Abstract

As part of the IMLS National Digital Platform initiative, The University of Georgia Alexander Campbell King Law Library, in partnership with the UGA Terry College of Business Department of Management Information Systems (MIS) and Georgia School of Law, propose to design an open-source software application providing data access and visualization of the U.S. legislative process focused on the passage of the Jumpstart Our Business Startups Act of 2012 (JOBS Act). An interactive, informative tool will aid citizens in becoming more engaged, allowing them to form their opinions in a quick, fact-based, and safe online environment.

Lead Applicant and Collaborators

- **Hani Safadi** (PI): Assistant Professor in the Terry College of Business
- **Usha Rodrigues** (Co-PI): Associate Dean for Faculty Development and M.E. Kilpatrick Chair of Corporate Finance and Securities Law at the University of Georgia School of Law
- **Carol Watson** (Co-Investigator): Executive Director of the Alexander Campbell King Law Library
- **Thomas Striepe** (Co-Investigator): Faculty Services Librarian at the Alexander Campbell King Law Library
- **Nicholas Berente** (Co-Investigator): Associate Professor with the Terry College of Business
- **Richard Watson** (Co-Investigator): Regents Professor and the J. Rex Fuqua Distinguished Chair for Internet Strategy in the Terry College of Business

To promote outreach we are partnering with several entities within the University System of Georgia: GALILEO (Georgia's Virtual Library), the UGA College of Education, the Willson Center Digital Humanities Lab, and the UGA Grady College of Journalism.

Specific Outcomes and Tangible Products

There are two main products: First, a database to integrate fragmented data sets around the JOBS Act. The data include the legislation, proposed and final rules, public comments filed with the SEC, and the text of speeches made by SEC officials. The database will serve as a repository for the documents. Second, software will be developed to access the database, analyze the data and make it available for citizens to interact and draw insights with an interactive visual interface.

Timeline

May 2017 – April 2018

Community Needs Addressed

With growing technological capabilities but shrinking civic engagement, laws, regulations, and policy agendas are pushed through Congress with little public oversight. To be more informed and engaged, the public needs to understand the relationships between the data—to tease out for example, how a corporate constituent's lobbying or PAC contributions trace through to legislative change, how a bank's numerous meetings with its regulator translate into rulemaking, or how feedback and comments on regulation proposals find their way into law. The platform will provide data access and visualization of the U.S. regulatory and legislative processes to the public.

Intended Audience

Contrary to popular belief, libraries are seeing a rise in attendance. Patrons are now putting free computer and internet access in the same category of importance as book. Young Americans are heavy technology users. The Digital Transparency Platform will be another valuable resource for civic engagement of these younger citizens and for library patrons in general.

University of Georgia Research Foundation, Inc.

Design and Evaluation of a 'Digital Transparency Platform' for Advancing Legislative and Regulatory Transparency

The University of Georgia Alexander Campbell King Law Library, in partnership with the UGA Terry College of Business Department of Management Information Systems (MIS) and Georgia School of Law, request $49,839 as part of the IMLS National Digital Platform initiative to design an open-source software application providing data access and visualization of the U.S. legislative process. An interactive, informative tool such as the proposed Digital Transparency Platform will aid citizens in becoming more informed and engaged, allowing them to form their opinions in a quick, fact-based, and safe online environment.

1 Statement of National Need

> "Openness in government strengthens our democracy, promotes the delivery of efficient and effective services to the public, and contributes to economic growth. As one vital benefit of open government, making information resources easy to find, accessible, and usable can fuel entrepreneurship, innovation, and scientific discovery that improves Americans' lives and contributes significantly to job creation." - President Obama Executive Order (May 9, 2013)

Democracy is more effective when there is an unimpeded flow of information between citizens and government and there is a high level of authentic citizen participation in the political process (Watson & Mundy, 2001). The open government initiative of the US federal government urges the implementation of three principles: transparency, participation, and collaboration (WhiteHouse.Gov, 2016). Transparency can be achieved by providing the public with information about what the government is doing, which promotes increased accountability. Open government agencies disclose information about their operations and decisions rapidly in forms that the public can readily find and use (Chun, Shulman, Sandoval, & Hovy, 2010).

