# America Contacts Congress Project
# Report 1

For the University of West Virginia
AB Consulting
2018 December 19

# Executive Summary

The *America Contacts Congress* project is a feasibility study for a tool to manage electronic correspondence and case files from offices of the U.S. Congress funded by the LYRASIS Catalyst Fund. The focus of Phase I (October-December 2018) was to establish needed functions and possible users for the congressional correspondence tool (hereafter "the tool") and to complete the detailed plan for the remainder of the project. During this phase, we completed workshop-style virtual sessions on functions and users with Advisory Board members and other experts; documented the tool's current functionality; and established plans for user and data testing in Phase II.

With strong engagement from identified experts, a solid plan for the remainder of the feasibility study, and key insights on the very challenging data at hand, the project has a strong start. With the results of the 2018 elections bringing many changes to Congress, the project is both timely and of interest to many institutions. A recent session at the Council for Networked Information (CNI) meeting presented by an Advisory Board member and a West Virginia University staff member was very well attended by deans and directors of major institutions.

We expect to finish on schedule and with useful, credible information in hand that will guide informed decision making.

# Methodology

This section describes the steps used to come to the recommendations in this report.

## Meetings with Project Advisory Board

AB Consulting met with the Advisory Board on October 16, December 4, and December 18 to discuss the scope of the project, clarify the Advisory Board's roles, and to work on key aspects of the project plan.

Advisory Board members:

- Danielle Emerling, Assistant Curator, Congressional and Political Papers Archivist, West Virginia University, Chair and Project Manager
- Brandon Pieczko, Processing and Digital Archivist, Russell Library, University of Georgia
- Alison White, Deputy Senate Archivist for Digital Archives, Senate Historical Office
- Hope Bibens, Political Papers Archivist, Drake University
- Nathan Gerth, Digital Archivist, University of Nevada Reno
- John Caldwell, Political Papers Archivist, University of Delaware
- Natalie Bond, Political Papers Archivist, University of Montana

## Site Visit to West Virginia University

AB Consulting made a site visit to Morgantown, WV, on November 7-8 to meet with key library staff, including Danielle Emerling; Jessica Tapia (Head, Web and Digital Services); David Davis (Software Engineer/Full Stack Developer) and Tracy McCormick (Junior Software Engineer); Paula Martinelli (Director of Development); WVU graduate students in history and political science; John Cuthbert (Director, West Virginia & Regional History Center); Karen Diaz (Dean, WVU Libraries); Bill Rafter (Head, Systems Infrastructure Group), Rodney Adlington (Professional Technologist 4, Systems Infrastructure Group), and Brian Sinsel (Professional Technologist 4, Systems Infrastructure Group).

The purpose of the trip was to meet key players in the tool's development, to document current tool functionality, to better understand the data the project is working with, and to discern how a tool like this one fits into IT in academic libraries. In addition, a number of useful conclusions resulted from the site visit, including:

- The project must be expansive when envisioning potential audiences in order to uncover all use cases. Our most fruitful subjects may come from outside traditional academics.
- We must keep it simple. It's possible that the tool is a bridge to another tool rather than an all-in-one system.
- The data is inconsistent and uneven. We need to know more about it, and should do data testing as soon as we can.
- Preservation/authenticity of data is key and must be documented at every step.
- We are working with sensitive data, and network security issues will be a huge issue.
- We believe that many academic libraries are moving away from locally developed bespoke tools to community developed tools. Our Phase II work should test these assumptions with Systems/IT personnel.

## Workshops

On November 13 and 14, AB Consulting led workshops with Advisory Board members and others with key knowledge of and familiarity with congressional correspondence:

- JA Pryse, Senior Archivist, Carl Albert Center, University of Oklahoma
- Rachel Hancock, Mikulski Papers Project Archivist, Eisenhower Library, Johns Hopkins University
- Elizabeth England, Digital Archivist, Eisenhower Library, Johns Hopkins University
- Mark Evans, Vice President, Archives and Collections Management, History Associates
- Rebecca Johnson Melvin, Coordinator of Special Projects, Manuscripts Librarian and Curator of the Joseph R. Biden, Jr. Senatorial Papers, University of Delaware
- Mike Crespin, Director and Curator of the Carl Albert Center, University of Oklahoma
- Patrice-Andre Prud'homme, Director, Digital Curation, Oklahoma State University Library
- Renee James, Curator for Distinct Collections, Arizona State University
- Jody Brumage, Archivist and Office Manager, Robert C. Byrd Center, Shepherd University

Prior to the workshops, attendees did pre-work on the tool functions and users that they anticipated. For functions, individuals gave a function a characteristic name, a short description, and added in any comments. For users, individuals named a potential user or audience member, described their needs, and added in any comments.

