America Contacts Congress
Providing Access to U.S. Congressional Correspondence Data

Prepared by Danielle Emerling, WVU Libraries
July 2019

PROJECT GOALS

Constituent correspondence with Congress provides unique opportunities to explore the perspectives of everyday Americans about the issues and policies affecting their lives. [1] Found in the congressional archives of libraries throughout the country, these records enable explorations of the influence of public opinion on elected officials and vividly illustrate the representative function of an office, particularly to constituent groups whose voices are often absent from the historical record. Scholars have long used paper constituent correspondence for numerous inquiries.

In the 1990s, congressional offices adopted electronic, proprietary systems to manage constituent correspondence, and soon after, libraries began to receive data exports from these systems. The flat data or relational database with a library of attached records is transferred when a member leaves office. At least two dozen libraries and archival repositories have acquired constituent correspondence data, and more receive such data at the close of each Congress. The complexity, format, and sensitive nature of the exported data have posed challenges for repositories. No repository has been able to process and provide access to these data in a replicable way.

In 2017, a report from the Society of American Archivists Congressional Papers Section cautioned that without a concerted effort by the congressional archival community, these important historical sources could be lost to the inherent threats of digital materials. [2]

AN ENTIRE BODY OF HISTORICAL DOCUMENTATION OF AMERICAN DEMOCRACY IS IN DANGER OF DISAPPEARING.

Archiving Constituent Services Data of the U.S. Congress: A report of the SAA Congressional Papers Section

In 2016, West Virginia University Libraries developed an open-source prototype that ingested and searched the constituent correspondence data from the office of former Senator Jay Rockefeller. With this tool, WVUL embarked on the LYRASIS supported feasibility study to determine whether the system was unique, warranted further development, and could be implemented at other libraries to process and provide access to these data sets. Specifically, the feasibility study sought to:

1. assess the system developed by WVUL;
2. engage the library community to define technical requirements for processing and accessing data;
3. determine and define the needs of potential users of the data;
4. and create a roadmap for developing cooperative governance and infrastructure.

---

[1] For an example of paper constituent correspondence found in congressional collections, see: Robertson, Cambell, "A Mirror of Today's Political Fray in Letters From 50 Years Ago," *The New York Times*, October 29, 2017, https://www.nytimes.com/interactive/2017/10/29/us/gore-letters.html

[2] Society of American Archivists Congressional Papers Section CSS/CMS Task Force, "Archiving Constituent Services Data of the U.S. Congress," November 2017, https://www2.archivists.org/sites/all/files/2017_CSS_CMS_Report.pdf.

PROJECT MODEL

In September 2018, WVUL hired AB Consulting to conduct the feasibility study. In addition to the congressional papers archivist at WVUL, the study benefited from an advisory board of experts, including Nathan Gerth, University of Nevada Reno; Natalie Bond, University of Montana/Montana State University; Brandon Pieczko, Richard B. Russell Library for Political Research and Studies at the University of Georgia; John Caldwell, University of Delaware; Hope Bibens, Drake University; and Alison White, U.S. Senate Historical Office. The study was broken into three phases, each of which built on the findings from the previous phase. The consultant filed a report at the end of each phase.

Phase I: Assess Existing Functionality (2018 October-December)

At an initial meeting at WVUL with the congressional papers archivist and systems development department, the consultant discussed and documented existing and potential tool functionality, including inherent limitations of the data and the system. The consultant also identified priority user groups and desired functionality/technical requirements of the congressional papers community by facilitating workshops with congressional and digital archivists.

Phase II: Develop User Requirements and Characterize Feasibility (2018 December-2019 March)

This phase included parallel work on data testing and testing preliminary ideas about users through focus groups and interviews. Two members of the Advisory Board conducted data testing at their institutions. The consultant held interviews with 10 potential users of the data (qualitative and quantitative researchers) and a focus group with congressional archivists to prioritize tool functions.

Phase III: Justification and Roadmap (2019 February-April)

Taking the work of the previous two phases, Phase III characterized the uses of and need for further development of the tool, created a development roadmap, and identified potential options for medium- to long-term administration and sustainability.

