

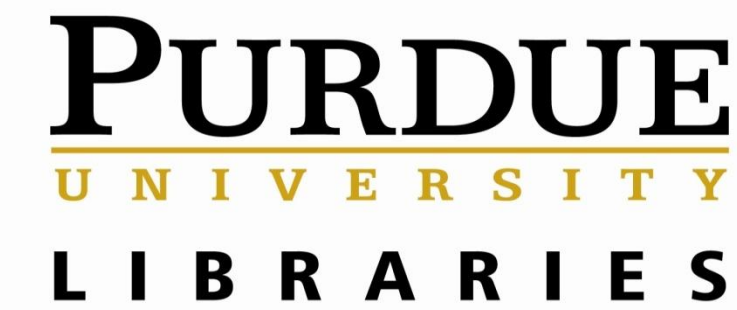
“Houston, we have a problem...” Designing a Digital Preservation System for a NASA Center for

Advanced Life Support

Jake Carlson – Data Research Scientist – Purdue University Libraries

/ Distributed Data Curation Center

Alexis Ramsey – Graduate Research Assistant – Purdue University Libraries



Access. Knowledge. Success.



Background

Purdue University Libraries

Recognizing that the nature of research and scholarly communication are changing rapidly and that the traditional roles and practices of the library have not kept pace with these changes, the Purdue University Libraries have begun an initiative to explore the potential of applying theories and practices of library science outside of the finished products of research (i.e. books and journals) to other sources of information such as unpublished works and data sets. Specifically, the Purdue Libraries actively seek out research collaborations between librarians and research faculty to develop solutions to real-world needs in the organization, management, discovery, sharing, and preservation of information resources as articulated by the researchers themselves.

ALS-NSCORT

The Advanced Life Support—NASA Specialized Center of Research and Training (ALS-NSCORT) was one of many interdisciplinary research centers at Purdue designed to develop holistic solutions to large scale challenges. Composed of scientists and engineers from Alabama A&M, Howard University and Purdue, the purpose of ALS-NSCORT is to explore and develop ways of supporting long-term human habitation in outer space.

Funding for the ALS-NSCORT center was redirected by NASA in 2007 before the project had reached its natural conclusion. Research had progressed to a point where the individual projects within ALS-NSCORT could have been integrated and tested as a single system. Facing the prospect of losing valuable information and having to start all over again, ALS-NSCORT approached the libraries for assistance.

Project Overview

Research question: Can the Libraries develop a digital preservation model based on user needs that would enable the Purdue Libraries to offer the preservation of research materials as a library service to interdisciplinary research centers and others?

The Libraries work in preserving material from ALS-NSCORT served as a pilot project to develop and test the preservation model. An important aspect of this project was to document the questions we encountered, the challenges we faced, the decisions we made as well as the resources that were required.

Objectives of the pilot project: The primary objective for ALS-NSCORT is to preserve their work in ways that will help them to revive their research projects and enable them to resume their work from where they left off if funding is reinstated from NASA or another agency.

