENHANCE to enhance the constituent experience and interaction with open governmental data portals.

Future Research and Limitations

Future research should examine ENHANCE in a case study or sandbox environment prior to wholescale integration by municipalities. For instance, ENHANCE proposes here certain anticipated citizen queries, however, further research should be conducted to understand what their primary concerns and tasks are when accessing open governmental data. This is crucial to determine the analytic techniques that must be included in the App Library. Conducting surveys and interviews, as well as examining citizens' comments and requests for information (possibly using text mining) can help achieve that. In addition, future research can investigate the likelihood of governmental entities to implement an ENHANCE framework, as well as the challenges associated with such an initiative.

Another direction for future research could address the role of transparency, or the lack there of, in open governmental data portals. Although these municipalities appear to be promoting data transparency, is this really the case when many of the data are not understandable to the average citizen? After all, the data are not useful if the consumers of the data cannot extract meaningful information. There may be numerous reasons why the goal of data transparency is not fully realized, and these obstacles should be illuminated.

Moreover, future research should examine the viability and real value of the concept of the armchair auditor. To truly realize the advantage of available data and provide real transparency, it is not enough for the data to be accessible and the analytics to be available and user friendly: the results of such frameworks should be actionable. In other words, to reach true transparency, these frameworks must enable the user to turn the results into actions. For example, a constituent would be able to

take her results and question her local government. A reporter would publish findings in the media, which would increase the likelihood of obtaining a response from local government officials.

REFERENCES

Alwan, L. C., and H. V. Roberts. 1988. Time-series modeling for statistical process control. *Journal of Business & Economic Statistics* 6 (1): 87–95. https://doi.org/10.2307/1391421

Arnott, D., and G. Pervan. 2008. Eight key issues for the decision support systems discipline. *Decision Support Systems* 44 (3): 657–672. https://doi.org/10.1016/j.dss.2007.09.003

Byrnes, P. E. 2015. *Developing Automated Applications for Clustering and Outlier Detection: Data Mining Implications for Auditing Practice*. Doctoral dissertation, Rutgers, The State University of New Jersey, Newark.

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

Dai, J., and M. A. Vasarhelyi. 2017. Towards blockchain-based accounting and assurance. *Journal of Information Systems* 31 (3): 5–21. https://doi.org/10.2308/isys-51804

Fraley, C., and A. E. Raftery. 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal* 41 (8): 578–588. https://doi.org/10.1093/comjnl/41.8.578

Fraley, C., and A. E. Raftery. 2002. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association* 97 (458): 611–631. https://doi.org/10.1198/016214502760047131

Frank, M., and A. A. Oztoprak. 2015. *Concepts of Transparency: Open Data in U.K. Local Authorities*. Proceedings of the CeDEM15, International Conference for E-Democracy and Open Government 2015, Danube University Krems, Krems an der Donau, Austria, May 20–22.

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

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

Jain, A. K., M. N. Murty, and P. J. Flynn. 1999. Data clustering: A review. *ACM Computing Surveys (CSUR)* 31(3): 264–323.

Lee, W., and D. Xiang. 2001. *Information-Theoretic Measures for Anomaly Detection*. Proceedings of the 2001 IEEE Symposium on Security and Privacy, Oakland, CA, May 14–May 16.

Mueller, D. 2009. *Public Sector Case Study: State of Oregon CAFR Project*. Proceedings of the XBRL-INT Conference, Paris, France, June 25.

Nassar, S., J. Sander, and C. Cheng. 2004. *Incremental and Effective Data Summarization for Dynamic Hierarchical Clustering*. Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data, Paris, France, June 13–18.

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

Pew Research Center. 2015. *Americans' Views on Data to Open Government*. Available at: http://www.pewinternet.org/2015/04/21/open-government-data/

Power, D. J. 2009. *Decision Support Basics*. New York, NY: Business Expert Press.

Provost, F., and T. Fawcett. 2013. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. Sebastopol, CA: O'Reilly Media, Inc.

Shim, J. P., M. Warkentin, J. F. Courtney, D. J. Power, R. Sharda, and C. Carlsson. 2002. Past, present, and future of decision support technology. *Decision Support Systems* 33 (2): 111–126. https://doi.org/10.1016/S0167-9236(01)00139-7

Sturgis, P. 2004. Analysing complex survey data: Clustering, stratification and weights. *Social Research Update* 43 (Autumn): 1–4.

Treiman, D. J. 1970. Industrialization and social stratification. *Sociological Inquiry* 40 (2): 207–234. https://doi.org/10.1111/j.1475-682X.1970.tb01009.x

Worthy, B. 2015. The impact of open data in the U.K.: Complex, unpredictable, and political. *Public Administration* 93 (3): 788–805. https://doi.org/10.1111/padm.12166

APPENDIX A

jeta-52096_Online Appendix: http://dx.doi.org/10.2308/jeta-52096.s01
jeta-52096_Cameron Speech 2009: http://dx.doi.org/10.2308/jeta-52096.s02