In preparation for the workshops, AB Consulting prepared two virtual whiteboards using RealTime Board and transferred the pre-work to virtual sticky notes.

The November 13 workshop focused on users. The group was able to better describe and consolidate the data from the pre-work to produce a very solid set of potential users that were both internal (library/archives professionals) and external (researcher end users).

The November 14 group focused on functions. After a review of the work of the November 13 group, they took the functions identified in the pre-work and matched them to the most likely users. As part of both of these workshops, the groups identified some user interface elements that should be present in any tool, including adherence with commonly accepted usability standards.

For more details on this process and the original data, please see Appendix A.

## Workshop Follow-Up Survey

After the workshops, AB Consulting took the users described and did some data cleanup to make descriptions of similar functions (or duplicates) uniform across users. They asked attendees to rank the relative importance of the functions associated with each one in a survey. For each user, respondents were asked to characterize functions:

- That the tool MUST have (If the tool doesn't have these, it is not successful)
- That it SHOULD have (The tool will be better if it has these)
- That it MAY have (Nice to have, but not essential)

They could also indicate that the function is not relevant to the user, or that they didn't know enough about this type of user to give a knowledgeable answer.

Respondents were also asked to rank the relative importance of different users, external and internal, as:

- Most important
- Moderately important
- Less important

Users had function lists of varying length, ranging from twelve to thirty-two items.

The survey was open for responses November 19-30, 2018. Respondents were asked to answer questions based on their expertise and experience with congressional correspondence management and research, giving their own opinions as an identified expert rather than those of their institution or the broader congressional papers/archives community.

Eleven of the fifteen workshop participants completed the survey.

## Survey Data Processing

Between December 1 and 16, AB Consulting worked with the workshop data to create this report and recommendations.

**Users and Functions**

For each user, AB Consulting converted each function ranking into a numeric value:

- Tool MUST have: 4
- Tool SHOULD have: 3
- That it MAY have: 2
- I don't know this user/not relevant to user: 0

Adding these numbers for each function created an overall score for each that was prioritized within each user.

AB Consulting calculated the functions most often ranked as MUST or SHOULD, weighting the score associated with each, in order to generate the list of functions common to most or all users that, depending on the outcomes of Phase II, could form part of a development roadmap.

**Prioritizing Users**

For each user, AB consulting converted the relative importance of each user into a numeric value:

- Most important: 3
- Moderately important: 2
- Less important: 1

Adding these numbers for each user generated an overall score for each. Using these scores, we calculated the relative importance of users within the internal/external categories and across them. These outcomes form the basis for the user testing recommendations.

For more details on this process and the original data, please see Appendix A.

# Preliminary User Personae

| **Administrator** | • Library/archives mid-level manager<br>• Dean/director of library or other cultural heritage institution |
|---|---|
| Individuals responsible for resource allocations and strategic decisions. Concerns include efficient operations, security, donor relations, and compliance with IT policies. Must prove to self and others that a tool is used and needed, and why a specialized tool is required. | |
| Top Functions:<br>• Passes security requirements<br>• Simple deployment<br>• Manage access<br>• Ease of maintenance<br>• Reporting | Constraints:<br>• Time: Low (competing priorities)<br>• Resources: Low (in most cases)<br>• Expertise: Low (unlikely that tool is in area of primary expertise) |

| **Government** | • Federal, state, or local government staff<br>• Agency staff<br>• Congressional staff |
|---|---|
| Scenarios: Researching issues that require agency response. Requires access to data of former members of Congress to inform current work. | |
| Top Functions:<br>• Search by congressional district<br>• Search by date/date range<br>• Subject search<br>• Browse by date/date range<br>• Search by address/zip code | Constraints:<br>• Time: Low<br>• Resources: Low<br>• Expertise: May vary depending on experience |