The promise of open government is hampered by inconsistent formats and file structures for publicly available data. For example, U.S. law provides boundless access to data about U.S. lawmaking to the public: the language of different versions of bills, their cosponsors, the citizens and PACs that donate to them, lobbying data, the meetings regulators have with the general public, and of course, public comment on administrative rulemaking. Unfortunately, these data are made available in different formats across repositories. Although all the information needed for public self-education is theoretically available, it is simply not sufficiently integrated to make the legislative process readily transparent. The release of abundant but separate data files does little to effectuate real knowledge and curbs what could be a very informed and engaged U.S. populous.

With growing technological capabilities but shrinking civic engagement (Bonnie, Stroud, & Breiner, 2015), laws, regulations, and full policy agendas are being pushed through Congress with little public oversight. To be more informed and engaged, the public needs to understand the relationships between and among the data—to tease out for example, how a corporate constituent's lobbying or PAC contributions trace through to legislative change, how a bank's numerous meetings with its regulator translate into rulemaking, or how feedback and comments on regulation proposals find their way into law. Indeed, a project funded by a 2013 Laura Bush 21st Century Librarian Program Grant recommended that public libraries adopt a focus on the demand side of open government, and fund and carry out a set of pilot projects focused on building new

understanding of preferred and best public library open government practices.[1] We propose to build an interactive and informative tool to provide data access and visualization of the U.S. regulatory and legislative processes to the public.

<u>Partnership and Outreach</u>

To achieve the above-mentioned goal, we partner with several institutions within and outside of the University of Georgia to ensure outreach to many segments of the society, and to develop an understanding of how technology built on open data connects to various aspects of civic life:

- Alexander Campbell King Law Library: The Alexander Campbell King Law Library, in keeping with its status as the oldest and largest public law library in the state, maintains a collection of over 500,000 digital and print titles, provides access to more than 100 online databases, and staffs ten faculty librarians, eight of whom have law degrees. In addition to providing professional expertise in legal research for the university community and the north Georgia region, the Law Library will contribute the technical and computing resources necessary to develop and sustain the Digital Transparency Platform.
- Georgia Library Learning Online (GALILEO): Lucy Harrison, Assistant Vice Chancellor for Academic Library Services and Executive Director of GALILEO, Georgia's virtual library database, sees the Digital Transparency Platform as a potentially invaluable tool for Georgians and citizens nationwide. Ms. Harrison is excited for the opportunity to work with us on piloting and promoting the tool in academic and public libraries across the state of Georgia. The GALILEO community comprises 2000 institutions and allows access to over 100 online databases and 10,000 full-text online journals.
- University of Georgia College of Education: Assistant Professor Kevin Burke is developing new and innovative curriculum for involving youth in their own communities to inspire local change. He and his colleagues in South Bend, IN; Buffalo, NY; Savannah, GA and Athens, GA are developing new models for youth participatory research (YPR) and youth civic engagement. Professor Burke finds the Digital Transparency Platform to have significant benefits for his research activities. He and his colleagues will pilot the prototype within their network of students and faculty.
- University of Georgia Willson Center Digital Humanities Lab (DigiLab): The Digital Humanities Initiative (DIGI) is an interdisciplinary hub for digital humanities projects and digital instruction space. DigiLab will help introduce the tool to a wide campus community and incorporate it as a component in courses associated with the Digital Humanities Undergraduate Certificate program.
- University of Georgia Grady College of Journalism and Mass Communication: The mission of Grady College is to, "assist the public in using the mass media to meet personal and community needs." In line with this mission, Grady specializes in research on political and policy communication, and the Dean of the Grady College, Charles Davis, has expressed strong support for the project and a willingness to collaborate going forward. This expertise will aid in gaining awareness for the tool with a broader audience.

---

[1] Award Number RE-00-13-0087-13 for the project titled The Role of Public Libraries in Improving Local Open Government Ecosystems

University of Georgia Research Foundation, Inc.

2 Project Design

We propose to create an online Digital Transparency Platform using state-of-the-art computational techniques to relate diverse data from multiple different sources over time, visualize the evolution of laws, and trace sources of influence in the lawmaking process. We postulate that with all the legislative data available at one data source and at their fingertips via this tool, citizens will become more educated about the legislative process and more engaged members of our democracy. The prototype will create a proof-of-concept and enable us to identify the resources and infrastructure requirements for scaling the project to handle future federal government legislation on an ongoing basis. The prototype will also be used to gauge what query and display services will meet citizens' needs. We outline below our plan for designing and building the prototype and our long-term plans for the project beyond this planning grant.

