A Vision of an ENHanced ANalytic Constituent Environment: ENHANCE

By

Stephen Kozlowski

A dissertation submitted to the
Graduate School - Newark
Rutgers, the State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
Graduate Program in Management
Written under the direction of
Professor Miklos Vasarhelyi
and approved by

_____
Dr. Miklos Vasarhelyi

_____
Dr. Won Gyun No

_____
Dr. Hussein Issa

_____
Dr. Daniel O'Leary

Newark, New Jersey

May, 2016

ABSTRACT OF THE DISSERTATION

A Vision of an ENHanced ANalytic Constituent Environment: ENHANCE

By Stephen Kozlowski

Dissertation Chairman: Prof. Miklos Vasarhelyi

Constituent demands for improved transparency in governmental reporting have been increasing since the recent (2008-2009) financial crisis in the U.S. that impacted the financial well-being of a number of governmental entities at both state and local levels. Since that time several governmental entities in the U.S. have launched open data portals. But to-date these initiatives do not incorporate robust analytic capabilities to satisfy constituent inquiries.

The purpose of this research is to present an ENHanced ANalytic Constituent Environment (ENHANCE) framework, facilitated by open government data, that fulfills the analytic requirements of the various governmental stakeholders, such as citizens, analysts, bond investors, creditors, vendors, auditors, and oversight officials. The first step in introducing this analytic capability is the availability of open governmental data, ideally presented in a standardized and easily usable format. The second step is that of providing a series of analytic 'apps' that can provide meaningful information to the entity's stakeholders. The third step is the development of the ENHANCE framework within which apps can function, acting upon the standardized government data, to support constituent analytic requirements. A fourth step encompasses developing an ecosystem to provide additional feature and function to the user interaction with ENHANCE and support ENHANCE in the evolving environment in which it operates.

This research contributes to academic literature by proposing an ENHanced ANalytic Constituent Environment (ENHANCE) framework where governmental stakeholders can create reports on demand to satisfy their analytic requirements. The concept of a robust analytic tool that provides meaningful analyses over open data is presented in this paper within the context of open government expenditure data. The ENHANCE concept provides capabilities to support an 'armchair auditor' activity (O'Leary, 2015).

This research describes the concept of an audit ecosystem, that is, a natural progression in the deployment of computer-based CA/CM tools. An initial definition for an audit ecosystem is provided, stated at this point as a holistic approach to the design and development of a technology-driven framework to provide overall management and control of the audit technology components employed, and coordination of the activities of the participants involved. It is feasible that an ecosystem tailored to embellish and support ENHANCE capabilities can be developed and deployed. This paper provides a background on Design Science Research Methodology and describes the initial work undertaken using DSRM to drive the design and development of the ENHANCE framework, the app recommender system which comprises a significant technical part of the framework, and an audit ecosystem.

In addition, this research provides a literature review of current practices with respect to open data initiatives and transparency in government financial reporting.

Acknowledgements

I wish to gratefully acknowledge the contributions and guidance provided by the members of my dissertation committee: Dr. Miklos Vasarhelyi, Dr. Won Gyun No, Dr. Hussein Issa, and Dr. Daniel O'Leary. I wish to especially thank Dr. Vasarhelyi for his mentorship during my four years at Rutgers. I also wish to thank Dr. Issa for his support and comments during the preparation of my dissertation.

I am most appreciative of and grateful for the love, support, and encouragement provided by my family during the past four years as I worked to develop the skills to allow me to undertake a new and fulfilling career.

Table of Contents

List of Tables

List of Illustrations

Chapter 1: Introduction

Governmental accounting differs from that of private sector accounting in that governmental accounting measures operating results from the perspective of fiscal responsibility, or proper stewardship of financial resources. Moreover, governmental accounting also emphasizes reporting financial results with respect to their compliance to budgetary controls. While new technologies that vastly improve financial reporting and analytic capabilities have been embraced by the private sector over the past several years these modernization efforts have not been fully adopted by the governmental sector. Constituent demands for improved transparency in governmental reporting have increased since the recent (2008-2009) financial crisis in the U.S. which impacted the financial well-being of a number of U.S.-based governmental entities at both state and local levels, such as Detroit, MI and Harrisburg, PA, for example. The public's demand for transparency is not merely limited to access to the financial data itself, but also for advanced data analytic tools that can provide meaningful insights into an entity's spending patterns. Several governmental entities in the U.S., for example the states of Ohio and Utah, and the cities of Chicago, Detroit, Las Vegas, Los Angeles, New York City, and Seattle, provide an open data environment for their constituents' use.

The development of these open data initiatives is in response to the public's demand for an ability to monitor governmental performance. These early initiatives provide the data that can support advanced analytics, but do not incorporate robust, if any, analytic capabilities. The average citizen will not likely have the tools or expertise required to produce sophisticated analytics on their own. As constituents increasingly rely on open data initiatives to allow monitoring of governmental activities how will they react to a

lack of proper analytics? Will citizens and other stakeholders find usefulness from access to data alone without analytic capabilities? Does the average citizen embody the capabilities to undertake sophisticated analytic analyses on their own? Do governmental entities have the skillsets, and particularly resources, to develop analytic tools, and in a timely fashion to satisfy constituent demands? What would be the incentive for third-party entities to develop such analytics? Would a lack of analytic tools be viewed as an overall failure of the open data initiative and lead to even greater levels of consumer distrust of government? Can a tool providing advanced analytic capabilities be developed to satisfy constituent demands?

Research has been conducted to define a number of the analytic apps which the ENHANCE framework can utilize to produce meaningful analyses for its users. Issa (2013) presents an advanced analytic app termed 'Exceptional Exceptions', and Byrnes (2015) presents advanced data clustering techniques, both of which may provide benefit to the ENHANCE user as part of the library of analytic apps that ENHANCE has access to use. Dai et al. (2014) propose an app recommender system that selects appropriate audit apps for audit engagements based on criteria specific to the auditor, client, and audit standards. As much as the analytic apps and especially an app recommender system are critical to support advanced analytic capabilities, there has not been research at this point to incorporate these components into a framework capable of producing sophisticated analytic results.

The primary purpose of this research is to develop an ENHanced ANalytic Constituent Environment (ENHANCE) framework, facilitated by open government data. The ENHANCE framework bridges the gap between readily accessible governmental

expenditure data and the lack of a tool to generate meaningful analyses from that data to support transparency initiatives. Such a framework, if implemented by government agencies, will help fulfill the reporting requirements of the various governmental stakeholders, such as citizens, analysts, bond investors, creditors, vendors, legislative/oversight officials, and auditors. The purpose of the ENHANCE framework is to provide advanced capabilities that are otherwise inaccessible to non-technical users. It is envisioned that the development of the ENHANCE framework will expand on the research noted above and provide a tool that can leverage the ability of the apps that are already defined.

The first step towards developing an advanced analytic tool is the availability of open governmental data, ideally presented in a standardized and easily usable format.

The second step is that of designing a series of analytic 'apps' that can provide meaningful information to the entity's stakeholders. For example, appropriate data analytics to help bond investors make sound decisions might include apps for: cluster analysis to group similar entities, fraud detection, pattern recognition, and time series analysis for identifying trends over time. Other apps include: benchmarking apps to help define what the normal behavior is expected to be, as well as key performance indicators (KPIs), 'Exceptional Exceptions' apps (Issa, 2013) that can identify and prioritize exceptions and anomalies, and cross-entity analyses to compare the efficiency of one governmental entity to another.

The third step is the development of the ENHANCE framework where apps can function, acting upon the standardized government data to support constituent reporting

requirements. Citizens may desire to understand how their tax money is being spent. Bond investors and creditors may want to ascertain the future financial health of an entity. Auditors may want to identify potentially fraudulent activity. What may provide the greatest value from an analytic framework such as ENHANCE will be realized when datasets across governmental entities are configured using common taxonomies, thus allowing for cross-entity analyses of comparable data, for example, expenditures for law enforcement activities across entities of similar size, jurisdiction, or other criteria.

In summary, this research presents a conceptual design for a technically sophisticated analytic tool to support user requests for a variety of analyses over governmental expenditure data. This analytic tool bridges the gap that exists today between governmental entities providing open data portals for access to raw data, and the lack of robust analytic tools that can provide meaningful information from the data, that is, tools that do not require significant technical expertise to launch. This research illustrates by example a number of analytic techniques that are to be included in a fully developed and deployed ENHANCE framework. Examples of the ENHANCE user interface and elements of each underlying screen are depicted. The design of the database to support the functioning of the ENHANCE user screens is provided. An example of what is anticipated to be a typical user interaction with ENHANCE is presented.

In addition this research describes the concept of an audit ecosystem, that is, a form of a digital ecosystem that supports Continuous Auditing/Continuous Monitoring (CA/CM) activities, including relevant features and attributes. The capabilities an ecosystem provides can support an analytic tool such as ENHANCE operating in an ever-changing environment with new sources of data and the availability of improved analytic tools.

This paper describes how Design Science Research Methodology supports the development of the ENHANCE framework, the app recommender system, and an audit ecosystem.

This research contributes to academic literature by expanding the literature on analytic techniques that support the private sector, such as Continuous Auditing and Monitoring (CA/CM), with research activities focused on the public sector. This research presents an ENHanced ANalytic Constituent Environment (ENHANCE) framework where governmental stakeholders can create reports on demand, based on governmental expenditure data, to satisfy their analytic requirements. This research describes the concept of an audit ecosystem and defines the components and attributes that comprise an audit ecosystem.

This research also presents a literature review of open data initiatives and transparency in government financial reporting.

Chapter 2: Literature Review

Chapter 2 presents a literature review of the topics of data transparency and open data initiatives in Section 2.1, and the topic of data analytics in Section 2.2.

2.1. Data Transparency and Open Data Initiatives

Data transparency is touted as a goal of the open data initiative (Peled, 2011). A definition of data transparency, in the context of a governmental setting, is that of an increased flow of timely and reliable economic, social, and political information (Vishwanath & Kaufmann, 1999). Additionally, the availability of information related to a governmental entity's service provision to its constituents, and its monetary and fiscal policy, is considered another aspect of data transparency (Vishwanath & Kaufmann, 1999).

Data transparency attributes include access, comprehensiveness, relevance, quality, and reliability (Vishwanath & Kaufman, 1999). The term access suggests not only the laws and regulations to require free access to all, equally, but also the dissemination of knowledge of the availability of the data (Vishwanath & Kaufman, 1999). Data relevant to the public includes information on current economic conditions and governmental policies, and there should not be an overabundance of data that confuses constituents (Vishwanath & Kaufman, 1999). Quality and reliability measures include timeliness, completeness, fairness, consistency, and a clear representation of the data (Vishwanath & Kaufman, 2009).

The need for data transparency at all levels of government has had several proponents after the recession of 2008-2009 that brought about financial crises in several

governmental entities. The State Budget Crisis Task Force was formed to specifically investigate financial reporting at the state level. In 2011 the Government Accounting Standards Board (GASB) (http://www.gasb.org) issued a research brief concerning the timeliness of state and local government financial reporting[1] in which they examined the timeliness of financial reporting by state and local governments during the period 2006-2008. The GASB research also includes a survey of financial information users that describe how timeliness impacts the usefulness of financial reporting. Although the survey respondents consider information made available within six months of the close of the financial year to be useful they consider it significantly less useful than information received within 45 days to three months after fiscal year-end (Attmore, 2011).

In addition to activities focused on data transparency initiatives as undertaken by any number of governmental entities there has also been research on the capabilities that open data can provide. Brito (2008) identified tasks that can be accomplished with data provided in an open and transparent environment. Several recent activities that have been facilitated by the internet can be applied to and enhance the value of open government data as entities strive toward meeting the objectives of 'Government 2.0'[2]. Of particular importance are 'mashups'[3] and 'crowdsourcing'[4] as they illustrate some of the activities that can be undertaken given the availability of open government data to provide analytic

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

[2] Government 2.0 refers to initiatives to create an open-source computing environment providing improved transparency and efficiency between governmental entities and their constituents.

[3] Brito (2008) defines mashups as the highlighting of hidden connections between different data sets that are identified when databases are combined. Such connections may also reveal patterns in the combined data that otherwise would not be detected.

[4] Brito (2008) defines crowdsourcing as a means a spreading the burden of sifting through significant amounts of data across a large number of individual participants each making a small contribution

capabilities. Brito (2008) recommends that governments provide their data online in an open, structured, and searchable fashion, but if they do not there should and probably will be external parties that will take whatever data is available and build unofficial databases for public access. The basis for generating useful analyses is that governments provide their data in a machine-readable format on accessible websites, and again if governments do not provide sufficient analytical tools then external parties will bridge this gap (Brito, 2008). Brito (2008) considers such external party activities a form of 'hacks'[5]. An example is that of The Washington Post's U.S. Congress Votes database (http://projects.washingtonpost.com/congress/) that includes disclosure forms for both current House and Senate members. The information for House members is prepared annually by the Office of the Clerk and made available on CD-ROM (Brito, 2008). For the Senate the information is available only in hardcopy form which must be scanned to create electronic copies (Brito, 2008).

Governmental entities need to be the source of their data for the following reasons (Brito, 2008): if the data comes directly from the governmental entity the integrity and quality of the data is ensured, the cost to governments to provide the data is undoubtedly less than that incurred by external parties to obtain the data via sophisticated 'hacks', and it is not possible to 'hack' data that has not been made accessible in any format.

Measuring data transparency is not necessarily clear-cut. While accounting information may be compared to the results of audit activities that have been undertaken this is not the case with other forms of information. In all cases the data must be of a level of quality to allow for transparency measures to be accurate. As transparency measures evolve to

[5] In this context Brito uses as the definition of hacks: 'a modification of a program or device to give the user access to features that were otherwise unavailable to them'

require greater sophistication, it may be appropriate to develop a transparency index using proxies for the transparency attributes noted above (Vishwanath & Kaufman, 1999).

Transparency can be impacted by the requirements, depending on the particular situation, for disclosure regulation. When providing data entails a high cost, disclosure regulation may be mandated. An estimation of the value of increased transparency to improving economic outcomes should be undertaken (Vishwanath & Kaufman, 1999). Where disclosure regulation is required there should be a measurement of the extent of disclosure needed that considers both the cost to produce the data and the benefits to disclose the data (Vishwanath & Kaufman, 1999). Disclosure regulation should be tailored to fit the governmental and social structure in place, for example, developing countries versus developed countries, and democratic, open forms of government versus more restrictive forms of government.

To support data transparency, specifically for accounting-related data, the data must adhere to normal accounting standards and principles in order to provide the expected reliability, comparability, and explanatory qualities of the data (Vishwanath & Kaufman, 1999).

While data transparency enables an informed and engaged constituent base, appropriate tools to provide the data are also needed for public engagement and involvement (Jaeger & Bertot, 2010). As Information and Communication Technologies (ICT) are used to develop open data environments, these changes will impact social, cultural, and technology factors (Bertot et al., 2010). There are several short-term activities that can

lead to long-term success (Bertot et al., 2010). The first requires developing assessments to measure the success of the transparency efforts. The second requires developing transparency readiness criteria specific for the country under study. The third requires an evaluation and modification of current IT systems to support open data. Although many IT system upgrades require the replacement of all components, in some instances updating existing systems may be more expedient and cost-effective than replacing them. Developing countries should investigate systems in use in countries with more advanced open data initiatives, and when feasible replicate them. The fourth activity, undertaking collaborative pilot projects, can leverage the abilities and resources of several entities to develop open data initiatives that can support all participants. There are international organizations such as the Open Government Partnership (www.opengovpartnership.org/) that set transparency goals and objectives and provide support to countries to achieve them.

In addition to providing the technological foundation to support transparency initiatives, ICT implementations can also reduce the potential for corruption by replacing and reducing human intervention in governmental procedures. In their review, Shim and Eom (2009) identified three types of anti-corruption factors: administrative reform, law enforcement, and social capital[6], with administrative reform being the most common, followed by law enforcement. In addition to minimizing self-serving behavior by government employees, ICTs can be configured to monitor employee behavior, provide information in a transparent manner to the public, and build social capital by providing

[6] For this research Shim and EOM employ the definition of social capital as provided by Putnam (1993): Social capital is includes the feature of social organizations that facilitate coordination and cooperation among the participants for mutual benefits. Putnam focuses on social norms, trust, and voluntary activities to develop communities and societies through the use of social capital

the site where citizens can interact with their governments. The information provided in the open data portals may include, in addition to expenditure-related data, policy-making decisions, and service delivery procedures.

Open data initiatives have been undertaken in the U.S. For example, a partnership between the Oregon State Controller's Division, the Association of Government Accountants (AGA), the Government Accounting Standards Board (GASB), and PriceWaterhouseCooper (PWC) formed to demonstrate the feasibility of interactive data for public sector reporting (Mueller, 2009).

In addition to governmental entities themselves, there are entities that report on government-related activities that provide a source of open government data. Trading in the secondary market for municipal securities has been shown to lack the data timeliness and frequency found in the corporate securities market, specifically price and information transparency (Reck & Wilson, 2006). Efforts have been made toward improving disclosure in the municipal market such as when the SEC in 1994 required (SEC Regulation 15c2-12) that brokers, dealers, and underwriters provide annual disclosure to the secondary market (Baber & Gore, 2008). Unfortunately there is no enforcement mechanism to require disclosure (Baber & Gore, 2008).

Governmental data transparency initiatives are not unique to the U.S. From an international perspective, data transparency and the right to access government information are regarded as an integral democratic right in many societies (Bertot et al., 2010). The Netherlands undertook improvements to standardized governmental reporting requirements as a step to enhance transparency and ease the reporting burdens on Dutch-

based companies. The Dutch National Taxonomy Project (NTP) was launched in 2004 with the objective of reducing the administrative burden for companies in meeting governmental reporting requirements (Bharosa, et al., 2011). Achieving this objective required, among other items, the standardization of data into a common language, that is, data taxonomy. One of the first steps in the project was that of a feasibility study on appropriate reporting structures that in 2006 presented the initial, required functionality for financial reporting based on the Extensible Business Reporting Language (XBRL) (Bharosa, et al., 2011). The NTP took a novel approach by reversing the perspective on the reporting chain from that of a government-centric systems view to that of identifying the businesses' financial administration as the starting point in the development (Bharosa, et al., 2011). This approach, considering business administration activities as the starting point, also required clear data definitions, and streamlined reporting processes by mandating the use of international open standards. In 2008 the NTP transitioned into the Standard Business Reporting (SBR) project, a joint public/private sector project with the primary goal still focused on reducing administrative burdens on businesses (Geijtenbeek and Lucassen, 2012). SBR replaced all paper-based filings for Dutch businesses. The SBR taxonomy was initially designed to accommodate financial statements, tax filings, economic statistics, and credit reports (Geijtenbeek and Lucassen, 2012). Subsequently, Healthcare, Education, Subsidies, Assurance, Tax compliance, and local government domains were added (Geijtenbeek and Lucassen, 2012). A voluntary filing program started in 2009 and mandatory filings started in 2013 with a phased implementation continuing through 2016.

A similar transparency-related project was undertaken by Australia when the government embarked on their SBR initiative (http://www.sbr.gov.au/) in 2006, again utilizing XBRL to simplify business reporting obligations. A standard reporting language such as XBRL was chosen was because it can minimize the regulatory reporting burden on businesses while maximizing the protection of public and private interests via regulation (Chen, 2012). XBRL was chosen as it provides the ability to capture and tag both financial and non-financial information in a standardized, digitized, and human-readable format. By standardizing the data definitions and reporting requirements fewer data elements are required (Chen, 2012). By reporting their data in a consistent manner to all governmental agencies businesses are able to provide timelier reporting as they will not need to transform their existing data sets to meet unique agency reporting requirements (OECD, 2009). The Ministry of Treasury has led the SBR initiative and worked and consulted with key stakeholders in expanding the program to include all levels of government and all significant stakeholders. Government agencies that have participated include the Australian Prudential Regulation Authority (APRA), the Australian Securities and Investment Commission (ASIC), the Australian Taxation Office (ATO), the Australian Bureau of Statistics (ABS), and State and Territory Revenue Offices (SROs) (Chen, 2012). Additional participants included members of the SBR Business Advisory Forum and representatives from the accounting profession, such as the Institute of Certified Bookkeepers and CPA Australia (Chen, 2012).

Kim and Cho (2005) documented transparency enhancements from the implementation of the OPEN (Online Procedures Enhancement from civil application) system by the city of Seoul, South Korea. The OPEN system provides online access for undertaking

administrative procedures such as housing and construction, sanitation, and building, for example. The intent was to avoid unnecessary delays in completing these types of transactions by constituents as well as to prevent inappropriate handling of these activities by civil servants (Kim & Cho, 2005). The researchers documented the positive impacts derived from the OPEN system. From an administrative perspective, an increase in administrative transparency was provided by the access to real-time information describing the procedures for the available online activities. Also, efficiency gains were realized as 657 procedures related to 54 services were defined and categorized into both approval-required activities and non-approval required activities (Kim & Cho, 2005). Forms related to 358 procedures in the current system were standardized for use by all governmental departments and agencies (Kim & Cho, 2005). From the user perspective the benefits include easier to use online access replacing the need to travel to city hall to transact business in person, as well as the ability to track the status of their requests online. The standardization and publication of the details of the 657 procedures assures that citizens will receive equal treatment by government officials and in turn will increase the citizens' trust in the processes (Kim & Cho, 2005).

2.2. Data Analytics

Data analytics, in the context of financial statement auditing, has been defined as the process of discovering and analyzing patterns, identifying anomalies, and extracting other useful information in the underlying data through analysis, modeling, and visualization (Stewart, 2015). Data analytics includes procedures to:

- Identify and analyze anomalous data patterns and outliers

- Map and visualize financial performance and other data across operating units, systems, products, or other dimensions for the purpose of focusing the audit on risks
- Build statistical or other predictive models that explain the data in relation to other factors and identify significant fluctuations from the model
- Combine information from disparate analyses and data sources in order to gain additional insights (Stewart, 2015)

While the use of data analytics by average citizens within the context of the ENHANCE framework may not include such specific purposes as required for formal, financial audits, it is expected that a constituent, acting in the role of an armchair auditor, has similar interests in their investigative actions.

Provost and Fawcett (2013), in presenting their conceptual approach to data science, include as one of three concept types that of 'thinking data-analytically', that is, identifying appropriate data and determining appropriate analytic procedures. Table 1 lists the analytic techniques as identified by Provost and Fawcett (2013).

Purpose	**Technique**
Identifying to which set of categories a new data belongs	Classification
Estimating or predicting the numerical value of some variable for new data – forecasting how much something will be	Regression
Grouping data in a population together by their similarity	Clustering
Characterizing the typical behavior of data either for an individual, group, or population	Profiling
Identifying similar observations based on data known about them	Similarity matching

Predicting connections between data items and estimating the strength of a connection (i.e., link)	Link prediction
Attempting to take a large set of data and replacing it with a smaller set of data that contains most of the important information from the larger set	Data reduction
Attempting to help understand what events or actions actually influence others	Causal modeling

Table 1: Analytic techniques as identified by Provost and Fawcett (2013)

When determining what analytics to include in the ENHANCE framework, a review of the above-mentioned analytics provides a basis from which a determination, at least for the conceptual design of the ENHANCE framework, can be made. For this research the intent is to provide a detailed illustration of a limited scope of usage of the ENHANCE framework. The intent is that the ENHANCE framework can support numerous analytic requirements, but for this research the scope is limited to an example that can be considered typical for an armchair auditor (O'Leary, 2015).

In support of an armchair auditor function it is anticipated that the user may be interested into grouping expenditure data into clusters. An example provided later in this paper illustrates the results of grouping transactions for a particular expense category by vendor in order to identify if there are any major (or very minor) vendors, how many vendors do provide the particular good or service, and the relative spread of expenditure dollars among the vendor population. Similarity matching has value when the user is focused on investigating for possible duplicate payments, a typical audit task. For states, and many larger municipalities, the sheer volume of expenditure transactions may pose a daunting problem in trying to comprehend any particular aspect of the data, such as a comparison of expenditures at a department or agency level. Data reduction analytics can shrink an

extremely large number of transactions into a variety of summarizations that can support the user's requirements. Future research may undertake to expand the role of the ENHANCE framework and may incorporate additional analytics not discussed in this present research. For example, the ability to match or contrast expenditures between similar entities can provide valuable information on the functioning of the entities. In this case similarity matching analytics provide the analyses desired.

Academic research also provides more information and guidance on analytics. Time series analytics, for example, may be used when pattern analysis is not sufficient, for example, when the data exhibits characteristics such as 'systematic nonrandom patterns' (Alwan & Roberts, 1988).

Grouping data into relevant thresholds can provide insights into the structure of the summarized data. The analysis of thresholds typically refers to the identification of the level at which an action or event will be triggered and undertaken (Granovetter, 1978).

Analytics that search for and identify anomalous behavior can also provide insights into underlying activities. Anomalous activities or behaviors are defined as those that do not coincide with established normal profiles (Lee & Xiang, 2001).

Cluster analysis is undertaken to identify subgroups within the data being analyzed (Fraley & Raftery, 1998) and can be considered a subset of pattern recognition. Unlike discriminant analysis which assigns items to pre-defined groups, cluster analysis identifies the appropriate groupings based on the data (Fraley & Raftery, 1998).

2.3. Discussion

This chapter describes the need for data transparency from the perspective of governmental entities, as well as a discussion of what transparency means. Open data and data transparency initiatives in the U.S. and internationally are presented. A definition of data analytics and examples of several analytic procedures are presented.

In comparing the research presented in this chapter with respect to the envisioned design, capabilities, and usage of the ENHANCE framework it is apparent there are gaps that this present paper addresses. The first is a gap between discussions of what transparency means, why it is important, and how to measure it, with an ability to define tools that can provide for or enhance transparency. The second represents a gap between governmental open data initiatives and an ability to generate relevant analyses over that data, in other words, to provide transparency. ENHANCE fulfills that need by providing an advanced tool that supports the analytic capabilities required by governmental constituents in order to understand how their government spends its money. Governmental open data initiatives provide the source and the capabilities of ENHANCE provide transparency into government expenditures. The third gap has to do with analytic tools themselves. At present a user typically has to purchase an analytic tool from the software provider. While current offerings provide robust analytic capabilities there are still limitations with any particular provider's tools with respect to the number and type of analytics available. Also, the user may have to reconfigure the source data upon which analytics are to be executed as the tools generally require the data to fit a particular format. The vision of the ENHANCE framework is that the embedded app recommender system will search across vendors to identify appropriate analytic apps, a task that most users would find

daunting if not impossible. ENHANCE also accommodates any data formatting that is required for the selected analytic apps to execute over the identified data, again a task that may surpass the technical abilities of many of the intended users.

References

Alwan, L. C., & Roberts, H. V. (1988). Time-series modeling for statistical process control. *Journal of Business & Economic Statistics*, *6*(1), 87-95.

Attmore, R. H. (2011, Summer). The Importance of Timely Government Financial Reporting. *Journal of Government Financial Management* http://www.gasb.org/cs/ContentServer?site=GASB&c=Document_C&pagename=GASB%2FDocument_C%2FGASBDocumentPage&cid=1176158583687

Baber, W. R., & Gore, A. K. (2008). Consequences of GAAP Disclosure Regulation: Evidence from Municipal Debt Issues. THE ACCOUNTING REVIEW, 565–591.

Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. Government Information Quarterly, 264–271.

Bharosa, N., van Wijk, R., Janssen, M., de Winne, N., & Hulstijn, J. (2011, June). Managing the transformation to standard business reporting: principles and lessons learned from the Netherlands. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times* (pp. 151-156). ACM.

Brito, J. (2008). Hack, mash, & peer: Crowdsourcing government transparency. The Columbia Science and Technology Law Review, 9, 119-157.

Chen, Y. C. (2012). A comparative study of e-government XBRL implementations: The potential of improving information transparency and efficiency. *Government information quarterly*, *29*(4), 553-563.

Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The computer journal*, *41*(8), 578-588.

Geijtenbeek, W., & Lucassen, H. (2012). Dutch approach to SBR assurance. In *Proceedings of the 25th XBRL International Conference, Yokohama, Japan, p. ASSR1. XBRL International*.

Granovetter, M. (1978). Threshold models of collective behavior. *American journal of sociology*, 1420-1443.

Issa, H. (2013). *Exceptional exceptions* (Doctoral dissertation, Rutgers University-Graduate School-Newark).

Jaeger, P. T., & Bertot, J. C. (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly*, *27*(4), 371-376.

Kim, S., & Cho, K. (2005). Achieving administrative transparency through information systems: A case study in the Seoul Metropolitan Government. In *Electronic Government* (pp. 113-123). Springer Berlin Heidelberg.

Lee, W., & Xiang, D. (2001). Information-theoretic measures for anomaly detection. In *Security and Privacy, 2001. S&P 2001. Proceedings. 2001 IEEE Symposium on* (pp. 130-143). IEEE.

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

O'Leary, D. E. (2015). Armchair Auditors: Crowdsourcing Analysis of Government Expenditures. *Journal of Emerging Technologies in Accounting*.

OECD. (2009, July). Forum on Tax Administration: Taxpayer Services Sub-Group, Guidance note, Standard Business Reporting www.oecd.org/ctp/administration/43384923.pdf

Peled, A. (2011). When transparency and collaboration collide: The USA open data program. *Journal of the American society for information science and technology*, *62*(11), 2085-2094.

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. "O'Reilly Media, Inc.".

Reck, J. L., & Wilson, E. R. (2006). Information transparency and pricing in the municipal bond secondary market. Journal of Accounting and Public Policy, 1-31.

Shim, D. C., & Eom, T. H. (2009). Anticorruption effects of information communication and technology (ICT) and social capital. *International review of administrative sciences*, *75*(1), 99-116.

Stewart, T. R. (2015). Data Analytics for Financial Statement Audits. AUDIT ANALYTICS, 105.

Vishwanath, T., & Kaufmann, D. (1999). Towards transparency in finance and governance. *Available at SSRN 258978*.

Chapter 3: Data Analytics for Government

Chapter 3 provides an overview of data analytics for government data, both U.S. and international initiatives. A background on U.S. open data is presented in Section 3.1 and a review of current open data initiatives is provided in Section 3.2. A review of recent legislation at the U.S. federal level is presented in Section 5.3. Insights into the topic of data transparency are provided in Section 3.4. Section 3.5 describes the current state of U.S. government data analytics, followed by a discussion of this chapter in Section 3.6.

3.1. Background on U.S. Open Data

3.1.1. The State Budget Crisis Task Force

The financial crisis of 2008-2009 brought about the most serious fiscal crises states have faced since the Great Depression, up to the point of bankruptcy in some cases. The State Budget Crisis Task Force (http://www.statebudgetcrisis.org) was formed to develop an understanding of the extent of the fiscal problems faced by U.S. states subsequent to the financial crisis of 2008-2009.

The task force was co-chaired by former Federal Reserve Chairman Paul Volcker and former New York Lieutenant Governor Richard Ravitch. The states of California, Illinois, New Jersey, New York, Texas, and Virginia were studied in detail. Among the more serious financial issues identified in these states included: the use of cash-based budgeting, the absence of relevant mid-year financial planning, and a lack of clarity regarding future financial obligations.

The initial task force report was issued in 2012 noted that the following expenditures were growing at rates faster than the anticipated revenues required to support them:

Medicaid programs, pension funds, and unfunded liabilities for healthcare benefits. It was also noted that several tax bases, such as sales and gasoline, were eroding, and that income taxes were becoming increasingly volatile after the financial crisis. The task force also observed that a lack of financial transparency at the state level hindered the public's ability to understand the critical nature of financial problems such as pension and other payment obligations. Obscure and delayed financial reporting, combined with substandard budgeting procedures, impeded the ability to address these financial problems.

The task force proposed that certain basic procedural approaches be introduced and conformed to by all states. The task force also urged states to give prompt attention to the financial relationships among all levels of government. Specific recommendations to satisfy the public's need for transparent and accountable government included: multi-year planning and budgeting techniques that should be part of normal fiscal planning, better tools to manage the governmental business cycle, and pension plans obligations must be clearly accounted for as well as the disclosure of the possible shortfalls and risks they face. Additionally, states must focus their attention to address the effects from federal deficit reduction efforts, and possible major changes to the federal tax system. Final recommendations include: tax reform at the state level to attain revenue inflows that are adequate, predictable, and have minimum volatility, and, due to the growing gap between states' financial resources and their financial obligations, there exists a need to re-examine the relationship between the federal and state governments.

In the closing statements in their 2012 report the task force noted that only a well-informed public can understand the financial issues faced by governments at all levels

and demand that appropriate actions be taken. The task force also predicted that current state spending, taxation, and administrative practices are not sustainable as the underlying problems are structural in nature and not merely cyclical.

The task force's final report, issued in 2014, included additional recommendations and concluded that the primary remedy to address state financial crises lies in the development of clear, concise, timely, and easily understandable state financial reports. Present financial reports are neither understandable nor prepared in a timely fashion to permit for public evaluation of the financial situation and allow for appropriate decision-making. Changes are needed to provide adequate financial information, and to achieve this requires employing the latest technologies. The task force notes that these issues are also shared at the federal government reporting level.

3.1.2. Government Accounting Standards Board

The State Budget Crisis Task Force has not been the only entity involved in a review of state financial reporting. The Government Accounting Standards Board (GASB) (http://www.gasb.org) issued a research brief in 2011 concerning the timeliness of state and local government financial reporting[7]. A summary of this brief was published in the Journal of Government Financial Management[8] (Attmore, 2011). 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.

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

[8] 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/Research-Publications/Journal.aspx

The findings reflect that 73% of the larger governments studied and 46% of the smaller governments issued their annual financial reports within six months following the close of the fiscal year. Although the survey respondents considered information made available within that timeframe to be 'useful' (the midpoint of the Likert scale used for the survey) they considered it significantly less useful than information received within 45 days to three months after fiscal year-end (Attmore, 2011). By comparison, the SEC requires publicly-held companies to file their annual reports within 60 days of their fiscal year-end[9], the federal government requires its agencies report with 45 days of fiscal year-end, and the government-wide consolidated federal financial report is due with 75 days of fiscal year-end, all significantly shorter timeframes than the six month period noted above.

3.1.3. Public Opinion on Open Data

The Pew Research Center conducted a nation-wide survey (http://www.pewinternet.org/2015/04/21/open-government-data/) 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. The results of the survey are applicable to federal, state, and local governmental levels.

[9] For a listing of SEC-mandated deadlines see: http://www.sec.gov/answers/form10k.htm

The survey results imply that most Americans are in general aware of recent governmental data initiatives and are optimistic that these efforts will make governments more accountable, but are less sure these initiatives will actually improve governmental performance. In summary, the survey results indicated the following (Pew Research Center 2015):

- 65% of Americans have accessed the internet in the past 12-month period to find government-related data or information, for example, how to undertake basic electronic government functions[10] such as conducting online transactions for fee, fine, and/or license payments
- A very small number of respondents indicated that their governments are highly effective in sharing their data (federal = 5%, state = 5%, local = 7%). Greater numbers believed their governments are 'somewhat' effective in sharing data (federal = 39%, state = 44%, local = 45%)
- A somewhat larger number (19% in both cases) can relate instances where their local governments either did or did not provide sufficiently useful information
- Relatively small numbers reported interacting with their government's data as defined by Government 2.0, that is, for monitoring governmental activities (locate student and/or teacher performance = 20%, identify hospital and/or healthcare performance = 17%, investigate government contracts with outside firms = 7%)
    - A majority of citizens believed that benefits will accrue from government data initiatives: 56% believed that open data will permit journalists to cover government activities more thoroughly and 53% believed that open

[10] Electronic government (or e-government) functions are generally defined as digital interactions between: citizens and government (C2G), government to government (G2G), government and citizens (G2C), government and employees (G2E), and government and businesses/commerce (G2B).

data will ensure greater accountability by government officials. An overall 66% of those surveyed believed open data will enhance government accountability. Less than half the respondents believed that open data will improve government service quality or decision-making. Half the respondents believed that open data will allow businesses to identify and create new products and services

- Only 23% of those surveyed believed the federal government acts appropriately 'most of the time', but many of the 23% of respondents believed that open data will provide benefits: 76% state that open data can make officials more accountable, 73% believed that open data will allow journalists to cover government activities better, 71% believed that open data results in better decision-making, 70% indicated that open data will allow citizens to better impact government activities, and 69% believed that open data will improve the level of government services

By analyzing the survey responses the participants were categorized into one of four groups, based on the respondent's level of engagement with government data and online applications, as well as their opinions about the positive impact on government derived from open data. 17% of respondents form the group 'Ardent Optimists' who have actively used online government resources and display an understanding of what open data means. This group also believed that open data will allow journalists to provide better information as well as improving governmental performance. 'Committed Cynics', who form 20% of the respondents, have also been steady users of online resources but were skeptical of any performance improvements that might come from an

open data environment. This group tended to show low levels of trust in government. 27% are 'Buoyant Bystanders' who believed open data will provide better accountability and improved services, but nonetheless indicated that they were not likely to use the online tools for data access. The remaining 36%, the 'Dormant Doubters', were also not likely to use online tools as they at present had used internet services on a very limited basis, and also did not believe open data will improve accountability or services. This group exhibits low levels of trust in government.

The report noted that an individual's perception of trust in government drove their belief about the value and success of government data initiatives in improving performance.

3.2. Current Open Data Initiatives

3.2.1. International Open Data Initiative

The Open Government Partnership (OGP) (http://www.opengovpartnership.org/) was launched in 2011 to provide a platform for world-wide domestic reformers committed to making their governments more open, accountable, and responsive to citizens. OGP initially consisted of eight founding governments (Brazil, Indonesia, Mexico, Norway, the Philippines, South Africa, the United Kingdom and the United States) and has since grown to 65 participating countries. In all of these countries, government and society are collaborating to develop and implement ambitious open government reforms. The OGP initiative aims to secure commitments from governments to promote transparency, empower citizens, fight corruption, and harness new technologies to strengthen governance. To participate in OGP a country must endorse an Open Government Declaration, deliver a country-specific action plan that has been developed with public input, and commit to an independent assessment, in the form of OGP's Independent

Reporting Mechanism (IRM) (OGP2), on their progress going forward. In total, OGP participants have made over 1,000 commitments to make their governments more open and accountable (OGP1).

OGP's vision is that governments will develop and maintain an environment of greater transparency, accountability, and responsiveness to their citizens, with an ultimate goal of providing a high-level of quality of governmental administration and services to their constituents (OGP2). There are three primary means for OGP to help ensure the proper conditions are in place for countries to undertake open government reforms: maintaining high-level political leadership and commitment to OGP, supporting domestic reformers with technical expertise and support, and fostering greater engagement in OGP by both citizen groups and civic organizations (OGP2).