| **IT Staff** | • IT/Systems staff from libraries/archives<br>• Archivists/librarians who do their own IT/ Systems support |
|---|---|
| Scenarios: Concerns include efficient operations, concerns about security and compliance with IT policies. May have expansive capacity and a desire to customize or very limited capacity to support specialized tool(s). May have capacity to contribute to development. Must prove to self and others that a tool is used and needed, and why a specialized tool is required. | |

| | |
|---|---|
| Top Functions:<br>• Manage access<br>• Passes security requirements<br>• Ease of maintenance<br>• Compatibility<br>• Document any changes to data | Constraints:<br>• Time: Low<br>• Resources: Low in most cases; may be high<br>• Expertise: Very high in IT/systems; moderate to high in structured data; low in congressional data specifically |

| **Journalists/Media** | • Local, state, national, international journalists<br>• Citizen watchdogs |
|---|---|
| Scenarios: Checking information veracity from congressional office, responsiveness of member, pulse of community. May be creating daily news or, in the case of watchdogs and in-depth journalists, may focus on a single story for long periods of time and be very tenacious and willing to invest time to get information desired. | |
| Top Functions:<br>• Have confidence that the data is legitimate<br>• Browse by date/date range<br>• Search by address/zip code<br>• Search by congressional district<br>• Fielded search | Constraints:<br>• Time: May be high or very low<br>• Resources: Low in most cases; may vary<br>• Expertise: Moderate to very high |

| **Peer Archivist** | Broader professional community of archivists and related professionals |
|---|---|
| Scenarios: Interest in broader ramifications of ACC project and re-use of tool for other purposes. May derive inspiration from project. Use for mediated reference and to surface connections between collections. | |
| Top Functions:<br>• Search by congressional district<br>• Sort data<br>• Subject browse<br>• Filter results<br>• Search by address/zip code | Constraints:<br>• Time: Low<br>• Resources: Low<br>• Expertise: Very High |

| **Policy** | • Policy groups and organizations<br>• Think tanks<br>• Public policy researchers<br>• Nonprofits focused on government |
|---|---|
| Scenarios: Search and analyze data aggregated from multiple members to track issues; use data to inform policy; looking on information veracity from congressional offices | |
| Ranked Functions:<br>• Search by congressional district<br>• Search by date/date range<br>• Browse by date/date range<br>• Fielded search<br>• Subject search | Constraints:<br>• Time: May be high (tenacious, do what it takes to get information desired)<br>• Resources: May be high (e.g. well funded think tank) or low (nonprofits)<br>• Expertise: Very High |

| **Political** | • Lobbyists<br>• Pollsters/Pundits<br>• Campaign Staff |
|---|---|
| Scenarios: Support polls, analyze events, predict future ones | |
| Top Functions:<br>• Search by date/date range<br>• Sort data<br>• Subject search<br>• Browse by date/date range<br>• Search by congressional district | Constraints:<br>• Time: Low (fast paced work)<br>• Resources: Moderate<br>• Expertise: High |

| **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 that data is authentic. | |
| Top Functions:<br>• Ingest data<br>• Identify PII<br>• Search by date/date range<br>• Browse by date/date range<br>• Document any changes to data | Constraints:<br>• Time: Low<br>• Resources: Low<br>• Expertise: Very High |

| **Professional** | Non-academic research: business, commercial, legal |
|---|---|
| Scenarios: Doing quantitative or qualitative research to support business, commercial, or legal needs | |
| Top Functions:<br>• Search by date/date range<br>• Search by congressional district<br>• Subject search<br>• Browse by date/date range<br>• Generate reports | Constraints:<br>• Time: Low; deadline driven<br>• Resources: May vary<br>• Expertise: High |

| **Public** | • Citizens<br>• Constituents<br>• Avocational historians |
|---|---|
| Scenarios: Conducting family history research, searching for known person by name. Particularly interested in case files. May be very tenacious. | |
| Top Functions:<br>• Search by date/date range<br>• Search by address/zip code<br>• Search by congressional district<br>• Search at a granular level for person/topic of interest.<br>• Subject search | Constraints:<br>• Time: High<br>• Resources: Moderate<br>• Expertise: May vary |