<u>Proof of Concept</u>

We plan to create the initial platform on the data surrounding the Securities and Exchange Commission (SEC) rules created in response to the passing of the Jumpstart Our Business Startups Act of 2012 (JOBS Act). Software will be developed for collecting and converting text files into database elements and next-generation text mining and relational event and social network analysis techniques will be used to automate the process of creating the relevant links between these various elements. An interface will be created to present the result of this analysis through intuitive visualizations to support public inquiry. The interface will be sufficiently friendly so that it may be utilized by K-12 through college students, teachers, journalists, lay citizens, activists, and government officials. In terms of operation, the platform will have the capability to field questions from all targeted groups. A citizen might ask: "Who in California lobbied for section A? Who did the regulators meet with and when? Who are the lawmakers in favor of and opposed to the idea in this subsection of the statute?" This platform will visualize the results and allow citizens to probe further and ask more questions.

There are two main steps in realizing the prototype. First and foremost, we plan to collect and integrate fragmented data sets around the JOBS Act into a single database. There is abundant online unstructured and semi-structured text-based data around the JOBS Act (SEC, 2016). The data include the legislation, proposed and final rules, public comments filed with the SEC, and the text of speeches made by SEC officials. The database will serve as a repository for the documents, and various tools will analyze the data and make it available for citizens to interact and draw insights (text mining, natural language processing, relational event network and social network analysis). A single database that links these elements requires that relationships be inferred from the text, e.g. a comment by person A relates to subsection B of regulation C, at time T. Computational text mining techniques, such as named entity recognition and topic modeling, have reached a level of accuracy that allows us to build a database linking the textual elements such as comments, speeches, and legislation.

Second, software will be developed to access the database and provide an interface to search and query it. Creating a useful digital platform is complicated by the overwhelming textual process of legislation and regulation development. Interactive visualization technologies increasingly enable non-technical people to readily access large amounts of data. Thus, when comments from constituents, for instance, overlap with topics identified by text mining in a proposed regulation, we can link the two, and present it visually.

University of Georgia Research Foundation, Inc.
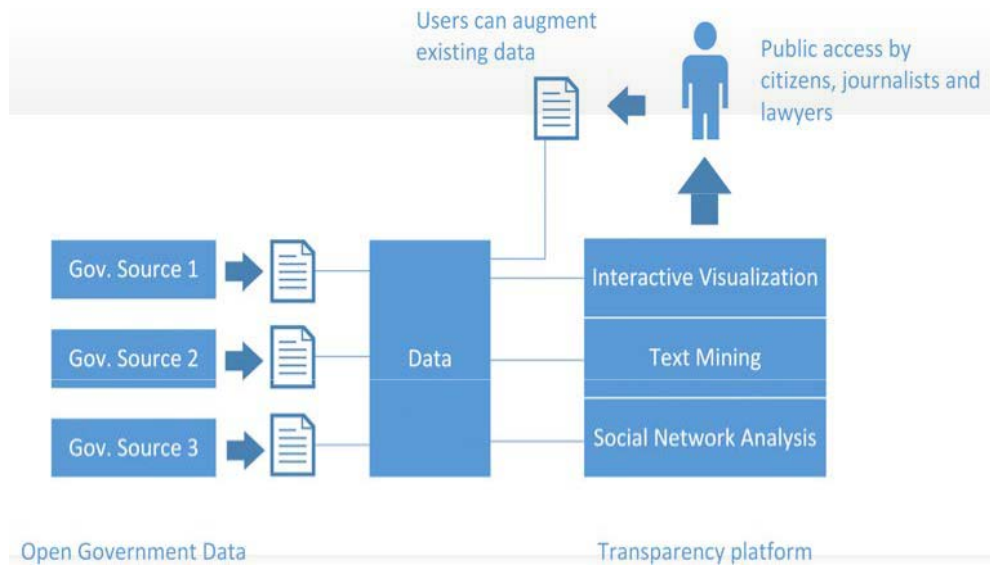
The figure above outlines the architecture of the proposed platform. We plan to experiment with several key technologies to identify concepts and entities from documents, associate them with each other over time, and interactively visualize these relationships to enable the exploration and sensemaking of open government data. As such, this platform will involve the innovative recombination of existing components in a civically useful way. Each of the components for the platform (text analysis, social and relational event networks, visualization) are evolving spaces and the contribution of this research will involve evaluating and integrating different technologies for the specific purpose of delivering on the promise of open government by enabling the public to make sense of volumes of government documents.