3.2.2. U.S. Open Data Initiatives

Several governmental entities in the U.S. have undertaken to 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 a number of governmental entities provide online access to their data, such as the examples in Table 2[11], the format that the data is provided in does not always processing the data for further analyses[12].

Governmental Entity	Website
U.S. Federal Government	http://www.data.gov
State of Ohio	http://www.ohiocheckbook.gov

[11] A listing of several state and municipal open data portal website addresses is presented in Table 49 in the Appendix. It would be difficult to identify all available data portals as at this time there is no repository or listing of all the governmental data portals available in the U.S. and governmental entities are regularly launching new portals

[12] Much of the available data is provided in PDF format which is difficult to manipulate for analytical purposes

State of Utah	http://www.utah.gov/transparency/index.html
City of Chicago	http://www.data.cityofchicago.gov
City of Detroit	https://data.detroitmi.gov/
City of Seattle	http://www.data.seattle.gov

Table 2: Examples of U.S. government data websites

The advances in technology that enable the re-engineering of governmental financial reports allow governments to integrate electronic reporting capabilities into many of their projects.

The federal government open data website (http://www.data.gov/) lists 131,461 datasets (as of the time of preparation of this paper). The data is available in a variety of formats: Excel, CSV, XML, JSON (JavaScript Object Notation), API (application programming interface), HTML, open XML, text, and PDF. The data covers topics that include: agriculture, business, climate, consumer, ecosystems, education, energy, finance, health, local government, manufacturing, ocean, public safety, science and research.

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 Government Accounting Standards Board (GASB), and PriceWaterhouseCooper (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. 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

was loaded into taxonomy building software that processed the data and converted it 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.

A partnership between Ohio Treasurer Josh Mandel and the finance visualization startup firm OpenGov (http://opengov.com/about/) is undertaking one of the most inclusive statewide data transparency efforts to date[13]. The initiative, announced in April, 2015, offers Ohio's 3,900+ local governments including townships, cities, counties, school districts and others a chance to post revenues and expenditures online free of charge through the state's budget transparency site OhioCheckbook.com (http://tos.ohio.gov/Transparency/Ohios-Online-Checkbook). The undertaking is the result of the state's financial transparency campaign that Treasurer Mandel launched upon taking office in 2007 when public salaries and state property data were made available online. As the rollout of the current initiative continues, which began in June, 2015, citizens will be able to track local government revenues and expenditures via interactive graphs that illustrate both high-level budget information and actual spending details check-by-check (Ohio). The data includes top earning government contractors, the highest paid officials, and revenue consumption by departments (OHIO). The OhioCheckbook.com website currently provides state expenditures for the past seven

[13] As presented in the April 22, 2105 edition of GOVTECH.COM: http://www.govtech.com/budget-finance/Ohio-Offers-Financial-Transparency-Statewide.html

years and the current initiative intends to bring transparency to all of Ohio's governmental entities. Since the December, 2014 rollout of the OhioCheckbook.com website, the state's financial transparency rating from the U.S. Public Interest Research Group (http://www.uspirg.org/) has jumped from 46th to first in the nation (Ohio).

The Mayor's Office of Data Analytics (MODA) (http://www.nyc.gov/html/analytics/html/home/home.shtml) supports New York City's ability to aggregate and analyze data from across city agencies to support numerous citizen needs such as an ability to more effectively address crime, public safety, and quality of life issues. MODA employs analytics tools for a number of initiatives such as risk prioritization, more efficient delivery of city services, more effective law enforcement, and an overall increase in transparency (NYCMODA). MODA's primary functions include collaborating with NYC agencies to support their implementation of data-driven solutions to address service delivery issues, developing a city-wide data platform to support data sharing, performing an oversight role for all city-wide data projects, and responsibility for implementation of NYC's Open Data Law. The initiatives cover five broad categories: supporting city operations, city-wide data sharing, disaster response and resiliency, economic development, and open data (NYCMODA). MODA partners with NYC agencies and external partners to implement data driven solutions to support governmental needs as well as undertaking data requests, training agencies in the use of the city-wide data sharing platform (DataBridge) and related best practices, and performing analyses for and implementations of enterprise IT solutions (NYCMODA).

Bloomberg Philanthropies announced on April 20, 2015[14] the launch of What Works Cities (http://whatworkscities.bloomberg.org/), a $42 million, three-year initiative that will assist mid-sized American cities to undertake data-driven projects to improve life for their residents. The initiative is seeking 100 cities with populations between 100,000 and 1 million residents to receive guidance from program partners including The Behavioral Insights Team, Harvard Kennedy School of Government Performance Lab, Johns Hopkins University's Center for Government Excellence, Results for America, and the Sunlight Foundation.

The program is designed to provide technology-minded mayors and IT leaders with comprehensive support in their undertaking of technology-driven projects[15]. Just prior to the official launch pilot programs were undertaken with about a dozen cities, including Chattanooga, TN, Seattle, WA, New Orleans, LA, Tulsa, OK, and Mesa, AZ. Bloomberg will invite 285 cities to participate. Results for America (http://results4america.org/) will provide overall program coordination of the efforts of all the partners and work to ensure an exemplary service experience for all the cities that participate. Results for America will instruct the participants in the opportunities that using better data provide. The Center for Government Excellence at Johns Hopkins University will work closely with city leaders to guide them through their projects.

---

[14] As presented in the April 22, 2105 edition of GOVTECH.COM: http://www.govtech.com/data/Bloombergs-What-Works-Cities-Initiative-Targets-100-Mid-Sized-Metros.html

[15] Example projects as provided on the What Works Cities website include: a travel planning app in Albuquerque, N.M. that reduced transit-related 311 calls by 25 percent, a New Orleans, La. data coordination software effort that reduced urban blight by 10,000 residences, and an Atlanta, Ga. performance management system that reduced backlogged housing code violation complaints by 70 percent.

Trading in the secondary market for municipal securities has been shown to lack the data timeliness and frequency found in the corporate securities market, specifically price and information transparency (Reck & Wilson, 2006). Efforts have been made to improving disclosure in the municipal market such as when as the SEC in 1994 required (SEC Regulation 15c2-12.) that brokers, dealers, and underwriters provide annual disclosure to the secondary market (Baber & Gore, 2008). Unfortunately there is no enforcement mechanism to require disclosure (Baber & Gore, 2008).

With its mission to protect investors, municipal entities, and the public interest, the Municipal Securities Rulemaking Board (MSRB) (http://www.msrb.org/) was established to promote a fair and efficient municipal market, to regulate firms engaging in municipal securities and advisory activities, and to promote market transparency. One of the initiatives that the MSRB undertook was to develop a free, online source of information that includes municipal disclosures, market transparency data, and educational materials. From this initiative was developed the Electronic Municipal Market Access (EMMA) (http://emma.msrb.org/) website. EMMA is considered to be one of the more significant changes that have occurred in the municipal bonds trading market. The EMMA website provides free access to two types of individual municipal bond data. The first is the information required when the municipality initiates a bond issue or prepares related primary market disclosures. This information includes official statements, refunding statements, continuing disclosures, and other related documents. The second is more relevant to the secondary market and includes information such as bond trading price, rate, and other historical trading data.

The MSRB requires that municipal securities dealers report current transactions within minutes of trade time so that the information can be available in as close to real-time as possible. A limitation with the data as available on the EMMA website is that the information is provided in PDF format and as such does not lend itself to analytical purposes, thus greatly limiting its usefulness.

3.2.3. State and Municipal Open Data Portals

Utah's transparency website (http://www.utah.gov/transparency/index.html) is dedicated to the transparency and accountability of state government finances. The website is administered by the Utah Division of Finance under the direction of the Utah Transparency Advisory Board. The Utah Transparency Advisory Board is comprised of eleven members knowledgeable about either public finance, or providing public access to public financial information. The data is searched by level, which includes state, cities and towns, counties, and higher education, for example. Within a level a specific entity is selected, then fiscal year, and finally expense or revenue. The results can be displayed at several levels of detail: organization (a one-line summary), category, fund, vendor name, and finally transaction level. The results can be printed.

Ohio's Online Checkbook data portal (http://www.tos.ohio.gov/Transparency/Ohios-Online-Checkbook) includes access to several datasets: the Online Checkbook allows taxpayers to understand how their tax dollars are being spent and hold public officials accountable for those expenditures. There are salary search tools at the state, local, teacher, higher education, and federal levels. Also included is a state properties map, a search for unclaimed funds tool, a policy statement on public records, the daily ledger report of state funds, and archived annual reports. With respect to the expenditure

details in the checkbook, the data is accessible by setting filters at several levels: the primary filter is for fiscal year, basic filters for agency and expense type, and advanced filters for appropriation line item, program, and fund. Some of the more popular datasets have pre-configured searches that include summary and detail level data, as well as pie-chart representations. Examples of these include payroll, travel, meal expenses, capital items, and state debt payments. Data can be output in a CSV format.

The elements of the detailed expenditure data include the following:

- A checkbook unique identifier for the specific transaction
- Transaction date
- Vendor name
- Voucher journal ID
- Check number
- Amount
- Entity code
- Business unit code
- Fiscal year
- Entity name

Other datasets available on this website include: listings of the largest expenses and also highest paid firms for last fiscal year, and lookups for salary information by name for teachers, higher education employees, federal employees in Ohio, and for local government employees. Analytic capabilities include comparisons between fiscal years, agencies, and expense types (including multiple values). There are 16 popular searches

available including payroll, roads & highways, and travel, among others. The data is displayed using pie chart graphics and summary/detail information is exportable to CSV.

Bond-related data that is available includes:

- Bond issuance:
    - Upcoming issues
    - Current issues
    - Past issues
- Bond programs are listed (Parks & Recreations, for example)
- Other issues also (Turnpike Commission, for example)
- Financial reports are provided:
    - Sinking fund reports
    - Chapter 154 reports

The Ohio Transparency website (https://ohio.gov/government/transparency/) includes a number of datasets relevant to governmental activities, and of those several include financial data that can be used to support the envisioned data analytics. Grants information, employee salaries, budget data, and tax collections can be reviewed on this website.

Grant information is presented in PDF format by department and fiscal year (2008-2015), and includes program, grantee, grantee address, amount, and grant period.

Employee salary information is provided in PDF format by pay period (two per month) for fiscal years 2010-2015, and includes pay period end data, agency, employee name, job title, and total gross amount. Overtime cost information is provided at an agency

summary level for fiscal years 2013-2015 by fiscal quarter. Overtime information is not easily downloadable. Monthly employee trend reports are also available, in PDF format, for fiscal years 2007-2015 that provide by agency the total employee count and number of full-time, part-time, and temporary workers. Users can compare reports for same month across fiscal years to note trends in employment patterns by agency.

Budget data is provided in PDF format for operating budgets for fiscal years 2006/2007 through 2016/2017 and capital budgets for fiscal years 2001/2002 through 2015/2016. Operating budget data, by fiscal year, is provided separately for the Main Operating budget, Transportation budget, and Ohio Industrial Commission and Bureau of Workers' Compensation. This information is provided for both Enacted Appropriations and Enacted Budget Bills. The detail information includes, for example for FY2014/2015: FY 2012, FY 2013 Estimate, FY 2014 Appropriation, FY 2013 to 2014 change in dollars and percent, FY 2015 Appropriation, and FY 2014 to 2015 change in dollars and percent. Capital budget appropriations are provided by fund, agency, fund and agency, agency and fund, and a new debt authorization document. This information is provided at a summary level, such as by agency, with a single amount only listed. There is a link from the Ohio Checkbook website to allow access to this budget information.

Tax collection information is available annually in PDF reports. The reports list by type of tax such as sales and use, income, resort, and corporation franchise, and provide, for example, gross tax collection for current and prior FY, net tax collections for current and prior FY, and a percent change in net collections prior-to-current FY.

The NYC Checkbook (http://www.checkbooknyc.com/spending_landing/yeartype/B/year/116) provides revenue, spending, payroll, and contract data. The primary filter is for either citywide agencies or other governmental entities and filters are available on all the data elements in the dataset. Data can be output in CSV format.

The cities of Chicago, Detroit, Las Vegas, Los Angeles, and Seattle also provide open data portals, all sourced on software provided by a single provider[16], and as such these portals all provide a similar level of access and capabilities while being tailored to the particular municipality. The Las Vegas and Los Angeles portals include checkbook-level data on expenditures. Checkbook search capabilities typically include by department, vendor, and expenditure type. In addition to the checkbook data the portals also typically include datasets representing other governmental activities: Las Vegas provides approximately 36 datasets that include public safety, planning, building safety, and checkbook data, for example. Seattle provides over 2,100 datasets, 21 of them finance-related, including information such as proposed and endorsed budgets, business licenses, and expenditures at a department/program level, to name a few. Data can be extracted from the portals in a number of formats including PDF, CSV, XLS, and XML.

3.3. U.S. Federal-Level Legislation and Executive Actions

In 2006 then-Senator Barack Obama co-sponsored the Federal Funding Accountability and Transparency Act (FFATA) (https://www.congress.gov/bill/109th-congress/senate-bill/2590) that required full disclosure to the public of all entities or organizations

[16] Socrata is a firm that markets a number of open data and transparency-related products to support local, state, and federal governments; see: http://www.socrata.com/industries/open-data-state-local-government/, http://www.socrata.com/industries/open-data-federal-governments/

receiving federal funds. Reporting requirements of federal funds recipients, such as state and local municipalities, were addressed in a three year pilot program. This pilot program was run by the Office of Management and Budget (OMB) and sought to present the full cycle of federal funds as provided to a diverse group of recipients, while ensuring the accuracy of the published data. The conversion of state and local reporting to a standardized uniform digital format was encouraged by the compliance requirements of the pilot program for grant recipients in the DATA Act.

In 2008, Senator Obama sponsored the Strengthening Transparency and Accountability in Federal Spending Act[17]. On December 8, 2009 the White House issued an Open Government Directive[18] that requires federal agencies to initiate actions to provide for greater transparency, participation, and collaboration.

The information that these federal-level directives and legislation intend to provide will be accessible by means of user-friendly websites such as Recovery.gov[19], USASpending.gov[20] and foreignassistance.gov[21]. The primary intent of this openness is to provide the public with technology-based tools to access information and gain the

[17] The 2008 Act amends the 2006 Act by requiring that the Office of Management and Budget's (OMB) federal funding database website allow the public to search and access all data in a machine readable format and that the website present information about federal awards and their recipients in a means to support users with different levels of understanding about government spending and also technical abilities. The 2008 Act also requires that the Director of OMB add specified data quality enhancements to the website. See: https://www.govtrack.us/congress/bills/110/s3077

[18] This memorandum requires executive departments and agencies to take the following steps toward the goal of creating a more open government: 1. Publish government information online, 2. Improve the quality of government information, 3. Create and institutionalize a culture of open government, and 4. create an enabling policy framework for open government See: https://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf

[19] Recovery.gov was created by the American Recovery and Reinvestment Act of 2009 and displays information about the Recovery Accountability and Transparency Board's activities.

[20] USAspending.gov is the publicly accessible and searchable website mandated by the Federal Funding Accountability and Transparency Act of 2006 to give the public access to information on how tax dollars are spent

[21] ForeignAssistance.gov provides access to U.S. Government foreign assistance funds and enables citizens to examine, research, and track aid investments.

knowledge which they can use to influence the governmental decision-making that impacts their lives.

In his first day in office in 2009 President Obama signed the Memorandum on Transparency and Open Government (www.whitehouse.gov/open/about) and (https://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_fy2009/m09-12.pdf). The intent of the President's action was to provide an open and accountable federal government that can link constituents to governmental activities. A primary objective of issuing this memorandum was to provide a means for constituents to track how their tax dollars, as well as funds from other revenue sources, were being utilized.

In 2011 President Obama expanded the effort for open government to a world-wide level when he launched the Open Government Partnership (OGP)[22] at the United Nations General Assembly meeting with another seven heads of state.

In 2014 President Obama signed into law the Data Accountability and Transparency Act (DATA Act) (https://www.congress.gov/bill/113th-congress/house-bill/2061) and (http://www.datacoalition.org/what-is-data-transparency/data-act/). A primary purpose of this legislation is to expand the Federal Funding Accountability and Transparency Act of 2006 by disclosing direct federal agency expenditures and linking federal contract, loan, and grant spending information to federal programs to enable taxpayers and policy makers to track federal spending more effectively. The legislation also mandates government-wide data standards for financial data and provide consistent, reliable, and searchable government-wide spending data that is displayed accurately for taxpayers and

[22] OGP is an international initiative intended for domestic reformers to make their governments more open, accountable, and responsive to citizens. OGP has grown from an initial 8 countries to the 65 participating countries. See: http://www.opengovpartnership.org/ and http://www.state.gov/j/ogp/

policy makers on the USASpending.gov website. Additionally, the DATA Act will simplify reporting for entities receiving federal funds by streamlining reporting requirements and reducing compliance costs while improving transparency, improve the quality of data submitted to USASpending.gov by holding federal agencies accountable for the completeness and accuracy of the data submitted, and apply approaches developed by the Recovery Accountability and Transparency Board to spending across the federal government. The legislation also amended the FFATA of 2006 to clarify the definition of a 'federal agency' to include executive departments, government corporations, and independent establishments.

The Financial Transparency Act of 2015 (also referred to as the Making All Data Open for Financial Transparency Act, or MADOFF Transparency Act) (http://www.datacoalition.org/what-is-data-transparency/financial-industry-transparency-act/) is the latest legislation (as of the time of preparation of this paper) to be introduced by the U.S. Congress. The legislation mandates both the development of data standards and publication of all data as required by law in a searchable format. The Act will require that the nine members[23] of the Financial Stability Oversight Council adopt and coordinate data standards for the information their respective agencies collect to ensure search abilities across agencies, as well as providing any other information as required by other laws, be made public in a searchable fashion. The legislation is intended to benefit

[23] The ten voting members of the FSOC include: Secretary of the Treasury, who serves as chairperson of the FSOC, Chairman of the Board of Governors of the Federal Reserve System, Comptroller of the Currency, Director of the Bureau of Consumer Financial Protection, Chairman of the Securities and Exchange Commission, Chairperson of the Federal Deposit Insurance Corporation, Chairperson of the Commodities Futures Trading Commission, Director of the Federal Housing Finance Agency, Chairman of the National Credit Union Administration. An additional independent member, with insurance expertise, is appointed by the President and confirmed by the Senate for a six-year term. See**:** http://www.treasury.gov/initiatives/fsoc/Pages/home.aspx

several constituent groups: first, by providing searchable data on investment opportunities to investors and investment analysts; second, to allow regulated entities to automate their currently manual compliance activities; and lastly, to permit financial regulators to match information as reported by a single firm to multiple agencies. This last benefit is specifically intended to prevent the occurrence of another investment scandal such as the Madoff investment fraud scandal.

3.4. Data Transparency Project

Many of the recent federal directives and legislation include 'transparency' in their title[24]. State initiatives that were discussed earlier also highlight transparency as a reason for and a result of their recent initiatives. Kim and Cho (2005) documented the transparency impact from the implementation of the OPEN (Online Procedures Enhancement from civil application) system by the city of Seoul, South Korea. The OPEN system provides online access to citizens to undertake administrative procedures such as housing and construction, sanitation, and building, for example. The intent was to avoid unnecessary delays in completing these types of transactions by citizens as well as to prevent unjust handling of these activities by civil servants (Kim & Cho, 2005).

The researchers documented the positive impacts derived from the OPEN system: from an administrative perspective there is an increase in administrative transparency as provided by the access to real-time information detailing the procedures related to the activities that are available online. Also, efficiency gains were realized as 657 procedures related to 54 services were defined and categorized into both approval-

[24] For example: Federal Funding Accountability and Transparency Act (2006), Strengthening Transparency and Accountability in Federal Spending Act (2008), Memorandum on Transparency and Open Government (2009), Data Accountability and Transparency Act (2014), and The Financial Transparency Act of 2015

required activities and non-approval required activities (Kim & Cho, 2005). The forms related to 358 procedures existing in the current government system were standardized for use by all governmental departments and agencies (Kim & Cho, 2005).

From the user perspective the positive impacts include ease of use as online access replaced the need to actually visit city hall to transact business, as well as being able to track the current status of their requested transaction. The standardization and publication of the details of the 657 procedures guaranteed the equal treatment by responsible government officials and increased the trust the citizens have in the processes (Kim & Cho, 2005).

International organizations such as the Open Government Partnership (http://opengov.com/about/) set transparency goals and objectives and provide support to countries in achieving them within their governmental operations.

3.5. The Current State of U.S. Government Data Analytics

There are a number of governmental entities that are now providing open data portals via the internet that make available any number of data items for viewing. These efforts have been assisted by the availability of software products designed to support open data initiatives. Several of these providers are listed in Table 3:

Provider	Website	Description
Socrata	http://www.socrata.com/industries/open-data-state-local-government/, http://www.socrata.com/industries/open-data-federal-governments/	Markets a number of open data and transparency-related products to support local, state, and federal governments
Streamlink Software	http://www.streamlinksoftware.com/products/grant-management-	Solutions tailored to grant management,

	system/sectors/government-grant-management/	specifically the grant-related activities of research, planning, activity management, performance, and reporting
Teradata	http://www.teradata.com/industry-expertise/government/?ICID=Sgovernment&LangType=1033&LangSelect=true	Database, Data Warehousing, and Business Intelligence (BI) solutions serve both public and private clients; also provides Tax Compliance products to identify tax fraud
BCL Technologies	http://www.bcltechnologies.com/solutions-industries.htm	Document conversion solutions that can take, for example, PDF documents and convert them into a more digital-friendly format, in support of a data transparency initiative
REI Systems	https://www.reisystems.com/MarketsSolutions/Pages/Open-Government.aspx	Supports transparency initiatives in federal agencies such as OMB and GSA in the design, implementation, and management of government-wide websites

Table 3: Selected Software Providers

REI supported New York City in the development of the city's Checkbook NYC 2.0 (http://www.checkbooknyc.com/spending_landing/yeartype/B/year/116) expenditure\transparency website. REI has also supported the development of the federal-level USASpending.gov (https://www.usaspending.gov/Pages/Default.aspx) and DATA.gov (http://www.data.gov/) websites. There are also firms that not only develop data analytic tools but also collect data and provide both data and analytics, typically on a

subscription basis. InsideGov (http://www.data.gov/) and Posiba (https://www.posiba.com/products) provide government-sourced data and appropriate analytics.

With respect to available open data initiatives two of the more robust and widely accepted are the 'Checkbook' data portals for New York City (NYC) and the State of Ohio that present detail-level expenditure data at the check level. The capabilities for each data portal are provided in Table 4 for NYC and Table 5 for Ohio.

NYC Checkbook:	
Data Available:	Revenue
	Spending
	Contracts
	Payroll
Data Feeds:	Budget
	Revenue
	Spending
	Contracts
	Payroll
Selection Criteria:	Citywide Agencies or
	Other Governmental Entities
Filters:	Agency
	Department
	Expense Category
	Spending Category
	M/WBE Categories
	Industry (Vendor)
	Payee Name Check Amount: From/To
	Contract ID
	Document ID
	Capital Project
	Date Filter: Year or Issue Date (From/To)
Data Column Selection (Output, in CSV Format):	Agency
	Associated Prime Vendor
	Vendor
	Calendar Year

	Capital Project
	Check Amount
	Contract ID
	Contract Purpose
	Department
	Document ID
	Expense Category
	Fiscal Year
	Industry
	Issue Date
	M/WBE Category
	Payee Name
	Spending Category
	Sub Contract Reference ID
	Sub Vendor (Yes/No)
Capabilities: Tools/Trends/Financial (Output in CSV Format):	At a summary level for the years 1995-2014: Changes in Net Assets, Fund Balances - Government, Changes in Fund Balances, General Fund Revenues and Other Financing Sources, General Fund Expenditures and Other Financing Uses, Capital Project Fund Aid Revenues, NYC Educational Constructional Fund
Capabilities: Published Financial Reports (Output in PDF Format):	Annual Reports, Budget Reports, Audits, Contract Administration, Financial Administration, Policy
Capabilities: Link to My Money NYC:	Online applications to help make NYC finances more transparent; links back to Checkbook NYC and M/WBE, Audit, and Pension information
Capabilities: Link to NYC Open Data	Nearly 1,300 datasets available listing governmental-related information
Capabilities: Link to NYC Finance Department	Information on Assessments, Taxes, Licenses, and Fines
Capabilities: Link to NYC Economic Development Commission	Lists Policy documents, annual reports, financials, and meeting minutes in PDF Format

Table 4: Characteristics of NYC Checkbook data portal

Ohio Checkbook:	
Primary Filter:	Fiscal Year
Basic Filters - Agency:	Type: (Agency, Boards & Commissions, Elected Official, Higher Education)
	Entity: (Accountancy Board of Ohio . . .) (144 Entities)
Basic Filters – Expense Type:	Expense Category: (Capital, Debt Service, Supplies . . .) (51 Categories)

	Expense Type: (Accounting/Auditing, Advertising . . .) (231 Types)
	Expense Code: (Advertising-Legal, Ammunition, Taxes . . .) (888 Expense Codes)
Advanced Payment Filters:	Appropriation Line Item: (Attorney Services, Health Care Programs . . .) (2,083 Line Items)
	Program: (Criminal Investigation, health & Safety . . .) (1,578 programs)
	Fund: (Building management, Higher Education Improvement . . .) (811 Funds)
Data Output (in CSV Format):	Checkbook Unique Identifier (9-digit numeric)
	Transaction Date
	Vendor Name
	Voucher Journal ID
	Check Number
	Amount
	Entity Code (alphanumeric)
Capabilities:	Comparisons, between: Fiscal Years, Agencies, Expense Types (multiple)
	16 popular searches available: (Payroll, Roads & Highways . . .)
	Pie Chart Graphics and Details provided
	Operational and Capital Budget Data available
	Bond-Related Data: Bond Issuance: Upcoming Issues, Current Issues, Past Issues)
	Bond Programs listed
	Financial Reports

Table 5: Characteristics of Ohio Checkbook data portal

From the information presented above for NYC and Ohio checkbook data it can be seen that although these portals provide a number of filtering and formatting capabilities over the data presented, at best there are only very basic analytic capabilities, such as the data comparisons that Ohio provides. An advanced analytic tool with capabilities such as those provided by the ENHANCE framework can fill a very apparent gap between open data and appropriate analytic tools.

3.6. Data Comparisons

Attention should be given to the data attributes provided by each entity with regard to their checkbook, or expenditure, data as these attributes both define what information is available for analysis and provide the limits as to what information can be gathered. A comparison and contrast of the data attributes for NYC and Ohio is provided in Table 6.

Data Attribute (Common Name or NYC/Ohio Specific Name for similar attribute)	NYC Checkbook	Ohio Checkbook
Agency/Entity	X	X
Fiscal Year	X	X
Check Amount/Amount	X	X
Department/Business Unit Code	X	X
Payee Name/Vendor Name	X	X
Document ID/Voucher Journal ID	X	X
Issue Date	X	
Transaction Date		X*
Checkbook Unique ID		X
Check Number		X
Associated Prime Vendor	X*	
Entity Code		X
Calendar Year	X	
Capital Project	X*	
Contract ID	X	
Contract Purpose	X*	
Expense Category	X	
Industry (Vendor)	X*	
M/WBE Category	X	
Spending Category	X	
Sub Contract Reference ID	X*	
Sub Vendor (yes or no)	X	
NOTE: X* indicates attribute is not populated for all records		

Table 6: Comparison of NYC Checkbook and Ohio Checkbook data attributes

The data attributes have been organized to display first those attributes in common between the two checkbook portals. Having a greater number of data attributes available

in a dataset will allow for analyses of the data from several perspectives and it appears the attributes in common between the selected checkbook portals will support a level of basic analyses. Although this example only includes expenditure data from two entities with rather robust data portals, it is expected that as other entities develop and deploy portals with transaction-level, expenditure data the attributes will be similar to the examples provided. For entities of sufficient size and complexity that operate either automated accounting systems or an Enterprise Resource Planning System (ERP) these data attributes should be readily available as they are typical of what is minimally incorporated in these systems.

The data attributes available in the examples given can support transparency initiatives and provide for analytic comparisons such as: across fiscal years, and/or filtered by transaction dates or periods, between similar agencies in different governmental entities, or between departments in agencies within one governmental entity or across entities. Vendor analyses within or across agencies, and at a more detailed level within or across departments within or across agencies are possible with vendor name criteria. A document or journal voucher ID allows for identifying unique documents and maintaining that level of detail in the analyses. The greater the number of attributes available implies the more varied analytic results that can be generated. For example, the department (the primary governmental unit) that the expenditure is charged to is generally provided. If the dataset includes the funding source and/or information on the underlying purchase contract, then analytics can output information from these perspectives as well as department. If multiple levels within department, such as

division, program, and/or activity, as available in the Austin, TX expenditure data, are provided then more sophisticated analyses can be generated.

As noted in Tables 4 and 5 the analytic capabilities of these data portals is somewhat restricted and thus limits transparency efforts. For the NYC portal the filters are essentially the data attributes of the selected dataset and cross-department or cross-dataset, such as budget versus actual, are not available in a pre-defined format. For those with some degree of technical expertise the NYC Checkbook data for budget, contracts, payroll, revenue, and spending is available in an Application Programming Interface (API) (http://www.checkbooknyc.com/api). Each API call is limited to 1,000 records however with the resultant data being sent and received using XML formats. Each API request includes one global parameter and two optional global parameters, but depending on the type of data selected there may be additional optional parameters. The API interface may not be of much value to the average citizen without sufficient technical expertise. In this case analytic capabilities come with a price, that is, required technical expertise.

The Ohio Checkbook site does not provide an API but does incorporate more sophisticated analytic tools that should be easily understandable by the average citizen. Comparison tools include selection of initial fiscal year, agency, and expense type with the next step being that of selecting the comparative filters, be they individual or multiple fiscal years, agencies, and/or expense types. The Ohio portal also provides 16 pre-defined, popular searches covering travel, payroll expenses, capital items, and state debt payments, for example. These searches cover datasets whose underlying data is deemed as critical to the constituency from a transparency perspective. The selected information,

for a specific fiscal year, is presented in a pie chart by either agency or expense type as well as by line item in a table at either the summary level by either agency or expense type, or also at the detail transaction level. The table data, either summary or detailed transaction level, is exportable in CSV format.

3.7. Discussion

In summary, the development and deployment of open data portals, in support of governmental data transparency initiatives to satisfy the needs of the constituency, have made great strides over the past few years, as technological advances have made these initiatives feasible and cost-effective. Efforts specifically focused on providing elements of transparency over that data, such as analytic capabilities, are not yet in place. Efforts to provide transparency have taken several different paths, such as providing official documents for online completion and submission by constituents, but not in providing robust analytics over transactional data. In many cases robust datasets are accessible to the public to satisfy inquiries about governmental spending and other governmental activities. The accessibility to robust analytics that present information in a meaningful manner to constituents is yet to come. Several data portals provide access to basic analytics or to development tools for the technically knowledgeable. The need exists to provide the public with access to robust analytics that provide useful and understandable results, all in support of greater governmental transparency. The intent in the design and development of the ENHanced ANalytic Constituent Environment (ENHANCE) is to close the gap between robust analytic tools and easy access to them, in order to leverage the open data initiatives in place and yet to come so that meaningful information can be drawn from them.

References

Attmore, R. H. (2011, Summer). The Importance of Timely Government Financial Reporting. *Journal of Government Financial Management* http://www.gasb.org/cs/ContentServer?site=GASB&c=Document_C&pagename=GASB%2FDocument_C%2FGASBDocumentPage&cid=1176158583687

Baber, W. R., & Gore, A. K. (2008). Consequences of GAAP Disclosure Regulation: Evidence from Municipal Debt Issues. THE ACCOUNTING REVIEW, 565–591.

Kim, S., & Cho, K. (2005). Achieving administrative transparency through information systems: A case study in the Seoul Metropolitan Government. In *Electronic Government* (pp. 113-123). Springer Berlin Heidelberg.

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

NYCMODA: http://www.nyc.gov/html/analytics/html/home/home.shtml

OGP1: http://www.opengovpartnership.org/about

OGP2: http://www.opengovpartnership.org/about/mission-and-goals

OHIO: http://www.govtech.com/budget-finance/Ohio-Offers-Financial-Transparency-Statewide.html

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

Reck, J. L., & Wilson, E. R. (2006). Information transparency and pricing in the municipal bond secondary market. Journal of Accounting and Public Policy, 1-31.

Chapter 4: A Vision of an ENHanced ANalytic Constituent Environment: ENHANCE

4.1. Introduction

The average citizen is not likely to have the tools or expertise required to produce sophisticated analytics on their own. As constituents increasingly rely on open data initiatives to allow monitoring of governmental activities how will they react to a lack of proper analytics? Will citizens and other stakeholders find usefulness from access to data alone without analytic capabilities? Do governmental entities have the skillsets, and resources, to develop analytic tools, and in a timely fashion to assuage constituent demands? What would be the incentive for third-party entities to develop such analytics? Would this lack of analytic tools be viewed as an overall failure of the open data initiative and lead to even greater levels of consumer distrust of government?

The remainder of this chapter is organized as follows: Section 4.2 provides an overview of Decision Support Systems, as ENHANCE supports decision making on the part of its intended users, governmental constituents who wish to understand how their government is spending its money. Section 4.3 presents the ENHANCE framework and the various apps that comprise it. A review of the elements of Design Science Research Methodology is presented in Section 4.4 as such an approach is appropriate in the actual development effort for ENHANCE. Section 4.5 reviews two presently available and sophisticated government data portals as models of the expenditure datasets available over which ENHANCE can undertake analytics. Section 4.6 describes a typical structure of governmental budget data and the benefits analytics over budget data can provide. Section 4.7 provides a review of a number of presently available analytics that are likely

to support a user's analytics requests. Section 4.8 illustrates a number of the analytic apps presented in Section 4.7 as executed over a sample of actual governmental expenditure data, and presents the output in formats that mimic what ENHANCE is intended to provide. Section 4.9 illustrates an example of using actual budget data in an analytic setting. Section 4.10 provides an example of the results of analytics when combing two data sources. It should be noted this scenario, while presenting the power of an analytic tool such as ENHANCE, is outside the scope of the ENHANCE framework design as described in this paper. A review of a specific setting in which ENHANCE can leverage its capabilities, the setting of an 'armchair auditor' function, is presented in Section 4.11. The design proposal for the ENHANCE user interface is detailed in Section 4.12 and the proposed data structure that underlies the ENHANCE user interface is presented in Section 4.13. Section 4.14 provides background information on what is envisioned to be one of the key components of the ENHANCE framework, an analytic app recommender system. This section describes prior research that defines an app recommender system for an audit application. A review of Design Science Research Methodology with respect to developing an app recommender system for ENHANCE is provided. The actual development of an app recommender system in support of the ENHANCE framework, as well as a complete DSRM template to drive that development, is outside the scope of this present research. Section 4.15 illustrates a hypothetical, but realistic, application of the analytic capabilities of ENHANCE to solve business-related inquiries. Conclusion, Limitations, and Future Research are provided in Section 4.16 and 4.17.

4.2. Decision Support System Overview

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 & Pervan, 2008). DSS provide decision makers with analytical capabilities and timely information to improve decision making (Power, 2009). Specifically, DSS (Power, 2009):

- 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 DSSResources.com: "An interactive, computer-based system 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".

DSS have been evolving from the 1970's and starting in the early 1990's four powerful technological developments combined to form the foundations for building DSS (Shim et al., 2002):

- The data warehouse
- On-line analytical processing (OLAP)
- Data mining
- The technology associated with the World Wide Web

Figure 1 depicts what is considered a typical model of the decision-making process in a DSS environment (Shim et al., 2002). The emphasis is on model development and problem analysis. Initially a problem is identified and/or recognized and then defined in a manner that supports the creation of models. Alternative solutions are developed and then models are developed to produce the analyses. The best choice resulting from the analyses is then identified and implemented. In actual usage the phases may overlap, comingle, and loop back to earlier stages in an iterative process as more information arises about the problem, or as proposed solutions fail (Shim et al., 2002).

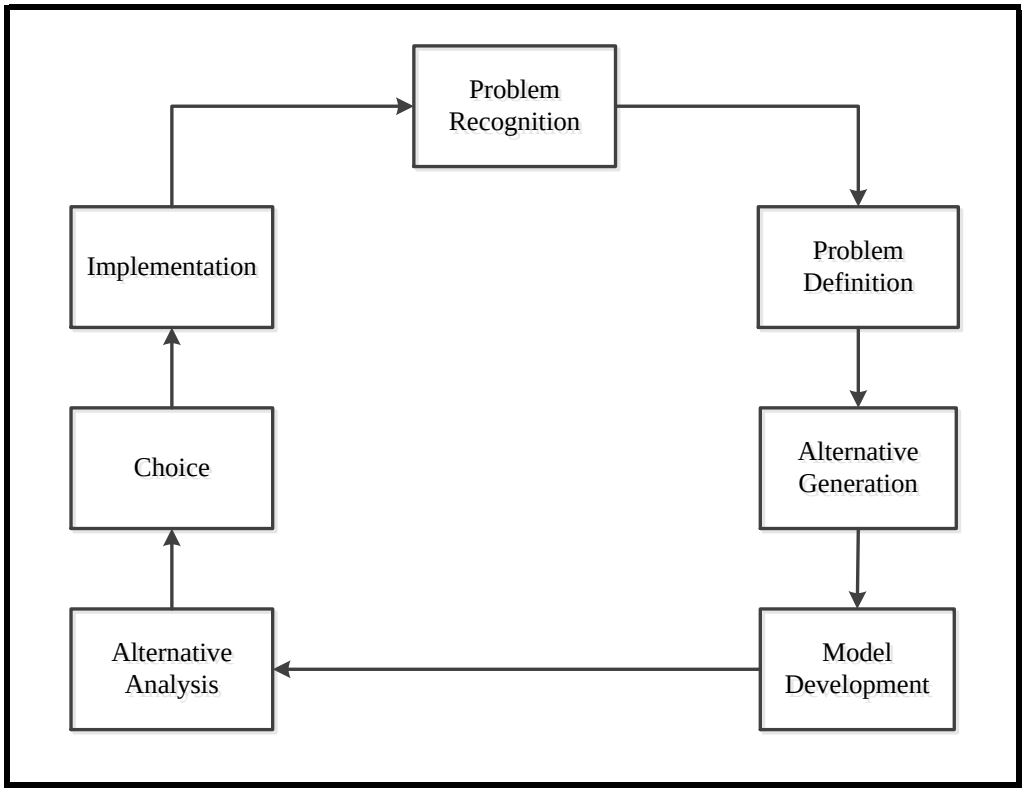

Figure 1: Decision-Making Process Model

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. With respect to the procedure the ENHANCE framework will support for user-requested data analytics, Figure 2 depicts a modification of Figure 1 for ENHANCE, based on the design criteria for ENHANCE as presented in this paper. The final action, the decision step, 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. In the first step the user initiates a request by accessing ENHANCE, and subsequently provides a detailed request based on feedback received from ENHANCE. Interacting with ENHANCE, the user selects the appropriate analytic apps and requests execution. ENHANCE presents the analytic results and the user determines next steps based on them.