| **Qualitative Researcher** | Faculty and graduate students in history, communication studies, geographical psychology, political science, and allied fields |
|---|---|
| Scenarios: Primary needs are search/access/read; manual search interface is the primary entry point; similar to working with paper materials | |
| Top Functions:<br>• Subject search<br>• Search by date/date range<br>• Subject browse<br>• Browse by date/date range<br>• Filter results | Constraints:<br>• Time: May vary, but generally motivated and tenacious<br>• Resources: Low to moderate<br>• Expertise: Very high |

| **Quantitative Researcher** | Faculty and graduate students in history, communications studies, linguistics, geographical psychology, data science, political science, and allied fields. |
|---|---|
| Scenarios: 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:<br>• Aggregate and access large amounts of data;<br>• Export structured datasets based on search/ browsing results<br>• Search and analyze CSS/CMS data<br>• Combine natural language and programmatic queries to explore the data<br>• Sort data | Constraints:<br>• Time: May vary, but generally motivated and tenacious<br>• Resources: Low to moderate<br>• Expertise: Very high |

| **Reference Archivist/Librarian** | Cultural heritage professional who focuses on reference and user services |
|---|---|
| Scenarios: Answer questions, prepare instructional materials in varied formats, work with researchers to seek answers and to understand resources that are available to support their research. | |
| Top Functions:<br>• Report usage<br>• Share outputs with other researchers<br>• Generate reports<br>• Search and analyze text of attachments to CSS/CMS data<br>• Search and analyze CSS/CMS data | Constraints:<br>• Time: Low<br>• Resources: Low<br>• Expertise: Very high |

| **Students** | • College/university undergraduates<br>• High school students<br>• K-9 students |
|---|---|
| Scenarios: Support teaching with primary sources, National History Day projects. Few or no programmatic approaches, main mode is search/access/read. May require mediated search or introductory/instructional materials. | |

| | |
|---|---|
| Ranked Functions:<br>• Search by date/date range<br>• Subject browse<br>• Browse by date/date range<br>• Search by congressional district<br>• Filter results | Constraints:<br>• Time: Low<br>• Resources: Low<br>• Expertise: Low |

# End Users Priority Ranking

Based on the responses to the follow-up survey to the workshops, overall ranking of end users (regardless of whether they are internal or external) is as follows:

| Users (Internal and External) | Overall 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 |

While it is not surprising to rank the internal processing archivist/librarian at the tool's development stage, any longer term development must focus as much or even more on the external user groups. This may help avoid the costly mistake of designing a tool for internal users only, which seldom results in either usability or success.

Within the internal and external categories, rankings are:

| Users (External) | Ranking |
|---|---|
| Quantitative Research (Academic) | 37 |
| Qualitative Research (Academic) | 36 |
| Students | 33 |

| | |
|---|---|
| Policy | 30 |
| Political | 30 |
| Government | 28 |
| Journalists/Media | 28 |
| Professional | 28 |
| Public | 28 |

| **Users (Internal)** | **Ranking** |
|---|---|
| Processing archivist/librarian | 39 |
| IT staff | 33 |
| Reference archivist/librarian | 32 |
| Administrator | 25 |
| Peer archivist | 23 |

# Tool Functions

## Current Functionality

During the site visit to West Virginia University, we reviewed the current functionality of the tool. The tool currently contains two data sets: the John D. (Jay) Rockefeller Papers, and the Nick Rahall Papers. The Rockefeller set is ASCII in the 32-field Archive Format; Rahall is tab delimited and largely in the House Interchange Standard (HIS).

Based on this small data sample in a demonstration, existing functionality is:
- Import large ASCII files
- Browse records
- Efficiently search by keyword

The development team has taken considerable pains to develop in an open-source framework and to create a tool that is simple to install and use in a variety of environments. Additionally, the search function is fast given the amount of data it searches, and they have made provisions for both the CSS/CMS data and the attached correspondence.

