Interim Performance Report Narrative
Identifying Number: LG-71-15-0174; Recipient: Internet Archive
Grant Period: 01/01/2016 to 12/31/2017; Reporting Period: 01/01/2016 to 12/31/2016

Project Title: Systems Interoperability and Collaborative Development for Web Archiving

Activities Completed During This Reporting Period
During Year One of the two-year project, Systems Interoperability and Collaborative Development for Web Archiving, all activities to be completed were successfully accomplished. At a high level, Year One work had the following outcomes:

- 12 conference/community presentations, forums, or discussion groups delivered
- 4 project online and communication channels created
- 3 community surveys conducted
- 2 draft papers circulated to the community for input and 4 blog posts written
- 2 Technical Working Group meetings convened
- 2 API specifications and data models drafted
- 2 APIs engineered and implemented in production environments
- 1 Affiliate Working Group formed (Tools Portfolio Group in IIPC)
- At least 3 additional institutions agreed to participate in testing project APIs in Year 2 and at least 2 affiliate APIs will be documented under the project's community umbrella

The project was given the working name the "WASAPI" project (Web Archiving Systems APIs) to help identify the project outcomes, limit verbosity, and to begin developing the social architecture and communication tools for building a larger community to support the grant-specific and post-grant API work that this project aims to catalyze. Overall, the first year of WASAPI focused on three areas of work within web archiving: community building and involvement, research and publication on collaborative technology development, and engineering two "production-ready" and partner-tested data transfer APIs as applied research. All Year One deliverables were met and progress was made in all areas of work -- in fact, the success of Year One work has allowed the team to include additional participating institutions in Year Two for testing project APIs and for additional API-based work to be organized within the nascent WASAPI community, both hoped-for outcomes in the original proposal. Project activities are organized around the grant's three main areas of research and development work.

1) What are the attributes of a community model that can support sustainable and broad based collaborative web archiving technology development?

During Year One of the project, significant accomplishments were made by staff in promoting the project's work, soliciting the involvement of the broader web archiving and digital library

community, and fostering its own community through feedback-driven forums intended to allow for an interactive, user-driven approach to the project's R&D work. Project staff gave 12 presentations at a variety of conferences, including:

- IIPC Web Archiving Conference (March 2016) - three sessions (see below)
- LDCX (March 2016)
- Archive-It Mid Atlantic Users Group Spring 2016 Meeting (March 2016)
- CNI Spring 2016 Meeting (April 2016)
- Texas Conference on Digital Libraries (May 2016)
- Archives Unleashed 2.0 (June 2016)
- Archive-It Annual Partner Meeting (August 2016)
- SAA Web Archiving Round Table Meeting (August 2016)
- IIPC Crawler Hackathon (September 2016)
- Dodging the Memory Hole (October 2016)

The participation of project staff at a number of these events went well beyond simply presenting about the project. For instance, at the IIPC Web Archiving Conference, the session included both a presentation by grant principals, but also an open session for discussion of project APIs and the potential for improved interoperability across the web archiving ecosystem. Prior to the session, grant staff distributed a draft paper on API-based interoperability (slated for publication in Year Two of the grant) to spur and frame the conversation. This led to an additional breakout group during the conference focused on APIs and provided the community feedback and input necessary to the grant team's work. In addition, a subsequent meeting was held with U.S. institutions attending IIPC that were interested in being involved in helping test the forthcoming project APIs. Similar feedback and input-oriented sessions were held at LDCX, the Archive-It Partner Meeting, and at the SAA Web Archiving Round Table. At each event, attendees were asked to help contribute use cases and user stories and to give feedback on development of the project APIs and other aspects of the WASAPI project's work, such as estimating local capacity for collaborative technology development and how APIs can help connect systems and serve researchers. Likewise, discussion of the WASAPI API draft specifications at the IIPC Crawler Hackathon allowed for the input of the international web archiving engineering community and helped the grant team focus project work and plan for future interoperability efforts.

For dissemination and reporting, the WASAPI team created a number of channels and portals for documenting the project's work and for encouraging the involvement of the larger community. The project's main working space, including links to publications, but especially code and technical documentation, is in Github, https://github.com/WASAPI-Community/data-transfer-apis. Project communication is done via a Slack team (https://wasapi.slack.com/), which has over 50 members. Additional outreach and dissemination tools include a Google Group and communication via multiple listservs (IIPC, SAA WA-RT, Archive-It). Overall, the Year One of

WASAPI was successful in forming and launching a community model focused on building API-based interoperable systems for web archiving and included the participation of curatorial, technical, research, and the larger digital library and archives community in its work.

2) What are the community needs and possibilities for the planned open API to facilitate transfer of web archive data between distributed systems and what other prospective APIs does it point to?

While some of the work towards researching and answering this question was undertaken as part of the community-focused events mentioned above and the technical work detailed in the next section, the WASAPI team did produce a number of deliverables to assess needs and help inform functional requirements for the API development. Two project-specific surveys were conducted to assess the state of preservation data transfer in the web archiving community and help frame use cases to guide API functionality. One survey was of the general web archiving community and one was of the Archive-It community. Both are published on the WASAPI Github page. Two of the WASAPI team members were also involved in conducting the annual NDSA Web Archiving Survey and help guide this survey and its corresponding report (published in 2017) to include content of value to WASAPI's work. Additional research was done on related surveys, such as the IIPC's 2015 survey on APIs.