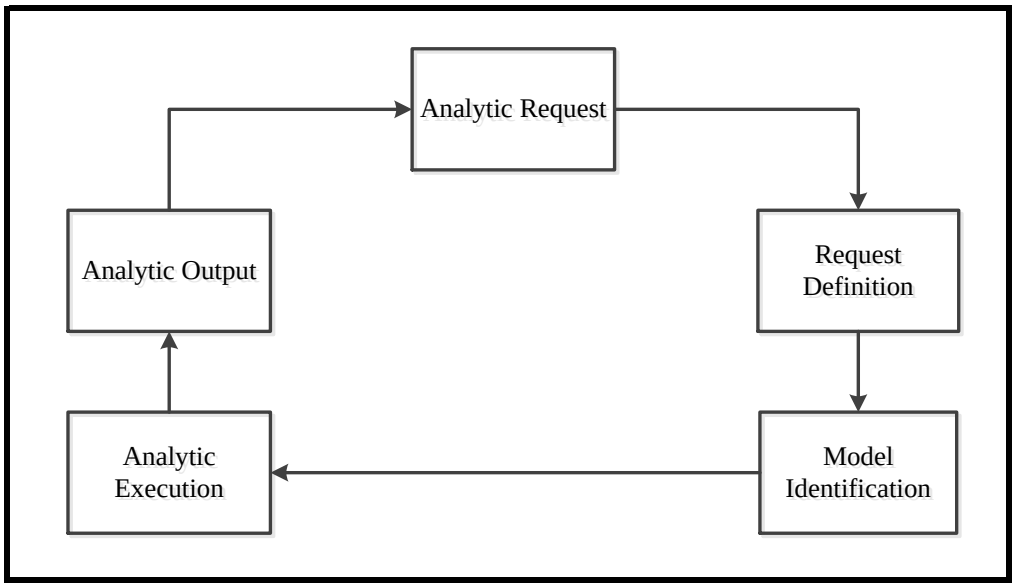

Figure 2: ENHANCE analytic procedure

To further understand the capabilities and applications of DSS, Arnott and Pervan (2008) describe ‘sub-fields’ of DSS:

- Personal Decision Support Systems (PDSS) are small-scale systems that are developed for a single manager or a small number of independent managers
- Group Support Systems (GSS) use a combination of communication and DSS technologies to support groups of people involved in decision making
- Negotiation Support Systems (NSS) supports groups undertaking negotiations between opposing parties
- Intelligent Decision Support Systems (IDSS) apply artificial intelligence techniques to decision support
- Knowledge Management-Based DSS (KMDSS) enhances decision making by supporting knowledge storage, retrieval, transfer, and application
- Data Warehousing (DW) systems provide the large-scale data infrastructure for decision support
- Enterprise Reporting and Analysis Systems are enterprise-focused DSS that include executive information systems (EIS), business intelligence (BI), and corporate performance management systems (CPM)

The ENHANCE framework, as it is intended for individual usage, falls within the first group, that of PDSS.

Sprague (1980) identifies the components, or subsystems, of a typical DSS:

- The data subsystem represents the data warehouse component
- The model subsystem corresponds to Online Analytical Processing (OLAP), knowledge discovery, and data mining tools

- The user interface subsystem enabling communication between the other subsystems and the decision-maker

The ENHANCE framework includes the analytics and the user interface, but the data is sourced from governmental websites and as such ENHANCE is not intended to provide data warehousing capabilities itself.

Characteristics and abilities of a typical DSS are provided by Power (2009):

- Facilitate and support decision making activities
- Interact with decision makers, or the users who control the interactions
- Provide ancillary support to decision makers but do not replace decision makers
- Intended for repeated use either on a routine basis or ad hoc
- Task-oriented in supporting the following: data analysis, identification and/or design of alternatives, a choice among alternatives, and implementation of the decision
- Identifiable system itself, or a specified subsystem of a larger, integrated IS
- Impacts the decision by improving the accuracy, timeliness, quality, and effectiveness of decision (or decisions)

By design the ENHANCE framework provides all the characteristics and abilities described above.

Power (2009) also describes the types of DSS based on the dominant architecture component:

- Communications-driven emphasize communicating, collaborating, and a shared decision-making support
- Data-driven emphasize access to and manipulation of data that can be internal to the organization, externally sourced, and/or data that is real-time in nature
- Document-driven is focused on document retrieval capabilities
- Knowledge-driven supply problem-solving expertise
- Model-driven provide access to and manipulation of quantitative models

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 apps for the user to select from.

The features of a data-driven DSS include (Power, 2009):

- Ad-hoc data filtering and retrieval where the filtering may offer drop down menus, pre-defined queries, and drill-downs from summary to detailed information
- Alerts and triggers when the underlying data changes
- Creation of data displays where the user can choose among scatter diagrams, bar graphs, and pie charts
- Data management and summarization that provides users the ability to view and/or create pivot tables, request custom aggregations, extract/download data
- View predefined data displays such as dashboards and scorecards

The ENHANCE framework incorporates most of the features described above; for this present research the use of alerts and triggers is not envisioned. Also, the exact

capabilities to be provided by the response app, which provides the results of the analytics to the user in a comprehendible manner, have not been defined and are outside the scope of this paper.

Having presented information detailing the definition, types, and capabilities of DSS, among others, a review of DSS research, and recommended improvements, is presented by March and Smith (1995) in order to provide guidance to this present research. March and Smith (1995) discuss what they consider a long-term issue in IS research: the 'tension' between academic rigor and professional relevance. They find that the emphasis of IS research has been on achieving rigor, which they describe as appropriate for a new discipline, but they quote Benbasat and Zmud (1999) who argue that the IS discipline is now relatively mature and it 'can afford to shift attention to relevance without undue concern about being criticized by others'. Although these findings were documented and published a number of years ago, this present research follows this line of thinking in presenting the ENHANCE framework. The ENHANCE framework provides a tool that is feasible to develop and highly relevant in the current environment that includes the availability of governmental open data portals and constituent demands for greater transparency and accountability over their government's financial activities.

4.3. The ENHANCE Framework

The purpose of this research is to present an ENHanced ANalytic Constituent Environment (ENHANCE) framework, facilitated by the availability of open government data, that fulfills the reporting and analytic requirements of the various governmental stakeholders, such as citizens, analysts, bond investors, creditors, vendors, auditors, and

oversight officials[25]. The first step in introducing this analytic capability is the availability of open governmental data, presented in a standardized and usable format. The second step is that of designing a series of analytic apps that can provide meaningful information to the entity's stakeholders. The third step is the development of the ENHANCE framework where apps can function, acting upon the standardized government data to support constituent reporting requirements. Data analytics as presented in this paper include apps 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

Future research can extend the capabilities of the ENHANCE framework by including:

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

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

The ENHANCE framework acts upon inputs from external agents to initiate robust analyses. The primary external agents include the user that is requesting the analytics, and the particular governmental dataset that will be analyzed. Upon receiving the user request for access to a particular dataset ENHANCE undertakes an initial analysis of the attributes and responds to the user. This allows the user to identify the attributes of interest for the analysis to ENHANCE. Based on this response ENHANCE provides a list of possible analytics that can be launched over the selected dataset. This list is based on the analytic apps accessible to the ENHANCE framework. The user responds with their selection of apps, requested output format, and applicable criteria for the app, such as fiscal years and periods under investigation. Once this information is transmitted to ENHANCE the apps can be launched and the resulting information formatted and provided to the user.

The ENHANCE framework incorporates a variety of apps that support its operation. The initial app, a structural app, performs an analysis of the structure and attributes of the user-requested dataset and provides that information to the user. The structural app also includes a search capability that identifies the specific website where the requested data resides, based on a description of the governmental entity, such as city or state name, as provided by the user. An app recommender system provides the user a list of relevant analytic apps based on user responses. The user's responses with selected apps, data selection criteria, and requested formatting are logged into the execution app which then launches the selected analytic apps. The analytic apps extract the appropriate data from the dataset and complete the requested analytics. The analytic apps pass their results to

the response app which formats the results into the user-selected formats and presents the results to the user. This process is presented diagrammatically in Figure 3.

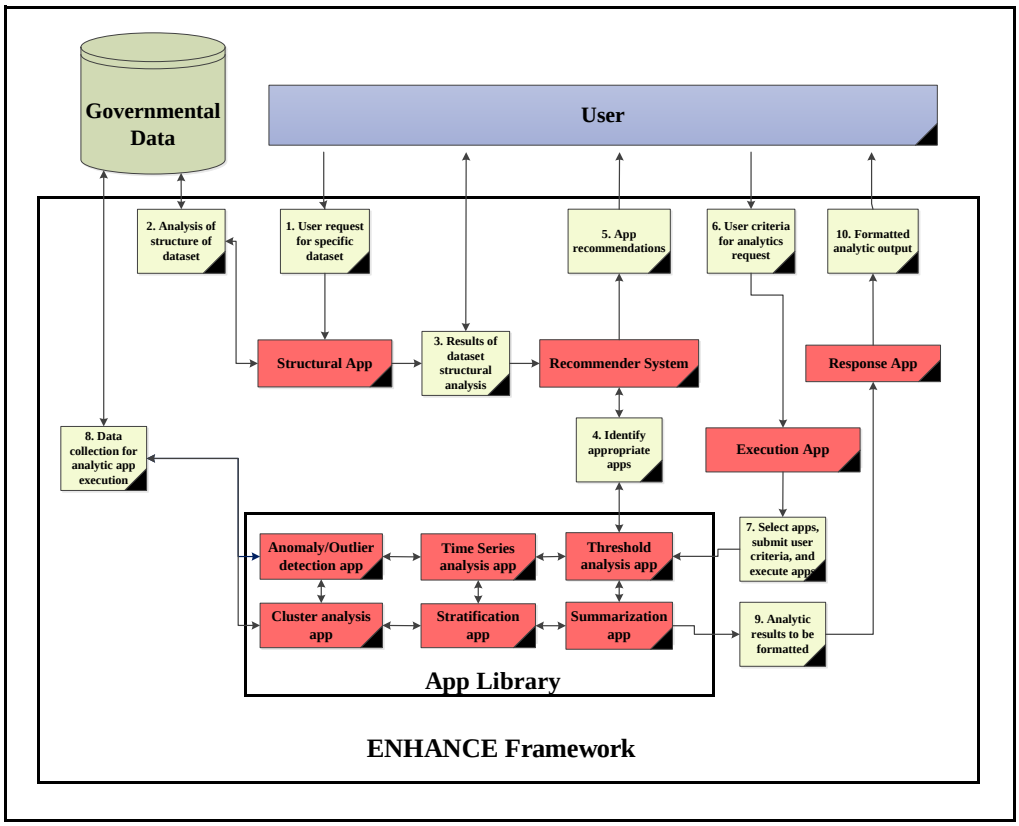

Figure 3: ENHANCE framework

4.4. Design Science Research Methodology to support the design of ENHANCE

The detailed design of the ENHANCE framework is outside the scope of this paper, but when that activity is undertaken, an appropriate tool to drive that development effort should be chosen.

Design science focuses on problem solving through the creation of artifacts that address either an unsolved problem in a unique or innovative manner, or a solved problem in a

more effective or efficient way (Geerts & O'Leary, 2014). A discussion of design science concepts is presented that can be used for the development of the ENHANCE framework.

Peffers et al. (2007) assert that there is a need for a generalized template to produce and present design science research. They suggest a template referred to as Design Science Research Methodology (DSRM), which consists of a sequence of six activities:

- Problem identification and motivation
- Definition of the objectives of a solution
- Design and development
- Demonstration
- Evaluation
- Communication

Table 7 summarizes the application of the DSRM template to include a proposed definition for each activity as appropriate for the development the ENHANCE framework, and the proposed knowledge tools as appropriate for each step (Geerts & O'Leary, 2014).

Development of a complete DSRM template is outside the scope of this present research; for this paper only the first step was undertaken. The Definition of the Problem Identification step is derived from the design concepts for ENHANCE as presented in this paper. The Knowledge Base relies on the literature review conducted for this paper.

The remaining five steps are shown in italics to indicate the information presented is not definitive but is only a proposal of what may be appropriate for each of the remaining

steps. The Definition of Objectives step requires application expertise to document the expected user interactions required to initiate analytics as well as technical expertise to translate these into technical terms. The Design and Development step requires technical expertise to undertake actual design documents. The Demonstration step will include development of the ENHANCE framework prototype and it is recommended this be presented to a focus group of target users for their review and comment. The final step, Communication, should include presentation of the completed system to a sample of target users and data providers for their review, testing, and approval.

DSRM Activity	**Definition**	**Knowledge Base**
Problem identification and motivation	Open government data initiatives provide constituents access to expenditure data but generally provide no or very limited analytic capabilities. A tool needs to be developed that can identify appropriate analytics to satisfy constituent requirements for relevant information from government data. The tool must access and format the selected government data in a manner that is acceptable to the chosen analytic apps. The tool must provide the results of the analytics in an understandable format for the user to comprehend.	Literature review of the government open data initiatives, what data transparency means and how to provide those capabilities to constituents, and the current state of data analytic capabilities
Definition of the objectives of a solution	*The ENHANCE framework needs to manage the user interaction process and from user inputs identify appropriate analytic apps, execute those apps over the user-selected dataset, and produce analytic results to*	*Develop an understanding of what user inputs into the ENHANCE framework are required to deliver appropriate analytics, and how the tool can translate user requests*

	satisfy the user's requirements	*into the technical activities required to deliver the analytic output, by means of detailed technical analyses and design documents*
Design and development	*Design a framework that can accommodate user requests for analytics and manage the underlying components*	*A thorough comprehension of the design of sophisticated IT systems*
Demonstration	*Definition of the specific logic to be incorporated into the ENHANCE framework in order to satisfy user requirements*	*Develop a prototype of the ENHANCE framework*
Evaluation	*Evaluation of the success of the ENHANCE framework in providing appropriate analytic apps, executing those apps, and providing appropriate output to satisfy the user's requirements*	*Compare the design documents for ENHANCE to the results of the testing of the prototype, and present to target user group*
Communication	*Present to appropriate audiences: data providers, ENHANCE users, and potential users*	*Data providers and anticipated users*

Table 7: Proposed DSRM template to guide development of the ENHANCE framework

Following a proven design methodology such as DSRM will facilitate the most appropriate design for the ENHANCE framework. Development of a complete DSRM template and the actual ENHANCE framework is an activity for future research as the information presented here is only an example.

4.5. Data Portals

4.5.1. Ohio Checkbook Portal

The Ohio Checkbook data portal shown in Figure 4 (http://www.ohiocheckbook.gov) incorporates a number of investigative and analytic tools to assist users in analyzing the available data. Upon accessing the portal two diagrams are provided: the first is a pie chart depicting 2014 (the most current full year) expenditures classified by 'expense type categories'. The diagram can be printed to PDF. The classification can be changed to reflect expenditures by 'expense type' (expense description), 'expense code' (numeric code for the particular expense), 'agency' (agency name), or 'agency type' (numeric code for the particular agency).

Figure 4: Ohio Checkbook data portal

The second diagram is a bar chart representation of state spending by fiscal year for the fiscal years 2008-2014, which can also be printed to PDF. The presentation can also be changed to that of a stacked bar chart or pie chart.

The primary portal page also provides two listings: the first is a listing of expense types for 2014 (most current full year available) from greatest amount expended to least and the second is a listing of highest paid companies (vendors) for 2014, again listed from greatest amount to least. These listings can be exported to CSV.

There are a number of tools available on the Ohio data portal to support user-defined analytics, and primary among them is the Compare tool. The Compare tool provides a spending comparison between agencies, the specific use of the funds, and how spending changes over time. Up to ten different items can be compared. The first selection allows for a multiple fiscal year selection for comparison. The second comparison option allows for selection by agency type, or specific agency. The third available comparison is by expense at the level of expense category, expense type, or expense code (detailed). In addition to these three filters there are advanced selection filters that include fund, program, appropriation line item, and again expense type.

The portal also provides for 16 predefined popular searches over financial and non-financial data. Financial data-related searches that may be of interest include the categories of travel, meals, capital items, payroll expenses, state debt payments, and office equipment.

A glossary of terms is provided to assist users in understanding the attributes of the data presented. The glossary includes acronyms, agency names, expense type category descriptions, and fiscal terms (what are appropriations, etc.).

The portal also connects to state budget information as provided by the Ohio Office of Management and Budget. Budget data is provided in PDF format for both operating

budgets for fiscal years 2006/2007 through 2016/2017 and capital budgets for fiscal years 2001/2002 through 2015/2016 (as of the time of preparation of this paper). For the operating budget, by fiscal year, both the Enacted Appropriations and Enacted Budget Bills are provided separately for the Main Operating budget, Transportation budget, and Ohio Industrial Commission and Bureau of Workers' Compensation budgets. Details provided include, using fiscal years 2014/2015 for example: FY 2012 budget, FY 2013 estimate, FY 2014 appropriation, changes between FY 2013 and 2014 in dollars and percent, FY 2015 appropriation, and changes between FY 2014 and 2015 in dollars and percent. Capital budget information includes capital appropriations by fund, agency, fund and agency, agency and fund, and also a new debt authorization document. The documents are at summary level, showing simply a total amount for each agency, fund, and so forth.

A link from the portal to the state procurement website provides for a review of contract information. There are search capabilities by contract name, contract type, or by a keyword search. Additional search criteria include a commodity category, date ranges, and program (open market vs. MBE, etc.). The resulting contracts list includes contract title, contract type, market type, contract number, effective and expiration dates, and vendor. The listing can be printed. Drill-down to a particular contract displays contract type, status, number, commodity categories, vendor name and vendor contract ID, effective dates, procurement program information, and dollar amount. This display can be screen-printed. There are also associated PDF files available that list bid tabulation and terms and conditions.

Detailed transaction-level expenditure data is accessible in the Ohio checkbook. The Ohio checkbook data includes the following search and filter capabilities as listed in Table 8.

Primary filter:	**Item:**	**Element:**
Fiscal year	Fiscal year	
Basic filters:		
Agency:	Type of Agency (summary level):	Agency, Boards and Commissions, Elected Officials, Higher Education
	Entity (detailed level – 144 entities in total):	Example: Accountancy Board of Ohio
Expense Type:	Expense Category (Broadest description - 51 categories):	Examples: Capital, Debt Service, Equipment
	Expense Type (231 types):	Examples: Accounting/Auditing, Advertising
	Expense Code (most specific description – 888 codes):	Examples: Advertising-legal, Ammunition, Taxes
Advanced Payment Filters:		
	Fund (811 funds):	Examples: Building Management, Higher Education Improvement
	Program (1,578 programs):	Examples: Criminal Investigation, Health & Safety
	Appropriation Line Item (2,083 line items):	Examples: Attorney Services, Health Care Programs

Table 8: Ohio Checkbook Data Filters

The transaction-level data found in the Ohio checkbook can be exported in CSV format and include the following ten attributes as listed in Table 9.

Attribute:	**Description:**
Checkbook Unique Identifier (9-digit numeric)	Unique identifier for the journal entry line item (system-generated?)
Transaction Date	Date of the transaction
Vendor Name	Name of the recipient of the

	payment
Voucher Journal ID	An agency-specific eight-digit number that records an authorized transaction with a vendor. There may be multiple transactions (vouchers) grouped into a single vendor payment (check).Unique identifier for each journal entry (Numeric and alphanumeric)
Check Number	Unique identifier for each payment (Numeric and alphanumeric)
Amount	Amount
Entity Code (alphanumeric)	Code of the entity for which the expenditure was made
Business Unit Code (alphanumeric)	Code for the Business Unit (or detailed department) within the entity for which the expenditure was made
Fiscal Year	Fiscal year in which the expenditure was made
Entity Name	Name of the entity for which the expenditure was made

Table 9: Ohio Checkbook Data Attributes

The analytic capabilities that Ohio supplies are commendable, but limited in functionality compared to those provided by the ENHANCE framework.

4.5.2. Austin Data Portal/Checkbook

The city of Austin, TX provides a robust open data portal (https://data.austintexas.gov/), (Figure 5) supported by one of the more popular portal providers[26]. The portal provides access to video tutorials on using the portal as well as a user-generated wiki to provide additional how-to information. There are also links to learn more about Austin's open data initiative and for a user to submit a suggestion for datasets not currently available.

[26] Powered by Socrata

The portal provides a dataset search and browse capability with the datasets listed in order of popularity.

A task bar provides a number of options for users. There is a list of View Types that includes the Socrata Data Lens that provides data visualizations, a list of datasets (as noted above), Charts, Maps, Calendars, Filtered Views (datasets with popular filters applied), External Datasets (such as U.S. Census data), Files and Documents (such as EMS incident counts), and Forms (not populated at this time). There several categories of filters that delimit the particular View Type selected: by City Department, by Category (such as Financial or Public Safety), and by Topic (key word search).

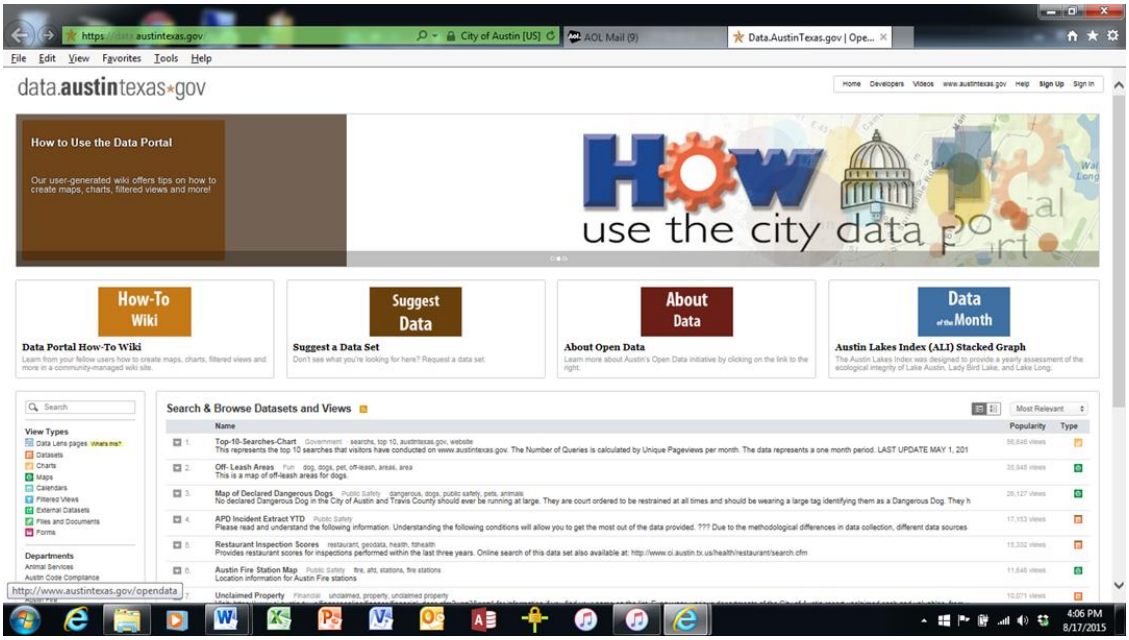

Figure 5: Austin Data Portal

In addition to the export capabilities available with the City of Austin checkbook data, the portal provider has made available the Socrata Open Data API (SODA) that include 'Official' libraries and Software Development Kits (SDK) that have been developed by Socrata and are fully supported by them. They include: Google Android, DataSyn SDK

(Java), Apple IOS, Java, PHP, Ruby, Scala, and Swift. There are also 'Community' libraries, as developed by the SODA developer community that are provided within SODA and include PHP, .NET, Julia, Python, and R. The technical expertise required to fully exploit the capabilities available with API's may be greater than that held by many if not most constituents. For the general public these API's may not be the answer to providing information beyond simple and straightforward data extracts. Analytic tools such as ENHANCE that can strike a balance between providing robust analytics while not requiring extensive technical expertise, can bridge the gap.

One of the datasets included is that of the city's checkbook data (https://www.austintexas.gov/financeonline/checkbook/index.cfm). The data can be extracted in numerous formats: CSV, CSV for Excel, JSON, PDF, RFD, RSS, XLSA, XLSX, and XML. The dataset include 33 attributes for each detailed record, with filters provided for each attribute. The data attributes are listed in Table 10.

Attribute:	**Description:**
Fiscal Year	Fiscal year in which the expenditure was made (YYYY)
Fiscal Period Number	Period within the fiscal year in which the expenditure was made
Department Code	Code for the department (entity) for which the expenditure was made
Department Name	Name of the department (entity) for which the expenditure was made
Fund Code	Code for the fund from which the expenditure was made
Fund Name	Name of the fund from which the expenditure was made
Division Code	Code for the division (within the department (entity)) for which the expenditure was made
Division Name	Name of the division (within the department (entity)) for which the expenditure was made
Program Code	Code for the program (within the division) for

	which the expenditure was made
Program Name	Name of the program (within the division) for which the expenditure was made
Activity Code	Code for the activity (within the program) for which the expenditure was made
Activity Name	Name of the activity (within the program) for which the expenditure was made
G/L Object Account Number	G/L financial account identifier
G/L Object Account Name	G/L financial account name
Vendor/Customer Legal Name	Vendor name
Vendor/Customer Code	Vendor identifier (alphanumeric)
Vendor/Customer Indicator (V)	Vendor indicator (V)
Referenced Document Code	Indicates type of source document for journal entry
Referenced Document Department Code	Code for the department responsible for source document
Reference Document ID Number	Unique identifier for source document
Commodity Code	Code for the good or service acquired
Commodity Description	Description for the good or service acquired
Check/EFT Issue Date	Date payment was issued (if payment completed)
Check Status (Paid)	'Paid' status (if payment completed)
Vendor Line Number	Line number of vendor document
Commodity Line Number	Line number of commodity document
Accounting Line Number	Line number of accounting document
Document Type Code	Journal Entry type code
Document Department Code (Number)	Department number
Document ID Number	Unique identifier for journal entry
Accounting Line Description	Description of good or service acquired
Amount	Amount
Calendar Year	Calendar year in which journal entry is recorded (YYYY)
Calendar Month	Calendar month is which journal entry is recorded (MM)

Table 10: Austin Checkbook data

As noted earlier, a greater number of attributes available in a particular dataset implies a greater potential for sophisticated and meaningful analyses. In addition to basic expenditure data such as amount, vendor, date, and department charged to (business unit) as provided by Ohio, Austin also provides funding source, division, program, and activity

attributes that allow for summarizations and other analyses and several levels of organizational rollup level, as well as comparisons at these levels.

4.6. Budget Data for Analytics

In addition to undertaking analyses over actual expenditure data, which is the focus of this paper with respect to the vision of the ENHANCE framework, another area of interest, and potential analytic requests, to constituents is that of the composition of governmental budget data as essentially all governmental financial activities are controlled and limited by an approved and authorized budget (typically prepared on a fiscal year basis). In addition to preparing analyses based on financial budget data, relevant governmental datasets for which analyses over budget data may be desirable include: Education, Transportation, and Healthcare. When analyzing governmental budget data the following terminologies listed in Table 11 are useful in understanding a budget from a particular governmental entity, and identifying where the funds are proposed to be disbursed to (GAO).

A presentation of analytics over budget data is not included in this research, but the development of analytics over budget data, and especially the development of analytics that can match budget data with actual expenditures, is an area that should be considered in future research.

Budget information:	Definition:
Anticipated resources	Future Revenue
Appropriations realized	Current Spending
Borrowing authority	A budget authority (BA) is enacted to permit an agency to borrow money and then to obligate against amounts borrowed. It may be definite or indefinite

	in nature.
Contract authority	These are obligations in advance of the appropriation. In this case the budget authority permits an agency to incur obligations in advance of appropriations, including collections sufficient to liquidate the obligation or receipts. The contract authority is unfunded, and a subsequent appropriation or offsetting collection is needed to liquidate the obligations.
Total authority	Includes all BA + BA from another appropriation + unobligated balances of the BA from the previous year that is still available for obligation
Un-apportioned authority	Unobligated resources not yet apportioned, and not available to obligate.
Apportionments	A distribution of amounts available for obligation, including budgetary reserves established in an appropriation or fund account. An apportionment divides amounts available for obligation by specific time periods (usually quarters), activities, projects, objects, or a combination thereof. The amounts so apportioned limit the amount of obligations that may be incurred. An apportionment may be further subdivided by an agency into allotments, sub-allotments, and allocations
Appropriations	A budget authority that can incur obligations and make payments for specified purposes. An appropriation act is the most common means of providing appropriations; however, authorizing and other legislation itself may provide for appropriations.
Allotments	The ability within an agency to incur an obligation within a specified amount.
Available authority	The amount available for the specified authority.
Budget details:	
Budget authority	The authority to enter into financial obligations that will result in immediate or future outlays. The basic forms of budget authority include (1) appropriations, (2)

	borrowing authority, (3) contract authority, and (4) authority to obligate and expend offsetting receipts and collections.
Percentage available authority at appropriation, apportionment, or allotment levels	Percentage of budget available at specified level
Budget spending	Amount spent for the particular budget/item
Commitments, obligations, expenditures	An administrative reservation of allotted funds, or of other funds, in anticipation of their obligation.
Total spending	Total spending for the specified budget
Percentage spending to Authority, at appropriation, apportionment, or allotment levels	Percent of total budget allocated to specified level
Re-appropriation	Legislation permitting an agency to obligate, whether for the same or different purposes, all or part of the unobligated resources.

Table 11: Example of budget data information

4.6.1. Incorporating Budget Data in Expenditure Analytics

The availability of budget data adds a new perspective to governmental data analyses. Expenditure data represents what has transpired, and although analyses can compare expenditures over time to identify trends and patterns, the inclusion of budget data can provide insights into the intent of governmental spending by those who developed the budgets prior to the occurrence of the actual expenditures.

4.7. Analytics Development

As noted with the Ohio and Austin, TX data portals described above, data availability is generally not an issue, but the availability of robust analytics that can provide meaningful results over that data has yet to be addressed. Without analytic capabilities the value of

the available data is limited. ENHANCE is intended to place that capability into the hands of the typical user of government data.

Academic research presents analytic apps that are appropriate to incorporate into the apps library that supports the ENHANCE framework, for example Issa's (2013) 'Exceptional Exceptions' app, and Byrnes (2015) data clustering techniques. Data analytics included in this present research have also been drawn from research by Provost and Fawcett (2013), and include:

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

Examples of the use of these apps over expenditure data as provided by an existing, municipal open data portal (Austin, TX) are presented in the following section.

4.8. Illustration of Data Analytic Apps

The initial technical conceptualization of ENHANCE begins with the identification of appropriate analytic apps which can provide the analytic capability. To illustrate the functionality incorporated into ENHANCE, the analytics presented here were undertaken using a specific, commercially available data analytics software product[27] (unless specifically noted otherwise). The analytic capabilities presented are not available in the Austin data portal and not typically available in other governmental data portals.

[27] CaseWare IDEA, version 9.2.0.630 (x86)

The data used for this tested was extracted from the City of Austin, TX data portal (https://data.austintexas.gov/) and is specifically from the checkbook, or expenditure, dataset. Data was extracted for fiscal years 2009 through 3rd quarter 2015, all departments, and the expense code for 'Books-Library' (code = 7486). The data was extracted into a Microsoft Excel format. A sample of the data extracted is displayed in Table 12. This data has been summarized for display purposes and does not include all 33 attributes available.

4.8.1. Expenditure Data

An extraction app as designed to function within ENHANCE can benefit user analysis as logic is included that truncates the data to exclude extraneous attributes that are not of interest to the user and thus enhance readability of the raw data. Once the user has identified to ENHANCE the dataset of interest and ENHANCE has responded with a list of data attributes in the particular dataset, the user can select those attributes of interest for further analytic purposes. In the example in Table 12 attributes for date, amount, vendor, payment, and organizational rollup levels have been maintained while other attributes have been masked or removed to simplify the presentation.

FY	Period	Department	Program	Activity	Expense	Vendor	Check Issued Date	Amount
2015	10	Library	Materials Management Services	Collection Support	Books-library	INGRAM LIBRARY SERVICES INC	7/2/2015	$13,181.86
2015	10	Library	Library Facility Impvs	Central Library Gen Fund	Books-library	INGRAM LIBRARY SERVICES INC	7/2/2015	$2,331.17
2015	10	Law	Opinions and Advice	Land Use and Real Estate	Books-library	WEST PUBLISHING CORPORATION	7/2/2015	$1,405.52
2015	10	Law	Advocacy and Dispute Resolution	General Litigation/Affirmative Action	Books-library	WEST PUBLISHING CORPORATION	7/2/2015	$1,405.53
2015	10	Law	Opinions and Advice	Municipal Operations	Books-library	WEST PUBLISHING CORPORATION	7/2/2015	$1,405.53
2015	10	Law	Opinions and Advice	General Counsel	Books-library	WEST PUBLISHING CORPORATION	7/2/2015	$1,405.53
2015	10	Library	Public Services	Youth Services	Books-library	INGRAM LIBRARY SERVICES INC	7/1/2015	$81.44
2015	9	Library	Materials Management Services	Collection Support	Books-library	OverDrive, Inc.	6/30/2015	$13,570.64
2015	9	Library	Library Facility Impvs	Central Library Gen Fund	Books-library	OverDrive, Inc.	6/30/2015	$13,392.25
TOTAL AMOUNT:								$48,179.47

Table 12: Example of Austin Checkbook Data for analytics (summarized for display purposes and does not include all attributes)

4.8.2. Descriptive Statistics

An analytic app can provide for a statistical analysis to be completed on appropriate data attributes in the dataset and Table 13 includes results from a statistical analysis of the Amount attribute in the dataset. This information can provide insights to the user as regarding the overall dataset such as: the total value of the expenditures, the number of total positive amount records and if any, negative amount records (for possible anomalous behavior), the largest positive and negative amounts, the number of zero amount records, and average amount value. A user can compare this statistical information for one specific entity, department, time period, and/or expenditure type to similar information for a comparable entity, department, or other criteria. At a very high level a user can contrast items such as minimum and maximum payment amounts (and

the payment range), average payment amounts, and number of payments between selected entities.

Statistic	Amount Field
Net Value	20,575,868.39
Absolute Value	20,577,010.73
# of Records	5,661
# of Zero Items	0
Positive Value	20,576,439.56
Negative Value	-571.17
# of Positive Records	5,655
# of Negative Records	6
# of Data Errors	0
# of Valid Values	5,661
Average Value	3,632.75
Minimum Value	-333.96
Maximum Value	136,269.09
Median Value	651.21
Mode Value	155.00
Record # of Minimum	4,206
Record # of Maximum	183
Sample Standard Deviation	8,809.16
Sample Variance	77,601,239.33
Population Standard Deviation	8,809.16
Population Variance	77,601,239.33
Population Skewness	5.402562
Population Kurtosis	44.477314

Table 13: Statistics generated for the Amount field

4.8.3. Data Summarization

The term summarization refers to presenting only the most important points, or describing just the key events, from the underlying information source

(www.vocabulary.com/dictionary/summarize). From a research perspective data summarization refers to the process of grouping data in order to transform similar items into more abstract conceptual representations (Yager & Petry, 2006). When dealing with large amounts of data, such as that referred to as big data, summarization is an important technique to assist in knowledge discovery, that is, the ability to extract useful information from large databases (Yager & Petry, 2006). Data summarization is a very sophisticated procedure that incorporates the techniques of generalization and compression as well as the use of a formal ontology that specifically describes the hierarchical summarization categories to be used (Yager & Petry, 2006). When undertaking data mining over dynamically changing data additional techniques must be applied in order to reflect an accurate summarization that takes into consideration the changing nature of the data, ideally using incremental techniques and not a complete reapplication of the summarization protocol (Nassar et al., 2004).

A summarization analytic app can be launched to summarize the dataset into a number of more easily comprehendible subsets. The expenditure data, for the expense category 'Books-Library', is summarized to the program level within each department, as displayed in Table 14. The results as illustrated have been sorted from highest expenditure amount to lowest. Given the data attributes selected, comparisons of spending activity by program within a department as well as for similar programs in different departments can be undertaken. The user can now easily comprehend and compare transactional volumes between entities. In addition to comparing transactional amounts between entities the transactional counts may indicate a situation where very high transactional volumes may overload the supporting administrative function. This

can indicate that one entity has embraced more efficient back office operations than another.

Fiscal Year	Dept.	Department Name	Program	# Records	Amount - Summarized
2014	85	Library	Materials Management Services	367	2,813,101.78
2014	85	Library	Library Facility Impvs	177	973,825.56
2014	41	Mayor & Council	Mayor/Council	2	97,998.19
2014	85	Library	Miscellaneous	47	56,990.82
2014	57	Law	Opinions and Advice	58	50,268.30
2014	57	Law	Advocacy and Dispute Resolution	35	35,456.96
2014	93	Emergency Medical Services	Professional Practice and Standards	72	22,300.59
2014	22	Austin Water Utility	Support Services	4	11,076.36
2014	58	Human Resources	Human Resources Management Services	10	8,969.36
TOTAL AMOUNT:					4,069,987.92

Table 14: Example of Data Summarization by Fiscal Year, Department, and Program

4.8.4. Time Series Analysis

A time series is described as an ordered sequence of values of a variable at equally spaced time intervals (NIST). Time series analyses are used to (NIST):

- Understand the underlying actions that produced the observed data
- Develop a model that represents the underlying actions and can be employed for forecasting or monitoring

Time series analyses provide the basis for many applications, including (NIST):

- Forecasting: economic and/or sales
- Analyses: budget, census, and/or stock market
- Operational analytics:
    - Yield Projections

- Process and Quality Control
- Inventory Studies
- Workload Projections
- Utility Studies

Time series analyses are an additional form of data summarization that can provide insights into spending trends and/or patterns, in this case, with the data presented in a matrix format. Time series analytics may be used when pattern analysis is not sufficient, for example, when the data exhibits characteristics such as 'systematic nonrandom patterns' (Alwan & Roberts, 1988). Table 15 displays the results for the expense category 'Books-Library' where department is presented in rows, columns represent fiscal year, and the expenditure amount is in the grid area. In this particular example one can identify specific departments where spending patterns vary significantly over time, such as the Mayor and Council department where some years incur no expenditures and others incur significant (almost $100,000) amounts. Other areas of interest can be departments, such as Government Relations, where there are very infrequent expenditures. Such activity can be indicative of fraud. Departmental and/or fiscal year comparisons are easily identifiable in this format.