We propose three steps for dealing with the challenges of making sense of public government data:

1. ***Concepts from Documents through Text Analysis***: Open government data are fragmented across multiple documents in a variety of formats. This is typically not sufficient for useful openness because standalone information rarely tells the whole story and those interested in understanding government activity typically need to explore the connections between elements of a variety of government documents created for different purposes. Different relevant documents and data sources need to be connected and integrated for useful concepts to enable the deeper awareness of these texts to emerge (Hellberg, 2014).

2. ***Concepts, Entities, and Events over Time through Network Analysis***: Understanding the presence, evolution, and shape of concepts in large sets of texts requires the linking and integrating of data from disparate sources as a precursor to understanding and making sense of it (Böhm et al., 2012; Heise & Naumann, 2012). Once concepts are identified through text analysis, they need to be linked and correlated. The relationships among entities will be represented with multidimensional social networks (Contractor, Monge, & Leonardi, 2011). In particular we will examine both cross-sectional and temporal relationships (Butts, 2008; Quintane, Conaldi, Tonellato, & Lomi, 2014).

3. ***Human Sensemaking through Visualization:*** Transforming raw data to high quality integrated information imposes new burdens on the public sector (Davies & Frank, 2013). This calls for both re-examination of the process of making government data available and improving technology to process it (Hellberg, 2014). Any technological solution should take into consideration engaging the public and enabling their participation and interactive

4

visualization is a key technological component (Chun et al., 2010; Graves & Hendler, 2013). We adopt the idea of external representation with interactive visualization for aiding cognition by evoking perception (Munzner, 2014).

To develop the graphical user interface, we will build on established scholarship for designing systems for text analytics (Stasko, Görg, & Spence, 2008). One key task in the visual presentation of textual data involves synthesizing concepts and ontologies from text (Gotz, Zhou, & Aggarwal, 2006). A suggested framework involves three views: navigation, knowledge, and data (Shrinivasan & van Wijk, 2008). The knowledge view allows the definition, discovery, and modification of concepts. The links between the concepts and evidence are presented in the navigation view using a graph-like representation. This allows the analyst to elicit hypotheses and evidence. Finally, references to relevant information from documents can be accessed in the data view (see Figure below).

We will also build on top of knowledge gained in developing visual systems in the law and open government disciplines such as The Legislative Explorer (Stramp & Wilkerson, 2015) and products of the Congressional Bills Project and Data.gov. These projects have developed an effective system for community engagement by analyzing explicit relationships like party membership, house and senate association, and voting. Building on this, we are looking to extract subtler relationships from textual data such as where an idea originated, who supported or opposed it, etc. Well-designed peripheral elements make our proposed Digital Transparency Platform different from what is currently available, and will theoretically enhance the level of public engagement and participation by lay citizens.
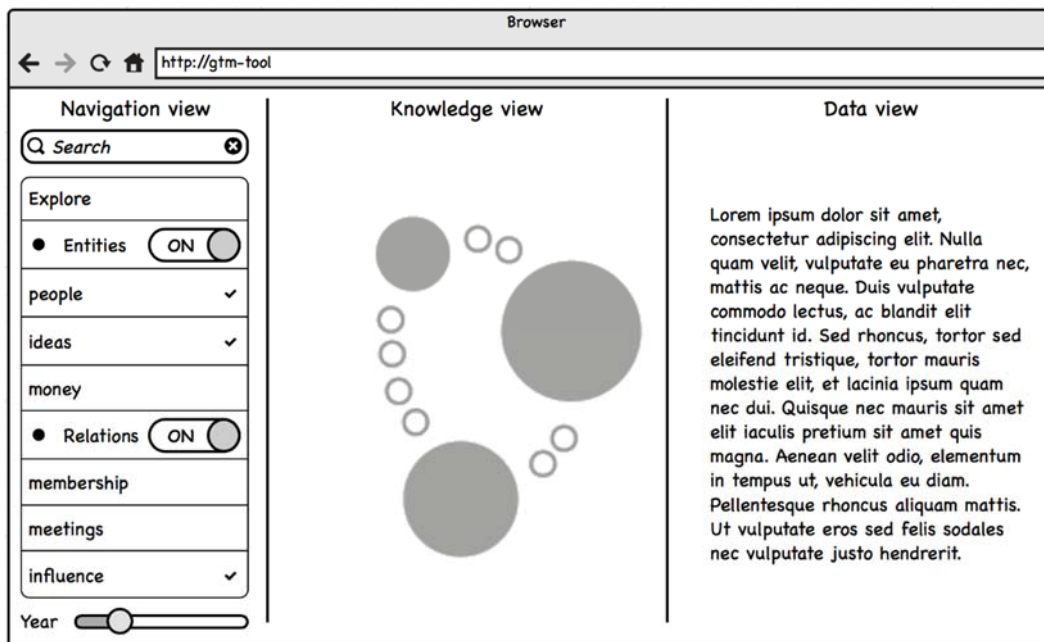
The outcomes of this phase include both a database around the JOBS Act and a platform to search and query it with an interactive visualization component. UGA law students, faculty, and staff will pilot the tool before going fully public with the platform. Once publicly available, anyone with internet access may use the tool. Because of our focus on building a platform that will be open to