Some other functions—geographic analysis particularly—are partly in place, but not wholly as the data is too inconsistent to support structured search in any useful way. Rockefeller is relatively simple and consistent, but Rahall is much less so. Given the findings reported in *Archiving Constituent Services Data of the US Congress*, this is unsurprising. In order to develop other functions—particularly for efficient normalization—the developers need to know more about the target data, and to derive that understanding from a representative sample. The project must test a broader range of sets in parallel with user testing in Phase II (See User and Data Testing for details).

## Functions Common to All Users

During the workshops, participants agreed that the following qualities are required for any tool, and form a set of assumptions:
- Meets generally accepted usability, navigability, and accessibility standards
- Ability to perform basic searches with some level of intuitive or user-friendly query support
- Unmediated search: an access point for research (e.g. not just a curatorial tool)
- Open source tool built on a framework designed to be flexible and extendable
- Browser based.

Curatorial functions most often rated as a "must have" or "should have" for processing archivists and IT staff:
- Document any changes to data
- Manage access
- Ingest massive amounts of ASCII
- Identify Personally Identifiable Information (PII)
- Normalize CSS databases
- Data review

Additional baseline functions, based on the functions most often rated as a "must have" or "should have" across ten or more user types, are:
- Auto-suggest in search;
- Boolean search;
- Browse by date/date range;

- Faceted search;
- Fielded search;
- File linking;
- Filter results;
- Free-text/natural language search;
- Generate reports;
- Search and analyze CSS/CMS data
- Search and analyze text of attachments to CSS/CMS data;
- Search by address/zip code;
- Search by congressional district;
- Search by date/date range;
- Sort data;
- Subject browse;
- Tagging/flagging

For more information on functions prioritized for specific users, please see Appendix D.

Depending on additional information gathered in Phase II, building stages for the tool could look like this:

| **Stage** | **Functions** |
|---|---|
| Minimum Viable Product | • Successfully import CSS/CMS data sets [up to XXX size—to be determined in Phase II] or larger in either ASCII, HIS, or SCDIF<br>• Data review/normalization/redaction<br>• Document changes to data |
| Basic Search Functions | • Free-text/natural language search;<br>• Search and analyze CSS/CMS data<br>• Search and analyze text of attachments to CSS/CMS data |

A development roadmap will be a central feature of the Phase III report.

# Function Decision Points for Phase II

In order to fully evaluate the potential of the tool to be relevant, successful, and sustainable, the project must address the following major decision points:

## What are the capacities and limitations of the data?

As noted in the *Archiving Constituent Services Data of the US Congress* report, the data at hand is known to be inconsistent in its completeness, structure, content, and documentation. We may conclude from the data review that the types of research that the data can support are more limited than what the priority users need.

Phase II will include reviewing a sample of as many data sets as possible to assess, at a basic level, the prospect that the data can support the desired research uses. Reviewing data sets that are known to be SCDIF will be of critical importance since present and future exports will be in that form.

## Is the future tool functionality basic or expansive?

Do the priority external users want this tool to do all of the common functions listed in the previous section (expansive)? Or, would they prefer that the tool primarily serve as a bridge to other tools that they already use for search, analysis, reporting, and visualizations?

To determine answers to these questions (which may include some combination not yet contemplated), Phase II user testing must focus on the tools that the researchers already use. This is particularly critical for those using sophisticated tools for quantitative research. Additionally, we need to know what forms(s) the data exports from the tool need to take to move as seamlessly as possible to other tools.

## What data is most critical to the priority users?

Is it the metadata (e.g. the 32- or 200+-field information created in the CSS/CMS), the data (e.g. the attached correspondence or case files?)? The Advisory Board feels that the metadata is the priority for the user audiences.

Testing that assertion with users will provide the insights that we need to put resources toward the data that is most important to priority users.

## What is the relative importance of quantitative versus qualitative research?

Is the most vital and passionately interested user audience primarily conducting quantitative or qualitative research? Both? The answer to this question relates closely to the three decision points listed above and is a central feature in a future development roadmap.

Both quantitative and qualitative users are among the highest priority users and will be part of user testing in Phase II.

# User and Data Testing

## User Testing

The work of the Advisory Board and AB Consulting in Phase I generated the most needed functions for the tool for a set of preliminary user personae and a sense of the highest priority users. Phase II gives us the opportunity to ask the highest priority internal and external users about their needs and desires for tool functions.