	Fiscal Year:							
Dept.	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**Total**
Austin Energy	33,112.43	7,920.26	4,697.41	21,487.19	2,386.62	1,880.99	106.49	71,591.39
Austin Water Utility	55,540.76	35,445.70	22,135.18	6,009.76	15,145.95	11,076.36	9,288.82	154,642.53
Transportation		2,471.50	175.42	397.80	170.00	189.03		3,403.75
Mayor & Council			18,631.96		33,184.73	97,998.19		149,814.88
Government Relations						111.25		111.25
Office of Real Estate Services					40.00			40.00
Management Services		95.00	1,057.02			1,200.00		2,352.02
Redevelopment Services	39.41	47.50		11,102.50	2,883.00		329.70	14,402.11
Communications & Technology Management	17,198.12	16.34	112.47	25,804.00				43,130.93
Law	80,757.31	74,355.39	93,809.66	103,472.83	91,541.69	85,725.26	75,795.03	605,457.17
TOTAL AMOUNT:	2,559,068.02	2,269,953.49	2,584,423.56	3,120,609.13	3,303,876.96	4,093,491.03	2,644,446.20	20,575,868.39

Table 15: Time Series analysis of Expenditures (Books-Library) by Department (row) and Fiscal Year (column)

Time series analyses also provide insights into trends in expenditures. Trending is defined as the general direction in which something tends to move, turn sharply, or change direction or orientation abruptly (www.vocabulary.com/dictionary/trend). This next example presents a summarization, at the department level, including period within fiscal year. Trends within a fiscal year, or across fiscal years if selected in the data set, can be identified at this level of display, as shown in Table 16. This particular data can also be sorted by period across fiscal years to search for common spending patterns by period over time. Data trends may be expected, such as higher expenditures for a road maintenance department during winter months, or it can also indicate unexpected patterns that might be indicative of anomalous behavior. Trending over time, especially increasing expenditures within a particular expense category may warrant closer investigation. Increased spending close to the end of the fiscal year may indicate an entity that is following a 'use it or lose' approach to keep entity spending close to their budget. This presentation in Table 16 has been sorted by department and period within

department to indicate spending trends over time, for a selected department (Law), expense category (Books-Library), and fiscal years 2013 and 2014 only.

Fiscal Year	Fiscal Period	Dept.	Dept. Description	# Records	Amount - Summarized
2013	1	57	Law	5	4,393.01
2013	2	57	Law	16	8,539.81
2013	3	57	Law	10	9,074.97
2013	4	57	Law	10	4,812.13
2013	5	57	Law	17	11,095.59
2013	6	57	Law	12	12,481.06
2013	7	57	Law	8	6,967.77
2013	8	57	Law	10	6,245.56
2013	9	57	Law	8	5,374.62
2013	10	57	Law	11	12,918.83
2013	11	57	Law	10	7,845.69
2013	12	57	Law	3	1,792.65
2014	1	57	Law	9	6,235.31
2014	2	57	Law	10	5,659.93
2014	3	57	Law	4	5,090.42
2014	4	57	Law	7	7,852.15
2014	5	57	Law	7	10,224.55
2014	6	57	Law	11	12,208.01
2014	7	57	Law	8	9,410.76
2014	8	57	Law	6	567.10
2014	9	57	Law	12	11,385.78
2014	10	57	Law	10	9,569.75
2014	11	57	Law	7	6,585.90
2014	12	57	Law	2	935.60
TOTAL AMOUNT:					177,266.95

Table 16: Example of Data Summarization by Fiscal Year, Fiscal Period, and Department (For selected department and fiscal years only)

4.8.5. User-defined Expenditure Thresholds

The term threshold represents either a point of departure or transition, or that of a limit or boundary (www.vocabulary.com/dictionary/threshold). An analysis of thresholds typically refers to the identification of the level at which an action or event will be triggered and

undertaken (Granovetter, 1978). The concept of a threshold has been incorporated into the design of research models. In one example models are developed to determine a threshold point in the study of collective behavior. The scenario involves individuals who have a choice of one of two alternatives with the costs and/or benefits of each choice dependent on the number of individuals who choose one of the alternatives (Granovetter, 1978). The threshold in this case represents the number of individuals who must choose a particular alternative before the individual under study will likewise choose that same alternative (Granovetter, 1978). The function of a threshold or thresholds supports additional modeling techniques, such as multivariate threshold models (Tsay, 1998), nonlinear time series models (Tsay, 1989), threshold autoregressive models (Chan & Tong, 1986), and more recently with probabilistic threshold querying (Cheng et al., 2004).

A threshold analytic app allows the user to identify the specific threshold level(s) they consider appropriate for the expense items under investigation. Table 17 displays the results of identifying records with a user-defined threshold amount equal to or greater than $10,000 for a specific department (Library), Fiscal Year (2014), and period (12). Given the attributes selected this data can also be sorted by, among others, program, activity, or vendor.

FY	Period	Department	Program	Activity	Vendor	Check Issued Date	Amount
2014	12	Library	Materials Management Services	Collection Support	INGRAM LIBRARY SERVICES IN	9/24/2014	60,091.72
2014	12	Library	Library Facility Impvs	Central Library Gen Fund	INGRAM LIBRARY SERVICES IN	9/24/2014	33,085.08
2014	12	Library	Library Facility Impvs	Central Library Gen Fund	RECORDED BOOKS INC	9/24/2014	24,892.60
2014	12	Library	Library Facility Impvs	Central Library Gen Fund	RECORDED BOOKS INC	9/19/2014	24,858.60
2014	12	Library	Library Facility Impvs	Central Library Gen Fund	RECORDED BOOKS INC	9/18/2014	63,458.48
2014	12	Library	Materials Management Services	Collection Support	RECORDED BOOKS INC	9/18/2014	28,457.60
2014	12	Library	Materials Management Services	Collection Support	INGRAM LIBRARY SERVICES IN	9/16/2014	22,597.13
2014	12	Library	Library Facility Impvs	Central Library Gen Fund	INGRAM LIBRARY SERVICES IN	9/16/2014	14,351.61
2014	12	Library	Materials Management Services	Collection Support	INGRAM LIBRARY SERVICES IN	9/16/2014	73,843.89
2014	12	Library	Library Facility Impvs	Central Library Gen Fund	INGRAM LIBRARY SERVICES IN	9/16/2014	33,492.04
2014	12	Library	Materials Management Services	Collection Support	RECORDED BOOKS INC	9/15/2014	18,122.86
2014	12	Library	Library Facility Impvs	Central Library Gen Fund	RECORDED BOOKS INC	9/15/2014	111,649.39
2014	12	Library	Materials Management Services	Collection Support	RECORDED BOOKS INC	9/10/2014	59,270.50
2014	12	Library	Library Facility Impvs	Central Library Gen Fund	RECORDED BOOKS INC	9/9/2014	11,398.16
TOTAL AMOUNT:							579,569.66

Table 17: Example of Single Purchase Amounts over $10,000 (summarized for display purposes and does not include all attributes)

4.8.6. User-defined Expenditure Stratifications

Stratification has been defined as ‘arranging something, or something that has been arranged, into categories’ (www.vocabulary.com/dictionary/stratification). Stratification may be most commonly associated with the categorization of society by any number of attributes. In the context of social research stratification may also be used to divide a sample into sub-groups from which random samples are selected and then these sub-samples are combined to form an overall sample, 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). The process of stratification, another aspect investigated by social researchers, refers to the principles or rules by which the individuals under study are distributed across the stratification (Treiman, 1970). Structure and process are inter-related in that a change in structure necessitates a change in process, as a change in process impacts structure (Treiman, 1970).

In addition to the stratification graph presented in Figures 7 and 8, other forms of presentation can be envisioned, such as a spatial representation of population attributes. Hoyle and Langley (2011) present in Figure 6 a spatial distribution of yellowfin tuna size by location (longitude and latitude). Depending on the attributes under investigation this form of presentation can be undertaken, for example, in analyzing an entity's Public Works inventory (or other), in order to provide a spatial representation of dollar value per physical location.

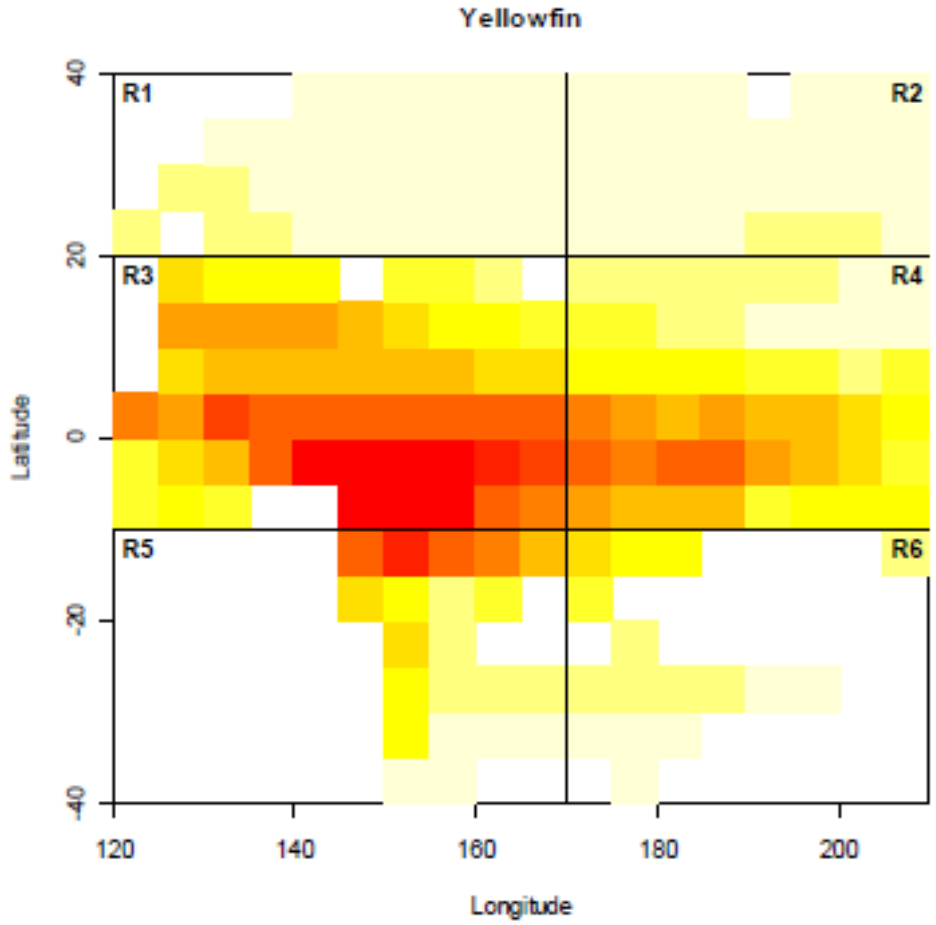

Figure 6: Average long-term (1960-1987) spatial distribution of yellowfin tuna CPUE (number of fish per 100 hooks). Darker colors represent higher CPUE (Hoyle and Langley, 2011)

A stratification app that classifies expenditure data into user-specified stratifications can provide insights into spending patterns in a visually understandable fashion, as displayed in Figures 7 and 8. Figure 7 presents expenditure amounts per stratification and Figure 8 presents the number of expenditure transactions per stratification. Comparing stratifications per department or fiscal year/period can indicate possible spending disparities. With respect to Figure 7 one might expect that the total dollars spent would increase with each increment in expenditure amount, but in this example the third stratification layer exceeds that of the fourth layer in total amount spent, which may merit

further analysis. This can indicate a situation where, due to purchase approval limits, multiple expenditures below the approval threshold occur in order to bypass required approvals. A review of Figure 8 might also confirm this scenario as the number of expenditures for the third stratification is rather large but unusually small for the fourth stratification.

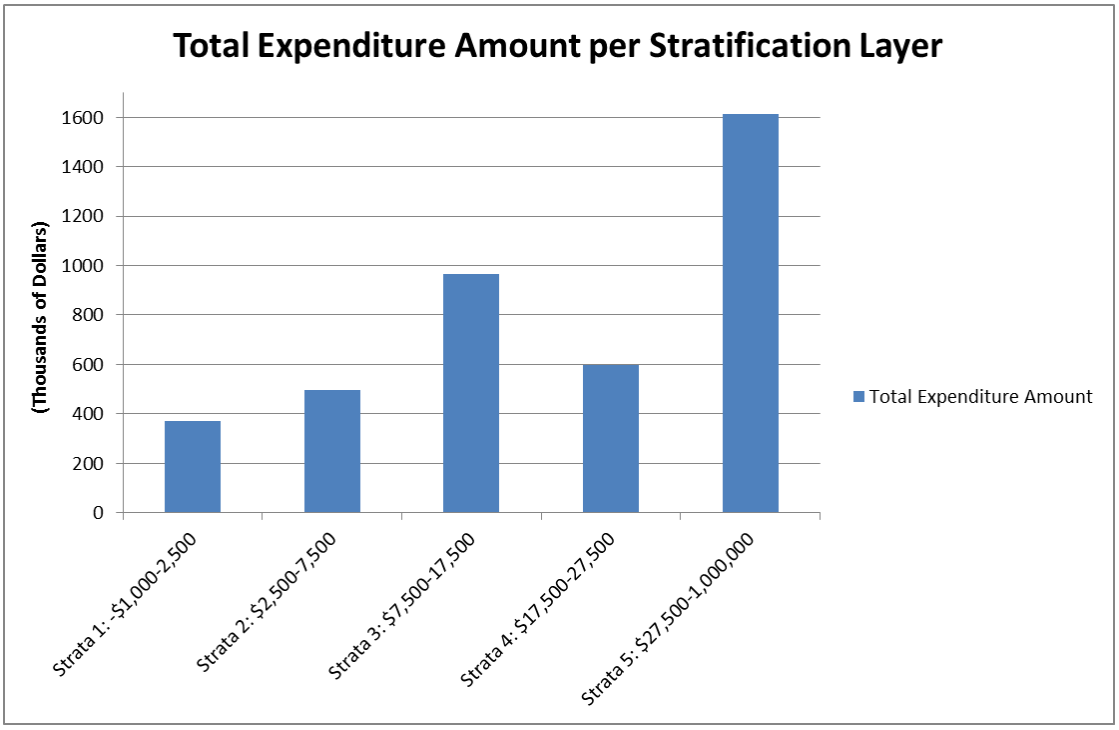

Figure 7: Total Expenditure Amount per Stratification Layer

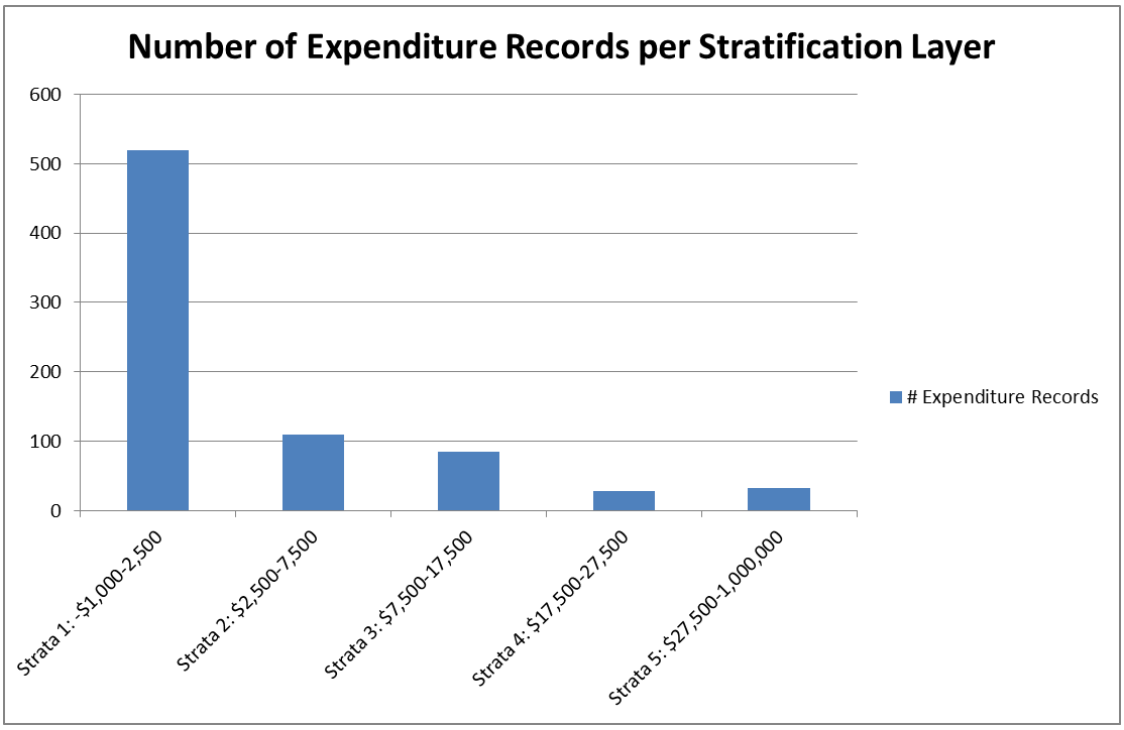

Figure 8: Total Number of Expenditure Transactions per Stratification Layer

4.8.7. Anomalous Spending Detection

Anomalous activities or behaviors are defined as those that do not coincide with established normal profiles (Lee & Xiang, 2001). In an analysis of expenditure data duplicate payment records can be considered as such a behavior. The specific attributes that define duplicate payments may change based on the situation at hand, with the exact criteria for duplicate payment investigations being user-defined. Table 18 presents the results of a duplicate record identification app where the amount is duplicated within a fiscal year, period, department and vendor. This test is based on 2014 data and the expense category 'Books-Library'. The analytic app identifies instances within a specific period (month) and department there have been payments submitted to the same vendor in the same dollar amount. The identified transactions, while presenting attributes similar

to those of duplicate payments, may also indicate spending anomalies such as employees splitting purchases into numerous, smaller dollar amounts to avoid required managerial approvals for larger dollar amounts. These smaller amount purchases can also indicate an overload to the administrative function to process the numerous orders. This data can also be sorted by department, program, and/or activity within department to search for possibly repeated duplicate payments within a department.

FY	Period	Department	Fund	Program	Expense	Vendor	Check Issued Date	Amount
2014	2	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	11/19/2013	10.00
2014	2	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	11/7/2013	10.00
2014	8	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	5/22/2014	13.00
2014	8	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	5/22/2014	13.00
2014	12	Library	Special Library	Miscellaneous	Books-library	INGRAM LIBRARY SERVICES	9/30/2014	23.45
2014	12	Library	General Fund	Materials Management Services	Books-library	INGRAM LIBRARY SERVICES	9/29/2014	23.45
2014	11	Library	General Fund	Materials Management Services	Books-library	PETTY CASH FUND 5081	8/5/2014	35.00
2014	11	Library	General Fund	Materials Management Services	Books-library	PETTY CASH FUND 5081	8/5/2014	35.00
2014	3	Library	General Fund	Materials Management Services	Books-library	INGRAM LIBRARY SERVICES	12/31/2013	74.26
2014	3	Library	General Fund	Materials Management Services	Books-library	INGRAM LIBRARY SERVICES	12/11/2013	74.26
2014	2	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	11/19/2013	80.00
2014	2	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	11/7/2013	80.00
2014	8	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	5/22/2014	80.00
2014	8	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	5/22/2014	80.00
TOTAL AMOUNT:								631.42

Table 18: Test for Duplicate Records (Amount) by Fiscal Year, Fiscal Period, Department, and Vendor (summarized for display purposes and does not include all attributes)

The duplicate record identification app also provides a more focused duplicate analysis is shown in Table 19, that is, for one specific vendor only. Issa (2013) describes in detail the duplicate record identification process. The analytic logic used is the same as for the previous example, that is, within a specific period and department identify payments of a duplicate amount, but in this case only for a specified vendor. This analytic request can result from the need for further research into a single vendor's activity as identified from the results of the previous analytic.

FY	Period	Department	Fund	Program	Expense	Vendor	Check Issued Date	Amount
2014	2	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	11/19/2013	10.00
2014	2	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	11/7/2013	10.00
2014	8	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	5/22/2014	13.00
2014	8	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	5/22/2014	13.00
2014	2	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	11/19/2013	80.00
2014	2	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	11/7/2013	80.00
2014	8	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	5/22/2014	80.00
2014	8	Law	Support Services Fund	Opinions and Advice	Books-library	BETTYE LYNN	5/22/2014	80.00
TOTAL AMOUNT:								366.00

Table 19: Test for Duplicate Records (Amount) by Fiscal Year, Fiscal Period, and Department for a Specific Vendor (summarized for display purposes and does not include all attributes)

4.8.8. Cluster Analysis

Clustering represents a gathering or cause to gather into a grouping a number of similar things (www.vocabulary.com/dictionary/cluster). Cluster analysis is undertaken to identify subgroups within the data being analyzed (Fraley & Raftery, 1998) and can be considered a subset of pattern recognition. Clustering is also considered as a procedure

within a data summarization activity (Nassar et al., 2004). Unlike discriminant analysis which assigns items to pre-defined groups cluster analysis identifies the appropriate groupings based on the data (Fraley & Raftery, 1998). From a research perspective clustering represents an unsupervised classification of patterns or observations that entails one activity in exploratory data analysis (Jain et al., 1999). Patterns can be represented, for example, as a vector of measurement or a point in multi-dimensional 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 and product groupings in retail databases, analyses of Web usage data, and image analyses for segmentation and quantization (Fraley & Raftery, 2002).

A cluster analysis app uses a variety of algorithms as there is no one definition of what a cluster is as it can take many different shapes and it also depends on the type of underlying data, such as either continuous or discrete, and whether the clustering involves identifying similarities or dissimilarities.

The following figures present an illustration of clustering analyses based on a popular, commercially available, visual analytic tool[28]. Figure 9 represents expenditures for a specific vendor across all departments and Figure 10 represents expenditures in the Library department across all vendors. This presentation format illustrates the relative magnitude of spending by the size of the circles for each vendor and may indicate the use of favored vendors that might not necessarily provide the most competitive pricing, or also indicate fraudulent activities. These diagrams are included to provide an example of what can be generated using clustering tools; further discussions of clustering analytics

[28] Visuals as presented in Figures 9 and 10 are modeled from output generated using Tableau 8.1

are outside the scope of this paper. Byrnes (2015) provides a detailed example of a sophisticated clustering technique that can be incorporated into ENHANCE's analytic apps library.

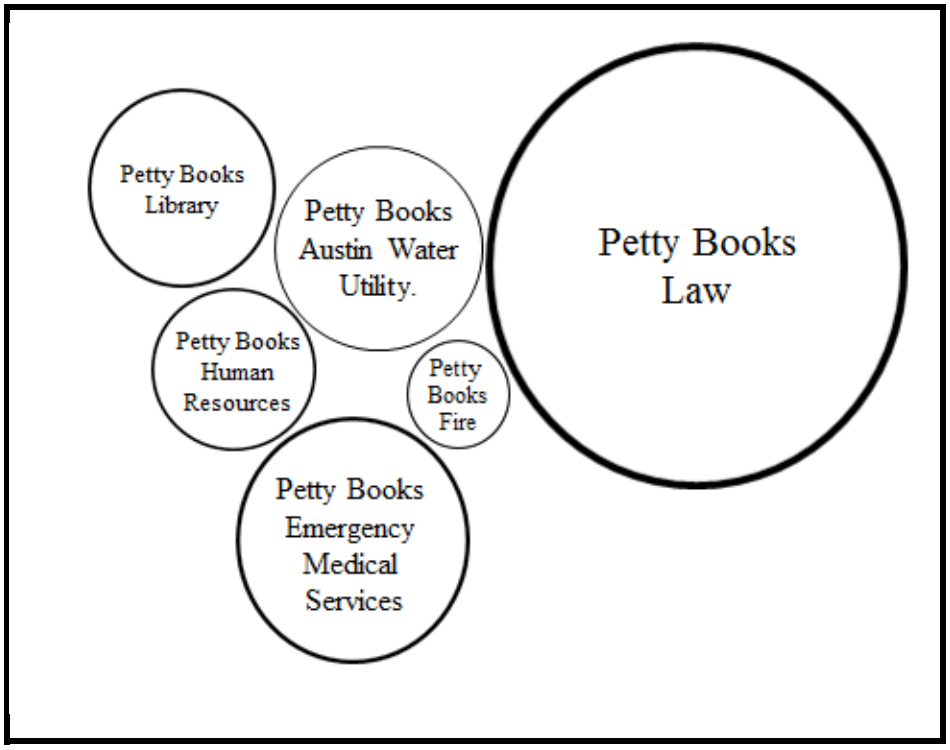

Figure 9: Expenditures to vendor Petty Books across all departments

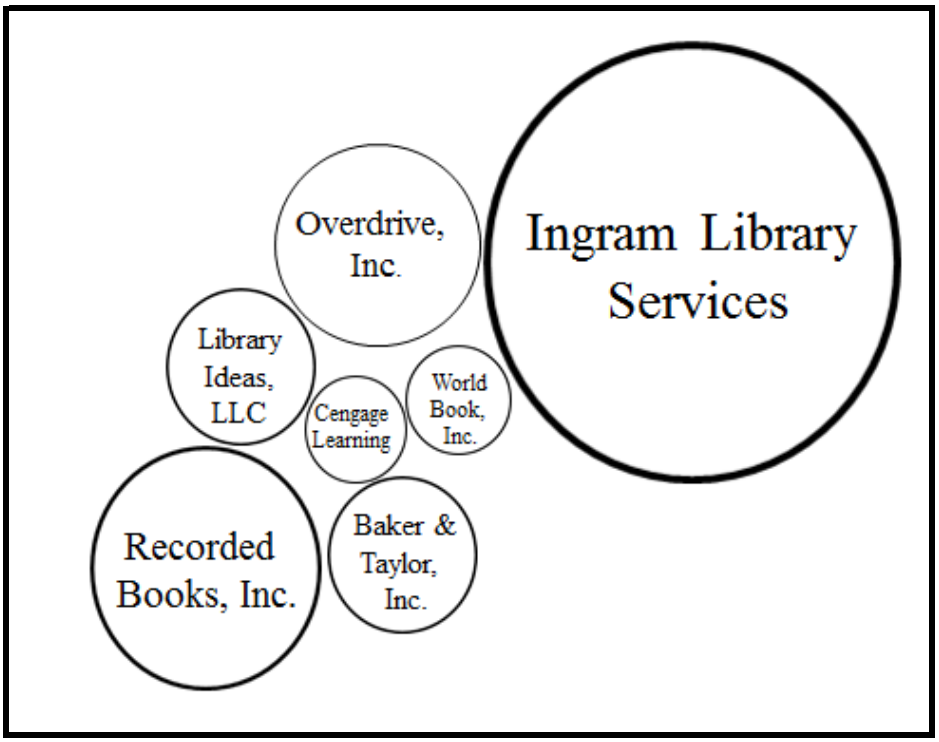

Figure 10: Library department expenditures across all vendors

4.9. Analytics Incorporating Budget Data

Although the Austin expenditure data does include a number of attributes that support the analytics as described above, there are even more informative analytics that can be generated by incorporating additional attributes and datasets into the process.

Budget data as provided by the Austin data portal[29] includes employee salary/wage figures by department and department rollup, and after review of this data with the expenditure data it was determined that the department rollup level in the budget data matched the department level in the expenditure data. A primary requirement when undertaking analyses utilizing two or more datasets is that there is at least one attribute in common between the datasets to allow for matching records, as is the case in this

[29] Budget data was extracted from the City of Austin dataset: 2015 Program_Budget_Operating_Budget_Vs_Expense_Raw_Data

example with the department attribute. Extracts of budget data for specific salary/wage related expense codes[30]. An example of a data extract with data for one expense code is displayed in Table 20.

Budget FY	Department	Program	Activity	Unit	Expense	Budget Amount
2015	Animal Services	Animal Services	Shelter Services	Vet Services	Regular wages - full-time	1,079,264.00
2015	Animal Services	Animal Services	Field Services	Animal Protection	Regular wages - full-time	806,161.00
2015	Animal Services	Animal Services	Shelter Services	Pet Placement	Regular wages - full-time	746,525.00
2015	Animal Services	Animal Services	Shelter Services	Kennel	Regular wages - full-time	569,644.00
2015	Animal Services	Support Services	Departmental Support Services	Administration and Management	Regular wages - full-time	464,879.00
2015	Animal Services	Animal Services	Prevention Services	Prevention	Regular wages - full-time	141,266.00
2015	Animal Services	Animal Services	Shelter Services	Ss Behavior Modification	Regular wages - full-time	97,247.00
2015	Animal Services	Animal Services	Shelter Services	Volunteer Services	Regular wages - full-time	32,938.00
2015	Austin Energy	Electric Service Delivery	Distribution Services	Distribution Construction	Regular wages - full-time	4,584,445.00
2015	Austin Energy	Electric Service Delivery	Distribution Services	Distribution Constr & Maint	Regular wages - full-time	4,464,404.00
2015	Austin Energy	Customer Care	311 City-Wide Call Center	Call Cntr City Wide Info Cntr	Regular wages - full-time	3,598,525.00
2015	Austin Energy	Customer Care	Customer Contact Center	Call Center	Regular wages - full-time	3,367,228.00
2015	Austin Energy	Support Services	Departmental Support Services	IT Operations	Regular wages - full-time	3,263,001.00
2015	Austin Energy	Support Services	Departmental Support Services	IT Engineering & Architecture	Regular wages - full-time	2,330,364.00
TOTAL AMOUNT:						25,545,891.00

Table 20: Example of Austin Budget Data for Payroll Expense (summarized for display purposes and does not include all attributes)

4.10. Analytics Combining Two Data Sources

While the analytics generated from the expenditure data as presented offers insights into a government's spending details, the inclusion of additional data sources can provide for analytics that offer alternate perspectives on government operations. The individual datasets must share at least one common attribute in order to join them for analytic purposes. Although outside of the scope of this present research with respect to the design of the ENHANCE framework, an example incorporating data from two sources to

[30] Expense codes 5001 (Salaries-Regular), 5004 (Shift-Differential), 5005 (Overtime), 5007 (Civil Service), and 5008 (Civil Service-Overtime)

arrive at insightful analytics is presented to describe what capabilities ENHANCE can provide in a future configuration. The inclusion of headcount by department figures allows for additional analytics, such as the ability to calculate departmental expenditures per employee. An analysis utilizing estimated departmental headcount information and applying that attribute to departmental expenditure data is described below.

In this example headcount information is not provided in the Austin data portal so a search for this information in other datasets was undertaken. A summary headcount figure for the entire city government (11,600) was identified (Austin Business Journal, 2015).

The derivation of headcount by department estimates is based on salary/wage information available in the Austin data portal. Budgeted Regular Wages (Full-Time) by department is created by a data summarization app and is shown in Table 21.

Department	Regular wages- full-time	Total
Austin Energy	104,136,256	104,136,256
Austin Water Utility	65,727,820	65,727,820
Police	32,137,101	32,137,101
Parks & Recreation	29,496,370	29,496,370
Planning & Zoning	22,191,100	22,191,100
Communications & Technology Management	21,056,563	21,056,563
Aviation	20,435,059	20,435,059
Solid Waste Services	19,614,566	19,614,566
Watershed	17,456,467	17,456,467
Library	16,559,732	16,559,732
Public Works - Transportation	15,320,431	15,320,431
Public Works	14,104,112	14,104,112
Financial Services	13,301,367	13,301,367
Health & Human Services	12,872,071	12,872,071
TOTAL AMOUNT:	404,409,015	404,409,015

Table 21: Example of Budgeted Regular Wages (Full-Time) by Department

Data is collected for all appropriate budgeted salary/wage codes in the same manner as shown in Table 21 and each are listed and totaled by department, as shown in Table 22.

Department	Salary-Reg 5001	Shift Diff 5004	Overtime 5005	Civil Svc 5007	OT-C Svc 5008	Total
Austin Energy	104,136,256	34,900	3,367,161	0	0	107,538,317
Solid Waste Services	19,614,566	17,313	1,201,265	0	0	20,833,144
Code Compliance	6,492,732	0	78,439	0	0	6,571,171
Austin Water Utility	65,727,820	160,660	4,084,628	0	0	69,973,108
Transportation	10,374,347	17,915	231,036	0	0	10,623,298
Mayor & Council	2,946,536	0	0	0	0	2,946,536
Government Relations	359,216	0	0	0	0	359,216
Office of Real Estate Services	2,346,844	0	0	0	0	2,346,844
Management Services	8,353,493	0	16,000	0	0	8,369,493
Office of the City Clerk	1,396,943	0	2,500	0	0	1,399,443
Municipal Court	8,599,850	21,500	15,000	0	0	8,636,350
Economic Growth & Redevelopment Services	4,093,641	0	1,607	0	0	4,095,248
Communications & Technology Management	21,056,563	9,000	230,061	0	0	21,295,624
Law	7,314,362	0	1,500	0	0	7,315,862
Human Resources	7,295,725	0	0	0	0	7,295,725
Communications & Public Information	1,698,134	0	8,000	0	0	1,706,134
Public Works	14,104,112	0	0	0	0	14,104,112
Public Works - Transportation	15,320,431	6,500	2,028,925	0	0	17,355,856
Watershed	17,456,467	0	411,310	0	0	17,867,777
Wireless	2,509,533	0	133,280	0	0	2,642,813
Office of Contract and Land Management	3,212,758	0	0	0	0	3,212,758
Planning & Zoning	22,191,100	0	156,115	0	0	22,347,215
Office of the City Auditor	1,922,846	0	0	0	0	1,922,846
Neighborhood Housing	1,594,401	0	0	0	0	1,594,401
Financial Services	13,301,367	0	10,698	0	0	13,312,065
Building Services	6,749,228	0	0	0	0	6,749,228
Small Minority Business Resources	1,941,051	0	964	0	0	1,942,015
Fleet Services	10,612,378	39,200	271,092	0	0	10,922,670
Aviation	20,435,059	161,836	991,300	0	0	21,588,195
Convention Center	12,416,109	119,598	521,665	0	0	13,057,372
Fire	6,127,313	0	40,872	89,116,755	7,554,527	102,839,467
Library	16,559,732	0	0	0	0	16,559,732
Parks & Recreation	29,496,370	0	258,227	0	0	29,754,597
Police	32,137,101	147,200	842,143	155,855,221	7,749,207	196,730,872
Health & Human Services	12,872,071	0	38,254	0	0	12,910,325
Animal Services	3,937,924	0	109,474	0	0	4,047,398
Emergency Medical Services	4,548,570	0	505,380	24,523,464	8,216,806	37,794,220
TOTAL AMOUNT:	521,252,949	735,622	15,556,896	269,495,440	23,520,540	830,561,447

Table 22: Example of Individual Salary/Wage budget items by department

The relative % of budgeted salary/wage amount for each department is calculated next. Departmental headcount estimates are derived as based on the budget dollars. The results are shown in Table 23.

Department	Salary-Regular 5001	Shift Differential 5004	Overtime 5005	Civil Service 5007	Overtime-Civil Service 5008	Total Budget Amount	% of Total Budget	Estimated Headcount Calculation
Austin Energy	104,136,256	34,900	3,367,161	0	0	107,538,317	0.13	1,502
Solid Waste Services	19,614,566	17,313	1,201,265	0	0	20,833,144	0.03	291
Code Compliance	6,492,732	0	78,439	0	0	6,571,171	0.01	92
Austin Water Utility	65,727,820	160,660	4,084,628	0	0	69,973,108	0.08	977
Transportation	10,374,347	17,915	231,036	0	0	10,623,298	0.01	148
Mayor & Council	2,946,536	0	0	0	0	2,946,536	0.00	41
Government Relations	359,216	0	0	0	0	359,216	0.00	5
Office of Real Estate Services	2,346,844	0	0	0	0	2,346,844	0.00	33
Management Services	8,353,493	0	16,000	0	0	8,369,493	0.01	117
Office of the City Clerk	1,396,943	0	2,500	0	0	1,399,443	0.00	20
Municipal Court	8,599,850	21,500	15,000	0	0	8,636,350	0.01	121
Economic Growth & Redevelopment Services	4,093,641	0	1,607	0	0	4,095,248	0.00	57
Communications & Technology Management	21,056,563	9,000	230,061	0	0	21,295,624	0.03	297
Law	7,314,362	0	1,500	0	0	7,315,862	0.01	102
Human Resources	7,295,725	0	0	0	0	7,295,725	0.01	102
Communications & Public Information	1,698,134	0	8,000	0	0	1,706,134	0.00	24
Public Works	14,104,112	0	0	0	0	14,104,112	0.02	197
Public Works - Transportation	15,320,431	6,500	2,028,925	0	0	17,355,856	0.02	242
Watershed	17,456,467	0	411,310	0	0	17,867,777	0.02	250
Wireless	2,509,533	0	133,280	0	0	2,642,813	0.00	37
Office of Contract and Land Management	3,212,758	0	0	0	0	3,212,758	0.00	45
Planning & Zoning	22,191,100	0	156,115	0	0	22,347,215	0.03	312
Office of the City Auditor	1,922,846	0	0	0	0	1,922,846	0.00	27
Neighborhood Housing	1,594,401	0	0	0	0	1,594,401	0.00	22
Financial Services	13,301,367	0	10,698	0	0	13,312,065	0.02	186
Building Services	6,749,228	0	0	0	0	6,749,228	0.01	94
Small Minority Business Resources	1,941,051	0	964	0	0	1,942,015	0.00	27
Fleet Services	10,612,378	39,200	271,092	0	0	10,922,670	0.01	153
Aviation	20,435,059	161,836	991,300	0	0	21,588,195	0.03	302
Convention Center	12,416,109	119,598	521,665	0	0	13,057,372	0.02	182
Fire	6,127,313	0	40,872	89,116,755	7,554,527	102,839,467	0.12	1,436
Library	16,559,732	0	0	0	0	16,559,732	0.02	231
Parks & Recreation	29,496,370	0	258,227	0	0	29,754,597	0.04	416
Police	32,137,101	147,200	842,143	155,855,221	7,749,207	196,730,872	0.24	2,748
Health & Human Services	12,872,071	0	38,254	0	0	12,910,325	0.02	180
Animal Services	3,937,924	0	109,474	0	0	4,047,398	0.00	57
Emergency Medical Services	4,548,570	0	505,380	24,523,464	8,216,806	37,794,220	0.05	528
TOTAL AMOUNT:	521,252,949	735,622	15,556,896	269,495,440	23,520,540	830,561,447	1.00	11,600

Table 23: Example of Derivation of Headcount Estimates Based on Budget Data by Department

An analytic app joined the headcount by department estimate to a summarization table of expenditure data (Books-Library) by department for fiscal year 2015 and the calculation of expenditure dollars per headcount was completed. The result is displayed in Table 24. As noted in the results the expenditures by headcount for Books-Library are significantly larger for the Library department than any other department. A presentation in this or similar format allows the user to review expenditures for the other departments to note possible discrepancies that may warrant further investigation.