Final Presentation and Discussion (2019 May 1-15)

On May 14, 2019, the consultant and members of the Advisory Board presented the findings of the Phase III report to the congressional community, including members of the Society of American Archivists' Congressional Papers Section and the Association of Centers for the Study of Congress.
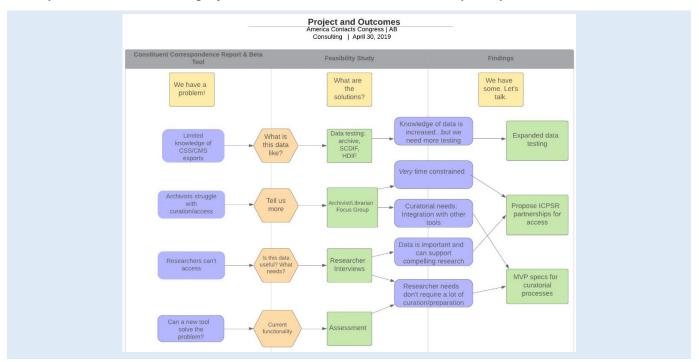
Overall, the project model allowed the consultant and Advisory Board to successfully address the goals of the project. One of the key strengths of this model was community engagement. The Advisory Board brought expertise to the project, as well as a sense of community ownership.

Workshops and focus groups with congressional archivists in both Phases I and II, as well as the final webinar and subsequent presentations at various conferences, not only provided valuable feedback to the project, but also were a step toward turning users into stakeholders. Further, the community engagement outlined above, as well as interviews conducted with potential end users of the data, challenged assumptions and led to the most important insights of the project.

Though no significant modifications were made to the overall project goals and model outlined in the proposal, it became clear early in the project that a better understanding of the data sets was necessary to evaluate what is, and is not, possible with the data. Thus, two elements were added to Phase II.

The first, data testing, was an effort to better understand the data to more knowledgeably address issues of scale and structure. In Phase I, the consultant noted that the tool's basic functionality was based on just two data sets. This shifted the project plan to approximately equal focus on user testing and data testing in Phase II.

Second, the consultant identified a need for Systems/IT input on the tool. The project needed a sense of potential interest and support after the grant in the form of funds, developer time, or both. The consultant distributed a brief survey to IT departments of Advisory Board members in December 2018, but unfortunately had a low response rate. We determined that these questions could be better addressed by library administration as the project moved forward after the feasibility study.

ACCOMPLISHMENTS

The project accomplished the four stated goals of the feasibility study, but most significantly, it created a realistic path forward for a 20-year-old problem. For the first time, congressional archivists have a better understanding of the constituent correspondence data and have concrete examples of its significance.

At a high level, the study assessed the data tool and defined minimum functions required by priority user groups, including processing archivists/librarians, qualitative researchers, and quantitative researchers. To do so, it engaged the library community to define technical requirements. Through interviews with researchers, the study refined the needs of potential users, and finally, it created a roadmap for the next steps of the project around areas of governance, development, partnerships, and community engagement.

LESSONS LEARNED

1. The tool is unique, and desired functionality is basic. For both processing archivists and researchers, the desired tool functionality is very basic. More important than expansive functionality is that it integrate well with tools used for curation and tools used for large-scale data analysis. Leading up to the project, it was assumed that the tool functionality would need to be more expansive, so this finding is incredibly important and will save the community resources.

2. The data sets present significant promise and manageable challenges. Testing confirmed that the data sets are large with some incompleteness and inconsistency. Data sets are structured as flat files, which the tool was developed to ingest, or relational tables, which will require additional development. Both formats present opportunities to study communication between offices and constituents, legislator responsiveness, and congressional history (political and social). The relational databases have the added potential for study of congressional staffing and office management of constituent communications.

3. Access to the data and attachments is transformative for qualitative and quantitative researchers. Research in this data is attractive for qualitative research because of keyword search and the potential to browse attachments. Structured metadata that is large-scale and searchable opens numerous doors for quantitative research. Further, incomplete and inconsistent data was neither particularly surprising nor a significant barrier for researchers. Again, this means the tool needs to do less than initially assumed.