the public, we choose to use open-source natural language processing toolkits and graph databases. Open-source software provides the flexibility of building on the transparent codebase and offering the developed software to others to build upon.

<u>Plan</u>

For the purposes of testing feasibility, the prototype proposed here is confined to a limited domain (JOBS Act). Following is a proposed budget for the prototype and a project timeline.

<u>Budget</u>

We request $49,839 for the IMLS Planning Grant. The breakdown of these funds are as follows: summer salary support for Professor Safadi ($13,907 plus $2,699 benefits). Summer support is also requested for Professor Rodrigues ($3,164 plus $601 benefits) who will be mentoring and managing the law student researcher. We also include stipends for one information systems/computer science graduate student research assistant ($16,293 plus $978 benefits) and one law student research assistant ($1,872 plus $28 benefits). $10,284 in indirect costs are requested at a rate of 26 percent MTDC.

<u>Timeline</u>

ID	Task Name	2017								2018			
		May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
1	1. Creating the database	████	████	████	████	████							
2	1.1. Data collection	████	████										
3	1.2. Text mining for data integration			████	████								
4	1.3. Database cleaning and tuning					████							
5	2. Developing the software					████	████	████	████	████	████	████	████
6	2.1 Developing the backend (computational linguistics)					████	████	████	████				
7	2.2. Developing the frontend (social network analysis and visualization)								████	████	████		
8	2.3. Usability testing									████	████	████	
9	3. Project report & sustainability plan												████

- Future plans include scaling the process to handle future federal government legislation.
- The Alexander King Law Library will permanently host the databases, providing ongoing technical support and maintenance. It is possible that the creation of a new librarian position may be needed to maintain integrity of the Digital Transparency Platform's data. These staffing needs will be explored in the planning phase.
- Another avenue of future work will involve usability testing and exploratory experimentation to iteratively enhance the platform in a way that attends to both industry best practice and extend relevant scholarship.
- Ultimately, an app will be developed particularly for those whose mobile devices are their primary avenue for internet access.
- Beyond open government, the activities in this project will move our knowledge forward for how to process and visualize text data for human sensemaking. This includes identifying visualization techniques and visual encoding idioms suitable for temporal textual data and creating algorithms to identify and correlate entities and concepts from text. The application

of such tools and algorithms extends beyond open government data to various forms of large textual datasets including news, legal documents, social media and online communities.

Team qualifications

The project team includes a legal scholar, with a recent specific focus on the JOBS Act, and information systems scholars, with skills in programming, database design and retrieval, text mining, computational linguistics, and the management of scientific projects. In addition, faculty services librarians from the Law Library, and two student researchers from information/computer science and law will be involved in the prototype development phase.

Hani Safadi is an Assistant Professor in the Terry College of Business at the University of Georgia. Before joining Terry, he was a visiting faculty in Stevens Institute of Technology. He holds a Ph.D. in Management Information Systems from McGill University. Dr. Safadi studies the role that technology plays in organizing work and producing innovations that are traditionally conceived within the boundaries of established firms and businesses. Methodologically, he brings skills in social network analysis and natural language processing. He is interested in mixed-methods research and the application of computational linguistics in studying text, documents and other qualitative data sets. Dr. Safadi will lead the project and in particular will be actively involved in every element of the development and evaluation of the platform.

Usha Rodrigues is the Associate Dean for Faculty Development and M.E. Kilpatrick Chair of Corporate Finance and Securities Law at the University of Georgia School of Law. She researches in the areas of corporate governance and securities law, and has a special focus on the JOBS Act. She has worked closely with individuals from the U.S. Securities and Exchange Commission and has acquired (or can acquire) the raw text files in digital format necessary for this project. A member of the prestigious American Law Institute, she has served as chair of the Business Associations Section of the American Association of Law Schools and as president of the Law and Entrepreneurship Association. Dean Rodrigues contributes an intimate knowledge of the JOBS Act (e.g. Rodrigues, 2016; Rodrigues & Stegemoller, 2014) and the processes of governing, and she will also assist with the evaluation of the platform and overseeing graduate students from the Law School.