2015 Expenditures				
Department	**# Records**	**Amount**	**Headcount (Estimated)**	**Expenditure $ per Headcount**
Austin Energy	1	106.49	1,502	0.07
Austin Water Utility	4	9,288.82	977	9.51
Economic Growth & Redevelopment Services	1	329.70	57	5.78
Law	68	75,795.03	102	743.09
Human Resources	4	4,701.20	102	46.09
Public Works	4	561.23	197	2.85
Public Works - Transportation	2	749.90	242	3.10
Watershed	9	432.99	250	1.73
Convention Center	3	322.55	182	1.77
Fire	2	3,579.00	1,436	2.49
Library	347	2,538,209.80	231	10,987.92
Parks & Recreation	1	40.00	416	0.10
Police	1	619.76	2,748	0.23
Emergency Medical Services	19	9,709.73	528	18.39

Table 24: Join of Headcount and Expenditure data with Calculation of Expenditure Dollar per Headcount

4.11. ENHANCE in Support of Armchair Auditing

The term, and concept 'Armchair Auditing' originated in the United Kingdom, with an early definition stated as[31]: "a website that uses open spending data provided by councils to dynamically generate reports for users on council spending according to various criteria". British Prime Minister David Cameron, in a podcast in 2010, committed to "extend transparency as far and as wide as possible. By bringing information out into the open, you'll be able to hold government and public services to account" (Cameron, 2010). In the same podcast he also stated "With a whole army of effective armchair auditors looking over the books, ministers in this government are not going to be able to get away with all the waste, the expensive vanity projects and pointless schemes that we've had in the past" (Cameron, 2010). If this form of government data analysis is undertaken on a large scale by a multitude of constituents 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. A tool such as ENHANCE can provide this capability and provide a consistent level of analytic sophistication to all the participants.

In addition to a lack of analytic tools to allow users to undertake sophisticated analytics, other issues with several of the early open data initiatives limited the abilities of armchair auditors. Some of the early data published did not include relevant date information for either the recording of the transaction or the occurrence of the underlying activity

[31] http://data.gov.uk/library/armchair-auditor

(O'Leary, 2015). Additionally, much of the early data was published in PDF or flat file formats, or even CSV format, that limit the ability analyze or link the data (O'Leary, 2015). Descriptions of the exact item purchased was also not necessarily provided, nor was there access to comparative cost information to support cost-benefit-analyses (O'Leary, 2015).

In their research studying one of the first authorities to adopt the U.K. Local Government Transparency Code[32], 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 may have been 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 also provided tools to present and interpret the data (Frank & Oztoprak, 2015). 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% of all users, followed by the media at 31%, the public (the targeted user group) at 21%, and non-governmental organizations at 7%. Businesses could use

[32] See: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/308185/Local_Government_Transparency_code_2014_Final.pdf. The UK Department for Communities and Local Government issued a transparency code that requires 400 of the larger UK authorities to publish a minimum set of open data for transparency reasons.

this available data to identify government services that could be outsourced, the media could, by publicizing the availability of the data, promote interest in the data, but who comprised the public group of users was not clear; were they political activists or possibly just curious citizens (Worthy, 2015)? It was envisioned that an army of armchair auditors would force the providers to maintain accountability for the data, but so far it appeared that the groups already formally or informally monitoring government activities, such as non-governmental organizations (NGOs), were fulfilling that role (Worthy, 2015). Greater participation in governmental monitoring is linked to greater transparency, but the primary participants may 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 itself 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, that is, an unstable activity dependent on typically very small groups (Worthy, 2015).

Another example that demonstrates where meaningful information is not provided is in an assessment of comparative costs for purchased items or even the details of what items were actually purchased (O'Leary, 2015). The ENHANCE framework provides the ability for spending comparisons across departments within a governmental entity, and if

comparable data taxonomies exist, to provide comparisons between different governmental entities. 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 allow users to comprehend the results of the analytics, but also may include the capability to undertake analyses that result in the output of only 'Exceptional Exceptions' (Issa, 2013), that is, the identification and prioritization of exceptions and anomalies.

In closing his research, O'Leary (2015) presents a number of concerns with respect to the armchair auditor concept that may need to be addressed by future research. Of those the following are noted which may be addressed if not fully at least to a great extent by ENHANCE:

Question: How can governments engage potential armchair auditors?

Answer: Providing a robust tool such as ENHANCE may encourage user participation

Question: How can technology engage potential armchair auditors?

Answer: The ENHANCE framework provides analytic capabilities that require minimal technical and/or analytic skills and the framework functions in a fashion that leads the user through the process in an interactive manner.

Question: What kinds of tools can be generated to facilitate armchair auditing?

Question: Would making tools available increase the number of armchair auditors?

Question: Would tools make their efforts easier to coordinate?

Question: Would tools increase the quality of the armchair auditor contributions?

Answer: The ENHANCE framework is intended to address all these concerns.

4.12. A Design Proposal for the ENHANCE User Interface

The development of the ENHANCE framework ultimately depends on which firm or organization commits to providing a commercially functioning application. Design elements that provide for a high level of user convenience and satisfaction are presented in this paper. It should be noted that although the ENHANCE user interface is intended to be as clear and understandable by the average, non-technically oriented user, an on-line help function with user documentation should be developed to clarify the user interface and ensure that the user can successfully undertake their desired analytics. The design and/or development of this documentation is outside the scope of the present research. The user interface consists of the following screens as described below.

4.12.1. User logon/create ID

The first screen provides the user initial access to the ENHANCE framework. Users can access ENHANCE by one of several means:

- Existing users can enter their ENHANCE user ID and password
- New users can create a user ID (email address) and password
- A user can also enter as a ‘Guest’

The use of a guest logon allows a user to interact with ENHANCE without the need to provide any personal information (i.e. anonymously), should that be a concern with the user. It should be noted that some of the functionality presented below that is tied to a

user ID may not be available to a guest user. For those users that have setup a user ID, once the user enters their user ID and password they will click one of two submit buttons to either create a new analytic or access previously defined analytics. This can be very helpful and efficient for users who require the same, or very similar, reports repeatedly. On the other hand, a guest will click to submit a new request every time. A listing of the information provided in this screen is provided in Table 25:

Item	**Characteristics**	**Automated or Provided by ENHANCE ?**	**Manual or Provided by User?**	**Req'd?**
User ID	User email address		Y	Y
Password	User created		Y	Y
Click to Submit a new analytic request	User prompt to create a new analytic request		Y	
Click to Access existing analytic requests	User prompt to access previously submitted analytic requests		Y	

Table 25: Attributes of the user logon screen (Screen 1)

An example of the initial logon screen is displayed in Figure 11:

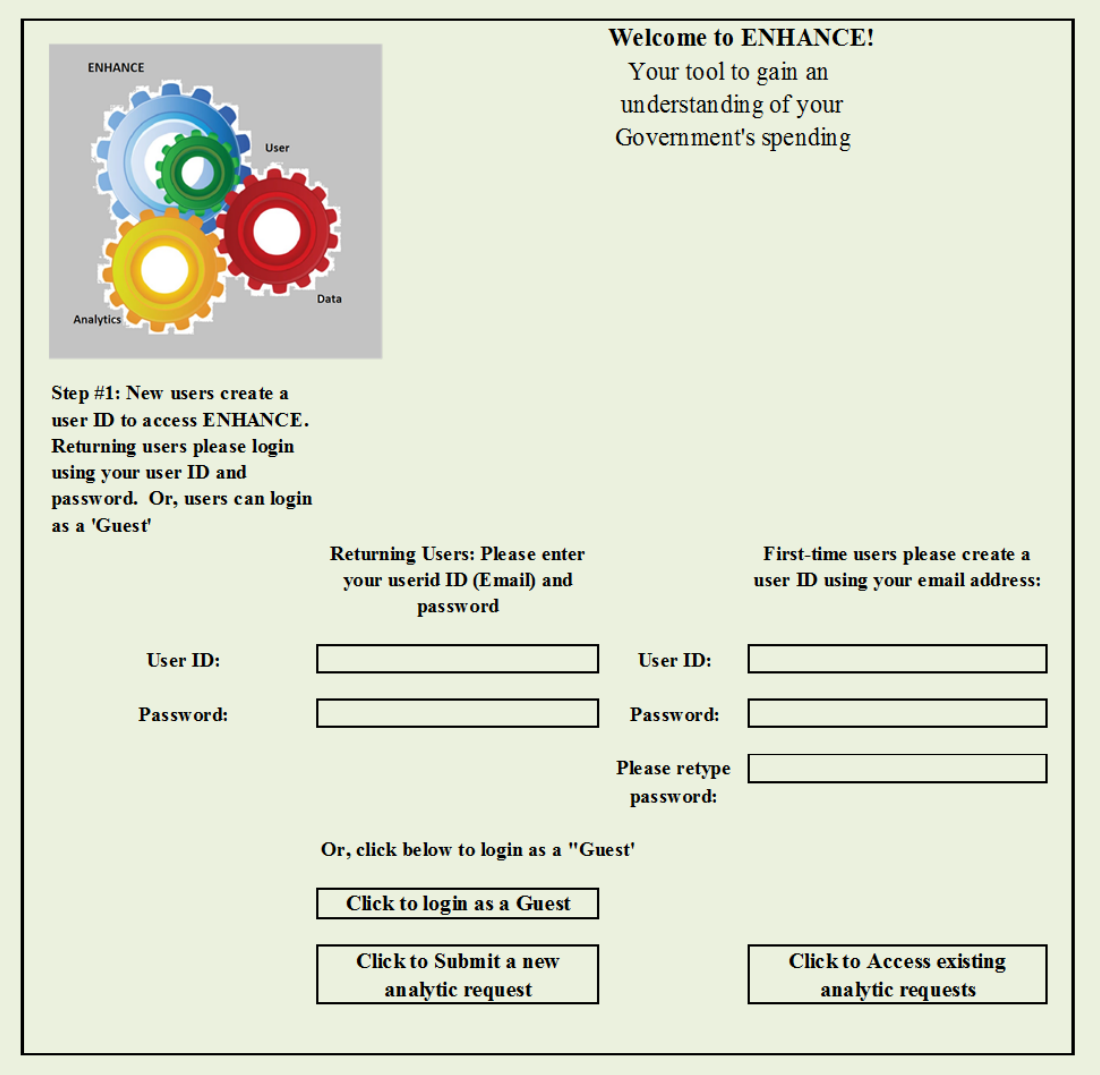

Figure 11: ENHANCE user logon screen (Screen 1)

4.12.2. Submit a target dataset request

The second screen (Figure 12) allows the user to submit an analytic request. A search engine capability (embedded in the structural app) is linked to this screen and will locate the website where the expenditure data exists. The user inputs the name of the governmental entity, for example, Austin, TX, (municipality) or Ohio (state) and the search engine will locate the website address for the expenditure data, for example, the Austin, TX expenditure data website is (https://data.austintexas.gov/Financial/Austin-Finance-Online-eCheckbook/8c6z-qnmj). The screen will display the website address to

verify a connection has been made. If the search is not successful an 'Error' message will be displayed and the user will need to modify the search term. Once the correct website is identified ENHANCE will create a new request record.

A date stamp (system date) is added to the record to indicate the submission date. A system-generated next (request) number is also added in order for the user to track their request, and should they submit multiple requests they are able to access each one individually. When the user has completed the input for the analytic request and clicks Submit ENHANCE accesses the requested dataset and prepares the information to present to the user as displayed on the third screen shown in Figure 13. Ideally, ENHANCE undertakes this step as a real-time, interactive process with the information as displayed in Figure 13 presented to the user within a very short period of time. Should this step require a more substantial amount of time to complete where it is not practical for the user to await the results interactively, ENHANCE emails the user when this step is complete and provides a link for the user to access the results. A listing of the information incorporated in this second screen is provided in Table 26:

Item	**Characteristics**	**Automated or Provided by ENHANCE?**	**Manual or Provided by User?**	**Req'd ?**
Government website address	Internet location where target data is accessed	Y		Y
Date Requested	System date	Y		Y
Request ID Number	System-generated next number	Y		Y
Submit	User prompt to submit interface values to ENHANCE for initial processing		Y	Y

Table 26: Attributes of the data request screen (Screen 2)

An example of the data request screen is displayed in Figure 12:

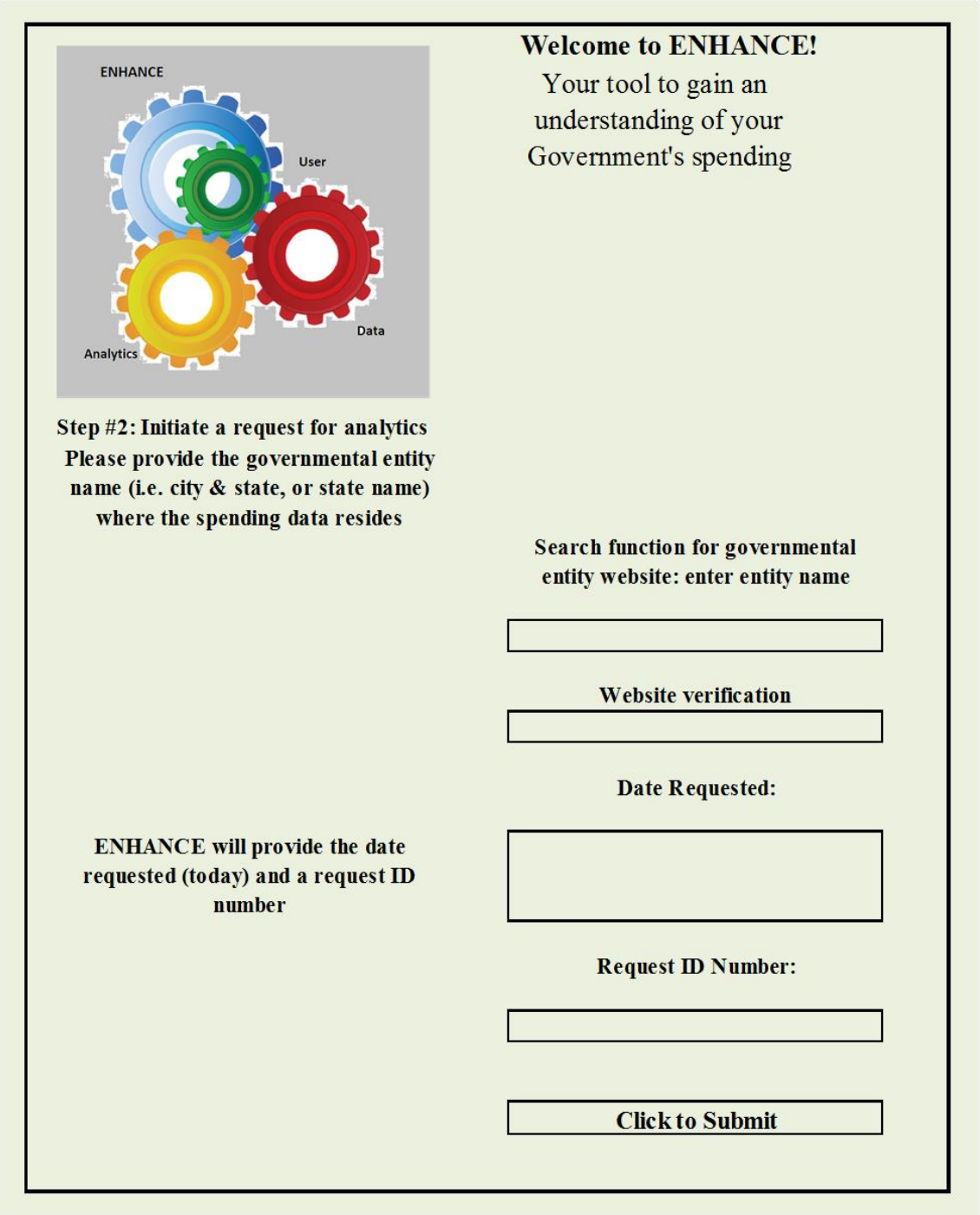

Figure 12: ENHANCE data request screen (Screen 2)

4.12.3. Dataset review and analytics selection

ENHANCE next lists the attributes for the selected dataset (Figure 13 displays an abbreviated attribute list for the example dataset selected). The description for each attribute, as provided by a glossary on the particular data portal, is displayed to assist the user in selecting attributes of interest. A sample value for each attribute, from the first record ENHANCE encounters in the dataset, is also displayed. ENHANCE prompts the user to indicate (Y = yes) if the user wishes ENHANCE to generate a listing of all valid values for a particular attribute or attributes. ENHANCE provides this information in the form of a drop-down list to accommodate data value lists of varying number. This information facilitates the user in identifying appropriate selection criteria for the analytics. For example, the user may be interested in analyzing expenditure data by department and when provided a list of all departments the user may identify specific departments to be included in the analyses. The data item list can also provide the user with an understanding of the extent of data items available, be it 50 departments or 5,000 departments for the particular governmental entity. For each attribute listed the user is prompted to identify those attributes of interest for inclusion in the analytics. The user has the ability to select particular attributes for inclusion in the requested analysis. The user can also alter the attribute selections on the following interface. A listing of the information incorporated in this third screen is provided in Table 27:

Item	Characteristics	Automated or Provided by ENHANCE?	Manual or Provided by User?	Req'd?
Attribute Name	Attribute name as provided by data source; one line per attribute displayed. Multiple screens may be required	Y		Y

Description	Description of the particular attribute	Y		Y
Example	Example of data item in first record encountered	Y		Y
Request a list of all attribute values for the selected dataset	User prompt for list of all values for a specific attribute (Y)		Y	
Select this attribute to include in the analytics	User prompt to include attribute for analysis (Y)		*Y*	
Submit	User prompt to submit screen values to ENHANCE for processing		Y	Y
Next Screen	Only available if data attributes require multiple screens to display		Y	
Previous Screen	Only available if data attributes require multiple screens to display		Y	

Table 27: Attributes of the attribute selection screen (Screen 3)

An example of the attribute selection screen is displayed in Figure 13:

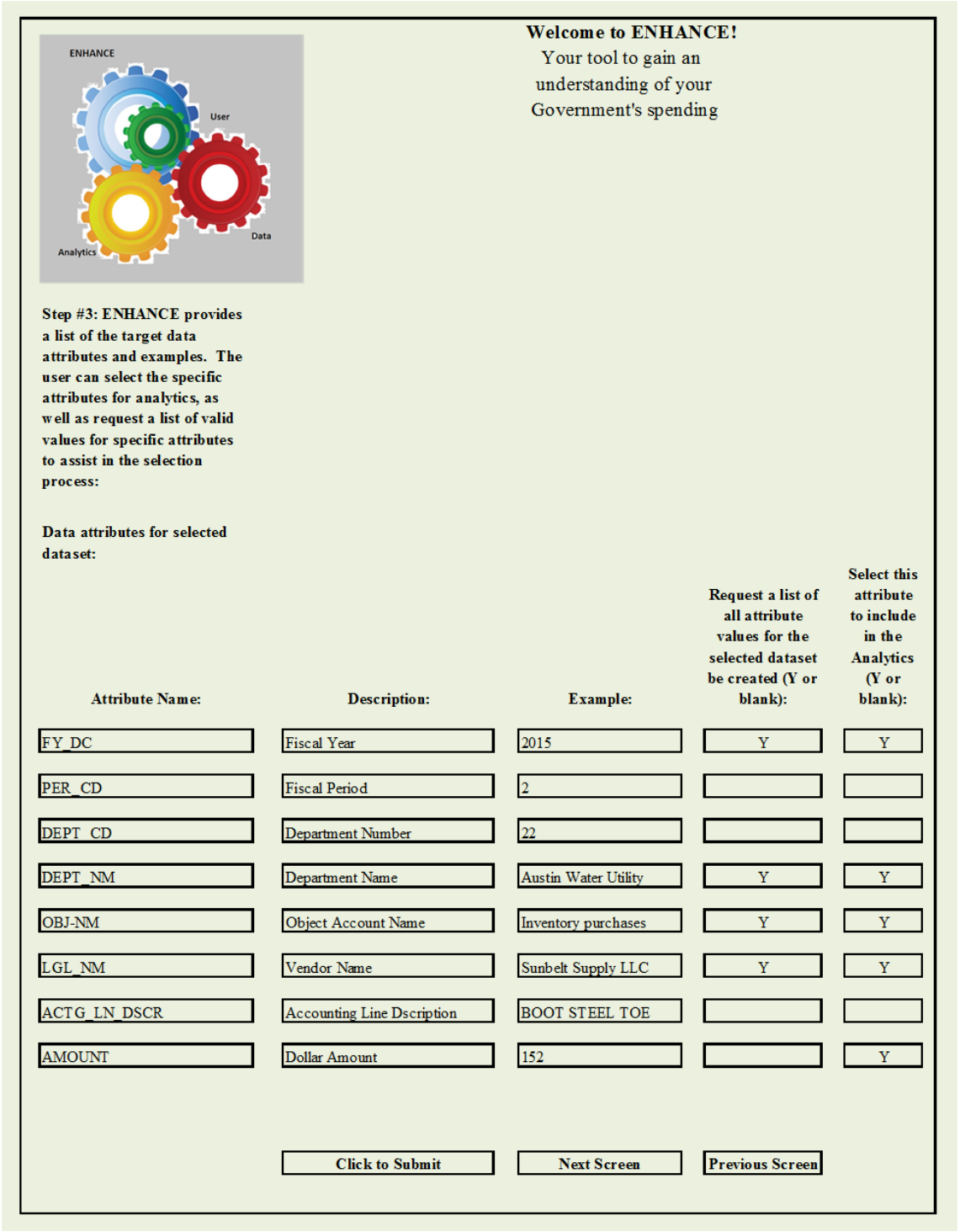

Figure 13: ENHANCE attribute selection screen (Screen 3)

4.12.4. Analytics definition

The next ENHANCE screen displays the attributes that the user selected in the previous screen for either display of valid values and/or inclusion in the analytics. The attribute name, description, and an example data value are provided, as also displayed on the previous screen (Figure 13). For each attribute that the user requests a list of valid values (Figure 13) a drop-down list is provided. Based on the user-selected attributes for analytics ENHANCE presents a list of potential analytics that can be launched. The list of proposed analytics is developed within the ENHANCE framework's recommender system. For each recommended analytic ENHANCE displays the most relevant attribute for the specific analytic, for example, for a Time Series analysis a date attribute such as fiscal year or period is most appropriate. The user selects the analytics they wish to execute (Y = yes) and also the output format for each analytic from a drop-down list. Depending on the type of output available for the specific analytic the user chooses from typical formats such as XLS, CSV, PDF, among others. ENHANCE lists the user-selected attributes with a drop-down list for each that allows the user to set selection criteria for the chosen analytics. The drop-down lists include all the valid values for the particular attribute. For numeric attributes Boolean logic is available for defining selection criteria as well as the ability for a user-specified list or range of values. The user has the ability to return to the previous screen (Figure 13) should they wish to alter their attribute selections. In this case ENHANCE reconfigures and redisplays the analytics selection screen (Figure 14) based on the revised criteria. Once the user completes their revisions on the analytic selection screen they click on the submit button

and ENHANCE launches the requested analytics and generates the appropriate output.

A listing of the information included in this fourth screen is provided in Table 28:

Item	Characteristics	Automated or Provided by ENHANCE?	Manual or Provided by User?	Req'd?
Attribute Name	Attribute name as provided by data source; one line per attribute displayed. Multiple screens may be required	Y		Y
Description	Description of the particular attribute	Y		Y
Example	Example of data item in first record encountered	Y		Y
Drop-down list of valid values	User prompt to view scroll-down list of all valid values for a specific attribute	Y		
Analytic Name	Based on user-selected attributes from the dataset the ENHANCE recommender system will select potential analytic apps; may be multiple items listed	Y		Y
Relevant Attribute	Attribute that is most relevant to the specific analytic	Y		Y
Select for analytics	User prompt to select analytics to execute		Y	
Select an output format	User prompt to select output format for presentation of the results of analytic, from a drop-down list	Y	Y	Y
Attribute name	For data selection for analytics	Y		Y
Description	Description of the particular attribute	Y		Y

Selection Criteria	User prompt to select attributes from a drop-down list, or Boolean logic for numeric fields such as amount, year, or period. The user can also define a list or range of values for selection.	Y	Y	
Return to Previous Screen	User prompt in order to reset selections on the attribute selection screen (Step 3)		Y	
Submit	User prompt to submit screen values to ENHANCE for processing		Y	Y

Table 28: Attributes of the analytic definition screen (Screen 4)

An example of the analytic definition screen is displayed in Figure 14:

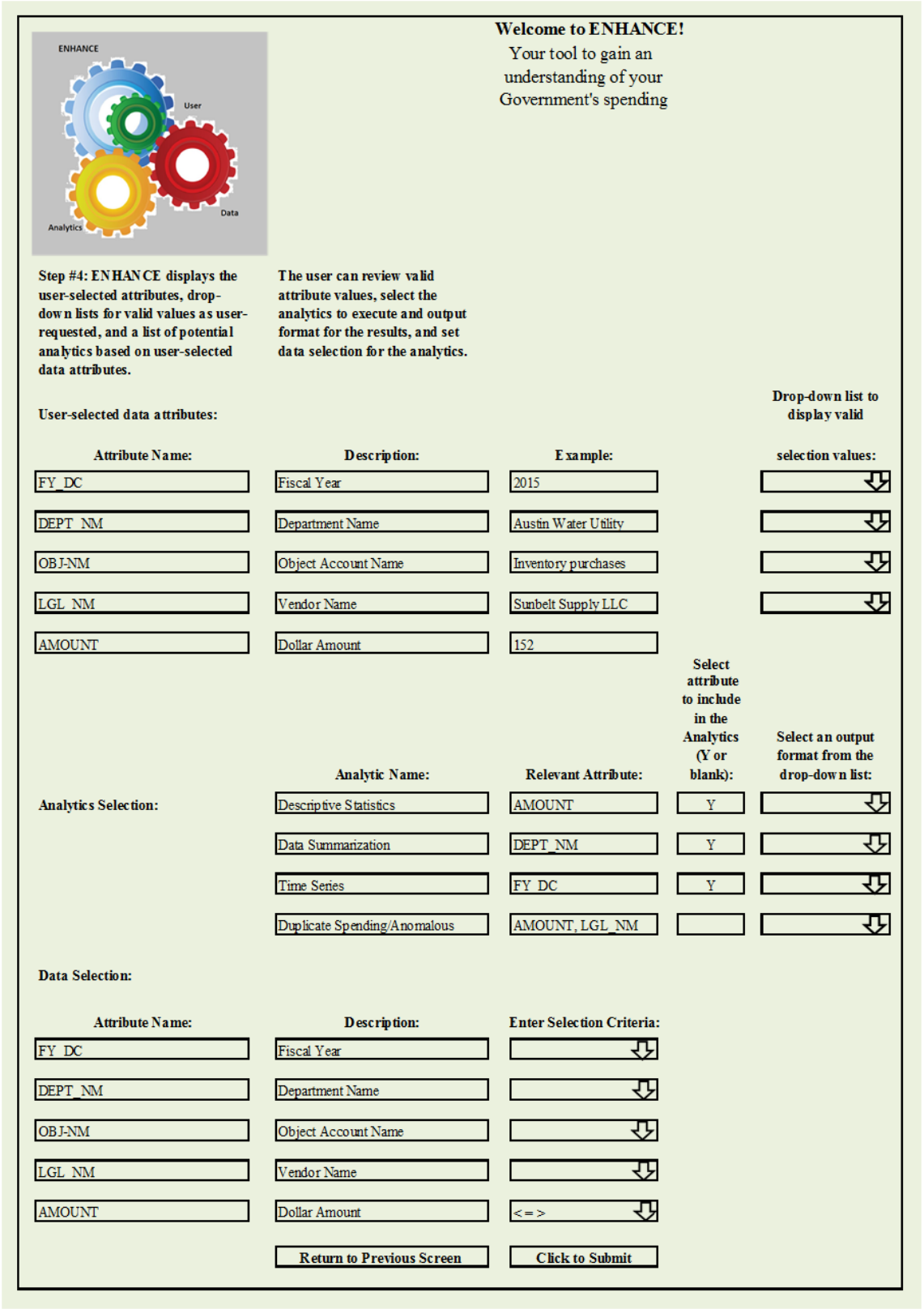

Figure 14: ENHANCE analytic definition screen (Screen 4)

4.12.5. Analytic output review

Once ENHANCE completes the requested analytics the results are prepared for user review. If it is not be feasible to complete the analytics in a real-time, interactive manner ENHANCE emails the user when the analytics are complete and provides a link to the user to access the results. The results screen displays the description of each analytic that was executed, and the output format, as requested by the user in the prior screen via a drop-down listing. The user is prompted to select which analytic results they wish to review (Y = yes). Once their selections are complete the user clicks Submit to prompt ENHANCE to display the results interactively. A listing of the information incorporated in this fifth screen is provided in Table 29:

Item	Characteristics	Automated or Provided by ENHANCE?	Manual or Provided by User?	Req'd?
Analytic	The description of the analytic as selected by the user in Step 4	Y		Y
Output format	The output format as selected by the user in Step 4	Y		Y
Review analytic output (Y = yes)	User select to review the output for a specific analytic		Y	
Submit	User prompt to submit screen values to ENHANCE for processing		Y	Y

Table 29: Attributes of the analytic output screen (Screen 5)

An example of the analytic output screen is displayed in Figure 15:

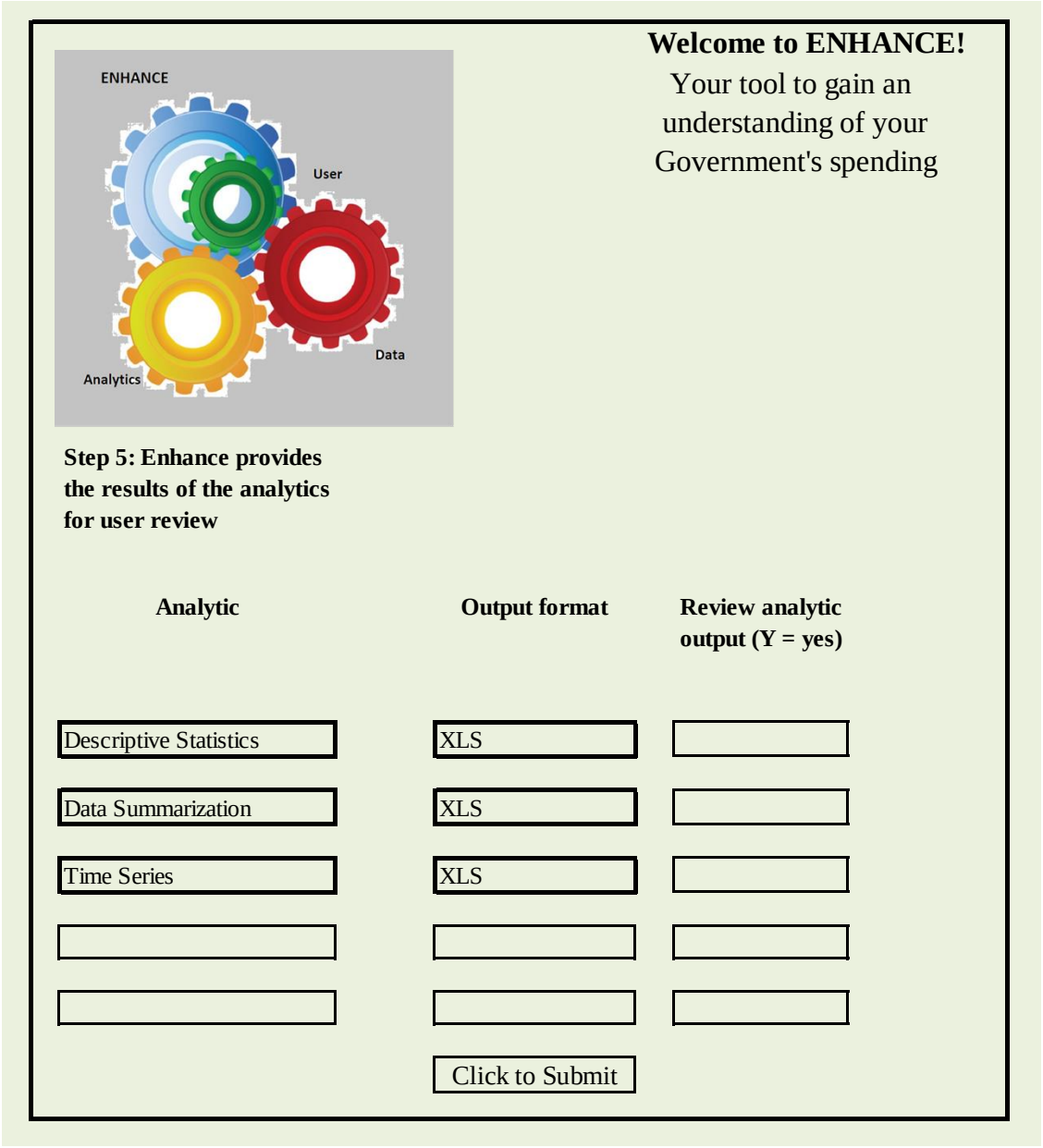

Figure 15: ENHANCE analytic output review screen (Screen 5)

4.12.6. Prior analytics review

ENHANCE provides the ability for the user to access previously submitted and executed analytics to review the criteria for the analytics and/or the resulting output. The starting point is for the user to identify which requests they wish to review. The Request ID

Number field provides a drop-down list of all requests for that particular user. Once a specific Request ID has been chosen the user accesses the Analytic drop-down list to identify the specific analytic to review. For each identified request the user has the option of reviewing the criteria for that request, that is, the specific analytics executed and the selection criteria used, or the user can review the output from the analytics for that request, or both. A listing of the information included in this screen is provided in Table 30:

Item	Characteristics	Automated or Provided by ENHANCE?	Manual or Provided by User?	Req'd?
Request ID Number	The ID number of the specific request the user wishes to review		Y	
Review user-selected analytics and criteria (Y = yes)	User prompt to review the analytics definition (Step 4) for a specific request		Y	
Review analytic output (Y = yes)	User prompt to review output (Step 5) for a specific analytic		Y	
Submit	User prompt to submit screen values to ENHANCE for processing		Y	Y

Table 30: Attributes of the access to prior analytics and results screen (Screen 5a)

An example of the prior analytics review screen is displayed in Figure 16:

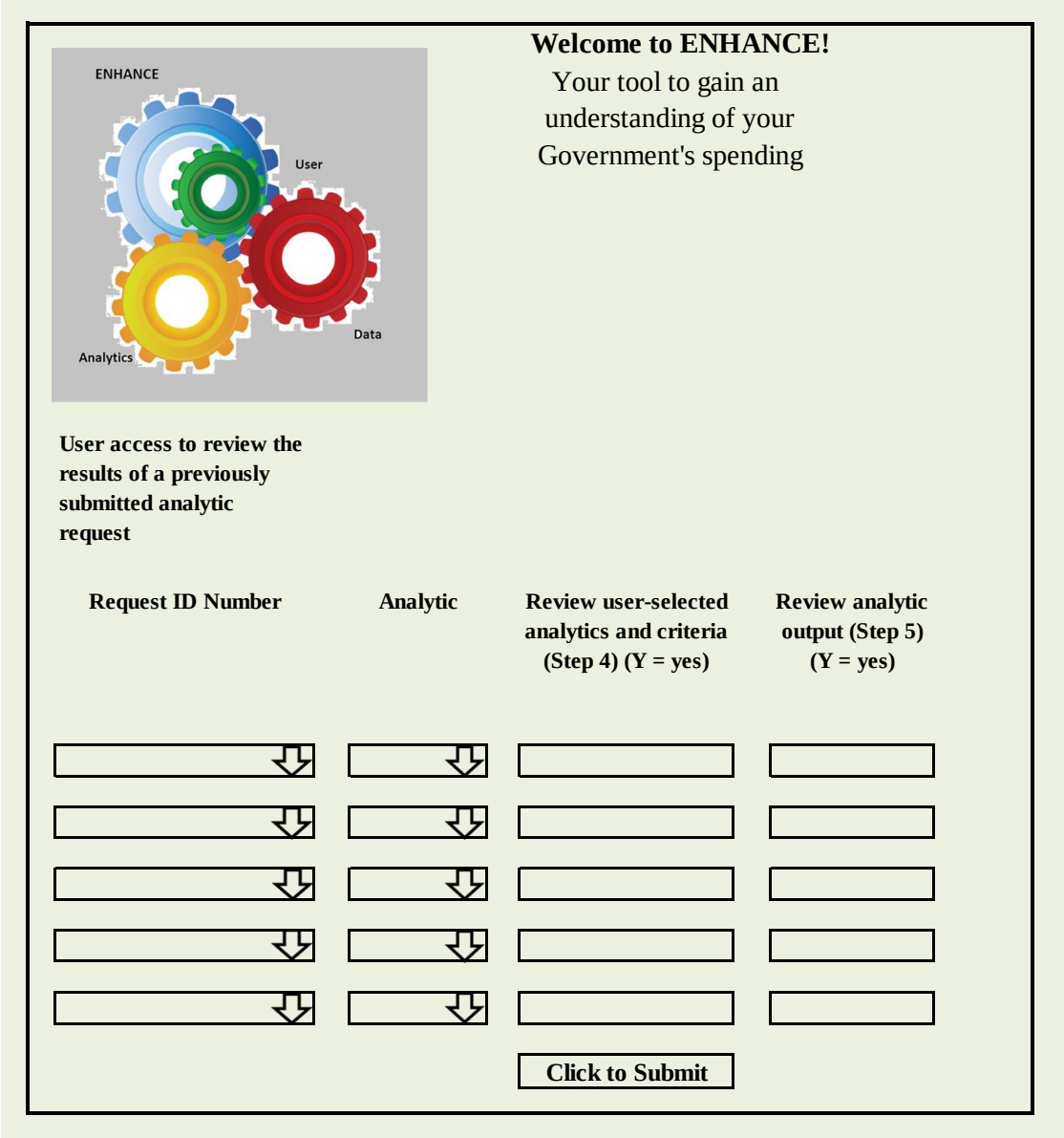

Figure 16: ENHANCE prior analytics review screen (Screen 5a)

4.13. ENHANCE Data Structure

To support the user interface attributes presented above, the data structure required for the ENHANCE framework is presented next. The table and attribute names are defined

using the CamelCase naming convention[33]. In addition to the attributes described for each table it is envisioned that each table also includes date and time last updated attributes (for each record), and the user ID of the person last updating the record. These particular attributes are updated by ENHANCE administrators and not accessible by the user community.

4.13.1. User ID Table

The UserIDTable (Table 31) stores the user ID's (UserID) and passwords (Password) for all registered ENHANCE users.