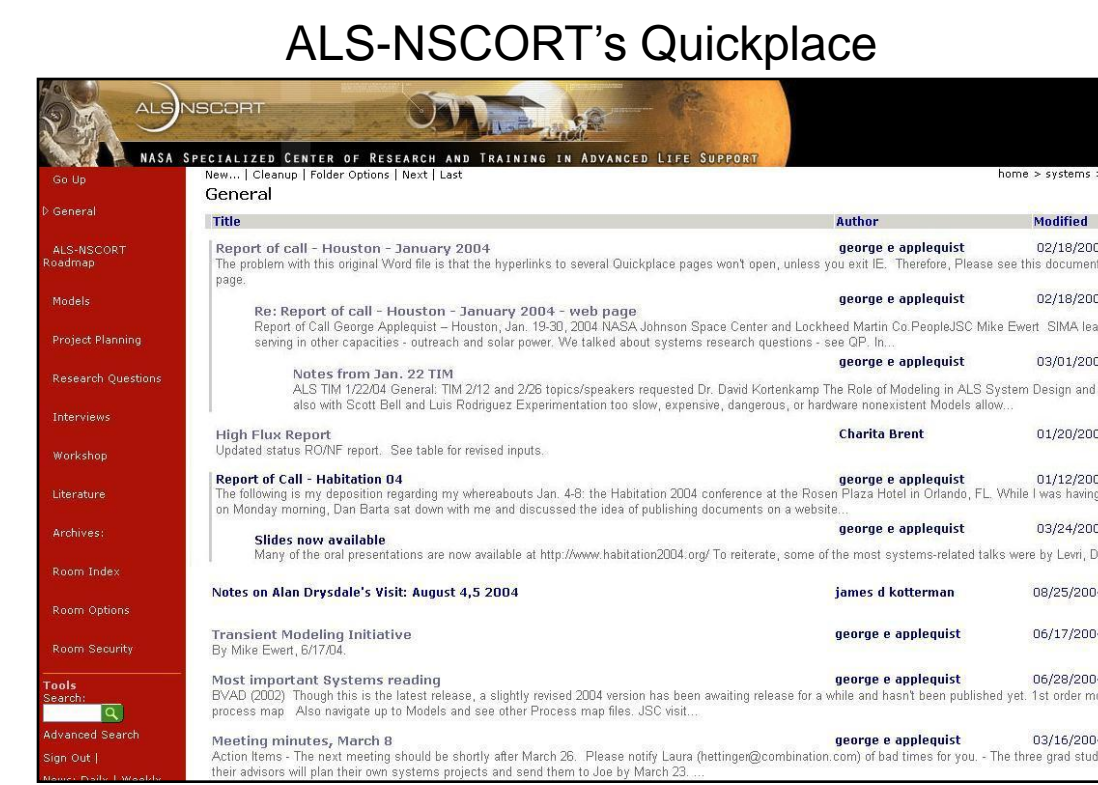
Their preservation system and the information it contains must be:

- Accessible
- Relevant
- Accurate
- Secure

The collection: ALS-NSCORT's research materials consisted of reports, documents, published articles, presentations, data sets, educational tools etc., in a variety of formats: Word, PowerPoint, Excel, PDF, JPEG, etc. These materials were made accessible to staff members and stored using QuickPlace communications software.

Phase 1 – Needs assessment, selection and appraisal

- Interview members of the ALS-NSCORT administration to identify specific user needs.
- Review ALS-NSCORT's research materials to develop an understanding of the nature of the materials and the potential challenges in preserving them.
- Conduct a selection and appraisal process with ALS-NSCORT to identify which items to include in the collection. Build an inventory spreadsheet to capture information about each item.