Carol A. Watson, J.D. is the Executive Director of the Alexander Campbell King Law Library. Ms. Watson has written extensively on institutional repositories, and received the 2015 American Association of Law Libraries SIS Outstanding Article Award. She also annually coordinates and speaks on internet legal research topics at the Institute of Continuing Legal Education in Georgia and frequently presents at American Association of Law Libraries annual meetings and The Center for Computer-Assisted Legal Instruction conferences for law school computing. Ms. Watson will lead with her knowledge of internet legal research regarding tool usability, target audience assessment and tool distribution and outreach.

Thomas J. Striepe, J.D. is the Faculty Services Librarian at the Alexander Campbell King Law Library. As head of the Digital Commons Team, he manages the Law School's institutional repository and serves as primary liaison to the faculty. Additionally, he supervises the library's Research Assistant Program and teaches courses in legal research. Mr. Striepe has presented on creating collaborations and using apps and technology for legal research at the American Association of Law Libraries annual conferences and Georgia Institute for Continuing Legal Education seminars. His experience in initiating the integration of technology into the Law School will be integral to the implementation and testing of the platform.

University of Georgia Research Foundation, Inc.

Nicholas Berente is an Associate Professor with the Terry College of Business at the University of Georgia. He received his Ph.D. from Case Western Reserve University and conducted his postdoctoral studies at the University of Michigan. He is a visiting fellow with the University of Liechtenstein, visiting faculty at the Indian School of Business, and Associate Editor for *Information Systems Research*. He is principal investigator of multiple National Science Foundation projects on advanced cyberinfrastructure. He studies how digital innovations are reshaping society's institutions through a variety of methods, including next-generation computational techniques. Dr. Berente will assist with all areas of development and evaluation, but also contributes an understanding human sensemaking processes to the collaboration.

Richard Watson is a Regents Professor and the J. Rex Fuqua Distinguished Chair for Internet Strategy in the Terry College of Business at the University of Georgia. He is the current Research Director for the Advanced Practices Council of the Society of Information Management and a former President of the Association for Information Systems. In 2011, he received the Association for Information Systems' LEO award, which is given for exceptional lifetime achievement in Information Systems. For several years, he has been a visiting researcher at Viktoria Swedish ICT engaged in establishing and applying Maritime Informatics to the European shipping industry. He is an Honorary Visiting Professor at Xi'an Jiaotong University. Dr. Watson is a specialist in data management and in development of data analytics approaches, and he also has a record of research into open government. He will work with the project primarily in an advisory capacity.

3 National Impact

Not only are libraries seeing a rise in attendance (contrary to popular belief), patrons are now putting free computer and Internet access in the same category of importance as book borrowing (Zickuhr, Rainie, Purcell, & Duggan, 2013). Young Americans are heavy technology users. The Digital Transparency Platform will be another valuable resource for civic engagements of these younger citizens and for library patrons in general.

The platform is a form of digital public good that amplifies the opportunity for public participation and engagement. Further, as we have seen over the last few decades, digitization is a progenitor of innovation, and the platform will be devised in such a way that it invites user contributions as well as extension and augmentation of the platform, thus actively sparking and encouraging further innovation. The Alexander Campbell King Law Library's development and hosting of the open source Digital Transparency Platform will enable widespread use of the tool and pave the way for other university and public libraries to feature the tool on their websites and home screens as a legitimate research tool.

We can only speculate about the eventual impact of having an interactive, visual representation of the lawmaking process. Accurate and reliable insight into the lawmaking process is now so difficult and manually-intensive – requiring that people trace through reams of government documents from different sources. Through the platform, anyone will be able to quickly and intuitively understand the main ideas and concepts of every bit of legislation and connect these to the sponsors and influencers – and opposing concepts - visually over time. All of this insight will be at any novice user's fingertips.

Although the prototype proof-of-concept is backward looking and focused on one law (the JOBS Act), one can anticipate that the platform could eventually cover all legislation processes, including dynamically evolving lawmaking processes in-progress. There are numerous potential applications for this platform. The initial use case would be to encourage engagement and interest of any typical

citizen in the lawmaking process. It could be used for K-12 students to understand how laws actually come about and to foster interest in the process. University students could use the platform in their coursework. Journalists could use the platform to understand laws and those who influence laws. For both students and journalists alike the platform could help them navigate the lawmaking process and look for patterns which would point to domains for further inquiry. Influencers, lobbyists, activists, and lawmakers themselves could look to see who is aligned with their goals and who is not, and affect how they form coalitions and negotiate.