Attribute Name	Description	Criteria
UserID	User ID to access Enhance is be the user's email address	Alphanumeric (Key)
Password	User-defined password should be defined using typical logic requiring a minimum length, alpha, numeric, lower-case, and upper-case characters	Alphanumeric

Table 31: Attributes of the UserIDTable

4.13.2. Data Website Address Table

The DataSiteTable (Table 32) stores the website address (DataSite) for the user-requested dataset. The table also holds the request ID number (RequestID) and requested date (RequestDate).

Attribute Name	Description	Criteria
RequestID	A unique, system-assigned next number to uniquely identify each submitted request. The RequestID is created in this step	Numeric (Key)

[33] CamelCase is the practice of writing compound words or phrases such that each word or abbreviation begins with a capital letter (from Wikipedia)

DataSite	Website address where requested government dataset is found	Alphanumeric
RequestDate	Date request was submitted (system date)	Date

Table 32: Attributes of the DataSiteTable

4.13.3. Attribute Table

The AttributeTable (Table 33) is populated by ENHANCE in response to the user's initial request for analytics (Figure 12). ENHANCE provides information on the particular dataset's structure to allow the user to select attributes of interest for analytic purposes.

Attribute Name	Description	Criteria
RequestID	A system-assigned next number to uniquely identify each submitted request	Numeric (Key)
AttributeLine#	A system-assigned next number, per Request ID, to uniquely identify each attribute	Numeric (Key)
AttributeDescription	Field name from the underlying dataset for each attribute	Alphanumeric
Description	Description of the particular attribute	
AttributeExample	Value for each attribute as stored in the first record encountered in the dataset	Alphanumeric
RequestAttributeList	User select to request a list of all valid values (Y = yes). ENHANCE creates and populates the valid values for the specific attribute to the AttributeValueTable	Y or blank
IncludeAttribute	User prompt to include this attribute in the analytics (Y = yes). Enhance includes this attribute in the subsequent analytic request	Y or blank

Table 33: Attributes of the AttributeTable

4.13.4. Attribute Value Table

For each attribute that the user requests a drop-down list of valid values, primarily to assist the user in setting appropriate selection criteria for the analytics, ENHANCE will populate the AttributeValueTable (Table 34).

Attribute Name	Description	Criteria
RequestID	A system-assigned next number to uniquely identify each submitted request	Numeric (Key)
AttributeLine#	A system-assigned next number, per Request ID, to uniquely identify each attribute	Numeric (Key)
ValueLine#	A system-assigned next number, per Request ID and attribute line #, to uniquely identify each valid value for a specific attribute	Numeric (Key)
AttributeValue	Valid value for a specific attribute	Alphanumeric

Table 34: Attributes of the AttributeValueTable

4.13.5. Selected Attribute Table

The SelectedAttributeTable (Table 35) provides the user with information on the attributes they selected for a specific analysis (Figure 13). This information is presented on the analytic selection screen (Figure 14). The user has the ability review the attribute value list for specific attributes that they requested in Screen 3.

Attribute Name	Description	Criteria
RequestID	A system-assigned next number to uniquely identify each submitted request	Numeric (Key)
AttributeLine#	A system-assigned next number, per Request ID, to uniquely identify each	Numeric (Key)

	attribute	
AttributeDescription	Field name from the underlying dataset for each attribute	Alphanumeric
Description	Description of the particular attribute	Alphanumeric
AttributeExample	Value for each attribute as stored in the first record encountered in the dataset	Alphanumeric
AttributeList	User prompt to access a list of valid values for a specific attribute (click to access AttributeValueTable)	(Click)

Table 35: Attributes of the SelectedAttributeTable

4.13.6. Analytic Selection Table

The AnalyticSelectionTable (Table 36) includes analytics that the ENHANCE recommender system has chosen, based on the users attribute selections in Figure 13. This information is presented on the analytic selection screen (Figure 14). The user specifies which analytics to execute (Y = yes) and which output format they prefer for the results of each analytic. The valid output formats are stored in the OutputFormatTable.

Attribute Name	Description	Criteria
RequestID	A system-assigned next number to uniquely identify each submitted request	Numeric (Key)
AnalyticLine#	A system-assigned next number, per Request ID, to uniquely identify each analytic	Numeric (Key)
AnalyticDescription	Analytic description, stored in the AnalyticTable	Numeric
RelevantAttribute	Attribute that is most relevant to the specific analytic, as determined by ENHANCE	Alphanumeric
AnalyticSelect	User prompt to select specific analytic to include in analysis (Y = yes)	Y or blank
OutputFormat	User prompt to select output	Numeric

	format for specific analytic from drop-down list (valid values are stored in the OutputFormatTable)	

Table 36: Attributes of the AnalyticSelectionTable

4.13.7. Analytic Master Table

The AnalyticMasterTable (Table 37) stores the description for every analytic app defined in ENHANCE and provides the description for screens that include analytic information.

Attribute Name	Description	Criteria
AnalyticMasterLine#	A system-assigned next number to uniquely identify each analytic	Numeric (Key)
AnalyticDescription	Analytic description as populated by ENHANCE for each analytic available in the analytic library	Alphanumeric

Table 37: Attributes of the AnalyticTable

4.13.8. Data Selection Table

The DataSelectionTable (Table 38) provides information to support the user in defining selection criteria for their requested analytics. This information is displayed on the analytic selection screen (Figure 14).

Attribute Name	Description	Criteria
RequestID	A unique, system-assigned next number to uniquely identify each submitted request	Numeric (Key)
AttributeSelectLine#	A system-assigned next number, per Request ID, to uniquely identify each attribute selection line	Numeric (Key)
AttributeDescription	Field name from the underlying dataset for each attribute	Alphanumeric
Description	Description of the particular	Alphanumeric

	attribute	
SelectionCriteriaLine#	A system-assigned next number, per Request ID, to uniquely identify each selection criteria line	Numeric (Key)
SelectionValue	User prompt to access drop-down list to input selection values for each attribute	Alphanumeric
BooleanCriteria	For numeric values user prompt to select Boolean logic for the specific selection value (Valid Boolean operators stored in BooleanOperatorTable	Alphanumeric

Table 38: Attributes of the DataSelectionTable

4.13.9. Output Format Table

The OutputFormatTable (Table 39) provides the description for all the available output formats in ENHANCE. The description is displayed on all screens that include output format.

Attribute Name	Description	Criteria
AnalyticMasterLine#	A system-assigned next number to uniquely identify each analytic (validated against AnalyticMasterTable)	Numeric (Key)
FormatLine#	A system-assigned next number to uniquely identify each analytic format	Numeric (Key)
FormatDescription	Description of specific format	Alphanumeric

Table 39: Attributes of the OutputFormatTable

4.13.10. Boolean Operator Table

The BooleanOperatorTable (Table 40) stores the valid values that are available to the user when setting the data selection on the analytic selection screen (Figure 14).

Attribute Name	Description	Criteria
BooleanLine#	A system-assigned next number to uniquely identify each Boolean operator	Numeric (Key)
BooleanOperator	Boolean operator	Alphanumeric

Table 40: Attributes of the BooleanOperatorTable

4.13.11. Output Results

The output results from any particular analytic depend on which output format option the user selects. In order to link the output to a specific request the filename for the output includes the Request ID # and Analytic Line #. Additionally, the file name includes Analytic Description and date. An example output file name, where the Request ID = 1000, the Analytic Line = 1, the Analytic Description = Time Series, and the date = June, 30, 2016, for an Excel output, would be: 001000_001_Time_Series_06.30.2016.xls.

4.13.12. Analytic Review Table

The information stored in the AnalyticReviewTable (Table 41) is displayed on the analytic output review screen (Figure 15). The user can select to review the output of the specific analytic on this screen (Y = yes).

Attribute Name	Description	Criteria
RequestID	A system-assigned next number to uniquely identify each submitted request	Numeric
AnalyticLine#	A system-assigned next number, per Request ID, to uniquely identify each analytic	Numeric (Key)
OutputFormat	Format in which output is configured	Alphanumeric
ReviewOutput	User prompt to review a specific analytic output (Y = yes).	Y or blank

Table 41: Attributes of the AnalyticReviewTable

4.13.13. Prior Analytic Review Table

The PriorAnalyticReviewTable (Table 42) is similar in nature to the AnalyticReviewTable in that it holds information to populate the prior analytics review screen which allows the user access to the output from prior analytic analyses (Figure 16). In addition to allowing the user to request access to the analytic output (Y = yes) this screen also allows the user to review the criteria underlying a prior analytic (Y = yes).

Attribute Name	Description	Criteria
RequestID	ENHANCE provides a drop-down list of all analytics for the specific user ID and the user can choose one request per line	Numeric (Key)
AnalyticLine#	Once the user has identified the specific request ENHANCE provides a drop-down list of all analytics for that request and the user can choose on per line	Numeric (Key)
ReviewAnalytic	User select to review the criteria for a specific request (Y = yes). ENHANCE to display the Step 4 interface for the specified request	Y or blank
ReviewOutput	User select to review a specific analytic output (Y = yes). A yes prompts ENHANCE to display the output	Y or blank

Table 42: Attributes of the PriorAnalyticReviewTable

The relationships between the ENHANCE user screens described earlier and the supporting data tables presented above are depicted in Figures 17 and 18. The ‘steps’ as

indicated in Figures 17 and 18 refer to the steps as outlined in Figure 3, the ENHANCE Framework.

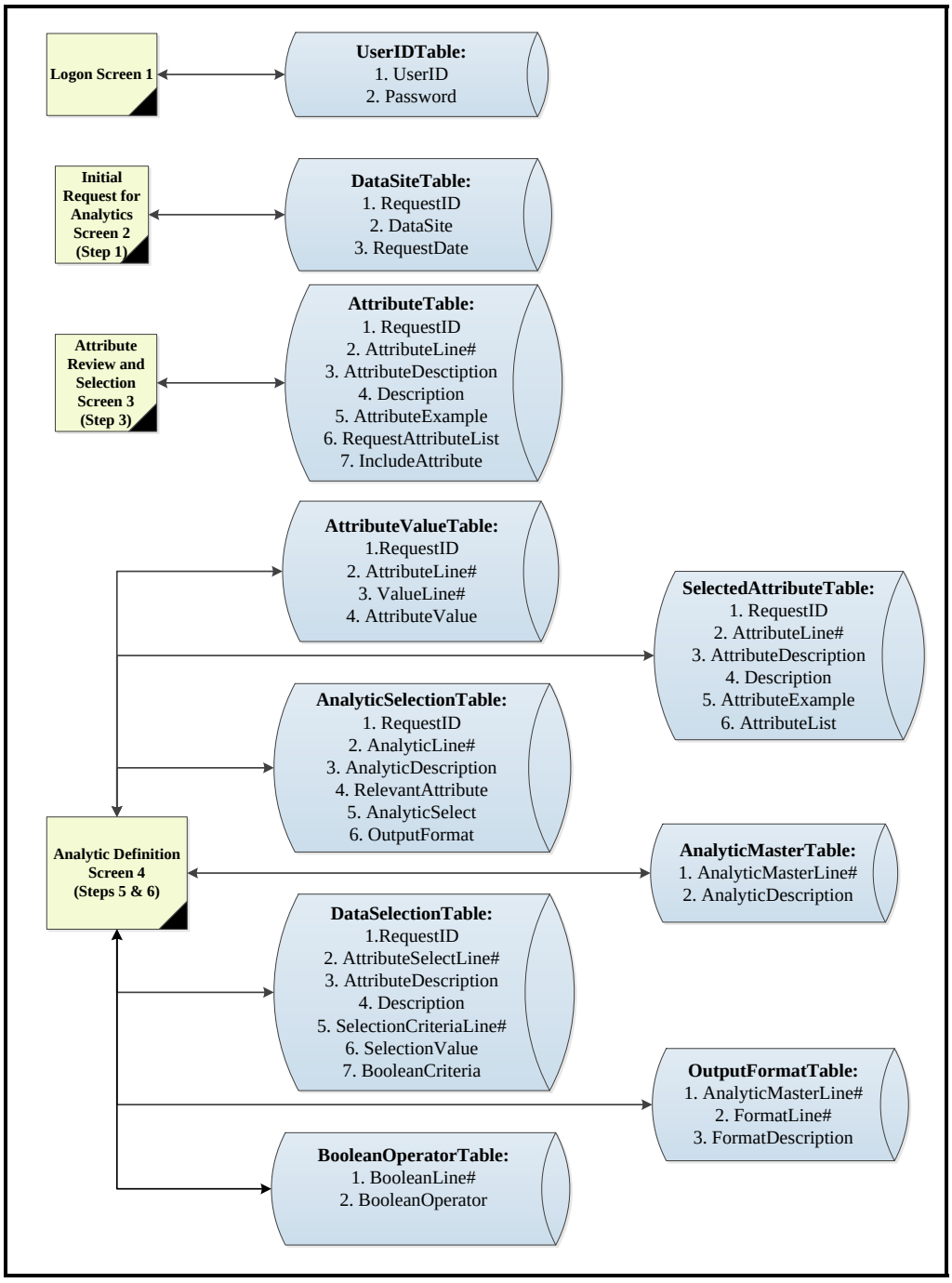

Figure 17: ENHANCE interface-data relationships

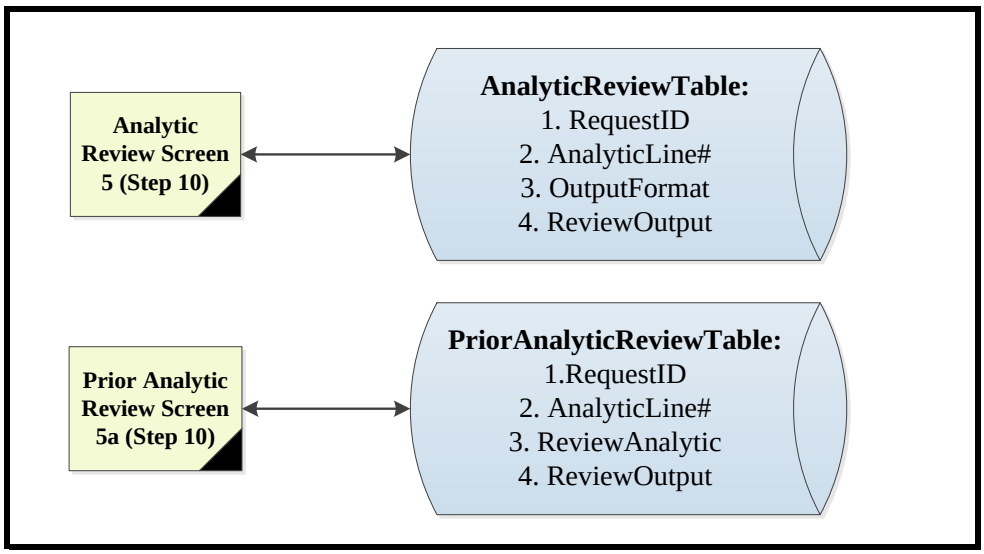

Figure 18: ENHANCE interface-data relationships (continued)

Figure 19 presents the ENHANCE framework from the perspective of the screens and the table structure as defined above, that is, how each screen and table relates to the apps that support the ENHANCE framework. The 'steps' as indicated in Figure 19 refer to the steps as outlined in Figure 3, the ENHANCE Framework.

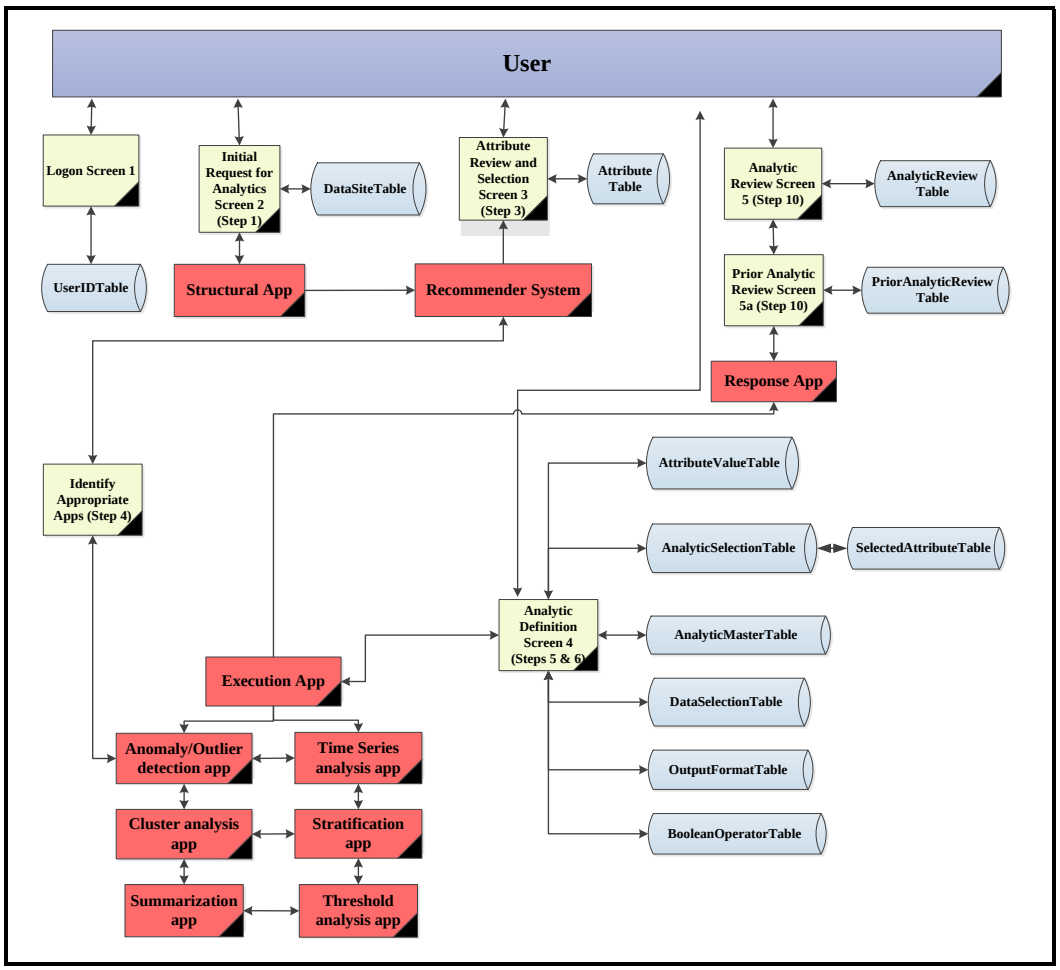

Figure 19: ENHANCE framework with screens and table structure

4.14. An App Recommender System: a key component of ENHANCE

The presentation of the ENHANCE framework up to this point has focused on a review of the capabilities of selected analytic apps and a conceptual design of the user interface. The ENHANCE framework encompasses a variety of apps, all in support of the analytic capabilities of the tool. The structural app performs an analysis of the dataset under investigation and provides information on the structure and attributes of the dataset to the user. The execution app captures the user's responses with respect to which analytic apps to execute as well as the dataset selection criteria. The app recommender system

provides the user a list of possible analytic apps based on their initial feedback. The user's responses with selected apps, data selection criteria, and requested formatting are logged in the execution app which launches the selected analytic apps that undertake the analytics. The response app formats the analytic results and presents them to the user. The specific design and configuration of the apps that support the ENHANCE framework is outside the scope of the present research, but a further discussion on the app recommender system is warranted as this is a key component of ENHANCE. Research has been conducted to define an audit app recommender system, a recommender system that is similar to the analytic app recommender system proposed for ENHANCE.

Audit apps have shown a recent rise in popularity, primarily due to analytics software providers such as CaseWare IDEA and ACL continually developing new audit apps (Dai et al., 2014). The availability of numerous analytics, primarily developed for use by the private sector, will support the analytic requirements by average citizens investigating governmental expenditure data.

IDEA presently provides 100+ routine audit apps that include Benford's Law, Fuzzy Duplicate identification, Gap Detection, Field Manipulation, Summarization, Stratification, and Sampling[34]. In addition to these analytic tools IDEA also offers SmartAnalyzer which consists of a number of queries and tasks that can support audit engagements to improve audit quality and consistency. SmartAnalyzer offers almost 60 routines for analyzing: General Ledger (17 routines), Accounts Receivable (seven routines), Inventory (13 routines), Fixed Assets (six routines), and Accounts

[34] See: http://www.casewareanalytics.com/products/idea-data-analysis

Payable (14 routines) transactions. An example of several General Ledger routines are listed below, that test for:

- Out-of-balance journal entries
- Duplicate journal entries
- Journal entries posted on specific days/dates/times
- Journal entries with large/rounded amounts or amounts that end in -999

ACL states that their analytics software incorporates "hundreds" of analytic ideas that have been developed since the firm's founding in 1987[35]. ACL groups their analytics into 12 parent categories and within these 23 sub-categories. The parent categories include:

- Financial Services
- Fixed Assets
- Gaming
- General Ledger
- Healthcare
- Information Technology
- Order-to-Cash
- Payroll
- PCards

[35] See: https://accounts.aclgrc.com/inspirations?utm_source=ACL.com&utm_medium=ACLAnalytics&utm_campaign=inspirations#byprocess/?utm_source=ACL.com&utm_medium=ACLAnalytics&utm_campaign=inspirations

- Public Sector
- Purchase-to-Pay
- Travel and Entertainment

Sub-categories within General Ledger include:

- Account Management
- Balances
- Journal Entries

ACL lists 28 specific analytic routines for the General Ledger category. A number of the GL analytics are listed below that include testing for:

- Suspicious weekend or holiday JE's
- Duplicate JE's: same dollar amount and GL account
- Benford analysis on JE amounts
- JE's entered close to the GL period cut-off
- Outlier JE amounts
- Even dollar amount JE's
- JE's with missing or invalid attributes
- JE validity: JE's over the stated approval limits, unauthorized JE's, JE's changes made by authorized user

As indicated by the number of analytics available from only two of the software providers in this marketplace, as well as an ever-increasing number of analytics being developed, the capabilities of an app recommender system are critical to ensuring an

appropriate analytic selection mechanism is provided to the user. When the sheer volume of available apps, or those that will be available in the future, is coupled with the potential audience for ENHANCE, that is, a literal army of armchair auditors each with possibly unique requirements, providing a highly sophisticated and capable tool to successfully undertake these analytic requests is critical.

Dai et al. (2014) propose an app recommender system (ARS) to support auditors, especially less experienced auditors, when selecting audit apps for specific audit client engagements. If an app recommender system is proposed to support (experienced) auditors then it seems appropriate to provide those capabilities to the (inexperienced) ENHANCE user, that is, a constituent who wishes to understand how their government is spending its money. An app recommender system is one of the elements described in the design of the ENHANCE framework (Section 4.2.). The development of an app recommender system to support ENHANCE may be the most critical element in the overall development effort. The success of the app recommender system in providing analytic apps that can produce with meaningful information may spell the success or failure of ENHANCE as an analytic tool.

4.14.1. Recommender Systems defined

Recommender Systems (RSs) are software-based tools and techniques that offer suggestions for users as to which 'items' (or services) that the user has expressed an interest may provide the greatest utility to the user (Ricci et al., 2011). Recommender systems support individuals who lack personal experience or competence when evaluating a number of alternatives (Ricci et al., 2011). The onset of internet shopping, where merchants can list literally an overwhelming number of items to choose from, was

one of the first areas that recommender systems were developed for. Recommendations are provided in the form of ranked lists based on the users' preferences and constraints. These recommender systems typically draw from explicit actions of the user, such as product ratings as provided by the user, or the user visiting specific websites. Recommendations are also be driven by observations of human behavior, such as where individuals make decisions by following suggestions made by colleagues, friends, or relatives.

Recommender systems gather different types of data in order to develop their recommendations. The more basic recommender systems rely on user ratings of items to submit recommendations. More sophisticated recommender systems utilize descriptive information relevant to both users and items, and any constraints between the users and items. The recommender system refers to three elements: users, items, and transactions, which represent the relation between users and items. A transaction consists of a record of the events that transpired between the user and recommender system. The transaction may include user feedback/ratings, typically expressed in numeric form, such as ratings on a scale of 1-10, ordinal, reflected as like-dislike or good-bad scales, or binary format, as a yes or no.

The goal of a recommender system is to predict if a particular item is worth recommending, that is, what level of utility does the item provide, as compared to other items of a similar nature? Six approaches to developing recommendations are identified as (Ricci et al., 2011):

- Content-based relies on items the user liked in the past to suggest new items at this time
- Collaborative filtering (CF) relies on what users with similar tastes liked in the past
- Demographic-based recommendations on the user's demographic profile
- Knowledge-based, that is, specific domain knowledge about how particular features meet the user's needs and liking, encompass two approaches:
    - Case-based uses a similarity function that estimates how closely the user's needs match the recommendations
    - Constraint-based, which is similar to case-based but leverages a pre-defined knowledge base containing explicit rules to relate the user's needs with item features
- Community-based recommendations are based on the user's friends' preferences, in other words, a social recommender system
- Hybrid-recommender systems are combinations of the above that are designed to offset limitations in any individual approach, such as exists with CF in that it cannot recommend new items that have no existing ratings

Early recommender systems used algorithms to leverage recommendations as provided by a user group whose members were identified as similar to the target user in terms of demographics and tastes. This is termed collaborative filtering (CF) and is based on the concept that if the target user previously mimicked the actions of a user group in the past with respect to item selections, then other recommendations from the community should be relevant and of interest.

4.14.2. An Audit App Recommender System described

The app recommender system proposed by Dai et al. (2014) suggests appropriate audit apps based on three components: audit standards, audit clients, and auditors' preferences on audit apps. Providing personalized suggestions for a particular auditor allows the auditor to more efficiently and effectively analyze the data.

The audit app recommender system is envisioned as an independent application that collects apps from the entire app marketplace and is not associated with any particular audit app vendor or vendors. Recommendations based on audit standards are generated by creating a structure that categorizes audit apps by industry, business cycle, accounts, audit assertions, and audit objectives. Recommendations based on audit clients are created using a two-stage collaborative filtering (CF) approach based on the premise that if an app has been used by auditors for similar clients, it is likely that app is appropriate for the new client. Recommendations based on auditor preferences take into consideration that auditors may prefer certain app vendors, versions, and/or user interfaces, that is, their historical preferences (Dai et al., 2014). Recommendations based on auditor preferences are developed using a CF approach similar to that used for recommendations based on audit clients. The app recommender system creates a final score for each audit app by combining the results of the previous steps and presents its recommendations to the auditor.

The proposed design of the app recommender system is based on an existing, constant number of audit apps from which it chooses. In reality this is not the case as new apps are being developed as analytics software vendors upgrade their products. And as the use

of audit apps grows in popularity, new apps will be developed with increasing frequency to meet new needs. The issue that arises is how to classify these new apps within the recommender system. One solution can be that of requesting app vendors to classify their new apps with respect to the audit objectives the app is intended to address, as they are introduced. This not only benefits the app recommender system and its users but also the vendors who by classifying their apps under such a scheme provides for increased visibility and use of the app.

Although the design of an app recommender system for ENHANCE is outside the scope of the present research, it is envisioned that the recommendation selection process would be based on the particular data attributes the user selects for the analysis, and supported by user preferences with respect to their previous analytic requests, similar to that presented above for recommendations based on auditor preferences.

4.14.3. Design Science Research Methodology to facilitate developing an App Recommender System for ENHANCE

The development of an app recommender system tailored to support both the ENHANCE framework and the needs of the user community is critical for the success of ENHANCE. The ENHANCE framework is envisioned to be robust in technical terms, incorporating state-of-the-art analytics, presentation tools and especially an advanced app recommender system. What is an appropriate procedure to drive development of an app recommender system to support ENHANCE? Design Science Research Methodology, as described earlier in this paper with respect to the design of the ENHANCE framework, can also support design of the app recommender system component of ENHANCE.

The development of an app recommender system to support the ENHANCE framework is outside the scope of the present research, but an example of how the process can occur using DSRM as a foundation is presented. Table 43 summarizes the application of the DSRM template to include a proposed definition for each activity as appropriate for the development of an app recommender system to support the ENHANCE framework, and the proposed knowledge tools as appropriate for each step (Geerts & O'Leary, 2014).

Development of a complete DSRM template is outside the scope of this present research; for this research only the first step was undertaken. The Definition of the Problem Identification step is derived from the design concepts for ENHANCE as presented in this paper. The Knowledge Base relies on the literature review conducted for this paper.

The remaining five steps are shown in italics to indicate the information presented is not definitive but is only a proposal of what may be appropriate for each of the remaining steps. The Definition of Objectives step may require questionnaires/surveys to elicit responses from potential users as to what they perceive as an appropriate app selection procedure, and then translate this into logic to support app selection drivers. The Design and Development step should be based on completion of a detailed definition/design document covering the anticipated app selection process, as well as a further literature review with respect to actual app recommender system design and usage. The Demonstration step will include development of a prototype app recommender system and it is recommended this be presented to a focus group of target users for their review and comment. The final step, Communication, should include presentation of the developed app recommender system to a sample of target users and data providers for their review, testing, and approval.

DSRM Activity	**Definition**	**Knowledge Base**
Problem identification and motivation	The ENHANCE framework requires a sophisticated app recommender system to support the users in fulfilling their analytic requests	Literature review of the capabilities of typical recommender and especially app recommender systems
Definition of the objectives of a solution	*Understand the search and selection process required to identify out of the total assortment of analytic apps available to choose from*	*Understand how user inputs into the ENHANCE framework translate into drivers for analytic app selection*
Design and development	*Design an app recommender system that can provide appropriate app selections to support user needs*	*Study examples of functioning recommender systems or appropriately-designed prototypes*
Demonstration	*Definition of the specific logic to be incorporated into the app recommender system in order to identify the most appropriate analytic apps*	*Develop a prototype app recommender system as it would operate within the ENHANCE framework*
Evaluation	*Evaluation of the success of the app recommender system in providing appropriate analytic apps to satisfy the user's requirements*	*Perform an analysis of the proposed analytic apps based on recommendations derived from the recommender system to determine if the proposed analytics provide reasonable and valuable analytical results to satisfy the user's requirements*
Communication	*Present to appropriate audiences: data providers, ENHANCE users, and potential users*	*Data providers and target users*

Table 43: Proposed DSRM template to develop an app recommender system for ENHANCE

Following a proven design methodology such as DSRM will facilitate the most appropriate design for an analytic app recommender system to support the ENHANCE

framework. Development of a complete DSRM template and app recommender system is an activity for future research.

4.15. An Example of ENHANCE in Practice

The following is an example of employing the ENHANCE framework to provide a user with tailored information from which they can undertake a specific action or decision (or not).

Arthur Armchair is a purveyor of computer products and services. He has an established business in Austin, TX and his firm does provide products and services to the City of Austin. Arthur is interested in understanding what other firms are providing similar services to the City of Austin. Armed with this information Arthur can identify if there are opportunities for his 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 has been made aware of the availability of the ENHANCE framework and the capabilities it encompasses. Arthur accesses ENHANCE and requests access to the Austin checkbook dataset. ENHANCE provides Arthur a list of the attributes available and from this list Arthur selects fiscal year (FY_DC), department name (DEPT_NM), object name (OBJ_NM), and amount (AMOUNT). 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 fiscal year = 2015, department name = Communications & Technology Management, object name = Computer Hardware, Computer Software, and Computer Supplies. Arthur wishes to obtain results that are summarized by vendor instead of detail line item data as is provided in the dataset, so he selects the Data Summarization analytic, by vendor, that 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 who supply over $100,000 of goods and services during 2015 so he first sorts the data by amount in descending amount and then truncates the output and the results are presented in Table 44:

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 AMONT:	5,005,525.99

Table 44: Results from ENHANCE summarization analytic for Austin, TX, as adjusted by user to reflect >$100,000 vendors

Arthur can use this information to support further investigations. He can investigate the listed firms to identify what types of goods and services they provide in general by reviewing their websites and/or social media postings to determine if he can compete with them to provide similar goods and services.

In another example Arthur is investigating whether to expand his business beyond Austin. He would like to open a business in Las Vegas, NV and hopefully provide goods and services to the City of Las Vegas. Arthur wishes to identify the vendors in Las Vegas that are providing goods and services to the city that are similar to what he provides. 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 selects fiscal year (FISCAL_YEAR) = 2015, department (DEPARTMENT) = Information Technology, and fund (FUND_TYPE) = Computer Services, and requests an analytic that summarizes amount by vendor. ENHANCE provides him with EXCEL-formatted output. Arthur is interested in the more significant IT vendors to the city and first sorts the output by amount in descending amount and then truncates the data to display only those vendors providing over $100,000 of goods and services, as shown in Table 45.

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 45: Results from ENHANCE summarization analytic for Las Vegas, NV, as adjusted by user to reflect >$100,000 vendors

In reviewing these results Arthur notes the top two vendors are not of interest to him as Oracle most likely is providing ERP or similar software to the city, and Verizon is most likely providing communication services to the city, and he removes these from the listing, resulting in Table 46.

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

Table 46: User-adjusted results from ENHANCE summarization analytic for Las Vegas, NV

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 somewhat 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[36] and found that Austin's population, at approximately 885,000 is about 50% greater than Las Vegas' at approximately 603,000. 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

[36] See: www.freebase.com for population information for 2013 for each city

population. Arthur will want to undertake additional analyses before committing to launch a new business venture in Las Vegas, 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.

Some governmental open data portals do provide access to contract information (Dai et al., 2015). Although the ENHANCE framework does not include analytic capabilities over contract data at this point in its design, a review of contract data in a scenario such as the one described does provide additional information on the vendor/governmental entity relationship and would assist a user such as Arthur in gaining a better understanding of that relationship.

In another possible scenario, if Arthur is acting in the role of a civic watchdog or armchair auditor this information on Las Vegas IT expenditures might encourage him to pursue more information on Las Vegas IT expenditures to determine if there is a possibility that the city is expending more resources than may be required.

4.16. Discussion

This chapter presents the concept of an ENHanced ANalytic Constituent Environment (ENHANCE) framework, facilitated by the availability of open government data, that fulfills the reporting and analytic requirements of the various governmental stakeholders, such as citizens, analysts, bond investors, creditors, auditors, and oversight officials. The ENHANCE framework acts upon the inputs from external agents to initiate the execution of analytic apps that produce the required analyses. The ENHANCE framework incorporates a variety of apps that support its operation, especially that of an app

recommender system to identify appropriate analytics apps to run. Data availability is generally not an issue, although the robustness of the data provided may limit possible analytics, as a number of governmental have launched data portals and others will be forthcoming In the cases investigated the availability of robust analytic tools that can provide meaningful analyses over that data have yet to be developed. Without proper analytics the value of the available data is limited. ENHANCE bridges that gap by providing that capability to the average user of government data in a form that does not require significant knowledge to undertake.

This research presents, by using existing technology, an example of the capabilities that a framework such as ENHANCE can provide the user. For example, a number of analytics have been executed over a representative sample of governmental expenditure data and the results presented in a manner that is easily understandable by the user.

The functional design of the ENHANCE user interfaces are defined and examples presented, as well as a definition of the underlying dataset and attributes that will support the ENHANCE user interfaces. Design Science Research Methodology is presented as one tool that can support the actual design of the ENHANCE framework.

As one of the more critical apps that allows the ENHANCE framework to function is that of an analytic app recommender system Design Science Research Methodology is again presented as one approach to drive the development of the app recommender system. A DSRM template is presented that includes a definition for each activity required for the development of an app recommender system to support the ENHANCE framework, and the knowledge tools as appropriate for each step in the development process.

The vision for the ENHANCE framework as presented in this research is limited as it is anticipated that the firm or organization that undertakes development of this sophisticated analytic tool will tailor the design to fit their business purposes and their anticipated user base. Topics will now be covered that are likely to be included in some form in the ultimate design of the ENHANCE framework, and if so included will greatly improve the usability of the tool.

An area for further discussion is related to the configuration of an ENHANCE user profile. The design presented in 4.12.1 provides for minimal information; only that required to identify a user and link that user to the analytic requests they have submitted. A much more robust user profile may be developed to identify relevant criteria for each ENHANCE user. Demographic data, including items such as level of education, job title, and field of employment (private, public, education), can benefit both the provider of the ENHANCE framework and the provider of the underlying dataset in understanding the participants who undertake an armchair auditor role with the ENHANCE framework and particular dataset. The ENHANCE provider can use this information to market the product to target user groups, and possibly further tailor ENHANCE and related products to fit the user. In this way ENHANCE can to an extent mimic online shopping sites such as Amazon that draw from user actions to recommend additional products. The governmental entity can likewise understand who is undertaking data analytics over their datasets and possibly identify high-level user profiles to describe their analytic users. Tracking analytic requests might also lead the data provider to identify additional datasets that may be of interest to their users. An additional benefit to the provider and users of the data is the identification of the most requested analytics, which can likely expedite the

users' analytic requests. Any tracking of user information should be balanced by a possible requirement for anonymity on the part of the user. The ENHANCE design presented in 4.12.1 includes an option for a user to access ENHANCE as a guest and provide no personal information. The decision on what level of user information to track, and how that information might be used, will be determined primarily by the firm or organization that develops and deploys the ENHANCE framework. The data providers may also request that specific user information be tracked.

With respect to the functionality provided by the ENHANCE framework and the app recommender system, user profile information as discussed above could be leveraged to improve the app recommender system. For the initial design of ENHANCE as described in Chapter 4, the app recommender system relies primarily on the user-selected data attributes for the specific dataset under analysis, and the particular apps the user has executed in past analyses to propose appropriate analytic apps for the current analysis. The recommendations can be greatly improved upon by including additional criteria.

A first step to accomplish this is to include historical analytic tracking information in the ENHANCE user profile. In addition to tracking the apps the user has requested in the past ENHANCE could also track the specific datasets accessed and attributes selected by the user. This information could supplant the previous app selections for the specific user by providing historical dataset and attribute information to assist the user in configuring their current request.

A second step in this process would be, using each user's previous dataset, attribute, and app selection criteria, to develop a dataset profile that would be provided to any user who

accesses the same target dataset for analysis. This information would indicate to these users the kind of analyses that prior users have requested from the specific dataset, that is, providing recommendations for analyzing this dataset. For a previous user of a specific dataset they would be provided not only their own historical preferences but also the preferences of other users of the same dataset and could be enticed into undertaking additional analyses.

A third step in providing sophisticated recommendations for ENHANCE users would be based on demographic data stored in the ENHANCE user profile. ENHANCE would match the demographic data of the user initiating an analysis with that of previous ENHANCE users that fit a similar demographic profile. A key consideration would be geographic location as it is anticipated that the ENHANCE user is interested in analyses of governmental entities in their geographic area such as state and municipalities located close to the user's physical location. Recommendations would be provided based on matching demographic data with previous ENHANCE users within the same geographic area. The user would be provided with datasets used previously, the attributes selected, and the analytics that were executed. An existing ENHANCE user would now be provided with three levels of recommendations based on their historical preferences, preferences of users of the same dataset, and preferences of users that fit their demographic and location profile.

