ASTROLINK ROADMAP FOR THE PRESERVATION, DIGITIZATION AND SERVICE OF HISTORICALLY SIGNIFICANT MATERIALS. M. A. Hunter[1] and A. E. Zink[1], [1]USGS Astrogeology Science Center (2255 N. Gemini Dr., Flagstaff, AZ 86005; mahunter@usgs.gov, azink@usgs.gov).

Introduction: Beginning in 2019, AstroLink was tasked with digitizing a subset of material identified by the NASA History Office as unique and of high interest to the planetary science community. That work was completed in 2020 and expanded to the rest of the collection, material amassed over decades by the Regional Planetary Information Facility (RPIF). In this process, materials are digitized, reviewed, corrected for text content, and documented using Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata (FGDC CSDGM) [1].

Completing this digitization will take years, so to facilitate public discovery of holdings the current spreadsheet-based inventories are being migrated to a web-based archive management platform.

Current Work: This current year's work focuses primarily on two tasks: 1) the migration of spreadsheet-based inventories and finding aids into CollectiveAccess, which uses open source catalog and publishing software to manage archives online, and 2) re-organization of materials to better facilitate long-term preservation.

Preservation. With the focus primarily on the digital service of collections, it is essential that physical materials are organized and archived according to best practices, to limit both degradation from exposure during storage and double handling. Much of the AstroLink collection is 50 years old or more, and even the best-preserved document mediums are at risk for permanent loss if they are not protected from light in climate-controlled facilities.

As documents are processed, they are also re-organized into acid-free file folders with uniform labels, within larger banker boxes, which block ultraviolet light and facilitate transportation as well as storage (**Figure 1**). Oversized materials, such as maps, are laid flat in map drawers and labelled according to the same organizational scheme.

Arguably the most difficult aspect of preservation work is identifying materials, whether they exist elsewhere and in what state, to determine whether they are unique to planetary science. AstroLink has discovered products published by a variety of sources, and they must be triaged if they are to be effectively processed, transferred to another institution, or disposed of according to the records management schedule. Storage space is increasingly difficult to justify and fund, so gaining efficiencies and modularity are necessary to protect collections into perpetuity.

Figure 1. Re-organized materials in protected boxes, away from direct sunlight.

Digitization. The digitization process is detailed in the Zink et al. 2021 *LPSC* abstract, which has progressed steadily despite on-site restrictions [2]. AstroLink has been able to continue digitizing operations in a de-centralized environment by developing and periodically updating guidelines. With the methodology and end-state as the focus of this work, it must be recognized that any online archive is only as useful as the content it offers. By using uniform specifications and open data formats, content can be generated concurrently with the development of the catalog infrastructure.

Increasingly, full texts are being indexed for content discovery, instead of relying strictly on keyword tags. Because many of the documents being digitized are

aggregations of topics, findings, or authors, they are all scanned with optical character recognition software and corrected to ensure embedded text is suitable for text search. Older, hand typed documents cannot be scanned reliably, and often require full over-writing of sections. Like creating metadata, this is a time-consuming step, but critical to ensure the long-term viability of the digital product.

Cataloging and Dissemination. This step is the primary focus of 2021 work, currently in progress (**Figure 2**). To this point, multiple spreadsheets had been used to track attributes unique to each product type (documents, maps, images and film, and artifacts). This resulted in disparate sources of information that were both inconsistently formatted (i.e., dates as text) and used (i.e., author names). By identifying overlap between spreadsheets, schema discrepancies have been resolved so that field data types will not collide during import. This has also been a useful exercise for assimilating like fields that were differently titled or used across inventories.

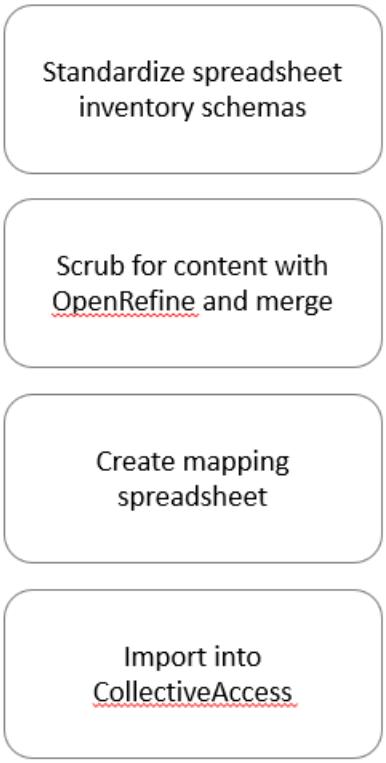

Figure 2. Major steps to migrating from spreadsheet inventories to a database.

The next step is to clean and complete spreadsheet inventories using OpenRefine, an open source web-based tool for manipulating tabular data. This will standardize the names of people, organizations, planetary bodies, missions, and instruments. This level of agreement is fundamental to enabling faceted search in downstream client applications, and it is at this level that community-scale development of controlled vocabularies is needed most.

The last steps are specific to migrating data into CollectiveAccess, which may not be directly applicable to other archive management platforms; however, these general steps are illustrative of mapping one data source to another. The mapping spreadsheet used for CollectiveAccess simply declares the mapping rule, source field, destination element, and options to parse, group or evaluate contents.

With the CollectiveAccess instance running and enhanced inventories complete, they will be imported and the CollectiveAccess profile will be further adjusted to support compatibility with major metadata standards, like FGDC or ISO (International Organization for Standardization), along with defined administrative/user roles, relationships and widgets to support archive management. It is the ultimate goal of this project that holdings can be imported or exported in sync with outside resources, and that the AstroLink profile will be available for others to use directly for setup.

Future Work: Looking ahead to the future, AstroLink plans to bring this archive management model to the greater planetary science community, and to empower custodians of historical planetary science collections of all sizes who wish to make their holdings public. Rather than offer a prescriptive end state that may not seem feasible for a small collection, this workflow is designed to address the most common problems for archive managers whose holdings are diverse and/or sparsely described.

Within the context of a larger Planetary Data Ecosystem (PDE) it will be necessary to solicit feedback on processes like this, as well as the standards and systems they are built on. Better understanding across which domains and at which levels interoperability are needed will drive many of the community's requirements. The underlying technology will continue to evolve but by maintaining interoperability across archives, a greater whole can be realized that is greater than the sum of its parts.

Acknowledgments: This work is funded by the FY21 USGS-NASA Planetary Spatial Data Infrastructure IAA.

References: [1] Zink, A. E. et al. (2020) *LPSC LI*, Abstract #2597. [2] Zink, A. E. et al. (2021) *LPSC LII*, Abstract #2548. [3] Skinner, J. A., et al. (2019) USGS Open-File Report 2019-1012.