America Contacts Congress Project

Report 3: Feasibility and Sustainability

For West Virginia University
AB Consulting
2019 April 30

Executive Summary

The *America Contacts Congress* project is a feasibility study for a tool to manage constituent correspondence and case files produced by offices of the U.S. Congress. The output from the systems that congressional offices have used to manage constituent correspondence since the 1990s is currently inaccessible to both archivists/librarians responsible for managing it and to researchers because it is so large and complex.

The foci of this third and final phase (March-April 2019) are:

- Characterizing the uses of and need for further development of the tool;
- Creating a development roadmap;
- Identifying potential options for medium- to long-term administration and sustainability.

The previous two phases, taken together, established the current state of the tool, needs of priority user groups, and gained further insight on the composition and structure of the CMS/CSS data. By the end of the second phase, it was clear that the tool is unique, needed by archivists/librarians and researchers, and that the data can support the desired inquiry. Additionally, for both archivists/librarians and researchers, the needed functionality is constrained rather than expansive.

A central question for this phase was how much we are focused on *solving a problem* (stewardship and researcher access to CSS/CMS) and how much we are focused on *developing a tool to solve a problem* (curatorial functions for CSS/CMS in repositories). The answer is *both*, with varying levels of investment in each choice. Accordingly, there is no single solution or next step; different strategies are appropriate to different elements.

Near-term needs are to communicate with stakeholder groups to share findings and to get their input on the following key decision points:

- Proposed model for partnership with ICPSR for access;
- Feasibility of Software as Service and open source development;
- Proposed model to perform more testing on SCDIF and HIS data;
- What organization (or combination of organizations) will govern the next stages of the project.

After those discussions in May and August 2019, the project will be on solid footing for its next stage of development.

For a graphical representation of the state of this project and proposed next steps, see Appendix: Project and Outcomes and Appendix: Next Steps.

Project Findings

Priority Users

During the first and second phase of this project, we characterized and determined the relative priority of and needs of users. We divided users into external (e.g. researchers) and internal (archivists/librarians and others in their institutions) and conducted a focus group and interviews with the top three: Processing Archivists/Librarians, Qualitative Researchers, and Quantitative Researchers.

For more details, please refer to the complete list and profiles of priority users in Appendix: Priority Users and User Profiles.

Functions Required by Priority Users

Processing Archivists/Librarians	**Qualitative Researchers**	**Quantitative Researchers**
Ingest data that is 1 GB and up efficiently Search and browse by date/date range Search and browse by subject Keyword search the CSS/CMS data (and, less critically, attachments) Generate reports on the status of the data Integrate with other tools for curation	Search and browse by subject Search and browse by date/date range Keyword search attachments Filter results to limit to records that have attachments	Aggregate and access large amounts of data across collections Export structured datasets in CSV or JSON for analysis in other tools Access attachments as text files for analysis Allow them to perform data cleanup

For more details on user needs and how they relate to this data, please see the Phase 2 Report.

Need for the Tool

By validating the user profiles during interviews and the focus group, we were able to flesh out pertinent details about the needs of priority users, the value of the data, and to confirm the need for the tool.

For its most basic function--to ingest and search CSS/CMS data--the tool appears to be unique, and no other tools appear to duplicate that specific functionality. Beyond that, the **desired tool functionality is very basic**. More important than expansive functionality is that it **integrates well** with both **tools used for curation** and **tools used for large-scale quantitative analysis**.

Priority researchers have different needs for the CMS/CSS data and attachments. ***Quantitative researchers*** feel they ***can do a great deal with the metadata*** exported from the CMS/CSS. ***Qualitative researchers*** imagined few to no uses for the metadata without attachments***, but relished the opportunity to efficiently search and read the attachments***.

The data and attachments are potentially valuable for both quantitative and qualitative research. The researchers we interviewed use a spectrum of tools to work with data like this. They felt strongly that this data can drive research questions for both types of research (or even both simultaneously) through different means. For all audiences, the transformative moment comes with access to the data. For more details, please refer to Appendix: Research Value of Data.

For the researchers we worked with, incomplete data was neither particularly surprising nor a significant barrier. As noted in the *Archiving Constituent Services Data of the US Congress* report, the data at hand is known to be incomplete, inconsistent, and sparsely documented, and that office practices vary widely. But ***so long as we can develop the ability to ingest the data, resolve the structure, and provide basic search, it should render this data useful for research to an unprecedented degree.***

Based on testing two additional data sets, we affirmed that the data are large and complex.

Last, we confirmed that the resources in the community and its associated stakeholders are limited for developing and sustaining this tool and for providing access to this data in general.