Team publications also contributed to meeting project goals in this area. As part of a number of the discussion and presentation events, a draft paper, "Interoperation Among Web Archiving Technologies," was distributed to attendees days before the events, with the goal of spurring discussion and guiding conversation. The paper is due to be published in Year Two of the project. An additional research paper, "Models for Collaborative Digital Library Development" has been drafted and is also slated for continued refinement and publication in Year Two. Four separate blog posts were published on the IIPC, Internet Archive / Archive-It, and Stanford University websites reporting survey results and project work. The WASAPI presentation and discussion at IIPC contributed to the organization of the IIPC Tools Portfolio group, which is intended to organize collaboration around APIs and whose proposal for a Compendium of Web Archiving Use Cases, was funded by IIPC to be created by WASAPI staff and whose authoring is a direct outcome of, and benefit it, the WASAPI project. As the grant proposal posited additional outcomes being an increased interest in, and work towards, affiliated projects in the community, the IIPC has agreed to fund work on an "Annotated Compendium of Web Archiving Use Cases." This need for this white paper emerged from the WASAPI project's work and will help provide additional background research to collaborative technology development and API design.

3) How can better interoperability of web archiving systems support new forms of access and research?

Project team organization of, and involvement in, the 12 community building and outreach events listed above included targeting diverse communities of users. Beyond the core web archiving community, presentations at CNI and LDCX involved the wider digital library community and talks and participation at Archives Unleashed and Dodging the Memory Hole involved the researcher and downstream user communities the project's API development is also intended to support. In addition, the Technical Working group met twice during Year One, for a full-day at Stanford in March and for a multi-hour meeting in Washington D.C. in December, to guide the project's work on API development. Working notes from the meetings are published on the WASAPI Github page. Project staff also met multiple times in Fall to plan the Year Two National Symposium event, determining format, attendees, and other details.

Lastly, the WASAPI project, though categorized as a research project, includes actual technology development, with key project outcomes being the creation of data transfer APIs consisting of working code running in multiple production environments. First, coming out of the Technical Working Group meetings, and with feedback and input via the aforementioned events and activities, a "general specification" for a data transfer API was developed. This was iteratively honed in accordance with community feedback and is intended to set bare minimum specifications for data transfer API development in any environment. Testing of production APIs in Year Two will help continue to refine this specification. Archive-It also published a "reference specification" related to the build of the Archive-It transfer API.

Both the general and reference specification are published in the WASAPI Github portal and includes user documentation and .yaml files that allow for viewing the specifications in the popular Swagger tool that allows uses to visualize and interact with an API without having an implementation in place. Working from these specifications, both Archive-It and LOCKSS developed data transfer APIs for their systems. These APIs will be tested in Year Two by all WASAPI grant partners and by additional institutions who came on-board the project in response to Year One's outreach activities. These APIs will facilitate the transfer of archived web data between systems and institutions and will be iteratively developed with community input throughout Year Two, allowing for improvement of both the general specification and local implementations. Both specification documents and both APIs are published open-source via the project's Github page: https://github.com/WASAPI-Community/data-transfer-apis

Changes

There were no major changes in key personnel, budget allocation, scope, or schedule during Year One of the project. The only notable events in this regard were the expansion of scope to include the work of overlapping projects within the grant partners' own institutions (no funding is needed or was requested for this addition). This includes documenting emergent affiliate APIs

by Internet Archive that complement the WASAPI work and help establish WASAPI as the hub of API development and web archiving systems interoperability collaboration.

Finding or Accomplishments During This Reporting Period
A number of findings and accomplishments came out of Year One of our work:

- The grant project team has been pleased to find significant interest in the national and international web archiving community around the project's work. This confirms the proposal's identification of the needs for a shared online space and larger community model for coordinating API-based and collaborative technology decisions.
- There is encouraging use of APIs in web archiving within some institutions or services, but this work remains poorly evidenced and largely uncoordinated. This is encouraging to grant staff, as it represents an opportunity that the WASAPI project is intended to address. This means the need for more coordinated effort on technology development, APIs, and systems integration is not just real but is growing.
- Capacity for technical development in web archiving at most institutions remains extremely low -- thus confirming the value of project's goals to align existing and emergent work at a national level to preclude engineering "silos" and maximize the impact of community work underway.
- Preservation functions for web archive data at most institutions remain minimal, again highlighting one of the project's goals, to facilitate local preservation and easier distribution of archived web data. There has been little change in the years prior to the grant's work in the number of institutions locally preserving their web data.
- The social architecture for supporting technical work in web archiving remains a critical need. Few web archiving events or groups at the national level are focusing on technical issues and groups in the larger digital library sphere that do focus on technical matters are involved with web archiving. As well, staffing and budgets for web archiving remain largely fractional and even many large research universities are lowering or flat-lining their staff/resource allocations for web archiving. Community coordinated and nationally aligned efforts are more vital than ever.