

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 Law, request \$49,818 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 digital transparency platform we are proposing aids users in becoming more informed and engaged citizens, allowing them to form their own opinions in a quick, fact-based, and safe online environment.

Statement of National Need

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, all of these data are made available in different formats across repositories. Although all of 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 and shrinking civic engagement¹, laws, regulations, and full policy agendas are being pushed through Congress with little public oversight. In order to be more informed, engaged citizens, 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.

Some open source projects, such as Legex and products of the Congressional Bills Project and Data.gov, 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 citizen.

Project Design

We propose to create an online digital platform using next generation 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. The assumption is that with all of the legislative data available at one data source and at the fingertips of citizens via this tool, citizens will become more educated 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 process to handle all future federal government legislation. The prototype will also be used to gauge what query and display services will meet citizens' needs.

We have every intention of creating a tool so user friendly that it may be utilized by K-12 through post-secondary college students, teachers, journalists, lay citizens, activists, and government officials. In terms of operation, the digital platform will have the capability to field questions from users. The user can ask: who in California lobbied for section A? And, where did the money go? This tool will visualize the results and allow the user to probe further and ask more questions.

¹ Committee on Improving the Health, Safety, and Well-Being of Young Adults; Board on Children, Youth, and Families; Institute of Medicine; National Research Council; Bonnie RJ, Stroud C, Breiner H, editors. Washington (DC): National Academies Press (US); 2015 Jan 27.

Creating a useful digital platform is complicated by the overwhelming textual process of legislation and regulation development. A single database that links these elements requires that relationships are 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 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. Interactive visualization technologies increasingly enable even 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 to the public.

For the prototype, the project team proposes 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). The database will include the legislation, proposed and final rules, public comments filed with the SEC, and the text of speeches made by SEC officials. Software will be developed for converting text files into database elements and text mining used to automate the process of creating the relevant links between these various elements. An interface will be created with intuitive visualizations to support general public inquiry.

UGA law students, faculty, and staff will pilot the tool before going fully public with the platform. Once public, anyone with internet access may use the tool. Ultimately, an app will be developed especially for those whose mobile devices are their primary avenue for Internet access. 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. The Alexander King Law Library will permanently host the databases, providing ongoing technical support and maintenance.

The IMLS Planning Grant team includes Thomas Striepe, Faculty Services Librarian; Associate Dean Usha Rodrigues, Professor of Law; Hani Safadi, Assistant Professor of Management Information Systems; and three student researchers, two from information/computer science and one from law. Modest summer salary and student support is required to complete the project planning and prototype development.

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 borrowing². Young Americans are heavy technology users. The digital transparency platform will be yet another valuable resource for patrons to frequent, while fostering a culture for civic engagement.

The platform is a form of digital public good that amplifies the opportunity for public participation and augmentation. As we have seen over the last few decades, digitization is a progenitor of innovation. The Alexander Campbell King Law Library's development and hosting of the open source digital platform will normalize the 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.

Budget

We would like to request \$49,818 for the IMLS Planning Grant. The breakdown of these funds are as follows: two months of salary support for Professor Safadi (\$25,007 plus \$4,854 benefits), who will devote 1.47 person-months over the summer. Summer support of 0.15 person-months is requested for Professor Rodrigues (\$3,164 plus \$614 benefits) who will be mentoring and managing the law student researcher. We also include stipends for three student researchers who will be working part time on collecting and identifying data and data sources for the prototype (\$5,872 plus \$28 benefits). \$10,280 in indirect costs are requested at a rate of 26 percent MTDC.

² Zickuhr, Raine, Purcell and Duggan; Princeton Survey Research Associates International for the Pew Research Center's Internet & American Life Project; October 2013