With a desire for relatively constrained functions and these limited resources, it is clear that **the tool does not need to solve every problem associated with curation of and access to this data.** The next stages project development consists of both tool development and using other means to address the larger problems.

Project Development Status and Needs

> *"Many potential tools and services wither, not due to shortfalls in demand or shortcomings in those products, but rather to a lack of attention to organization and community building."* (*Katherine Skinner et al, Community Cultivation Field Guide)*[1]

Development is much more than technical development. As Skinner observes, it also includes governance, resources, and community. Thus, an assessment of development status and needs must include all four areas.

It Takes a Village: Open Source Software Sustainability, A Guidebook for Programs Serving Cultural and Scientific Heritage (ITAV) offers a framework for doing just this. [2] ITAV looks at governance, technology, resources, and community engagement in a three-stage framework of development. For each of the four areas, a ten-point scale helps to assess the current phase it is in.

Current State

The ITAV framework uses a scale of 0-10 to assess what phase a project is in: Phase 1 (scores of 0-3), Phase 2 (scores of 4-6), or Phase 3 (scores of 7-10). According to this assessment, the ACC Project is solidly in Phase 1, with some elements on the cusp of Phase 2.

Area	ITAV Score
Governance: score of 2 (Phase 1)	Advisory Board with specific roles for this phase of project (but not beyond); project sponsor with outcomes to deliver; WVU as project manager
Technology: Score of 3 (Phase 1)	Beta tool that has had some degree of multi-site testing and has allowed us to increase our understanding of the data. However, data testing has been limited. There needs to be more testing prior to the next stage of development.
Resources: Score of 3 (Phase 1)	Grant funding for feasibility study with consultant, but nothing is solidly committed beyond May 2019.

[1] Skinner, Katherine, et al. *Community Cultivation: A Field Guide*. Atlanta: Educopia Institute Publications, 2018. https://educopia.org/cultivation/

[2] *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.

Community Engagement: Score of 3 (Phase 1)	Good engagement from a small group of people in workshops, user studies; congressional papers (CP) community is aware of the project and will become more so over the next few months; reaching out to broader stakeholders (researchers, organizations with related interests). Clear plan to engage CP community May-August 2019, which will give us some feedback on most likely courses of action.

Priorities for Next Steps

Continuing the use of the ITAV framework, here are the tasks needed to move most areas toward Phase 2. Governance needs to complete developmental steps of Phase 1.

Area	Steps
Governance: Continue Phase 1	• Explore one or more partnerships • Determine what needs to be governed • Consider options for governance models • Decide and form partnerships • Communicate to stakeholders
Technology: Move toward Phase 2	• Perform more data testing on CSS/CMS data • Develop a Minimum Viable Product for curatorial functions • Build transparent processes for development
Resources: Move toward Phase 2	• Identify some near-term resources to fund data testing, some development, and to consider and develop agreements with any partners • Explore one or more partnerships that can support governance, technology, and community engagement
Community Engagement: Move toward Phase 2	• Communicate with CP community • Increase engagement with associated systems community • Reach out to researcher community that was part of testing (ACSC et al) once plans are tangible enough for them to engage • Ensure that the congressional offices/Sergeant at Arms are drawn into the conversation

Near-Term Tool Development

This phase is designed to move toward a Minimum Viable Product (MVP) for curatorial functions and a constellation of potential partnerships over the course of 2019-2020. Its focus areas include:

- Data testing and an inventory of institutions that hold CSS/CMS data;
- Development of MVP for curatorial functions.

During Phase 2, two Advisory Board members set up an instance of the tool at their institution and ingested data. Doing so increased our knowledge of the contents and structure of the data, but required a great deal of their time. Clearly, asking volunteers to test data is not a viable method for getting a valid sample, but the size of and privacy concerns about the data mean that a different testing model is not a simple matter. A different model will be required but requires community discussion.

Based on our current level of knowledge, a high-level specification for an MVP is as follows:

Stage	Functions
Minimum Viable Product for Curatorial Functions	• Successfully import CSS/CMS data sets that are 1 GB or larger in either ASCII, HIS, or SCDIF • Facilitate review of data for content and structure • Generate reports for curators on the status of the data: Success of import and source of any blockers • Free-text/natural language search of CSS/CMS data and attachments

MVP development will require a much more detailed specification than this brief version.