4. Resources in the community and its associated stakeholders are limited. Libraries and archives lack resources for developing and sustaining this tool as an open-source project and for providing access to this data in general.

5. Partnerships are key moving forward. For tool development, open source is likely not the sole sustainability model, so membership supported and software as a service models may play a role. To facilitate research access, the community should consider a partnership with an allied community with data expertise and infrastructure that archives lack. Finally, the community must determine the best ways in which to partner and govern preservation, curation, and access to these data sets in the long term.

NEXT STEPS

Using the framework articulated in the LYRASIS report, *It Takes a Village: Open Source Software Sustainability, A Guidebook for Programs Serving Cultural and Scientific Heritage,*[3] the consultant created a roadmap to address the areas of governance, technology, resources, and community engagement. Phase IV of this project, which will conclude in June 2020, will include the following activities.

1. Governance

West Virginia University Libraries is the manager of this project, but governance is shifting toward the Association of Centers for the Study of Congress.[4] The ACSC is an existing non-partisan membership organization that includes both curators and researchers, and this initiative fits within their mission to preserve the records of Congress and to promote the study of Congress. An ACSC Constituent Services Data Task Force is reviewing ways in which to codify the charge of coordinating and overseeing the data preservation and access program, allow for community feedback and interaction, manage partnerships, and obtain further resources.

2. Tool Development

The tool began as an open-source project, but sustaining this requires a committed IT community. While open-source development will likely play a role, it is not the entire answer. The ACSC Task Force will consider a software as a service model, in which the tool would be hosted by a vendor, and a membership supported model. Development would benefit from greater demand for a data curation tool, so the Task Force will also investigate whether there is a larger audience for the tool in the archives profession.

---

[3] *It Takes a Village: Open Source Software Sustainability, A Guidebook for Programs Serving Cultural and Scientific Heritage*, LYRASIS, February 2018. Available at https://www.lyrasis.org/technology/Documents/ITAV_Interactive_Guidebook.pdf.

[4] The Association of Centers for the Study of Congress (ACSC) was founded in 2003 as an independent alliance of organizations and institutions which promote the study of the U.S. Congress. Find out more at https://www.congresscenters.org/.

3. Access Partnership

Generally, archives lack the infrastructure to provide access to large data sets containing sensitive or private information, so the ACSC Task Force will work toward a partnership with the Inter-University Consortium for Political and Social Research (ICPSR)[5] at the University of Michigan. In a pilot project, a few institutions will create agreements with ICPSR and transfer data. This will create a model for other repositories. Further, it will bring together multiple constituent correspondence data sets, a great benefit to researchers.

4. Community Engagement

So far, Advisory Board (AB) members have delivered presentations and received community feedback at the Association of Centers for the Study of Congress annual meeting (May 2019), the final project webinar (May 2019), and a meeting of the Advisory Committee on the Records of Congress (June 2019). A representative of ICPSR will join members of the AB for a workshop at the Congressional Papers Section preconference day at the Society of American Archivists annual meeting (August 2, 2019), and members of the AB will present an education session (August 4, 2019). The ACSC Task Force is making plans for involving more community members in decision making.

PROJECT INFORMATION

America Contacts Congress: Providing Access to U.S. Congressional Correspondence Data Project Summary and Reports webpage

    Phase 1 Report, December 19, 2018
    Phase 2 Report, March 22, 2019
    Phase 3 Report, April 30, 2019
    Findings and Next Steps Webinar, May 14, 2019

LYRASIS Member Summit video and presentation slides, October 2018

Constituent Correspondence Data Tool GitHub

Association of Centers for the Study of Congress Constituent Services Data Task Force

---

[5] The Inter-University Consortium for Political and Social Research (ICPSR) is an international consortium of more than 750 academic institutions and research organizations that provides leadership and training in data access, curation, and methods of analysis for the social science research community. Find out more at https://www.icpsr.umich.edu/icpsrweb/.