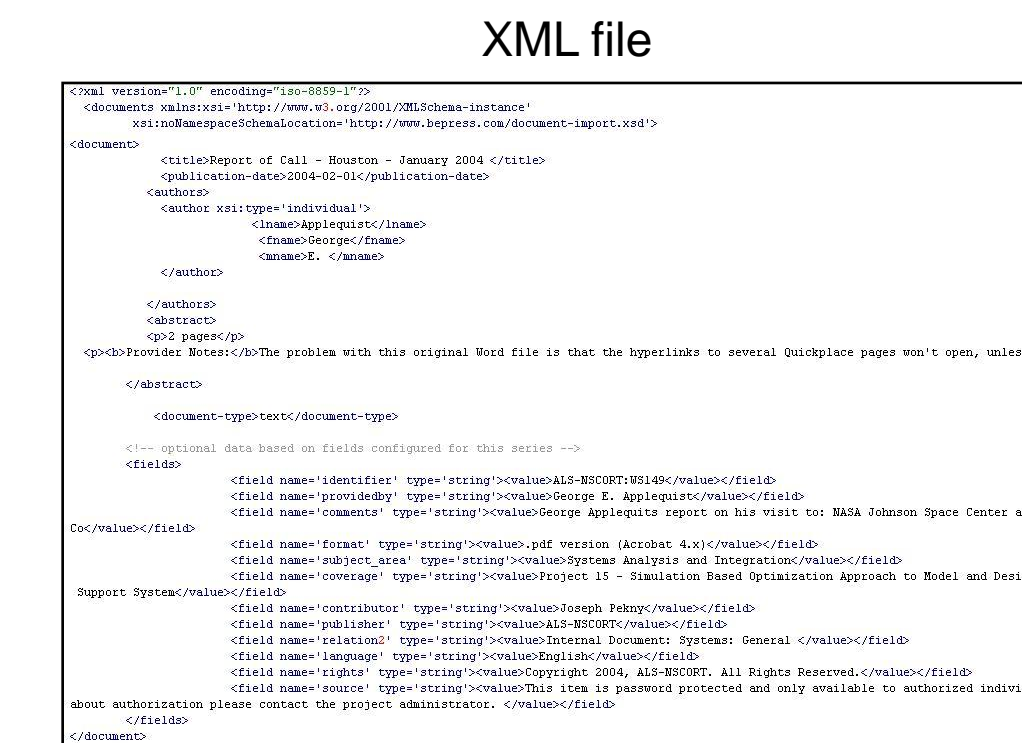


Phase 2 – Analysis and design of metadata

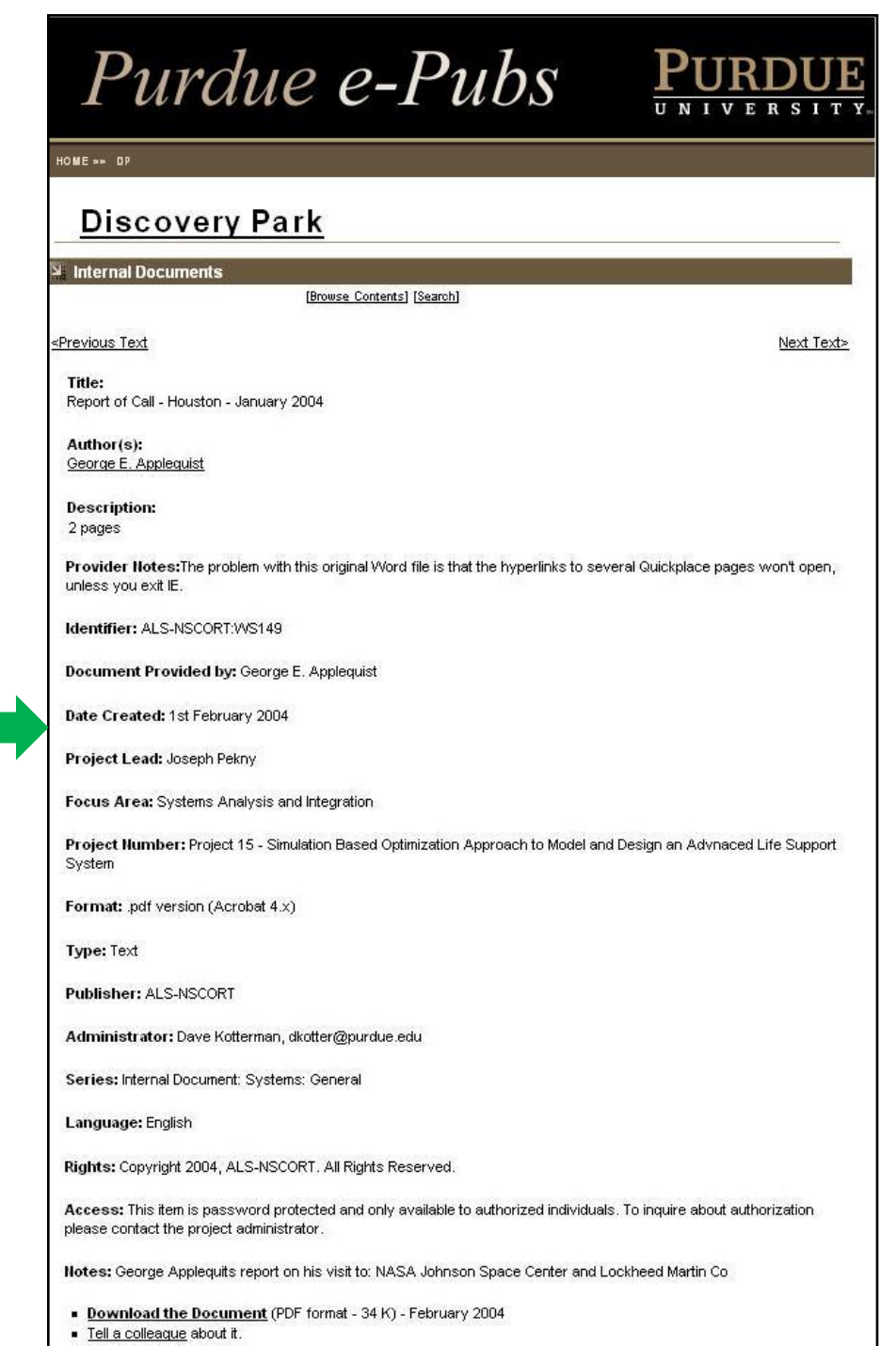
- Conceptualize the metadata that would be needed to satisfy the functional requirements of the system as articulated by ALS-NSCORT.
- Review existing documentation to determine what metadata already exists and what will need to be created.
- Explore relevant metadata standards and select a standard to employ for the ALS-NSCORT collection (Qualified Dublin Core).
- Update the inventory spreadsheet to reflect the metadata fields selected for the collection. Create the metadata and determine how to address missing or incomplete information.