Elements of Sustainability and Recommendations

The project is most likely to develop and sustain governance, technology, resources, and community engagement in the near and longer term using a number of elements as a combination. ***These options are not mutually exclusive.***

Open Source Development	
Overview	Open Source software is non-proprietary and is openly shared through utilities like GitHub. Developers create open source projects with the expectation that other developers who are interested will contribute to the software and share it freely. Many organizations and individuals

	have strong commitments to open source and prefer it to vendor-provided software that may require significant advocacy to effect change.
Suitability	The developers at West Virginia University created the tool as open-source and gave careful attention to its architecture so that it would be viable for other developers to work with and contribute to it. The development discussion have not yet extended beyond West Virginia University, so we don't yet know the potential for engagement with other developers.
Recommendation	**Open source is likely not the sole sustainability model, and the project should consider how viable this option is after discussions with the CP community.** An essential element of successful open source software is an engaged technical community ready to contribute to development. For this tool, the most engaged individuals are archivists, who are extremely time constrained. Although we may hear differently during conversations with archivists over the next few months, what we know so far suggests that they have little or no capacity to contribute to development and limited capacity to advocate for development resources in their organizations.[3]
Area of Support for Project Development	May contribute to a portion of MVP development.

Software as Service	
Overview	Software as Service (SaaS) is a model where software is licensed as a subscription and centrally hosted, and is commonly used for a variety of applications in use by organizations and individuals. SaaS is most likely to be successful with a substantial user community that can provide enough subscription and hosting fees to support software maintenance and development along with the required infrastructure. SaaS can host software with or without the data. A variation is a subscription for central hosting of open-source software. A number of companies perform this service for archives and libraries, including LYRASIS, Atlas Systems, and LibraryHost,

[3] This is not unique to this project and has been observed for other open source software for archives. See Schaefer, Sibyl, "Challenges in Sustainable Open Source: A Case Study." Code4lib Journal 9, March 22, 2010. Available online: https://journal.code4lib.org/articles/2493.

	most commonly for collections management software like ArchivesSpace.
Suitability	This is a specialized tool for a niche market. The tool is very specific to the structure of CSS/CMS data and thus to the repositories that currently hold, or are about to receive, that data. We know of roughly 25 institutions that hold CSS/CMS data, but believe that number has increased with the 2018 election outcomes. Communications with the CP community over the next 3-4 months will allow us to gain more current information on potential numbers of users. Because of the restrictions on and sensitivity of the data, institutions may be hesitant to consider either a full or limited SaaS model, and a viable model may require only the tool to be hosted with data remaining at the institution. There are also significant differences in local and cloud-based hosting; the latter may require fundamental architectural changes to the software, and there are also differences in storage demands.
Recommendation	**It's possible that there would be sufficient demand to merit central hosting** and that data security concerns could be mitigated. A full SaaS model seems unlikely given the relatively small number of likely customers. If discussions with the CP community show interest in either a full or limited model, discuss with one or more organizations that offer SaaS. For this community, the strongest answers may come experientially: A small handful of institutions could participate in a pilot and share their experiences. The governance function to support this would be to identify and evaluate organizations that offer SaaS, create the necessary agreements, and manage a pilot.
Area of Support for Project Development	Would play a part in MVP development. Need drives governance function. Would address constraints of librarians and archivists.

Membership Supported	
Overview	Membership support models have become increasingly common in cultural heritage over the last fifteen to twenty years. They are a

	means for a group of institutions to come together to support something that they need and are willing to support. They have also proliferated to such an extent that there is a widely acknowledged sense of "membership fatigue."
Suitability	As Roger Schonfeld recently observed on *The Scholarly Kitchen*, every good idea does not require a new organization.[4] There are a number of existing membership organizations among the stakeholders. One that has had involvement since the beginning of this project is the Association of Centers for the Study of Congress (ACSC), an alliance of organizations that promote the study of Congress headquartered at the Robert C. Byrd Center.
Recommendation	**Partner with ACSC to facilitate the next stages of tool development,** either wholly with ACSC funds or with ACSC collecting a modest contribution from a group of stakeholders. This tool is specialized and has a particular stakeholder group, and it is prudent to allow those stakeholders to invest in it. This type of relationship could facilitate not only the next stage of development, but beyond. The governance function to support this would be to establish the relationship with ACSC, define decision making processes, and create clear agreements for using or collecting fees and providing oversight of developer staff.
Area of Support for Project Development	Provides near-term support for data testing and inventory without a lengthy grant application process. Need drives governance function.