This third level of recommendations could most assuredly support an armchair auditor presence. Consider an example where publicity concerning spending patterns for a local municipality has attracted the attention of one or a few individuals. These initial armchair auditors, most likely existing EHNANCE users, will access the municipality's

dataset, configure, and launch appropriate analyses. Following these initial analyses additional citizens, possibly less familiar with ENHANCE, can nonetheless access ENHANCE and mimic the analyses previously executed. Now there exist a number of citizens each having successfully completed analyses and are able to judge for themselves either the propriety or impropriety of the expenditures. There are likely now additional citizens who have gained expertise in using ENHANCE and will continue their armchair auditor activities.

These potential refinements to the ENHANCE recommendation capabilities will not only improve the user interaction with ENHANCE but also demonstrate to the user the sophisticated nature of an analytic tool designed specifically for their use.

4.17. Limitations and Future Research

This present research activity has some limitations. This paper presents as examples of ENHANCE capabilities the results for a select number of analytics. Additional analytics need to be identified and incorporated into the ENHANCE framework in future research. Such analytics may include: Exceptional Exceptions analytics (Issa, 2013) that identify and prioritize exceptions and anomalies when faced with a large number of records, Cross-entity analytics that at present are limited by a lack of a common data taxonomy as states and localities have to support differing reporting requirements, KPI development which may be of greatest benefit to the governmental entities themselves, and Pattern Recognition analytics that would include development of mathematical formulas defining the data patterns

The detailed design and development of the ENHANCE framework itself has yet to be undertaken. A complete development of the DSRM template to support the design of the ENHANCE framework and app recommender system will need to be undertaken, using the examples presented above. The use of DSRM, as presented for the development of the ENHANCE framework and app recommender system, will likely need to be replicated for other elements of the ENHANCE tool. The proposed design for the user interfaces and underlying datasets will need to be reviewed for accuracy and completeness.

What may be the greatest capability of and benefit from a robust analytic tool such as ENHANCE may be that of providing comparative analyses across governmental entities. As noted in the comparison of detailed, expenditure data as provided by two governmental entities the data does vary significantly in the number and type of elements provided. For analytic analyses within a particular entity this would not be an issue as the data is consistent within the entity, but cross-entity analyses can be severely hampered. A formal taxonomy that can be used by all governmental entities for presenting data on their respective portals would greatly enhance the ability to perform comparative analytics across entities. There are other research activities[37] that are undertaking the design of a prototype taxonomy for governmental open data.

Future research can investigate incorporating analytic capabilities into the ENHANCE framework to allow for comparing actual, historical expenditures to budget data prepared in advance of the expenditure cycle under investigation. As budgets are not prepared at a

[37] See: A tale of two standards: A Pan-American study on the feasibility of XBRL data standards for U.S. and Brazilian federal and local governments (working paper) By Deniz Appelbaum, Hussein Issa, and Stephen Kozlowski

detailed transactional level, and the attributes may differ from those provided with the actual expenditures, a sophisticated matching app will likely need to be developed.

The ENHANCE framework is presented in this research as a tool that provides analytic capabilities over governmental expenditure data. There are numerous other datasets provided in many of the open data portals, and for those whose data is presented using a defined taxonomy, the ENHANCE concept can be applied, for example, governmental purchase contracts (Dai et al., 2015).

One aspect relevant to an eventual development of a robust analytic tool such as the ENHANCE framework is whether it will be provided to the public. Ideally such a tool would be available for use without charge, but that may not be the case. The perspective of the organization deploying the tool may impact its projected usage; if deployed by a government watchdog group, for example, the focus may be on identifying fraud or instances of favoritism in contracting. In this instance the intended usage would align with that of the armchair auditor movement. If governmental entities themselves deploy the tool then the focus may be on simpler analytics that do not present examples that might raise constituent concerns. In this case the intended usage would be similar to that for the underlying open data itself, that is, more altruistic in nature by satisfying a constituent desire.

References

Alwan, L. C., & Roberts, H. V. (1988). Time-series modeling for statistical process control. *Journal of Business & Economic Statistics*, *6*(1), 87-95.

Arnott, D., & Pervan, G. (2008). Eight key issues for the decision support systems discipline. *Decision Support Systems*, *44*(3), 657-672.

Austin Business Journal: http://www.bizjournals.com/austin/news/2015/02/02/database-searchcity-of-austin-salaries.html

Benbasat, I., & Zmud, R. W. (1999). Empirical research in information systems: the practice of relevance. *MIS quarterly*, 3-16.

Byrnes, P. E. (2015). *Developing Automated Applications for Clustering and Outlier Detection: Data Mining Implications for Auditing Practice* (Doctoral dissertation, Rutgers University-Graduate School-Newark).

Cameron, D. (2010). PM's Podcast on Transparency. Available at: https://www.gov.uk/government/news/pms-podcast-on-transparency

Chan, K. S., & Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of time series analysis*, *7*(3), 179-190.

Cheng, R., Xia, Y., Prabhakar, S., Shah, R., & Vitter, J. S. (2004, August). Efficient indexing methods for probabilistic threshold queries over uncertain data. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30* (pp. 876-887). VLDB Endowment.

Dai, J., Krahel, JP, & Vasarhelyi, M. (2014). Which Audit App(s) Should Auditors Use? An Exploratory Study of Using Recommender Systems for Audit App Selection. Rutgers Working Paper.

Dai, J., Li, Q., & Vasarhelyi, M. (2015). Design of Audit Apps for Armchair Auditors to Analyze Government Procurement Contract. Rutgers Working Paper.

Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The computer journal*, *41*(8), 578-588.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, *97*(458), 611-631.

Frank, M., & Oztoprak, A. A. (2015). Concepts of Transparency: Open Data in UK Local Authorities. In *CeDEM15: Conference for E-Democracy and Open Government* (p. 185). MV-Verlag.

GAO: www.gao.gov/new.items/d05734sp.pdf

Geerts, G. L., & O'Leary, D. E. (2014). A supply chain of things: The EAGLET ontology for highly visible supply chains. Decision Support Systems, 63, 3-22.

Granovetter, M. (1978). Threshold models of collective behavior. *American journal of sociology*, 1420-1443.

Hoyle, S. D., & Langley, A. D. (2011). Spatial size data stratification for length-based stock assessments. *WCPFC SC7 SA IP-9, Pohnpei, Federated States of Micronesia*, 9-17.

Issa, H. (2013). *Exceptional exceptions* (Doctoral dissertation, Rutgers University-Graduate School-Newark).

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, *31*(3), 264-323.

Lee, W., & Xiang, D. (2001). Information-theoretic measures for anomaly detection. In *Security and Privacy, 2001. S&P 2001. Proceedings. 2001 IEEE Symposium on* (pp. 130-143). IEEE.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision support systems*, *15*(4), 251-266.

Nassar, S., Sander, J., & Cheng, C. (2004, June). Incremental and effective data summarization for dynamic hierarchical clustering. In *Proceedings of the 2004 ACM SIGMOD international conference on Management of data* (pp. 467-478). ACM.

NIST: (www.itl.nist.gov/div898/handbook/pmc/section4/pmc41.htm)

O'Leary, D. E. (2015). Armchair Auditors: Crowdsourcing Analysis of Government Expenditures. *Journal of Emerging Technologies in Accounting*.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45-77.

Power, D. J. (2009). *Decision support basics*. Business Expert Press.

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. "O'Reilly Media, Inc.".

Ricci, F., Rokach, L., & Shapira, B. (2011). *Introduction to recommender systems handbook* (pp. 1-35). Springer US.

Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision support systems*, *33*(2), 111-126.

Sprague Jr, R. H. (1980). A framework for the development of decision support systems. *MIS quarterly*, 1-26.

Sturgis, P. (2004). Analysing complex survey data: Clustering, stratification and weights. *Social research UPDATE*, (43).

Treiman, D. J. (1970). Industrialization and Social Stratification*. *Sociological Inquiry*, *40*(2), 207-234.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, *84*(405), 231-240.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association*, *93*(443), 1188-1202.

Worthy, B. (2015). The Impact of Open Data in The UK: Complex, Unpredictable, and Political. *Public Administration*.

Yager, R. R., & Petry, F. E. (2006). A multicriteria approach to data summarization using concept ontologies. *Fuzzy Systems, IEEE Transactions on*, *14*(6), 767-780.

Chapter 5: Introducing an Audit Ecosystem

5.1. Introduction

Computer-based audit analytic tools were first designed and put into practice over twenty years ago (Vasarhelyi & Halper 1991) and since then many computer-based continuous audit (CA) tools have been developed and implemented (Kogan et al. 1999). Concurrent with the initial development of CA tools there were optimistic predictions about how CA would transform auditing which have not been realized, even though the technology supporting CA is more advanced than that envisaged in 1991 (Alles, et al.2008).

The development of CA tools is one of few instances where a significant innovation in accounting practice has been driven by the academic community (Alles et al. 2008). Academic researchers continue to play an important role in CA development. Academic researchers who create a conceptual model of CA ensure that it becomes a true audit methodology, and not simply a collection of disparate technologies. Also, academic researchers can conduct innovative implementations without facing the challenges practitioners will have as they turn to CA, for example, the process of reengineering the audit practice to adapt it to CA (Alles et al. 2008).

The purpose of this chapter is to provide an initial definition for an audit ecosystem, stated at this point as a holistic approach to the design and development of a technology-driven framework to provide overall management and control of the audit technology components employed, and coordination of the activities of the participants involved. An audit ecosystem is intended to address ongoing changes to the environment in which the automated audit functions operate as that environment continues to evolve. The audit

ecosystem is simply a configuration of a ‘digital’ ecosystem to support computer-driven audit techniques. Zuiderwijk et al. (2014) present an overview of the characteristics of digital ecosystems. The design characteristics of a digital ecosystem include, among others (Zuiderwijk et al., 2014):

- A functioning whole
- Structured as a multi-level and multi-dimensional entity
- Developed primarily through local specializations and adaption
- Influenced by:
    - An information, communication, and networking technology infrastructure
    - e-government, e-business, and e-society
    - Human resource development
    - A policy and regulatory environment

The development of an audit ecosystem is the natural progression in the deployment of computer-based Continuous Auditing/Continuous Monitoring (CA/CM) tools, and as with earlier CA development efforts this activity is preferably undertaken in the academic community.

The capabilities that an audit ecosystem provides can also be incorporated with the ENHANCE framework to benefit the user community and the evolving nature of the environment in which it operates. As the analytic capabilities that ENHANCE provides form the basis for constituent audits of governmental expenditures, the incorporation of the capabilities of an audit ecosystem are a natural fit.

5.2. Why an Audit Ecosystem?

The auditing profession has reached a critical point in its existence due to not only advances in information technology but also the new approaches many businesses are undertaking to remain competitive in today's real-time economy (Byrnes et al., 2015). Auditing has not kept pace with the real-time economy by adhering to approaches and techniques that appear outdated (Byrnes et al., 2015). Byrnes et al. (2012) also note that organizations are not reaping all the benefits that CA/CM technologies can provide, particularly external audit applications. Bumgarner & Vasarhelyi (2015) propose that the dramatic advances in technology and analytics necessitate changes to processes, and ultimately the human element, in the audit environment. In discussing one aspect of the audit procedure, the traditional, discrete risk assessment procedure, and the fact that risks are not monitored in an on-going manner, Stewart (2015) proposes a more adaptable approach to both monitoring risks and expediting changes to audit procedures and testing routines to provide for higher quality audits.

In light of these assessments of the current state of the audit profession there appears a need to develop a new protocol to not only support the audit practice but ensure its viability into the future. Businesses have recently been described as a type of ecosystem that include information flows, participants, and IT infrastructure, all operating in an evolving timeframe (Bumgarner & Vasarhelyi, 2015). Mike Wons, CTO of the State of Illinois, in describing the Illinois "FIRST"[38] strategy states that it "helps us create a

[38] See: http://www.govtech.com/state/Accelerate-Illinois-Modernization-Day-6-Previewing-the-Illinois-FIRST-IT-Strategy.html?utm_medium=email&utm_source=Act-On+Software&utm_content=email&utm_campaign=5%20Tips%20to%20Help%20State%2C%20Local%20Governments%20Combat%20Cyberthreats%20%7C%20Open%20Data%20Census%20Provides%20a%20To-Do%20List%20for%20States&utm_term=Accelerate%20Illinois%5Cu2019%20Modernization%2C%20D

unique *ecosystem* that allows agencies across the state to improve its operations and get closer to the resident and business customer with minimal overhead and maximum reach". The advent of cloud computing has provided the technical foundation or a number of businesses to develop cloud-based ecosystems, namely Amazon, Google, and Apple (Bumgarner & Vasarhelyi, 2015). Given that many consider firms such as these among the forefront of technology innovators, should an ecosystem concept be considered to bolster the evolution of the audit profession to leverage the available technology?

Developing a technology-based ecosystem to support a technology-driven audit function provides a much-needed infusion of state-of-the art technology into the audit practice in order to better support clients functioning within today's real-time economy. An ecosystem incorporates current approaches and techniques and evolves them as business scenarios change. An ecosystem can support CA/CM technologies and better manage their usage than is provided with simply human intervention. An ecosystem also provides the adaptability that Stewart (2015) proposes, and not only for risk assessments but other audit functions as well. Byrnes et al. (2015) present possible adjustments that may impact future audit approaches, and among them changes in the timing and frequency of audits, increased education in technology and analytic methods, and adoption of full population examination in place of sampling. Utilizing an ecosystem provides the basis for real-time, continuous audits, acts as a buffer to the more complex

---

ay%206%20%5Cu2014%20Previewing%20the%20Illinois%20%5Cu201CFIRST%5Cu201D%20IT%20Strategy

"The Illinois "FIRST" strategy is based on operating as a preferred one stop, one entity and one voice across the state for Information Technology. Focused on "SMART Citizen" enablement we must collectively improve services and experiences of all residents, businesses, visitors and "everything" in, around and across the state".

technology underlying the ecosystem and thus lessen the requirement for significant technical education, and also allows for audits of complete data populations.

The remainder of this chapter is organized as follows: Section 5.3 presents a literature review describing digital ecosystems and software agents. Section 5.4 describes the audit ecosystem design proposal including attributes, features, and agents, and Section 5.5 presents audit ecosystem components represented in diagrammatic form. Section 5.6 presents the ENHANCE ecosystem in diagrammatic form. Section 5.7 describes an Open Data Ecosystem, a form of a business ecosystem, from which attributes to support the ENHANCE framework can be modeled, Section 5.8 describes applying Design Science Research Methodology to guide development of an ecosystem for the ENHANCE framework, and Section 5.9 provides a Discussion.

5.3. Literature Review

5.3.1. Digital Ecosystems Defined

A digital ecosystem has been defined to include interconnected and interrelated, independent agents existing in a digital environment, who interact as a functional unit and are linked together through related actions, information, and transaction flows (Zuiderwijk et al., 2014). Digital ecosystems are attributed to be robust, scalable, self-organizing, and encompass decentralized peer-to-peer networks that are comprised of distributed agents (Zuiderwijk et al., 2014).

The concept of a digital ecosystem originated in the early part of the 21st century, triggered by the European Commission-sponsored Go Digital initiative, whose aim was to boost the adoption of Information and Communication Technologies (ICT) by

European small and medium-sized enterprises (SMEs) as ICT was considered to be a major contributor to economic growth and efficiency. (Nachira et al., 2007).

Digital ecosystems have been defined as:

- Loosely-coupled, demand-driven, domain clustered, agent-based collaborative environments where each participant is proactive and responsible for its own benefit (Chang & West, 2006).
- Distributed, adaptive, open socio-technical systems, with properties of self-organization, scalability and sustainability (Briscoe & De Wilde, 2009).

Digital ecosystems also encompass decentralized peer-to-peer networks that form an underlying tier of distributed agents (Zuiderwijk et al., 2014).

Uden et al. (2007), in their research with respect to e-learning ecosystems, describe characteristics of a digital ecosystem that can also be incorporated into an audit ecosystem definition:

- An ability to utilize new forms of electronic interaction to provide high connectivity and electronic handling of numerous forms of information
- The capacity to offer multiple channels for buying and selling of services, for example, and the ability to capture and utilize business intelligence from various information sources
- An integration of business and human efforts supported by the advanced information systems within a digital ecosystem to facilitate close interaction between the participants

- A provision for cross-disciplinary interaction and engagement, that is, a mix of expertise that preserves and enhances productivity (Uden et al., 2007).

Service-Oriented Computing (SOC), as described by Sacha et al. (2007) in their research on a service-oriented peer-to-peer architecture, is an example of software applications modeled as collections of loosely coupled, interacting services that communicate using standardized interfaces, data formats, and access protocols. The primary advantage of SOC is that it enables interoperability between different software applications running on a variety of platforms and frameworks, potentially across administrative boundaries (Sacha et al., 2007). They continue by defining a Service Oriented Architecture (SOA) in terms of three elements: a service provider that publishes and maintains a service, a service consumer that uses the service, and a service registry that allows service discovery by prospective consumers. An audit ecosystem may very likely include an SOA component to support interactions with users and providers.

The definition of a software agent, as provided by Briscoe and De Wilde (2009), is that of a software component that acts autonomously for a user as an agent to meet the objectives the agent was designed for. In their research on digital ecosystems Briscoe and De Wilde (2009) continue by defining a multi-agent system (MAS) as a system composed of several software agents that when acting collectively can achieve goals that are otherwise unachievable individually. Within the design of an audit ecosystem, and when considering agents as referring to CA/CM agents (tools), this multi-agency concept can be applied.

Pranata et al. (2011) present a distributed mechanism to provide for resource protection in a digital ecosystem. Interaction and collaboration between digital ecosystem entities challenge the ability to protect both resources and information and to address this a robust mechanism is required to ensure only authentic entities can access the appropriate resources, and this mechanism must also maintain the confidentiality and integrity of resources over an untrusted network (Pranata et al., 2011). Security concerns within an audit ecosystem also need to be addressed, possibly using a similar structure.

Foon and Yen (2011) elaborate on the development of a corporate knowledge ecosystem designed to create and propagate knowledge as well as encourage learning among all the entities in the ecosystem. Such knowledge ecosystems can create, utilize, and capitalize on the knowledge resources in their habitats as well as creating knowledge in their domains by conducting research and experiments, utilizing it via the sharing of knowledge and commercialization of new concepts, theories and innovations, and capitalizing on it by gaining competitive advantage (Foon & Yen, 2011).

5.3.2. Software Agents Research

Software-based agent research is generally accredited with beginning in the 1980's (Nwana & Ndumu, 1999). The goal in the development of agent-based software was to create software with the ability to interoperate, that is, programs that can exchange information and services with other programs and thus together solve problems that neither can address individually (Genesereth & Ketchpel, 1994). The application programs that were developed comprised software components that communicated with their peers, and communicated by exchanging messages in an agent communication language (Genesereth & Ketchpel, 1994). Maes (1994) describe computer agents that

engage in a cooperative process with human counterparts where both initiate communication, monitor events and perform tasks. These agents provide the ability to mask the complexity of difficult tasks, perform tasks on behalf of the user, train or teach the user, assist users to collaborate, and monitor events and procedures (Maes, 1994).

As agent development progressed the following classification scheme emerged:

- Gopher agents: Execute simple tasks utilizing pre-specified rules and assumptions
- Service performing agents: Execute a well-defined task as requested by the user
- Predictive agents: Volunteer information or services to a user without being prompted (Jennings & Wooldridge, 1996)

Wooldridge and Jennings (1995) describe the following attributes of an agent:

- Autonomous: Solve problems without direct human or agent intervention and are able to exhibit a problem-solving capacity
- Social ability: Interact with other agents and humans to complete their own problem solving and assist others as required
- Responsiveness: Perceive their environment and react to changes in that environment in a timely fashion
- Proactiveness: Take the initiative and exhibit goal-directed behavior when appropriate

Jennings and Wooldridge (1996) define two additional characteristics that differentiate software agents from other IT disciplines such as AI , distributed computing, or object-oriented systems:

- Autonomously complete high-level tasks that are delegated to them
- Determine the objective of a newly delegated task, evaluate how to meet the objective effectively, and subsequently perform the required actions

An ‘agent-based collaborative environment’ is one of the attributes defining a digital ecosystem as presented by Chang and West (2006) and they expand on the concept of a collaborative environment by noting that human agents and software agents together effect both interaction and knowledge sharing within an ecosystem. The inclusion of software agents in ecosystem design is quite accepted, and software agents likewise have a significant role in the definition of an audit ecosystem.

Papazoglou (2001) describes using software agent technology to help bridge the gap between a flexible design and usable e-business applications, and proposes that in a multi-agent e-business environment it is necessary to organize agents into categories depending on functionality and competencies:

- Application agents: Each agent is specialized to single area of expertise, provides access to the available information and knowledge sources in that domain, and works cooperatively with other agents to solve a complex problem in that domain
- Personal (or interface) agents: Works directly with users to help support the presentation, organization, and management of a user profile, requests, and information collections. A personal agent provides its user easy and effective access specialized services and information widely distributed on the Web. The user’s agent observes and monitors the actions taken by the user in the interface and suggests better ways to perform the task. These agents can assist users in

forming queries, finding the location of data, and explaining the semantics of the data, among other tasks

- General business activity agents (Papazoglou, 2001): Perform a large number of general commerce support activities including search agents that navigate effectively through fragmented online electronic information and services in order to find trading partners and items of interest, negotiation agents that negotiate on behalf of a buyer or seller, billing agents, marketing agents to market product and services on the Internet, or legal agents that advise on issues surrounding privacy, taxes, export controls, and also:
    - Information brokering agents (Papazoglou, 2001): Provide facilities such as locating information on Web sources or other agents that are required to solve a common problem. Brokering agents have the ability to maintain, update, and access distributed directory services that list products and business services as well as employing navigation services such as maintaining hyperlinks, advanced keyword and context search agents, including:
        - Search agents: Access the network looking for particular kinds of information, filter it, and return it to their users. Search agents are designed to mitigate the information overload caused by the availability of large amounts of poorly catalogued business data
        - Information agents: Perform the role of managing, summarizing, manipulating or collating information from many

diverse distributed sources. The information is then collated and sent back to the user

- Negotiation and contracting agents (Papazoglou, 2001): Negotiate the terms of a business transaction with regards to exchange and payment. Terms may cover delivery, refund policies, arranging for credit, installment payments, copyright or license agreements, usage rights, and distribution rights, among others. Contract terms can address liabilities, forms of payment, terms of payment, billing and payment instructions, delivery instructions, return policies, methods of dispute resolution, and so on. Contracts can be negotiated with respect to prices, terms of payment, penalties, necessary documentation, credit checks, required insurance, or collateral or margin

- System-level support agents (Papazoglou, 2001): Function on top of the distributed objects infrastructure and are provided with access not only to other application objects but also to such facilities as transaction processing, permanent object storage, and event services, among others. Some of the advanced functionality that is required to be provided in support of e-commerce are described here:

    - Planning and scheduling agents: Form a plan that specifies the future actions and interactions for each agent. Typically, in e-business applications an agent may act as the group planner for a cluster of agents surrounding an application agent, for example, the procurement agent. The planning agent forms a plan, which it uses to coordinate the

other agents. The plan specifies how agents coordinate in a multi-agent planning system and also identifies all actions and interactions of agents. To work cooperatively in dynamic situations, planning agents must be able to cooperate with other agents despite their inconsistent views of planned actions and interactions, and rely on artificial intelligence (AI) technologies such as fuzzy systems, neural networks, and genetic algorithms.

- Interoperation agents (Papazoglou, 2001): Support e-business processes that require legacy systems and transactional software components to interwork with new applications and systems. For example, in a value-chain workflow application information that may be acted upon by the various application agents may originate from different information systems that provide such services as:
    - Entry to an enterprise resource planning (ERP) system checking inventory for the products described in the procurement order
    - Entry to a distribution/shipping system using the customer address and delivery condition information to schedule a delivery. A translation agent can be used here to provide a common information model based on the underlying information models of each application as well as translating business rules and transactions for partners to incorporate within their own systems

- Business transaction agents (Papazoglou, 2001): Collect, manage, analyze, and interpret commercial data to make more intelligent and effective transaction-related decisions. Examples include collecting business references, coordinating and managing marketing strategies, determining new product offerings, granting/extending credit, and managing market risk. When applied to e-commerce transactions, business agents can simplify the processing, monitoring and control of transactions by automating a number of activities. Agent support for e-commerce business may, for example, include controlling the workflow governing a set of electronic transactions or monitor and enforce the terms and conditions of electronic contracts
- Security agents: Provide the security services required for the conduct of e-business, such as, the collection of commercial data only from trusted and controlled sources. Agent support for secure e-business can be segmented into five distinct categories: authentication, authorization, data integrity, confidentiality, and non-repudiation (Papazoglou, 2001).

5.4. Ecosystem Proposal

5.4.1. Attributes

Troubleyn et al. (2013) define the requirements for a flexible QoS Framework and these are also descriptive of attributes of an audit ecosystem:

- Adaptive: The audit ecosystem must accommodate data from various sources and differing structures, and support the CA/CM agents in completion of their tasks without interruption

- Scalable: The audit ecosystem must compensate for varying database sizes due to client size and scope of investigation, including big data, and support the CA/CM agents in performing in a most effective manner regardless of data size
- Distributed Approach: In order to complete analyses in a reasonable amount of time control activities must be decentralized so that the CA/CM agents can operate independently and undertake decision-making locally
- Support heterogeneity: Due to data variabilities in format and structure as well as various communication technologies and CA/CM agents with unique capabilities

5.4.2. Design Features

A definition of audit ecosystem features provides comprehension of what an audit ecosystem encompasses. Concepts presented by Barraca et al. (2013) describe several of these features:

- Collaborative and cooperative communication: Between CA/CM agents, especially when conducting multiple, related tests simultaneously which may be required when analyzing big data
- Autonomous (autonomic) control loops: Allow for the coordination of simultaneous activities as may occur when analyzing big data
- Management mechanisms: To support CA/CM agents in completing their tasks while reacting to unpredicted events or data conditions
- Policies (audit rules): That are continually updated and distributed so that all CA/CM agents have access to the latest versions. Rules may be altered as a result of specific data situations or anomalies encountered during the analysis

- Distributed knowledge: Allow for CA/CM agents to consider local rules as well as neighboring rules (for other CA/CM agents) which allows agents to act in coordination

The concept of swarm-bots (Swarm-bots), as noted by Arumugam et al. (2013) and as developed by the Future and Emerging Technologies program of the European Commission, does present another feature to include in the design of an audit ecosystem: swarm bot-like processes for large-scale investigations, such as with big data, so that multiple analyses/tests can occur simultaneously.

Based on the work of Uden et al. (2007) several of the digital ecosystem characteristics described can be incorporated into audit ecosystem features:

- Ability to utilize new forms of electronic interaction as available to allow for robust connectivity and electronic handling of information between CA/CM agents and the data under investigation
- Provides multiple channels for buying and selling of (audit) services, captures and uses applicable knowledge and business intelligence relevant to the audit activities from numerous sources
- Integrates business and human endeavors to facilitate close interaction between the participants in the audit process
- Incorporates cross-disciplinary interaction and engagement by offering a mix of expertise from other disciplines

The concepts of Service Oriented Computing (SOC) and a Service Oriented Architecture (SOA) as described by Sacha et al. (2007) can apply to an audit ecosystem design.

Consider that the audit ecosystem is available for ad-hoc usage by an audit client: the design includes an SOA element allowing a client to procure use of the audit services. The elements of an SOA as defined for an audit ecosystem are: the CA/CM tools provider and audit firm, the audit client uses the service, and an audit service registry that allows service discovery by prospective consumers (Sacha et al. 2007).

Briscoe and De Wilde's (2009) definition of multi-agent systems (MAS's) applies directly to the proposed audit ecosystem design in that it is envisioned that multiple CA/CM agents interacting with the specific data under investigation at a point in time with each agent not only achieving the objectives outlined for it but also in acting in concert all agents can achieve greater goals than can be achieved individually.

Distributed evolutionary computing (DEC) techniques, also presented by Briscoe and De Wilde (2009) can provide the foundation to enhance the learning capabilities of audit processes when new data situations are encountered.

The distributed resource protection mechanism presented by Pranata et al. (2011) is quite relevant to audit ecosystem requirements to ensure only appropriate entities are able to access the resources (data and agents). Such a mechanism is also required to maintain the confidentiality and integrity of resources when audit ecosystem activities occur over an untrusted network (Pranata et al., 2011).

Foon and Yen's (2011) description of a corporate knowledge ecosystem designed to create, utilize and capitalize on knowledge resources can be applied to provide storage of audit test criteria and evolutionary updates to those criteria.

5.4.3. Software Agents

The various agents presented by Papazoglou (2001) for use in a digital ecosystem are also applicable to an audit ecosystem:

- Application agents: CA/CM agents that are specialized to a single area of expertise and work in cooperation with other agents to solve complex audit problems are but one example of the many application agents that encompass an audit ecosystem
- Personal: (or interface) agents: Work directly with users, primarily client and provider staff, to help support the presentation, organization, requests, and information collections, such as providing user access to audit results
- General business activity agents: Perform a large number of general support activities such as search agents that navigate effectively through fragmented online electronic information in order to provide guidance to the CA/CM agents
    - Information brokering agents: Provide facilities such as locating information on Web sources or other agents that are required to solve a common problem, such as specialized agents to support CA/CM agents in addressing data anomalies, for example:
        - Negotiation and contracting agents: Negotiate the terms of a business transaction as regards to exchange and payment, as is required when transacting for audit services
- System-level support agents: Provides objects with access not only to other application objects but also to such facilities as transaction processing when acquiring audit services

- Planning and scheduling agents: a multi-agent plan is formed that specifies the future actions and interactions for each agent. Typically, an agent may act as the group planner for a cluster of agents surrounding an application agent such as to support multiple CA/CM agents analyzing big data simultaneously, for example
- Interoperation agents : Audit processes may require accessing information from legacy systems and CA/CM agents from separate providers
- Business transaction agents: Can support the process in acquiring and deploying new CA/CM apps into the audit ecosystem
- Security agents: Provide security measures for information, communications and data to/from the audit ecosystem (Based on Papazoglou, 2001).

5.5. Audit Ecosystem Components

Having described the characteristics of an audit ecosystem, these characteristics are now presented in a diagrammatic fashion. Figure 20 presents the many characteristics of an audit ecosystem in a single view, including attributes, features, and software agents as described above.

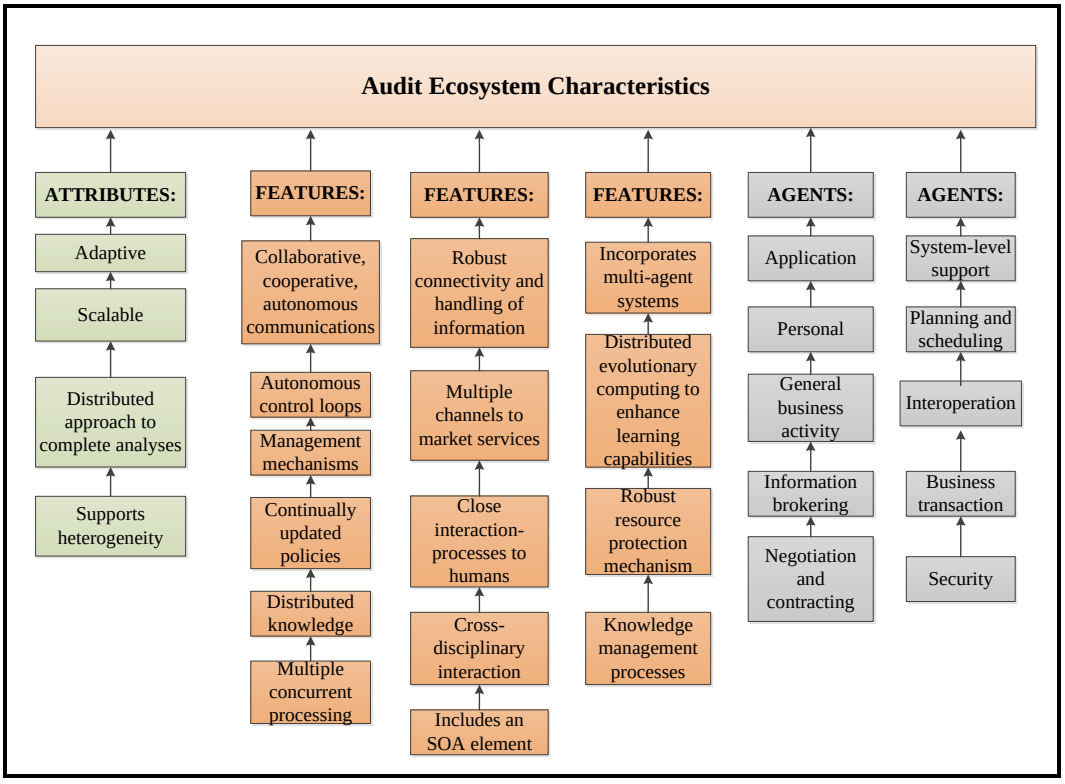

Figure 20: Ecosystem characteristics

A diagram of the external influences (participants) to an audit ecosystem, identified for this research as auditor, auditee, auditee data, audit standards, and audit analytic results/outputs, is presented in Figure 21. These represent the participants that are involved in a traditional audit that focuses on a single client and auditor/firm conducting the audit, the client's data which is the subject of the audit, and the results of the audit activity. An audit ecosystem supports the automated CA/CM tools that replace manual auditor activities with automated procedures that not only provide capabilities beyond that afforded by manual procedures, such as an audit of all the data and not just selected items, but also in a much more efficient manner. Given that technology is driving the new audit process, inputs are provided in a digital format: the data must be available in a

machine-readable format, and the resultant audit findings from the CA/CM tools are provided in a digital format. Most importantly, the inputs, referred to as auditee profile, auditor profile, and audit standards, drive the audit app selections via an audit app recommender system (Dai et al., 2014). An audit ecosystem will manage changes to these external elements as they evolve over time, primarily to accommodate new and enhanced audit apps but also changes to auditee and auditor profile information, audit standards, and auditee data.

An analytic app recommender system was presented earlier in this paper as a key component to the ENHANCE framework and it is envisioned that an audit ecosystem will include an audit app recommender system, as proposed by Dai et al. (2014), that will manage the audit app selection process in a manner analogous to the analytic app recommender system for ENHANCE.

The auditee profile will likely contain information similar to that incorporated into the app recommender system proposed by Dai et al. (2014) where recommendations are first based on audit standards that create a structure classifying audit apps by industry, business cycle, accounts, audit assertions, and audit objectives. For the second step recommendations are based on audit clients themselves by using a two-stage collaborative filtering (CF) approach that follows the premise that if an app has been used for similar clients by auditors, it is likely that app is appropriate for the new client.

For the third step recommendations are based on auditor preferences taking into consideration that auditors may prefer certain app vendors, app versions, and/or user interfaces that reflect their historical preferences (Dai et al., 2014). Recommendations

based on auditors' preferences are developed using a CF approach similar to that used for recommendations based on audit clients. The app recommender system creates a final score for each audit app by combining the results of the previous steps and presents its recommendations to the auditor (Dai et al., 2014). This type of information will be sourced in the auditor profile.

Figure 21 depicts the agents, as described in Section 5.4.3 that are incorporated in an audit ecosystem. General business agents, and in particular agents that undertake the negotiations for CA/CM agents, support the recommender system. Similarly, business transaction agents support the acquiring and deploying of CA/CM agents as identified by the recommender system. Application agents, planning and scheduling agents, and system-level support agents support the operation of the CA/CM agents. Application agents also support the issue resolution process. Personal agents support the auditor and auditee participants and the presentation of the results of the CA/CM activities for these participants. Information brokering agents and interoperation agents locate information as required with respect to the identification of auditee data and its characteristics. Security agents act to protect the auditee data from incursion during the transmittal process from auditee to audit ecosystem.

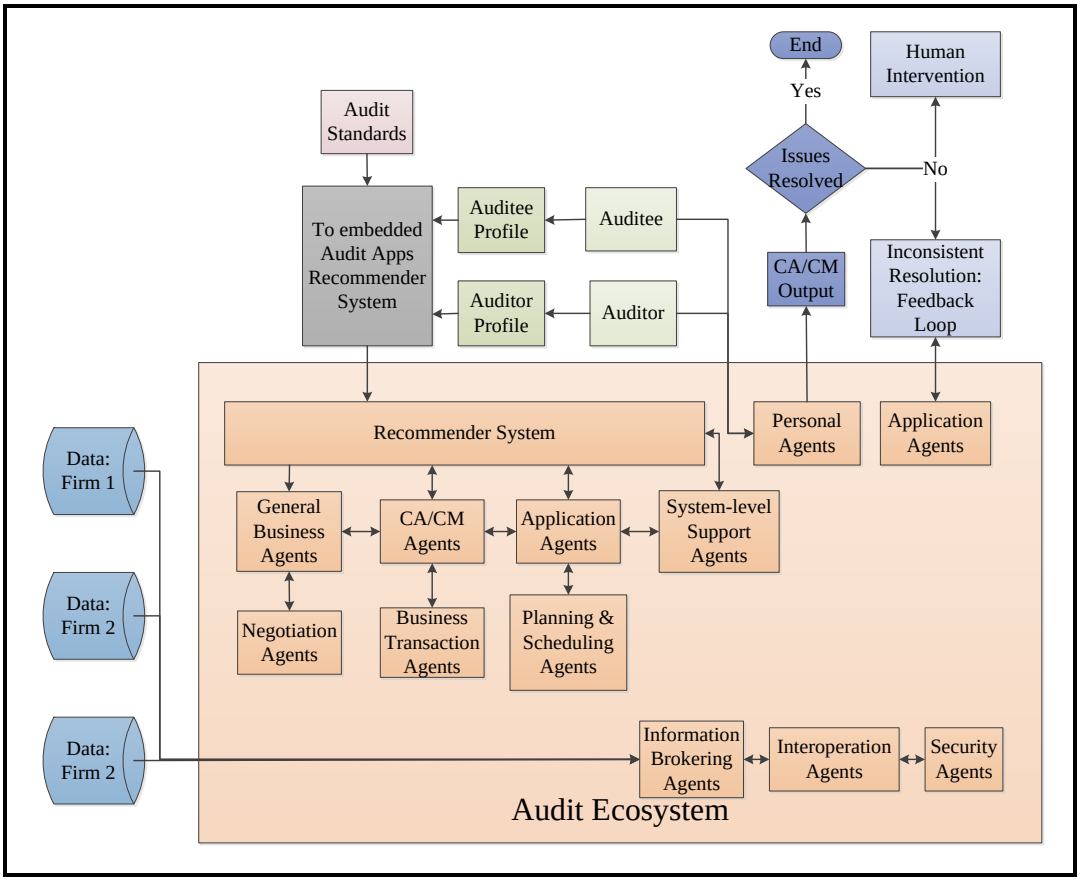

Figure 21: External influences

5.6. ENHANCE Ecosystem Components

A diagram of the digital agents, operating within an ecosystem configuration in support of the ENHANCE framework is presented in Figure 22. The external influences are identified as the ENHANCE user and the governmental dataset. The functionality that the ecosystem agents can provide to the ENHANCE framework apps is described below. In the example presented, the agents allow ENHANCE to work with a greater variety of data structures, analytic apps, and analytic output formats than is envisioned in the initial design of ENHANCE.