Based on the ranking of the relative importance of identified user groups, and given the need to conduct data testing in parallel, user testing will focus on the following groups:

- Processing archivists/librarians
- Qualitative researchers
- Quantitative researchers
- Systems/IT administrators

The testing will be primarily through focus groups/interviews conducted remotely by AB Consulting, as follows:

- Processing archivists/librarians: Focus group 1
- Systems/IT administrators: Focus group 2
- Qualitative researchers: Individual interviews
- Quantitative researchers: Individual interviews

All focus group or interview participants will receive a $25 gift card of their choice from a constrained list of options.

Focus groups are the choice for users who are less difficult to schedule and who are most likely to engage in conversation with other participants in the group.

**Focus Group 1**, with processing archivists/librarians, will include individuals with curatorial responsibilities for CSS/CMS data from members of Congress. A group will be composed of 6 to 10 subjects, AB Consulting as moderator, and members of the Advisory Board as additional interlocutors. Recruiting will be drawn from members of the Society of American Archivists Congressional Papers Section who were not among those who participated in the November workshops.

The process for Focus Group 1:

- AB Consulting prepares focus group materials
- Advisory Board reviews questions and requests any revisions
- Invitation and screening survey out to Congressional Papers Section members that includes the date and time of the focus group
- AB Consulting conducts short screening interviews with respondents to ensure that their interests and expertise are consistent with the needs of the project. If so, they receive an invitation to the focus group the week of February 4 from AB Consulting
- The focus group is conducted via Zoom
- AB Consulting and attending Advisory Board members will summarize the discussion and findings. AB Consulting will integrate them into the Phase II report
- AB Consulting sends thanks and any follow-up with participants

**Focus Group 2**, with Systems/IT administrators, will include the respondents to the Systems/IT administrators brief survey conducted in December. It will be primarily an opportunity to follow up

on their responses and to get a sense of how the institutions might (or might not) be able to contribute to a longer-term development plan.

The process for Focus Group 2:
- AB Consulting prepares focus group materials based on responses to the Systems/IT administrators survey
- Advisory Board reviews questions and requests any revisions
- AB Consulting sends invitations to the focus group, the week of February 4
- The focus group is conducted via Zoom
- AB Consulting and attending Advisory Board members will summarize the discussion and findings. AB Consulting will integrate them into the Phase II report
- AB Consulting sends thanks and any follow-up with participants

Interviews are the choice for the users who are time constrained and difficult to schedule: the academic faculty and graduate students who make up the Qualitative and Quantitative researchers groups. Effective recruiting will be essential for the interviews to yield useful results: subjects must be already engaged in and passionate about congressional correspondence as a central part of their research. Advisory Board members and other individuals with significant expertise in this area will need to reach out to subjects whose feedback merits the use of their time and ours.

All subjects, whether qualitative or quantitative researchers, should have the following characteristics:
- Already doing work that includes congressional communications, either analog or digital, or public opinion analysis
- Willing to describe their existing and future work
- Available in February 2019

The process for the interviews:
- AB Consulting prepares focus group materials
- Advisory Board reviews questions and requests any revisions
- Invitation and screening survey out to individuals and any other venues (e.g. listservs, social media, blog posts) recommended by Advisory Board members and other experts, including the Association of Centers for the Study of Congress
- AB Consulting conducts short screening interviews with respondents to ensure that their interests and expertise are consistent with the needs of the project. If so, they receive an invitation from AB Consulting to choose an interview time the weeks of February 11 or 18
- AB Consulting will conduct the interviews via Zoom, ideally with at least one Advisory Board member who can probe answers based on their subject knowledge
- AB Consulting and attending Advisory Board members will summarize the discussion and findings. AB Consulting will integrate them into the Phase II report
- AB Consulting sends thanks and any follow-up with participants

All interviews and focus groups must be complete by Friday, February 22, so that data analysis and report writing can be completed by March 22. **This timeline is not flexible.**

## Data Testing

The tool's basic functionality is currently based on just two data sets, one of which has significant inconsistencies. Clearly, testing a broader set of data is essential in order to understand the target data and what is—or is not—possible to do with it. This is a shift from the original project plan,

which called for a primary focus on user testing in Phase II, to an approximately equal focus of user testing and data testing.