Partnership with ICPSR	
Overview	The work with processing archivists/librarians confirmed that archivists are very time constrained and lack the administrative and technical infrastructure to manage both curatorial functions and access to CMS/CSS. That suggested a need to look outside of the archives community for existing expertise and infrastructure.
Suitability	Inter-university Consortium for Political and Social Research (ICPSR) is a membership-supported consortium of 750 academic institutions and

[4] Schonfeld, Roger. "Learning Lessons from DPLA." *The Scholarly Kitchen*, November 13, 2018. Available at https://scholarlykitchen.sspnet.org/2018/11/13/learning-lessons-from-dpla/.

	research organizations that is a unit within the Institute for Social Research at the University of Michigan. They maintain and facilitate access to a data archive that has more than 250,000 files, including highly sensitive data, for social science and other research. They have in-depth expertise in creating and maintaining agreements with data providers (which include academic, government, and other institutions) and have established models for mediated access to sensitive or restricted data. A partnership with ICPSR could facilitate researcher access to the data. We would still need the tool, but it would need to be only the Minimum Viable Product for curatorial functions. Institutions that wished to facilitate access to their CSS/CMS data would make their own agreements with ICPSR to manage access. A continuation of this project could set up a framework for making agreements between institutions and ICPSR that ensure mutual long-term benefits for all parties. Preliminary conversations indicate that ICPSR is very interested in working with CSS/CMS data. The size and contents of the data are similar to what they already manage, though the relational structure of SCDIF may require some expanded capabilities on their part. As a membership-supported consortium that works with academic institutions, there is good potential for administrators to feel comfortable with the organization--or even to already have a relationship with them. This may also be a very challenging idea for the congressional papers community. Congressional collections are intertwined with donors, development relationships, and institutional identity. Near-term discussions with the community will provide important indications of the viability of this option.
Recommendation	**The project will discuss this option with the CP community in May.** Depending on that discussion, the project should compose a small group of institutions to **pursue a pilot project with ICPSR**. **This is an opportunity for archives to facilitate use** by partnering with an allied community with the expertise and infrastructure that archives lack.

	This option still requires developing and sustaining the Minimum Viable Product for curatorial functions, but does not require building out the end user access functionality. The governance function to support this is to create and sustain guidelines for institutions negotiating with ICPSR and donors to ensure mutual long-term benefits for all parties.
Area of Support for Project Development	Supports researcher access in a way that acknowledges the constraints of archivists and librarians. Helps constrain tool development needs since neither the tool nor associated infrastructure will need to provide researcher access.

Partnership with Vendor	
Overview	Vendor partnerships related to systems and data offer appealing efficiency: Rather than forming solutions in response to a vendor's methods, stakeholders can influence the activities and priorities of the vendors. They are a common model with Integrated Library System vendors, either through co-development agreements or organized user groups.
Suitability	One of the central problems driving the need for this project is a disconnect between the vendors of the CSS/CMS systems and the repositories that take long-term responsibility for the data. The congressional office are their customers, and the repositories are unlikely to become customers. The advent of the SCDIF format shows that it is possible for the congressional papers community to influence both the offices and the vendors, but it is not in the vendors' interest to form a partnership.
Recommendation	Based on past experience, **this direction is likely not worth pursuing.** It has been explored in the past at the direction of the Clerk of the House, and vendor response was not useful. The vendors created an Access database option for House exports, but it didn't scale up sufficiently.
Area of Support for Project Development	n/a

Dissemination and Discussion

A near-term step (and a bridge from the current project to the next stage) is to disseminate and discuss the project findings with stakeholders, particularly the congressional papers community to gauge initial reactions that are likely to include both excitement and concerns. The plans for doing so are as follows:

- Association of Centers for the Study of Congress (ACSC) annual meeting, May 8-10, Washington, DC
    - This group includes archivists, researchers, and others with a subject interest in Congress. Project director Danielle Emerling and Advisory Board members John Caldwell, Alison White, and Hope Bibens will present the project and its findings for discussion.
- Project findings webinar, May 14
    - We invited members of the Congressional Papers Section and those on the ACSC list to this one-hour program. Project director Danielle Emerling and consultant Jodi Allison-Bunnell will present the project and its findings for discussion. Nathan Gerth and Brandon Pieczko will discuss their experiences with data testing, and other Advisory Board members will join for Q&A. We will both facilitate discussion during and invite brief feedback after the webinar.

From these processes, we anticipate that we will gain insight on the community's reaction to the project's findings, and that we will get feedback on the following:

- What constellation of stakeholder groups will guide the next stage of development;
- Partnerships with ICPSR for providing researcher access;
- Options for next-stage data testing and inventory;
- Viability of SaaS for hosting;
- Viability of open source development.

This will allow the project director and consultant to firm up plans for near-term work as the project concludes. For a visual representation of decision points and impacts, please see Appendix: Next Steps.

In June, the project director will give a presentation to the The Advisory Committee on the Records of Congress in D.C.