Finally, it is important to note the impact that the Digital Transparency Platform can have on academic research. Legal scholars, psychologists, sociologists, political scientists and the like now spend hours and hours poring over documents to gain insight into the lawmaking process and associated institutions and organizations. With interactive visualization of a common database and next generation text and network analysis embedded in the platform, future researchers can accomplish in minutes what has taken months in the past. Further, those interested in studying platforms themselves – information scientists, librarians, information systems researchers, and computer scientists – will have a novel and extensible platform through which they can explore and develop ideas around data visualization, human-computer interaction, user-generated content, sensemaking, knowledge sharing, and a host of other topics and their relationship to policymaking and institutional change. Although the prospects of any open-ended digital platform are wide and various, and the shape of its evolution will be difficult to predict, the Digital Transparency Platform will provide an extensible environment that can be adapted for a host of purposes.

References

Böhm, C., Freitag, M., Heise, A., Lehmann, C., Mascher, A., Naumann, F., … Schmidt, M. (2012). GovWILD: integrating open government data for transparency. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 321–324). ACM.

Bonnie, R. J., Stroud, C., & Breiner, H. (2015). *Investing in the health and well-being of young adults*. National Academies Press.

Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology*, *38*(1), 155–200. http://doi.org/10.1111/j.1467-9531.2008.00203.x

Chun, S. A., Shulman, S., Sandoval, R., & Hovy, E. (2010). Government 2.0: Making connections between citizens, data and government. *Information Polity*, *15*(1), 1.

Contractor, N. S., Monge, P. R., & Leonardi, P. M. (2011). Multidimensional networks and the dynamics of sociomateriality: Bringing technology inside the network. *International Journal of Communication*, *5*(1), 682–720. http://doi.org/1932–8036/20110682

Davies, T., & Frank, M. (2013). "There"s no such thing as raw data': exploring the socio-technical life of a government dataset. In *WebSci '13: Proceedings of the 5th Annual ACM Web Science Conference* (pp. 75–78). http://doi.org/10.1145/2464464.2464472

Gotz, D., Zhou, M. X., & Aggarwal, V. (2006). Interactive visual synthesis of analytic knowledge. In *IEEE Symposium on Visual Analytics Science and Technology 2006, VAST 2006 - Proceedings* (pp. 51–58). http://doi.org/10.1109/VAST.2006.261430

Graves, A., & Hendler, J. (2013). Visualization Tools for Open Government Data. In *Proceedings of the 14th Annual International Conference on Digital Government Research* (pp. 136–145). New York, NY, USA: ACM. http://doi.org/10.1145/2479724.2479746

University of Georgia Research Foundation, Inc.

Heise, A., & Naumann, F. (2012). Integrating open government data with stratosphere for more transparency. *Web Semantics: Science, Services and Agents on the World Wide Web*, *14*, 45–56.

Hellberg, A.-S. (2014). Policy, Process, People and Public Data. In *13th IFIP WG 8.5 International Conference on Electronic Government* (Vol. 8653, pp. 265–276). Retrieved from http://link.springer.com/chapter/10.1007/978-3-662-44426-9_22

Munzner, T. (2014). *Visualization Analysis and Design*. CRC Press.

Quintane, E., Conaldi, G., Tonellato, M., & Lomi, A. (2014). Modeling Relational Events: A Case Study on an Open Source Software Project. *Organizational Research Methods*, *17*(1), 23–50. http://doi.org/10.1177/1094428113517007

Rodrigues, U. (2016). Dictation and Delegation in Securities Regulation. *Indiana Law Journal, Forthcoming*.