The structural app is supported by information brokering agents and interoperation agents that assist when dealing with numerous and varied datasets to identify the location of the dataset, the structure of the data, and the attribute descriptions. Security agents maintain the integrity of the data from incursion as it is transmitted from source to ENHANCE.

The response app, which formats the analytic results for presentation to the user, is supported by application agents that allow the app to incorporate a wide variety of output formats, and personal agents that identify to the app the preferred formats for each user.

The recommender system, execution app, and especially the underlying analytic apps are supported by a number of agents. General business agents, particularly in the form of negotiation agents, undertake the acquisition of new analytic apps from a variety of sources. System-level support agents, in the form of application agents, support the execution of highly sophisticated analytic apps that may require a number of inputs to launch. Planning and scheduling agents, another form of system-level support agents, coordinate the execution of a number of analytic apps, should there be a need to launch several apps to complete a sophisticated analytic request.

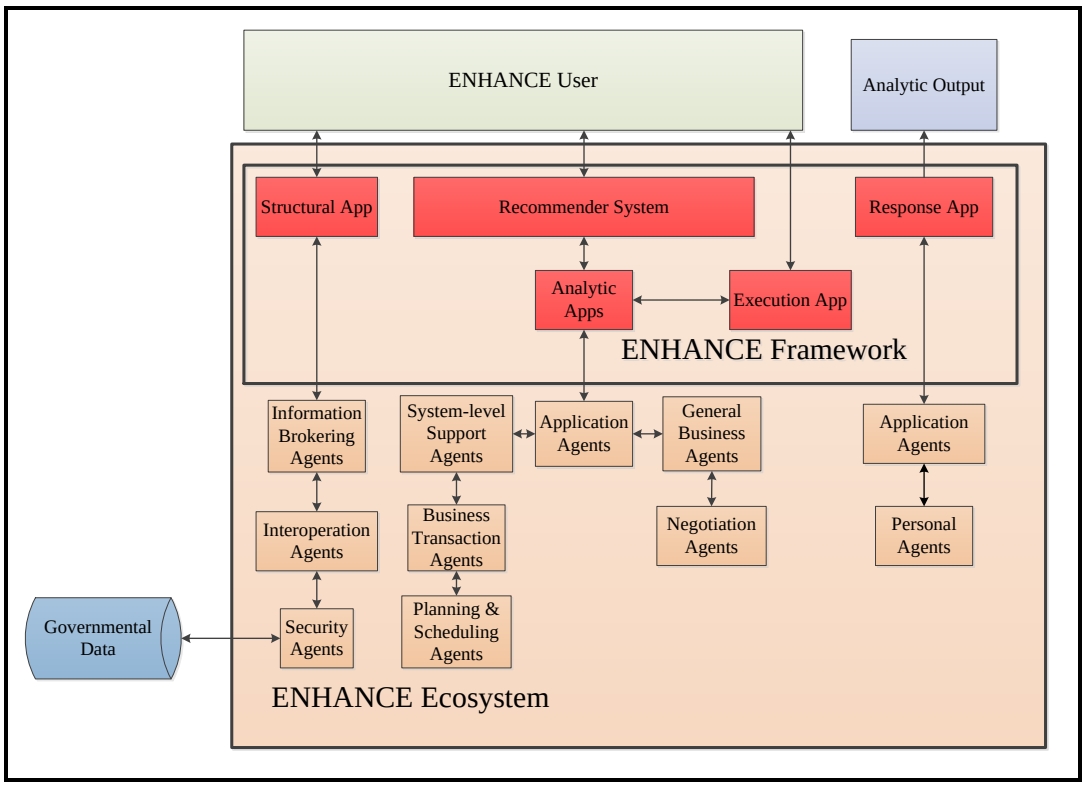

Figure 22: ENHANCE Ecosystem

5.7. An Open Data Ecosystem

The ecosystem presentation to this point has focused on design elements for an audit ecosystem, a form of a digital ecosystem that supports to use of automated CA/CM technologies. As the ENHANCE framework is intended to support investigations, or audits, of governmental expenditure data by constituents, the concepts of an audit ecosystem are consistent with the theme of this research. This research now introduces concepts that form the basis of a digital ecosystem to support the ENHANCE framework from the perspective of the governmental open data environment for which it is envisioned to operate. Ecosystem concepts have been presented that can fulfill any number of requirements in any number or environments. For example, Zuiderwijk et al.

(2014) present characteristics of digital ecosystems that function specifically in business environments. The concept of a business ecosystem has been defined to include:

- An economic community supported by a foundation of interacting organizations and individuals (Moore, 1996)
- Loosely interconnected participants who depend on each other for mutual effectiveness and survival (Iansiti & Levien, 2004)
- Suppliers and customers (Moore, 1996)

Attributes of a business ecosystem, as well as digital ecosystems in general, include autonomous operations and self-organizing capabilities (Zuiderwijk et al., 2014).

The concept of an open data ecosystem, another form of a business ecosystem, is envisioned to simplify both the publishing and usage of data; this data can well include the recent governmental open data initiatives upon which the ENHANCE framework functions (Zuiderwijk et al., 2014). The benefits of an open data ecosystem include increasing user participation as well as motivating innovation by both providers and users as they draw meaningful information from the data, as the ENHANCE framework provides (Zuiderwijk et al., 2014).

The essential functions of an open data ecosystem include the ability to (Zuiderwijk et al., 2014):

- Release and publish open data on internet
- Search, find, evaluate, and view the data
- Cleanse, analyze, enrich, combine, link, and visualize the data

- Interpret and discuss the data, and provide feedback to the data publisher.

It should be noted that the analysis function includes tools such as statistical analyses and other capabilities that can provide new insights into and understanding of the data. Ecosystem capabilities provide for integration of the various technologies required to provide and derive benefit from open data, and adjust for the evolving nature of these technologies. The ENHANCE framework is designed to include elements two through four; ENHANCE is not intended to publish data at this point but only access governmental open data portals to source data for advanced analytics.

If the ENHANCE framework already includes capabilities as provided by an open data ecosystem, why design such an ecosystem for ENHANCE? The ENHANCE framework at this point is intended to provide robust analytics, as available presently, over rather straightforward datasets, that of governmental expenditure data as recorded at the transaction level. As noted earlier, software firms that provide products with data analytic capabilities are continually increasing their offerings as demand for analytic apps increase (Dai et al., 2014). It is not unimaginable that highly sophisticated analytic apps will be developed in the future whose requirements may exceed the capabilities of ENHANCE as envisioned today. These highly sophisticated analytic apps may require more inputs than simply a dataset to function. An ecosystem approach may be the solution to address the evolving nature of available analytic apps, among other items existing in the environment in which ENHANCE operates.

Analytic requests may become more sophisticated from the perspective of the datasets under analysis. It was noted earlier that combining related datasets can provide the

ability to identify discrepancies between them. The development of standard data taxonomies is still in its infancy, but the process is underway and will be evolving over time. Again, an ecosystem approach may better address the evolving nature of data structures and data combinations rather than constant modifications to the ENHANCE framework.

As noted earlier with respect to audit ecosystems, ecosystems incorporate current approaches and techniques and evolve them as business scenarios change and support underlying technologies and better manage their usage than is provided with simply human intervention. Utilizing an ecosystem can also act as a buffer to the more complex technology underlying the ecosystem and thus minimize the requirement for a user to acquire significant technical education, one of the goals in developing ENHANCE. Incorporating an ecosystem in support of the ENHANCE framework will likely require significantly less development effort than constantly upgrading the framework itself as technology and the environment within which it operates changes.

There have been defined additional elements that permit integration of the open data ecosystem elements and allow them act as a unified whole (Zuiderwijk et al., 2014):

- User instructional aids (‘pathways’ as referred to by the author) to show how open data can be used
- A quality management system to assess the quality of the data
- A provision for different types of metadata to connect the elements, ensuring interoperability and data processing

These elements are not addressed in the initial design of the ENHANCE framework but can be incorporated by means of a supporting ecosystem.

Open data ecosystems, besides being classified as business ecosystems, can also be considered within the overall classification of another type of digital ecosystem, an information ecosystem. Information ecosystems have been defined as complex systems, each existing in a specific environment, that include people, work to be done, value to be received from the work, technology to support the people and the work, and the required interconnections between them (Zuiderwijk et al., 2014). Information ecosystems place a particular emphasis on the behavior of the human participants as they are supported by the technology components (Nardi & O'Day, 1999). Typical components of an information ecosystem include information systems, databases, workflows, people, and infrastructure (Zuiderwijk et al., 2014).

Open government ecosystems, a subset of open data ecosystems, are intended to promote the interrelationships between the data, open data providers, open data users, physical infrastructures, and institutions with the participants considerably interdependent with one another (Zuiderwijk et al., 2014). Interdependent disciplines in an open government ecosystem include (Harrison et al., 2012):

- Government policies and practices
- Innovators who undertake to combine technology, business and government
- Users, civil society, and business

The intent is that open government ecosystems can stimulate user participation in public affairs or policy making processes and that government open data ecosystems can facilitate decision-making and planning.

Open data ecosystems may include a special class of users that have been referred to as infomediaries, or intermediate consumers that add value to the open datasets by cleaning, analyzing, and integrating them. The infomediaries publish what they produce, as well as develop better ways to publish and share data (Pollock, 2011). Infomediaries can be considered potential users of ENHANCE as it can assist them in completing their tasks.

The characteristics described for business ecosystems in general and open data ecosystems in particular lend themselves to support the ENHANCE framework, particularly as the one of the key drivers behind the development of ENHANCE is the availability of open government data upon which ENHANCE functions. In developing an ecosystem to support the ENHANCE framework it seems appropriate to investigate and incorporate characteristics and attributes of both audit and open data ecosystems that can evolve as the environment in which ENHANCE operates evolves. An ecosystem can support the technologies underlying ENHANCE and better manage their usage than is provided with simply human intervention, especially as the technologies become progressively more sophisticated. Utilizing an ecosystem can act as a buffer to the more complex technology underlying the ecosystem and thus lessen the requirement for significant technical education to operate the ENHANCE framework. An Ecosystem can support a much more robust analytic tool, ENHANCE to the power of two: ENHANCE^2.

5.8. Design Science Research Methodology to guide an audit ecosystem design

The attributes and characteristics described for both the audit ecosystem and open data ecosystems can lend themselves to supporting and enriching the ENHANCE framework. The ENHANCE framework is intended to provide users the ability to collect and analyze open data in such a manner that permits the user to 'audit' governmental expenditures. Activities such as armchair auditing are supported in the design of ENHANCE. The ENHANCE framework is envisioned to be robust in technical terms, incorporating state-of-the-art analytics, recommender system, and presentation tools. What is an appropriate procedure to drive development of an ecosystem to support ENHANCE? Design Science Research Methodology is presented as a tool that can facilitate design of an audit ecosystem.

Developing a complete DSRM template to support the design of an audit ecosystem, as well as an adaptation of an audit ecosystem to support ENHANCE is outside the scope of this paper, but an example is provided in Table 47. It is envisioned that the DSRM template will be similar for both an audit ecosystem as well as an ENHANCE ecosystem so the descriptive information in the table reflects both CA/CM and ENHANCE technology. For the present research only the first step was undertaken, that of Problem identification and motivation. Definition is derived from the discussion above that presented a number of ecosystem capabilities. Knowledge Base is derived from the literature review included in this paper.

The remaining five steps are shown in italics to indicate the information presented is not definitive but is only a proposal of what may be appropriate for each of the remaining steps. The Definition of Objectives step may require questionnaires/surveys to elicit

responses from potential users as to what they perceive as appropriate ecosystem capabilities to support the underlying technology, and then translate this into logic to support CA/CM and/or ENHANCE. The Design and Development step should be based on completion of a detailed definition/design document covering the anticipated ecosystem process, as well as a further literature review with respect to actual ecosystem design and usage. The Demonstration step includes development of a prototype ecosystem and it is recommended this be presented to a focus group of target users for their review and comment. The final step, Communication, should include presentation of the developed ecosystem to a sample of target users and data providers for their review, testing, and approval.

DSRM Activity	**Definition**	**Knowledge Base**
Problem identification and motivation	An audit ecosystem is intended to support CA/CM (ENHANCE) capabilities within an evolving landscape . How can ecosystem elements improve CA/CM (ENHANCE) capabilities?	Literature review of the capabilities of a typical digital ecosystem, and also those referred to as 'business', 'information', and 'open data' ecosystems
Definition of the objectives of a solution	*Understand what specific functions in the CA/CM architecture (ENHANCE framework) can be enhanced by ecosystem elements*	*Understand from a user and data provider perspective the most valuable improvements that can be made to CA/CM techniques (ENHANCE). It is expected these improvements would focus on enriching the user experience with CA/CM (ENHANCE)*
Design and development	*Design an ecosystem that can enrich the capabilities of the CA/CM (ENHANCE framework)*	*Study examples of functioning ecosystems or well-designed ecosystem prototypes*
Demonstration	*Definition of the specific*	*Develop a prototype of a*

	ecosystem elements that can enrich CA/CM (ENHANCE)	*CA/CM ecosystem (ENHANCE$^{2)}$ as it would be envisioned to operate*
Evaluation	*Evaluation of improvements to CA/CM (ENHANCE) as provided by the addition of ecosystem elements*	*Perform a gap analysis to evaluate the proposed changes with respect to the recommendations provided by users and data providers*
Communication	*Present to appropriate audiences: CA/CM users (data providers, ENHANCE users, and potential user)s*	*Data providers and target users*

Table 47: Proposed DSRM template to develop an ecosystem to support ENHANCE

Following a proven design methodology such as DSRM will facilitate developing the most appropriate ecosystem design that can enrich the capabilities of either CA/CM tools and/or the ENHANCE framework (as stated earlier as ENHANCE^2). Development of a complete DSRM template and ecosystem to support CA/CM and/or ENHANCE is an activity for future research.

5.9. Discussion

The purpose of this research, starting with a review of recent and significant research in the areas of digital ecosystems and software agents, is to define a specific form of digital ecosystem, an audit ecosystem, that is, an environment in which computer-based CA/CM agents can operate with the greatest efficiency and effectiveness in order to provide the greatest benefit to both client and provider. The development of an audit ecosystem is the natural next step in the deployment of computer-based CA/CM agents.

The characteristics attributed to digital ecosystems in general, and more specifically audit ecosystems, lend themselves to application in support of the ENHANCE framework, as

the primary function of ENHANCE is to provide the user with an analytic tool to undertake personalized 'audits' over open government data. In developing an ecosystem to support the ENHANCE framework, understanding the characteristics and attributes of both audit and open data ecosystems and incorporating them in a blend that enriches user interaction with ENHANCE will provide an enriched user experience and an analytic tool that can evolve as the environment in which it operates changes.

This research describes an accepted approach to drive the development of an ecosystem tailored to support CA/CM and/or ENHANCE, that of Design Science Research Methodology. A DSRM template is presented that includes a definition for each activity required for the development of an ecosystem environment capable of enriching the CA/CM and/or ENHANCE experience, and the knowledge tools as appropriate for each step in the development process.

5.9. Limitations and Future Research

There are limitations with this present research activity. This paper presents examples of an audit ecosystem's design elements, specifically features and attributes, but this is by no means all-inclusive. In adapting an audit ecosystem concept to ENHANCE additional characteristics may be identified. An ecosystem can benefit the underlying IT applications by adapting the applications as the environment changes over time. An assessment of potential elements driving change in the environment need to be assessed, for example, newly developed audit apps (for an audit ecosystem) and/or analytic apps (for an ENHANCE ecosystem) that may require more sophisticated inputs in order to function.

The detailed design and development of the audit ecosystem and adaptation to support the ENHANCE framework has yet to be undertaken. A complete development of the DSRM template to support the design of the audit ecosystem and an ecosystem to support the ENHANCE framework needs to be undertaken, using the examples presented above.

References

Alles, M. G., Kogan, A., & Vasarhelyi, M. A. (2008). Putting continuous auditing theory into practice: Lessons from two pilot implementations. *Journal of Information Systems*, *22*(2), 195-214.

Arumugam, S., Ritesh, K. K., & Prasad, A. R. (2013) Wireless Robotics: Opportunities and Challenges. *Wireless Personal Communications* Published online: 22 March 2013

Barraca, J. P., Sadeghi, R., & Aguiar, R. L. (2013). Collaborative Relaying Strategies in Autonomic Management of Mobile Robotics. *Wireless Personal Communications*, 1-20. (2013) 70:1077–1096

Briscoe, G., & De Wilde, P. (2009, October). Computing of applied digital ecosystems. In *Proceedings of the International Conference on Management of Emergent Digital EcoSystems* (p. 5). ACM.

Bumgarner, N., & Vasarhelyi, C. M. A. (2015). Auditing—A New View. AUDIT ANALYTICS, 3.

Byrnes, P. E., Ames, B., Vasarhelyi, M., & Warren, D. (2012). The Current State of Continuous Auditing and Continuous Monitoring.

Byrnes, P. E., Al-Awadhi, C. A., Gullvist, B., Brown-Liburd, H., Teeter, C. R., Warren Jr, J. D., & Vasarhelyi, M. (2015). Evolution of Auditing: From the Traditional Approach to the Future Audit. AUDIT ANALYTICS, 71

Chang, E. & West, M. (2006). Digital Ecosystems and Comparison to Existing Collaboration Environment, *WSEAS TRANSACTIONS on ENVIRONMENT and DEVELOPMENT*, vol. 2, pp. 1396 – 1404.

Foon, L. S., & Yen, T. F. (2011). Creating and Developing a Corporate Knowledge Ecosystem *Annual Summit on Business and Entrepreneurial Studies (ASBES 2011) Proceeding 486*

Geerts, G. L., & O'Leary, D. E. (2014). A supply chain of things: The EAGLET ontology for highly visible supply chains. Decision Support Systems, 63, 3-22.

Genesereth, M. R., & Ketchpel, S. P. (1994). Software agents. *Commun. ACM*, *37*(7), 48-53.

Harrison, T. M., Pardo, T. A., & Cook, M. (2012). Creating open government ecosystems: A research and development agenda. *Future Internet*, *4*(4), 900-928.

Iansiti, M., & Levien, R. (2004). *The keystone advantage: what the new dynamics of business ecosystems mean for strategy, innovation, and sustainability*. Harvard Business Press.

Jennings, N., & Wooldridge, M. (1996). Software agents. *IEE review*, *42*(1), 17-20.

Kogan, A., Sudit, E. F., & Vasarhelyi, M. A. (1999). Continuous online auditing: A program of research. *Journal of Information Systems*, *13*(2), 87-103.

Maes, P. (1994). Agents that reduce work and information overload. *Communications of the ACM*, *37*(7), 30-40.

Moore, J. (1996). The Death of Competition: Leadership & Strategy. *The Age of Business Ecosystems, HarperCollins, New York*.

Nachira, F., Dini, P., & Nicolai, A. (2007). A network of digital business ecosystems for Europe: roots, processes and perspectives. *European Commission, Bruxelles,* Introductory Paper.

Nardi, B. A., & O'Day, V. (1999). *Information ecologies: Using technology with heart*. Mit Press.

Nwana, H. S., & Ndumu, D. T. (1999). A perspective on software agents research. *The Knowledge Engineering Review*, *14*(2), 125-142.

Papazoglou, M. P. (2001). Agent-Oriented Technology in Support of E-Business Enabling the Development of "Intelligent" Business Agents for Adaptive, Reusable Software. *Communications of the ACM Volume: 44 Issues: 4 Pages: 71-77 Published: APR 2001*

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45-77.

Pollock, R. (2011). Building the (Open) data ecosystem. *Ideas and musings, Open Data*.

Pranata, I., Skinner, G., & Athauda, R. (2011).A Community Based Authentication and Authorisation Mechanism for Digital Ecosystem *5th IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST 2011), 31 May -3 June 2011, Daejeon, Korea*

Sacha, J., Biskupski, B., Dahlem, D., Cunningham, R., Dowling, J., & Meier, R. (2007). A Service-Oriented Peer-to-Peer Architecture for a Digital Ecosystem *2007 Inaugural IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST 2007)*

Stewart, T. R. (2015). Data Analytics for Financial Statement Audits. AUDIT ANALYTICS, 105.

Swarm-bots: http://www.swarm-bots.org/.

Troubleyn, E., Moerman, I., & Demeester, P. (2013). QoS challenges in wireless sensor networked robotics. *Wireless Personal Communications*, Volume: 70:1059–1075

Uden, L., Wangsa, I. T., & Damiani, E. (2007) The future of B-learning: B-learning ecosystem *2007 Inaugural IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST 2007)*

Vasarhelyi, M. A., & Halper, F. B. (1991). The continuous audit of online systems. *AUDITING-A JOURNAL OF PRACTICE & THEORY*, *10*(1), 110-125.

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *Knowledge engineering review*, *10*(2), 115-152.

Zuiderwijk, A. M. G., Janssen, M. F. W. H. A., & Davis, C. B. (2014). Innovation with open data: Essential elements of open data ecosystems.

Conclusion

Constituent demands for improved transparency in governmental reporting have increased since the recent (2008-2009) financial crisis in the U.S. that impacted the financial well-being of a number of U.S.-based governmental entities at both state and local levels. Since that time several governmental entities in the U.S. have lead the effort to provide an open data environment but these early initiatives do not incorporate robust analytic capabilities to satisfy constituent inquiries.

This paper describes the development and deployment of the open data portals that support governmental data transparency initiatives in order to satisfy the needs of the constituency. Many of these portals provide robust datasets that are accessible to satisfy constituent inquiries about governmental spending and other governmental activities. The accessibility and availability of robust analytics to present information in a meaningful fashion to constituents are yet to come. The need exists to provide to the public robust yet easy-to-use analytics that provide useful and understandable results, all in support of greater governmental transparency. The intent of the ENHanced ANalytic Constituent Environment (ENHANCE) framework is to provide just that and bridge the gap between open data initiatives and the capability to provide advanced analytics over that data.

This paper identifies what is meant by data transparency, what efforts have been undertaken to provide transparency over governmental activities, and how the ENHANCE framework, coupled with open governmental data, can provide transparency over this data to constituents in a manner not presently available.

A review of recent U.S. legislation describes open data and transparency efforts currently underway at the federal level. The legislative efforts, in promoting open data, are setting the stage for an advanced analytic tool to support constituent demands for meaningful information. Such a tool is not presently available.

The primary focus of this research entails the presentation of an ENHanced ANalytic Constituent Environment (ENHANCE), facilitated by open government data, that fulfills the reporting requirements of the various governmental stakeholders, such as citizens, analysts, bond investors, creditors, and oversight officials and auditors[39]. ENHANCE's capabilities bridge the gap existing between current governmental open data initiatives and a constituent demand for robust analytics that provide meaningful information from the raw data. The attributes of Decision Support Systems (DSS) are discussed, as the capabilities that ENHANCE provides classify this tool as a DSS.

This research presents, by using existing technology, an approximate rendition of the capabilities that a framework such as ENHANCE is able to provide to the user. A number of analytics have been executed over a representative sample of governmental expenditure data and the results presented in a manner that is easily understandable by the user. A design for the ENHANCE user interface is proposed. An example of ENHANCE in practice to support a constituent need is described.

Of the components that encompass the ENHANCE framework, that of an analytic app recommender system is key and likely the most sophisticated component within the

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

framework. A background on recommender systems in general, and recent research specifically focused on an audit app recommender system, is presented.

This research describes the structure of an audit ecosystem, that is, a natural progression in the deployment of computer-based CA/CM tools, and applies this technology to the ENHANCE framework to provide the user a technology-rich analytic capability.

This research provides an initial definition for an audit ecosystem, that is, a holistic approach to the design and development of a technology-driven framework to provide overall management and control of the audit technology components employed, and coordination of the activities of the participants involved. Characteristics of open data ecosystems and open government ecosystems are also included, as attributes for an audit ecosystem as well as these ecosystems all support the capabilities of the ENHANCE framework. An ecosystem approach can provide the ENHANCE user with additional feature and function over and above that provided with the basic framework, with the intent to provide a very capable tool in support of an 'armchair auditor' activity (O'Leary, 2015). An ecosystem approach provides the ability to manage the ever-changing environment within which ENHANCE operates.

Design Science Research Methodology is presented as an appropriate tool to support the design effort for not only the ENHANCE framework but also the app recommender system and ecosystems to support both CA/CM and/or ENHANCE technologies. Using a tool such as DSRM assures that the most robust design is achieved in order to provide the capabilities envisioned for all these technologies.

This research contributes to academic literature by proposing an ENHanced ANalytic Constituent Environment (ENHANCE) where governmental stakeholders can create reports on demand to satisfy their analytic requirements. This research applies existing concepts, particularly DSS and DSRM, to the development of the ENHANCE framework, an app recommender system, and related ecosystems as described in this paper and extends the literature on these items.

In addition, this research provides a literature review of current open data practices and transparency efforts in government financial reporting.

References

Alles, M. G., Kogan, A., & Vasarhelyi, M. A. (2008). Putting continuous auditing theory into practice: Lessons from two pilot implementations. Journal of Information Systems, 22(2), 195-214.

Arnott, D., & Pervan, G. (2008). Eight key issues for the decision support systems discipline. *Decision Support Systems*, *44*(3), 657-672.

Arumugam, S., Ritesh, K. K., & Prasad, A. R. (2013) Wireless Robotics: Opportunities and Challenges. Wireless Personal Communications Published online: 22 March 2013

Attmore, R. H. (2011, Summer). The Importance of Timely Government Financial Reporting. *Journal of Government Financial Management* http://www.gasb.org/cs/ContentServer?site=GASB&c=Document_C&pagename=GASB%2FDocument_C%2FGASBDocumentPage&cid=1176158583687

Austin Business Journal: http://www.bizjournals.com/austin/news/2015/02/02/database-searchcity-of-austin-salaries.html

Baber, W. R., & Gore, A. K. (2008). Consequences of GAAP Disclosure Regulation: Evidence from Municipal Debt Issues. THE ACCOUNTING REVIEW, 565–591.

Barraca, J. P., Sadeghi, R., & Aguiar, R. L. (2013). Collaborative Relaying Strategies in Autonomic Management of Mobile Robotics. Wireless Personal Communications, 1-20. (2013) 70:1077–109.

Benbasat, I., & Zmud, R. W. (1999). Empirical research in information systems: the practice of relevance. *MIS quarterly*, 3-16.

Bertot, J. C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. Government Information Quarterly, 264–271.

Bharosa, N., van Wijk, R., Janssen, M., de Winne, N., & Hulstijn, J. (2011, June). Managing the transformation to standard business reporting: principles and lessons learned from the Netherlands. In Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times (pp. 151-156). ACM.

Briscoe, G., & De Wilde, P. (2009, October). Computing of applied digital ecosystems. In Proceedings of the International Conference on Management of Emergent Digital EcoSystems (p. 5). ACM.

Brito, J. (2008). Hack, mash, & peer: Crowdsourcing government transparency. The Columbia Science and Technology Law Review, 9, 119-157.

Bumgarner, N., & Vasarhelyi, C. M. A. (2015). Auditing—A New View. AUDIT ANALYTICS, 3.

Byrnes, P. E., Ames, B., Vasarhelyi, M., & Warren, D. (2012). The Current State of Continuous Auditing and Continuous Monitoring.

Byrnes, P. E. (2015). *Developing Automated Applications for Clustering and Outlier Detection: Data Mining Implications for Auditing Practice* (Doctoral dissertation, Rutgers University-Graduate School-Newark).

Byrnes, P. E., Al-Awadhi, C. A., Gullvist, B., Brown-Liburd, H., Teeter, C. R., Warren Jr, J. D., & Vasarhelyi, M. (2015). Evolution of Auditing: From the Traditional Approach to the Future Audit. AUDIT ANALYTICS, 71

Cameron, D. (2010). PM's Podcast on Transparency. Available at: https://www.gov.uk/government/news/pms-podcast-on-transparency

Chan, K. S., & Tong, H. (1986). On estimating thresholds in autoregressive models. Journal of time series analysis, 7(3), 179-190.

Chang, E. & West, M. (2006). Digital Ecosystems and Comparison to Existing Collaboration Environment, WSEAS TRANSACTIONS on ENVIRONMENT and DEVELOPMENT, vol. 2, pp. 1396 – 1404.

Chen, Y. C. (2012). A comparative study of e-government XBRL implementations: The potential of improving information transparency and efficiency. Government information quarterly, 29(4), 553-563.

Cheng, R., Xia, Y., Prabhakar, S., Shah, R., & Vitter, J. S. (2004, August). Efficient indexing methods for probabilistic threshold queries over uncertain data. In Proceedings of the Thirtieth international conference on Very large data bases-Volume 30 (pp. 876-887). VLDB Endowment.

Dai, J.,Krahel, JP, & Vasarhelyi, M. (2014). Which Audit App(s) Should Auditors Use? An Exploratory Study of Using Recommender Systems for Audit App Selection. Rutgers Working Paper.

Dai, J., Li, Q., & Vasarhelyi, M. (2015). Design of Audit Apps for Armchair Auditors to Analyze Government Procurement Contract. Rutgers Working Paper.

Foon, L. S., & Yen, T. F. (2011). Creating and Developing a Corporate Knowledge Ecosystem Annual Summit on Business and Entrepreneurial Studies (ASBES 2011) Proceeding 486

Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. The computer journal, 41(8), 578-588.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. Journal of the American statistical Association, 97(458), 611-631.

Frank, M., & Oztoprak, A. A. (2015). Concepts of Transparency: Open Data in UK Local Authorities. In *CeDEM15: Conference for E-Democracy and Open Government* (p. 185). MV-Verlag.

GAO: www.gao.gov/new.items/d05734sp.pdf

Geerts, G. L., & O'Leary, D. E. (2014). A supply chain of things: The EAGLET ontology for highly visible supply chains. Decision Support Systems, 63, 3-22.

Geijtenbeek, W., & Lucassen, H. (2012). Dutch approach to SBR assurance. In Proceedings of the 25th XBRL International Conference, Yokohama, Japan, p. ASSR1. XBRL International.

Genesereth, M. R., & Ketchpel, S. P. (1994). Software agents. Commun. ACM, 37(7), 48-53.

Granovetter, M. (1978). Threshold models of collective behavior. American journal of sociology, 1420-1443.

Harrison, T. M., Pardo, T. A., & Cook, M. (2012). Creating open government ecosystems: A research and development agenda. *Future Internet*, *4*(4), 900-928.

Hoyle, S. D., & Langley, A. D. (2011). Spatial size data stratification for length-based stock assessments. WCPFC SC7 SA IP-9, Pohnpei, Federated States of Micronesia, 9-17.

Iansiti, M., & Levien, R. (2004). *The keystone advantage: what the new dynamics of business ecosystems mean for strategy, innovation, and sustainability*. Harvard Business Press.

Issa, H. (2013). Exceptional exceptions (Doctoral dissertation, Rutgers University-Graduate School-Newark).

Jaeger, P. T., & Bertot, J. C. (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. Government Information Quarterly, 27(4), 371-376.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. ACM computing surveys (CSUR), 31(3), 264-323.

Jennings, N., & Wooldridge, M. (1996). Software agents. IEE review, 42(1), 17-20.

Kim, S., & Cho, K. (2005). Achieving administrative transparency through information systems: A case study in the Seoul Metropolitan Government. In Electronic Government (pp. 113-123). Springer Berlin Heidelberg.

Kogan, A., Sudit, E. F., & Vasarhelyi, M. A. (1999). Continuous online auditing: A program of research. Journal of Information Systems, 13(2), 87-103.

Lee, W., & Xiang, D. (2001). Information-theoretic measures for anomaly detection. In Security and Privacy, 2001. S&P 2001. Proceedings. 2001 IEEE Symposium on (pp. 130-143). IEEE.

Maes, P. (1994). Agents that reduce work and information overload. Communications of the ACM, 37(7), 30-40.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision support systems*, *15*(4), 251-266.

Moore, J. (1996). The Death of Competition: Leadership & Strategy. *The Age of Business Ecosystems, HarperCollins, New York*.

Nachira, F. (2002). Toward a network of digital business ecosystems fostering the local development, http://www.europa.eu.int/information_society/topics/ebusiness/godigital/sme_research/index_en.htm

Nachira, F., Dini, P., & Nicolai, A. (2007). A network of digital business ecosystems for Europe: roots, processes and perspectives. European Commission, Bruxelles, Introductory Paper.

Nardi, B. A., & O'Day, V. (1999). *Information ecologies: Using technology with heart*. Mit Press.

Nassar, S., Sander, J., & Cheng, C. (2004, June). Incremental and effective data summarization for dynamic hierarchical clustering. In Proceedings of the 2004 ACM SIGMOD international conference on Management of data (pp. 467-478). ACM.

NIST: (www.itl.nist.gov/div898/handbook/pmc/section4/pmc41.htm)

Nwana, H. S., & Ndumu, D. T. (1999). A perspective on software agents research. The Knowledge Engineering Review, 14(2), 125-142

NYCMODA: http://www.nyc.gov/html/analytics/html/home/home.shtml

OECD. (2009, July). Forum on Tax Administration: Taxpayer Services Sub-Group, Guidance note, Standard Business Reporting www.oecd.org/ctp/administration/43384923.pdf

OGP1: http://www.opengovpartnership.org/about

OGP2: http://www.opengovpartnership.org/about/mission-and-goals

OHIO: http://www.govtech.com/budget-finance/Ohio-Offers-Financial-Transparency-Statewide.html

O'Leary, D. E. (2015). Armchair Auditors: Crowdsourcing Analysis of Government Expenditures. Journal of Emerging Technologies in Accounting.

Papazoglou, M. P. (2001). Agent-Oriented Technology in Support of E-Business Enabling the Development of "Intelligent" Business Agents for Adaptive, Reusable Software. Communications of the ACM Volume: 44 Issues: 4 Pages: 71-77 Published: APR 2001

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45-77.

Peled, A. (2011). When transparency and collaboration collide: The USA open data program. *Journal of the American society for information science and technology*, *62*(11), 2085-2094.

Pew Research Center. (2015, April). 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*. Business Expert Press.

Pranata, I., Skinner, G., & Athauda, R. (2011).A Community Based Authentication and Authorisation Mechanism for Digital Ecosystem 5th IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST 2011), 31 May -3 June 2011, Daejeon, Korea

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. "O'Reilly Media, Inc.".

Reck, J. L., & Wilson, E. R. (2006). Information transparency and pricing in the municipal bond secondary market. Journal of Accounting and Public Policy, 1-31.

Ricci, F., Rokach, L., & Shapira, B. (2011). *Introduction to recommender systems handbook* (pp. 1-35). Springer US.

Sacha, J., Biskupski, B., Dahlem, D., Cunningham, R., Dowling, J., & Meier, R. (2007). A Service-Oriented Peer-to-Peer Architecture for a Digital Ecosystem 2007 Inaugural IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST 2007)

Shim, D. C., & Eom, T. H. (2009). Anticorruption effects of information communication and technology (ICT) and social capital. International review of administrative sciences, 75(1), 99-116.

Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision support systems*, *33*(2), 111-126.

Sprague Jr, R. H. (1980). A framework for the development of decision support systems. *MIS quarterly*, 1-26.

Stewart, T. R. (2015). Data Analytics for Financial Statement Audits. AUDIT ANALYTICS, 105.

Sturgis, P. (2004). Analysing complex survey data: Clustering, stratification and weights. Social research UPDATE, (43).

Swarm-bots: http://www.swarm-bots.org/ .

Treiman, D. J. (1970). Industrialization and Social Stratification*. Sociological Inquiry, 40(2), 207-234.

Troubleyn, E., Moerman, I., & Demeester, P. (2013). QoS challenges in wireless sensor networked robotics. Wireless Personal Communications, Volume: 70:1059–107Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. Journal of the American Statistical Association, 84(405), 231-240.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, *84*(405), 231-240.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. Journal of the American Statistical Association, 93(443), 1188-1202.

Uden, L., Wangsa, I. T., & Damiani, E. (2007) The future of B-learning: B-learning ecosystem 2007 Inaugural IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST 2007)

Vasarhelyi, M. A., & Halper, F. B. (1991). The continuous audit of online systems. AUDITING-A JOURNAL OF PRACTICE & THEORY, 10(1), 110-125.

Vishwanath, T., & Kaufmann, D. (1999). Towards transparency in finance and governance. Available at SSRN 258978.

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. Knowledge engineering review, 10(2), 115-152.

Worthy, B. (2015). The Impact of Open Data in The UK: Complex, Unpredictable, and Political. *Public Administration*.

Yager, R. R., & Petry, F. E. (2006). A multicriteria approach to data summarization using concept ontologies. Fuzzy Systems, IEEE Transactions on, 14(6), 767-780.

Zuiderwijk, A. M. G., Janssen, M. F. W. H. A., & Davis, C. B. (2014). Innovation with open data: Essential elements of open data ecosystems.

Appendix

Municipal Website	Description
https://data.cityofboston.gov	The de facto place for Boston's municipal data
https://data.seattle.gov	Seattle data portal
https://data.cityofchicago.org	The City of Chicago's Data Portal is dedicated to promoting access to government data and encouraging the development of creative tools to engage and serve Chicago's diverse community.
https://data.southbendin.gov	Welcome to the South Bend Open Data Portal. Freely accessible and transparent municipal data from the City of South Bend
https://data.detroitmi.gov/	The City of Detroit has launched its Open Data Portal initiative in order to increase public access to valuable data and information concerning City government operations and service delivery
https://data.austintexas.gov/	Austin's open data portal

State Website	Description
https://data.illinois.gov	Welcome to the State of Illinois Open Data Portal Initiative. This is a clearinghouse of various data sets in a standard format that is readable by virtually all computer systems.
https://data.mo.gov	Financial data relating to the purchases of goods and services by the state
https://ohio.gov/government/transparency/	Ohiocheckbook.com empowers taxpayers to follow their money and hold public officials accountable
http://www.utah.gov/transparency/index.html	Dedicated to the transparency and accountability of government finances

Table 48: Examples of municipal and state data portal websites