The next opportunity comes three months later with the Society of American Archivists (SAA) annual meeting, July 31-August 6, Austin, TX:

- America Contacts Congress: The Project to Save Congressional Correspondence Data, August 4. Project director Danielle Emerling and Advisory Board members John Caldwell, Alison White, Nathan Gerth, and Brandon Pieczko will present the project and findings, along with any new developments, for discussion.
- Congressional Papers Section, August 2. Discussion of this project will be one part of a full-day program for the section.

SAA is after the conclusion of the present project but is a prime time to seek affirmation from and to engage stakeholders in the work ahead. It will provide an opportunity for in-depth discussion of the project's outcomes and work accomplished in June and July with archivists, including both those highly focused in this area and those who manage congressional papers as part of other responsibilities.

Appendix: Priority Users and User Profiles

During the Phase 1 workshops, we developed the following list and priority ranking of users:

Users (Internal and External)	Priority Ranking
Internal: Processing archivist/librarian	39
External: Quantitative Research (Academic)	37
External: Qualitative Research (Academic)	36
External: Students	33
Internal: IT staff	33
Internal: Reference archivist/librarian	32
External: Political	30
External: Policy	30
External: Government	28
External: Public	28
External: Professional	28
External: Journalists/Media	28
Internal: Administrator	25
Internal: Peer archivist	23

For more details on these user types, please see the Phase 1 report.

For user testing in Phase 2, We selected end users for special focus based on the priority ranking:

- Processing archivists/librarians
- Quantitative Researchers
- Qualitative Researchers
- IT Staff

Although the Phase 1 ranking for students was the same as that for IT staff, we chose them rather than students for special focus in Phase 2 in order to split our inquiry equally between internal and external users. We also felt that by including graduate students in the Researchers category, we were focused on the most likely users for this complex data. For more details on rankings of potential users, please see the Phase 1 report.

After the Phase 2 user testing, we produced the following brief user personae:

Processing Archivist/Librarian	Cultural heritage professionals who focus on management and curation of unique and special content
Scenarios: Prepare collections for researcher use by arranging and describing. Protect rights of privacy and other restrictions placed by agreement or statute. Ensure data is authentic.	
Top Functions for Tool: ● Ingest data that is 1 GB and larger efficiently ● Search and browse by date/date range ● Search and browse by subject ● Keyword search the CSS/CMS data (and, less critically, attachments) ● Generate reports on the status of the data ● Integrate with other tools for curation	Constraints: ● Time: Very Low ● Resources: Low ● Expertise: Very High

Qualitative Researcher	Faculty and graduate students in history, communication studies, geographical psychology, political science, and allied fields
Primary needs are search/access/read; manual search interface is the primary entry point; similar to working with paper materials	
Top Functions: ● Subject search/browse metadata ● Date/date range search/browse metadata ● Keyword search attachments	Constraints: ● Time: May vary, but generally motivated and tenacious ● Resources: Low to moderate ● Expertise: Very high

• Filter results to limit to records that have attachments	

Quantitative Researcher	Faculty and graduate students in history, communication studies, linguistics, geographical psychology, data science, political science, and allied fields.
Big data people who use emerging technologies and data science to do research that is not feasible from analog records. Entry point is programmatic. Aim and expertise is to decipher patterns not discernible at any other scale. Create outputs that may include text, visualizations, or cutting edge approaches to presentation.	
Top Functions: • Aggregate and access large amounts of data across collections • Export structured datasets in CSV or JSON for analysis in other tools • Access attachments as text files for analysis • Allow them to perform data cleanup	Constraints: • Time: May vary, but generally motivated and tenacious • Resources: Moderate to high • Expertise: Very high

For more details on the process used to produce these conclusions, please see the Phase 2 Report.

Appendix: Stakeholders

West Virginia University (WVU)

As the original developer of the tool and the manager of the current project, WVU has identified a strong interest in ensuring that their investment yields something of value and that they receive acknowledgement for their foundational role. WVU is very supportive of many options for project sustainability. The developers are eager to engage in more work on the tool, but have limited time to do so.

America Contacts Congress Advisory Board Members (AB)

Advisory Board members have invested considerable time and effort in this project. Two members performed data testing that was not anticipated in the original project plan that was both time-consuming and (by their reports) a valuable learning opportunity. AB members remain engaged and invested in moving the project to its next stage of development, and have considerable subject knowledge about the project. However, they have limited time to devote to continued volunteer efforts.

The Advisory Board will either need to re-develop in the next stages of the project and take on new roles, or its functions will move elsewhere. Retaining some degree of current members' involvement will be essential for continuity, but others will need to step into these leadership roles over time.