Rodrigues, U., & Stegemoller, M. (2014). What all-cash companies tell us about IPOs and acquisitions. *Journal of Corporate Finance*, *29*, 111–121. http://doi.org/http://dx.doi.org/10.1016/j.jcorpfin.2014.07.003

SEC. (2016). SEC.gov | Jumpstart Our Business Startups (JOBS) Act. Retrieved from https://www.sec.gov/spotlight/jobs-act.shtml

Shrinivasan, Y. B., & van Wijk, J. J. (2008). Supporting the analytical reasoning process in information visualization. *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08*, 1237. http://doi.org/10.1145/1357054.1357247

Stasko, J., Görg, C., & Spence, R. (2008). Jigsaw: supporting investigative analysis through interactive visualization. *Information Visualization*, *7*(2), 118–132. http://doi.org/10.1057/palgrave.ivs.9500180

Stramp, N., & Wilkerson, J. (2015). Legislative Explorer: Data-Driven Discovery of Lawmaking. *PS: Political Science & Politics*, *48*(1), 115–119.

Watson, R. T., & Mundy, B. (2001). A strategic perspective of electronic democracy. *Communications of the ACM*, *44*(1), 27–30. http://doi.org/10.1145/357489.357499

WhiteHouse.Gov. (2016). Transparency and Open Government. Retrieved from https://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment

Zickuhr, K., Rainie, H., Purcell, K., & Duggan, M. (2013). *How Americans value public libraries in their communities*. Retrieved from http://libraries.pewinternet.org/2013/12/11/libraries-in-communities/

Schedule of Completion

Project Period: 05/01/2017 – 04/30/2018

ID	Task Name	2017								2018			
		May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
1	1. Creating the database	█	█	█	█	█							
2	1.1. Data collection	█	█										
3	1.2. Text mining for data integration			█	█								
4	1.3. Database cleaning and tuning					█							
5	2. Developing the software					█	█	█	█	█	█	█	█
6	2.1 Developing the backend (computational linguistics)				█	█	█	█	█				
7	2.2. Developing the frontend (social network analysis and visualization)							█	█	█	█		
8	2.3. Usability testing										█	█	
9	3. Project report & sustainability plan												█

DIGITAL PRODUCT FORM

Introduction
The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded digital products (i.e., digital content, resources, assets, software, and datasets). The products you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products can be challenging. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions
You must provide answers to the questions in Part I. In addition, you must also complete at least one of the subsequent sections. If you intend to create or collect digital content, resources, or assets, complete Part II. If you intend to develop software, complete Part III. If you intend to create a dataset, complete Part IV.

PART I: Intellectual Property Rights and Permissions

A.1 What will be the intellectual property status of the digital products (content, resources, assets, software, or datasets) you intend to create? Who will hold the copyright(s)? How will you explain property rights and permissions to potential users (for example, by assigning a non-restrictive license such as BSD, GNU, MIT, or Creative Commons to the product)? Explain and justify your licensing selections.

A.2 What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

A.3 If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

Part II: Projects Creating or Collecting Digital Content, Resources, or Assets

A. Creating or Collecting New Digital Content, Resources, or Assets

A.1 Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and format you will use.

A.2 List the equipment, software, and supplies that you will use to create the content, resources, or assets, or the name of the service provider that will perform the work.

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to use, along with the relevant information about the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

B. Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance. Your plan may address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

C. Metadata

C.1 Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata. Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

C.2 Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

D. Access and Use

D.1 Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

D.2 Provide the name(s) and URL(s) (Uniform Resource Locator) for any examples of previous digital content, resources, or assets your organization has created.

Part III. Projects Developing Software

A. General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

A.2 List other existing software that wholly or partially performs the same functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your software and explain why you chose them.

B.2 Describe how the software you intend to create will extend or interoperate with relevant existing software.

B.3 Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

B.4 Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

B.5 Provide the name(s) and URL(s) for examples of any previous software your organization has created.

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under open-source licenses to maximize access and promote reuse. What ownership rights will your organization assert over the software you intend to create, and what conditions will you impose on its access and use? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain and justify any prohibitive terms or conditions of use or access and detail how you will notify potential users about relevant terms and conditions.

C.2 Describe how you will make the software and source code available to the public and/or its intended users.

C.3 Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository:

URL:

Part IV: Projects Creating Datasets

A.1 Identify the type of data you plan to collect or generate, and the purpose or intended use to which you expect it to be put. Describe the method(s) you will use and the approximate dates or intervals at which you will collect or generate it.

A.2 Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

A.3 Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

A.4 If you will collect additional documentation, such as consent agreements, along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

A.5 What methods will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

A.6 What documentation (e.g., data documentation, codebooks) will you capture or create along with the dataset(s)? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

A.7 What is your plan for archiving, managing, and disseminating data after the completion of the award-funded project?

A.8 Identify where you will deposit the dataset(s):

Name of repository:

URL:

A.9 When and how frequently will you review this data management plan? How will the implementation be monitored?