Phase 3 - Developing the system and ingesting the materials

- Prepare a collection space for custom metadata and access requirements in the Libraries' institutional repository.
- Convert the inventory spreadsheet into XML files.
- Ingest the XML files and ALS-NSCORT materials into the digital preservation system.
- Document the model including the challenges faced and the decisions made to address them.



Ingest into the IR



Needs/Challenges/Decisions

Repository

Need: A digital repository to house ALS-NSCORT's collection of research materials.

Challenge: Developing a repository specifically for ALS-NSCORT would require resources beyond what was available in-house.

Decision: After exploring different options we decided to use the Libraries' institutional repository to host the ALS-NSCORT collection. Our vendor, bepress, provides ongoing bit-level preservation of all items in the repository. Working with bepress we modified the repository to enable discovery and access to items in the collection in ways that would satisfy the needs and requirements of ALS-NSCORT.

Preservation Format

Need: Preservation of the content of items within the collection.

Challenge: Whether to keep items in their original format or change them into PDF format. bepress will ensure the continued functionality of PDF files but has not extended this assurance to other file formats.

Decision: Items in formats that were primarily text based (.doc, .rtf, .ppt, etc.) were transformed and ingested into the repository as PDF documents. Items that were not primarily text based (.jpg, .xls, etc.) were ingested in their native format. All items in the collection will be burned to DVDs in their native format and given to ALS-NSCORT as back-up.

Copyright

Need: ALS-NSCORT wanted to include their publications in the collection.

Challenge: ALS-NSCORT authors did not retain copyright over their publications.

Decision: The Libraries could not legally ingest these publications into its repository. Instead we provided a record for each publication and included a pointer to direct users to where they could potentially obtain a copy of the item – publisher's website, library database, a physical copy in the library, etc.

Access Control

Need: Some of the materials within the collection are still "active" and cannot be distributed beyond ALS-NSCORT staff or trusted colleagues.

Challenge: The preservation system needs to support controlled access – allowing access to those ALS-NSCORT deem as trusted users while preventing access for others.

Decision: We investigated possible repository access controls and settled on the password protection for restricting access to ALS-NSCORT's "active" materials. The records for all ALS-NSCORT items are openly available to enable discovery. However, if the link to download a restricted item is clicked the user is asked to enter their account login and password as authentication for access.

Conclusion

Immediate outcome: We developed a functional short-term digital preservation system that meets the needs of ALS-NSCORT.

- Accessible: The collection is discoverable and accessible to ALS-NSCORT staff and others as appropriate through the Libraries' institutional repository.
- Relevant: The collection contains items identified by ALS-NSCORT as having enduring value for their needs.
- Accurate: The metadata describes the nature, purpose and context of each item within the collection as a whole.
- Secure: ALS-NSCORT has control over who can access their research materials.

Broader outcome: The Libraries are analyzing what we have learned from this pilot project and will assess the potential for using this model as the basis of offering digital preservation as a library service to other research centers.

Further investigations and explorations in this area will include:

- The reliability of using our institutional repository vendor.
- Ways to reduce the amount of labor involved.
- Approaches to securing the funding necessary to support the service.