Society of American Archivists (SAA), Congressional Papers Section (CPS)

The Congressional Papers Section of SAA, one of a number of subject-focused groups that provide advocacy, communication, perform work, and develop members of the Society, has a vital interest in this project and in solving the problem of providing curation of, and access to, CSS/CMS data. The section formed the CSS/CMS Task Force in 2016, and the report they issued in 2017 forms the genesis of this project: It called for the community "to develop a technological solution for processing, preserving, and providing access to constituent data that will benefit both large and small repositories."[5]

Association of Centers for the Study of Congress (ACSC)

ACSC is a membership organization that has a strong interest in facilitating access to CSS/CMS data. Members played a role both in developing this project and contributed substantially to the Phase I workshops and to recruiting researchers for Phase 2 interviews. As one of the primary curatorial and researcher audiences for the tool and project, their support is essential. ACSC could play a role in next-phase governance and funding.

---

[5] *Archiving Constituent Services Data of the U.S. Congress: A Report of the CSS/CMS Task Force*. Society of American Archivists, November 15, 2017. Available online: https://www2.archivists.org/groups/congressional-papers-section/archiving-constituent-services-data-of-the-us-congress-a-repor-0

LYRASIS

LYRASIS is a non-profit membership organization that supports "enduring access to the world's shared academic, scientific, and cultural heritage." This project is supported by a Catalyst Fund grant that supports new ideas and innovative projects by LYRASIS members, and the project development plan has used the *It Takes a Village* framework developed by LYRASIS. The organization has several potential interests in the next stages of the project, including feasibility of a SaaS service model and funding.

ICPSR

Inter-university Consortium for Political and Social Research (ICPSR) is a membership-supported consortium of 750 academic institutions and research organizations that is a unit within the Institute for Social Research at the University of Michigan. They maintain and facilitate access to a data archive that has more than 250,000 files, including highly sensitive data, for social science and other research. They have in-depth expertise in creating and maintaining agreements with data providers and have established models for mediated access to sensitive or restricted data.

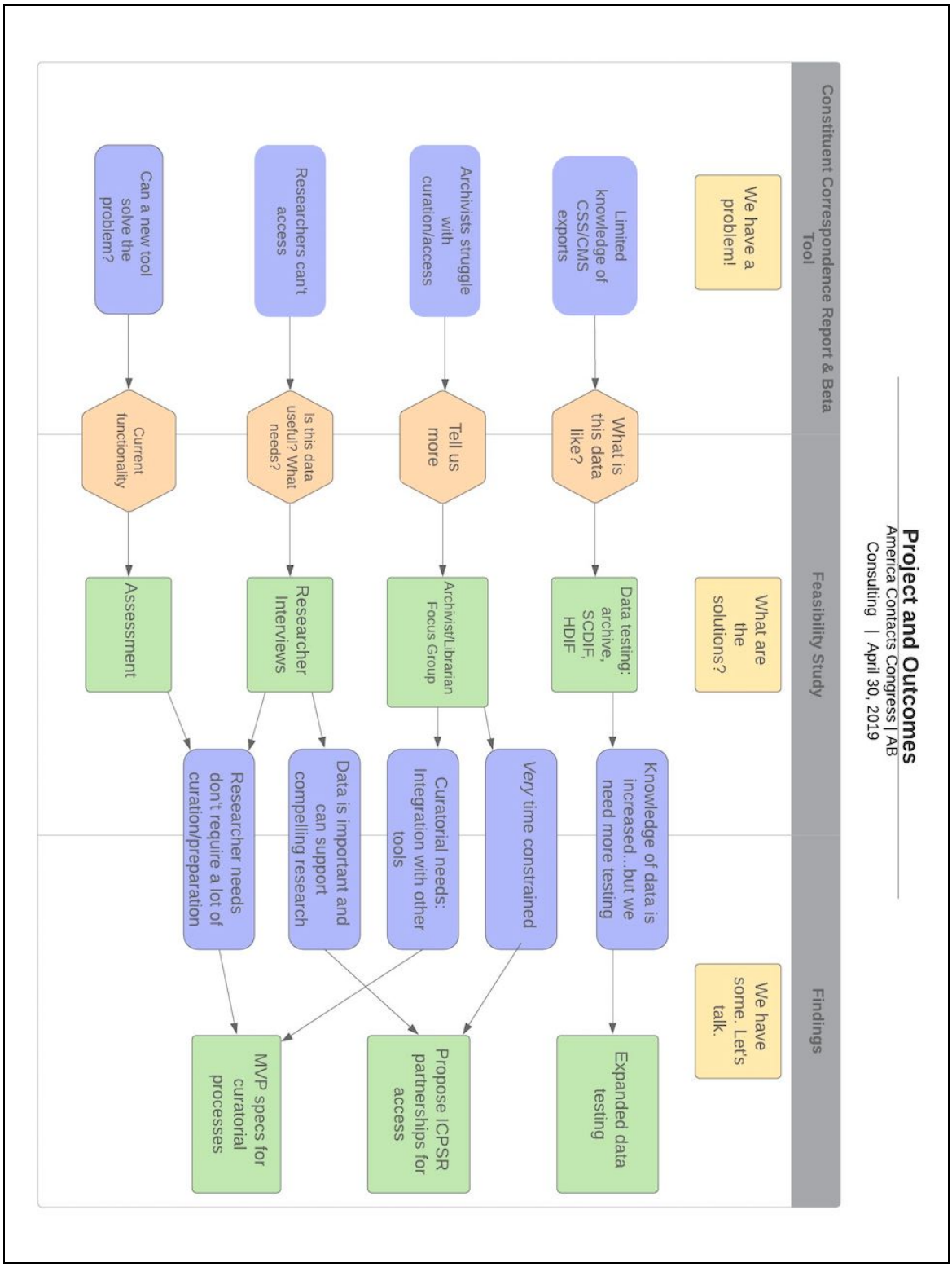
As part of this project, we had several conversations with them about their level of interest in offering long-term curation and managing end user access to this data. They are very interested and will be part of dissemination efforts in May 2019.