Data testing is under way with members of the Advisory Board through February 28. It is deliberately narrow in scope to ensure that the process yields the information that the project needs and does not inadvertently result in either a drain on members' time or testing an emergent tool not ready for that type of evaluation.

Advisory Board members with access to this data are testing one or more sets in order to yield the following:

- Vendor/system used in original office (if known)
- Name of set
- Approximate size of set (in MB, GB, or TB)
- Approximate dates associated with set
- Format of data set (flat file, CMS output, Archive Format, SCDIF)
- Whether the respondent was able to ingest the data set with the tool
- What the column headers were (e.g. name, date, subject)

This testing data will allow us to characterize:

- The size of data set that the tool will need to be able to ingest and search;
- What formats existing data sets are in;
- How the data is structured. This is a particularly critical point; evidence to date suggests that the data structures in this material are very inconsistent. In order to meet potential user needs (e.g structured search, large-scale data analysis, efficient normalization/de-duplication/redaction), we must be able to discern data patterns.

Together with the user testing—which will validate perceived user needs—data testing will allow the project more accurately characterize the relationship between user needs and the data's capacity to meet those needs.

# Project Plan and Timeline

## Phase I: Assess Existing Functionality, 2018 October-December

With the acceptance of this report, which characterizes the existing functionality of the tool and articulates the detailed project plan, this phase is complete.

## Phase II: Develop User Requirements and Characterize Feasibility, 2018 December-2019 March

This phase includes parallel work on data testing and testing preliminary ideas about users through focus groups and interviews.

**Data testing: 2018 December through 2019 February**

- The data testing plan is described in the previous section
- For the data testing instructions and reporting form, see Appendix B
- Testers will report results no later than February 28, 2019

**Focus groups and interviews: 2019 January-February**

- User testing is described in the previous section.
- Results will be summarized no later than February 28, 2019

**Second Report**

The report will be reviewed and approved by the project director and Advisory Board by 2019 March 22.

## Phase III: Justification and Roadmap, 2019 February-April

This phase takes the work of the previous two phases to characterize the uses of and need for further development of the tool, creates a development roadmap, and identifies potential options for medium- to long-term administration and sustainability. The Advisory Board will continue to meet regularly to advise on how we characterize the tool and strategize development.

**Development Roadmap: 2019 February-April**

Work begins with Focus Group 2 (IT/Systems) in early February and identifies potential options.

**Administration Roadmap: 2019 February-April**

This begins with the uses of and need for the tool and its further development, then identifies how it could be administered and sustained.

**Third Report**

The report will be reviewed and approved by the project director and Advisory Board by April 30, 2019

## Final Presentation/Discussion, 2019 May 1-15

In this final phase, the consultant will present the Phase III report to the advisory board and other audiences (including, but not limited to, SAA's Congressional Papers Section, and contacts from the Association of Centers for the Study of Congress). The final presentation will be scheduled between May 1 and 15 and delivered via Zoom.

For more details, please see Appendix E for a Gantt chart of the Phase II-final project plan

# Appendix A: Advisory Board Workshops and Data Analysis

Workshop Materials

- Instructions for Participants
- Introduction/Icebreaker
- Functions Worksheet
- Users Worksheet
- Workshop Script
- Download of workshop board: November 13
- Download of workshop board: November 14
- Merged users and functions from workshops

Workshop Follow-Up

- Framing message for survey
- Survey Text
- Raw Data
- Data Analysis

# Appendix B: Data Testing Instructions and Expected Data

Instructions
Survey

# Appendix C: List of Expected User Testing Materials

Archivists/Librarians Focus Group

- Invitation for AB use
- Screener
- Invite/not communications script
- Focus Group Questions
- Follow-up communications script

IT/Systems Focus Group

- Invitation
- Focus Group script and questions
- Follow-up communications script

Quantitative/Qualitative Researchers

- Invitation for AB and ACSC use
- Screener
- Follow-up communications
- Interview script and questions
- Follow-up communications script

# Appendix D: Functions By User

https://docs.google.com/spreadsheets/d/1kX4qTUF7Xfa2djq0Ra6TJGqYiq_H-VOzKSDPNdL-NoK4/edit?usp=sharing

# Appendix E: Project Plan Gantt Chart