Appendix: Research Value of Data

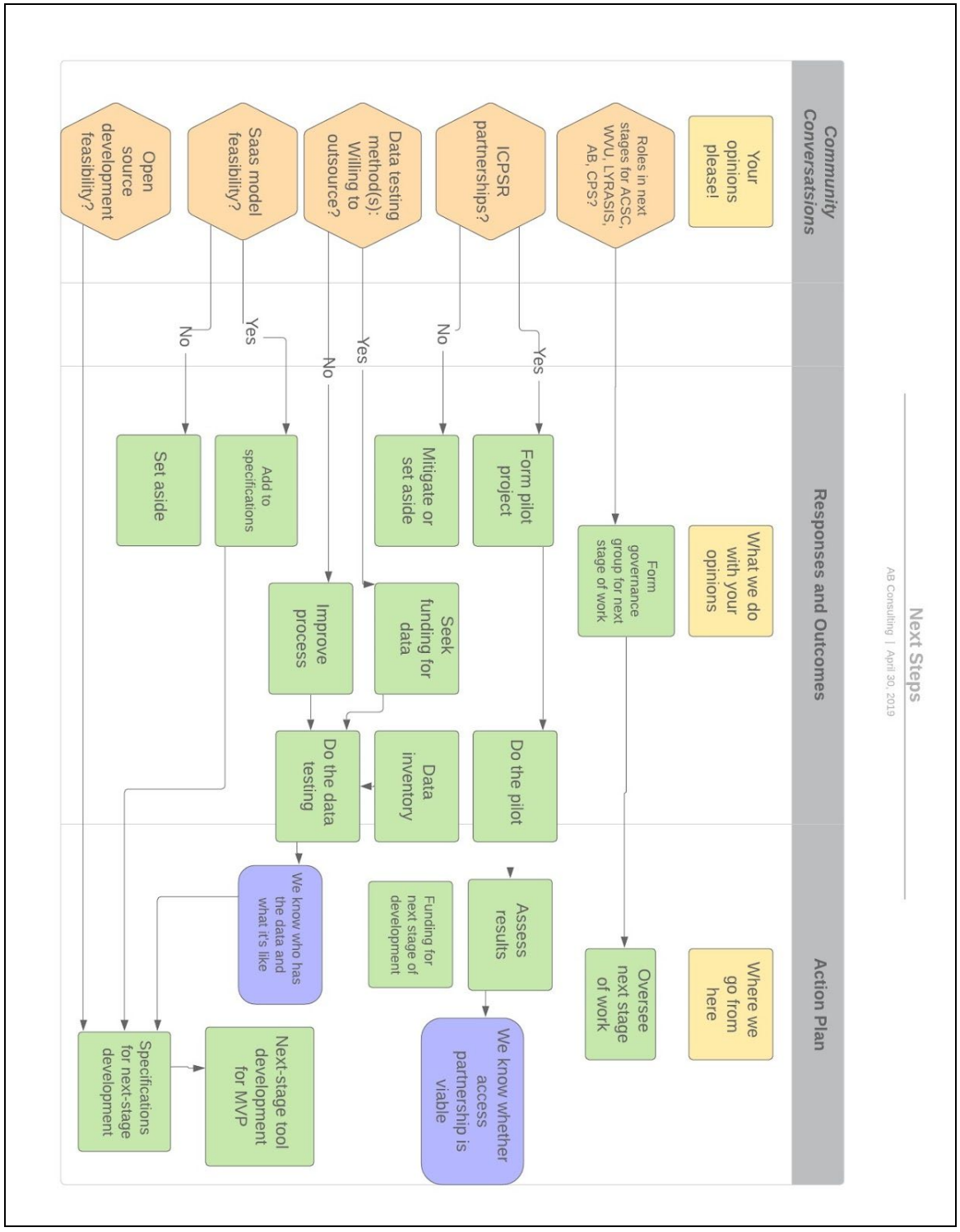
Looking broadly at the potential research value of the data compared with the stated research interests of the researchers interviewed during Phase 2, it appears that the data could support quantitative and qualitative inquiries in the following ways:

- Congressional staffing and the revolving door with lobbying firms:
    - ***Significant potential in SCDIF and HIS, but not in Archive format.*** The tool could increase ability to search efficiently on staff names and to identify the specific roles of individuals so long as the SCDIF or HIS 8A table is intact. This information is not present in the Archive format.
- Large-scale analysis of broadcast communications from congressional offices (and stated desire to have a comparative data set):
    - ***Considerable potential in both formats.*** All three can support analysis of correspondence received and sent during a specific time period and/or on a topic and to compare with broadcast communications from that same office.
- Congressional leadership:
    - ***Little Potential in any format***, as leadership decisions are not likely to be documented in constituent correspondence. If the set is SCDIF and the office used the 7 tables for schedule data, or HIS used 7A table for schedule data, the potential could increase somewhat.
- Legislator responsiveness and representation:
    - ***Considerable potential in all formats.*** This data (Archive format, SCDIF, and HIS) fills a known gap in knowledge of the relationship between constituent communications and legislative action.
- Congressional office management of constituent communications:
    - ***Considerable potential in SCDIF, some in HIS, but not in Archive format.*** SCDIF data that has the 1, 2, 3, and 8 table present (and, for completeness, the 6 table) and with some data in the 1A, 1B, 1F, 2A, 2C, 3A, and 8A fields can support significant insight into office processes. In the HIS data, table 3A identifies staff working on casework files.
- Congressional history, both political and social:
    - ***Considerable potential in all formats.*** The ability to search congressional correspondence by keyword, subject, and date rather than manually reading through files arranged chronologically is a significant advance that could vastly increase the use of these underutilized resources. However, this potential depends on the ability to identify, search, and read the attachments.

Appendix: Project and Outcomes



Appendix: Next Steps



Appendix: Key Terms[6]

Archive Format: A 32-field data export format used for transferring data from Constituent Services Systems to repositories beginning in about 1990.

CSS/CMS: Constituent Services Systems and Correspondence Management Systems. Proprietary systems used in (respectively) the U.S. Senate and U.S. House to receive, store, index, and send correspondence with constituents. Some offices also use these systems for scheduling and casework.

House Interchange Standard: A data export format for CMS data systems. It became an export choice for House offices that are preparing to send their CMS to a repository, but is not as well documented as the Senate format.

Processing Archivist/Librarians: Throughout this report, terms for cultural heritage professionals who focuses on management and curation of unique and special content.

Qualitative Researchers: Throughout this report, refers to faculty and graduate students in history, communication studies, geographical psychology, political science, and allied fields whose primary mode of interaction with sources is search/access/read.

Quantitative Researchers: Throughout this report, refers to faculty and graduate students in history, communication studies, linguistics, geographical psychology, data science, political science, and allied fields whose primary mode of interaction with sources is large-scale and computational.

Repository: Throughout this report, a cultural heritage institution that manages congressional papers after a U.S. Congress member leaves office.

SCDIF: Senate CSS Data Interchange Format. A data export format for CSS systems with over 200 fields. Originally designed for migration between CSS systems, it became an export choice for Senate offices that are preparing to send their CSS to a repository in 2016.

[6] For a more complete list of terms associated with this data and further details, please see the *Archiving Constituent Services Data of the U.S. Congress* report, https://www2.archivists.org/sites/all/files/2017_CSS_CMS_Report_0.pdf.

Appendix: Methodology Overview

The project has used the following methodologies to come to its conclusions:

- Phase I workshops with stakeholders to create preliminary notions of users and their needs:
- Phase I assessment of tool's current state;
- Phase II data testing to better understand the structure and content of the data;
- Phase II focus group and interviews with priority users;
- Throughout, Advisory Board member oversight and input.

For more details, please refer to the Methodology sections in the reports:

- Phase I report
- Phase II Report