National Library of Medicine

NDSR Project   NLM-Developed Software as Cultural Heritage

Project Summary

The Resident will create a pilot workflow for the curation, preservation, and presentation of a historically valuable software product, developed by the National Library of Medicine (NLM), which is deemed to be historically noteworthy due to its usage by a user community and/or its distinctive technical properties that are at risk of being lost due to obsolescence. The Resident will undertake a comprehensive review of software products developed by the NLM over the past 50 years with the goal of defining a software curation process, including preparation of files for preservation and developing effective workflows for long-term access to software products developed at the National Library of Medicine. The Resident will also research other efforts in the field of software preservation to determine if existing practices and tools could be brought to bear in the project. The work of this project may serve as a case study for the development of software preservation workflows and will likely be of great interest to other institutions considering such activities within their own organizations.

Specific Goals / Objectives

For this project, the Resident will:
- Survey other cultural heritage organizations and groups such as NIST to identify software selection criteria and best practices for the project.
- Survey NLM's software development corpus
- Identify candidates which would be particularly well-suited for preservation and presentation as historical artifacts.
- Select a NLM software product to be used for the pilot project.
- Conduct interviews with NLM staff and possibly others involved with the design, development, manangement and marketing of the selected software
- Conduct interviews with users of the selected software.
- Create a draft curation workflow informed by best practices.
- Work with NLM's Digital Repository group to prepare and ingest the software files.
- Work with NLM staff to prepare a public online presentation of the software, providing the historical context in textual or audiovisual form.

Timeframe & Deliverables

Months 1-4
- Survey other cultural organizations and groups such as NIST to identify software selection criteria and best practices for the project.
- Survey of NLM's software development.
- Select a NLM software product to be used for the preservation and presentation pilot project.

Months 5-8
- Conduct interviews with NLM staff involved with the design, development, manangement and marketing of the software along with users of the software.
- Write a narrative history contextualizing the artifacts.
- Develop a curation workflow, including appraisal and description of the artifacts.

Months 9-12
- Prepare and ingest the software (e.g., the binaries) into the NLM repository.
- Develop a public-facing web presentation, providing the historical context in textual or audiovisual form, and, if feasible, an interactive simulation.
- Create a final report summarizing the lessons of the overall project and providing recommendations for future software preservation activities.
- Present the findings and demonstrate the project outputs to NLM staff.

Project Deliverables
- Inventory of NLM software development including the historical significance of the

software projects according to NLM staff.
- A document stating the criteria used to identify the NLM software products that are well-suited for preservation and accompanying historical presentation.
- A report documenting the curation development process and repository-facing workflow used to ingest the content.
- Recommendations for further development of software curation workflows, long-term preservation of the software, and presentation of software as historical artifact at NLM.
- Software assets ingested into the repository.
- An online presentation of the software product providing its historical significance.

Resources Required

Mentors: Ben Petersen, Head, Preservation & Collection Management, Public Services Division, and Rebecca Warlow, Head, Images and Archives, History of Medicine Division.

Access to select staff within the National Library of Medicine, including the Office of Computer and Communications Systems (OCCS) and Library Operations (LO) units directly involved with the chosen software, the Digital Repository Working Group, the History of Medicine Division.

As needed, contacts with other institutions that have demonstrated interest and experience in software curation and preservation.

Context

NLM's over 50 years of software development to support biomedical research and medical librarianship has resulted in programs and data sets mailed to end-users in floppy disks and tapes, pioneering mainframe-based data services, and unique web-based tools and services. For example, Grateful Med[1], client-side software for preparing queries against NLM datasets; the original, mainframe-based DOCLINE interlibrary loan requesting service[2]; and the MeSH Browser[3], which provides web-based look-up of the MeSH vocabulary .
The archival community has seen a growing awareness of the need for software and hardware preservation, recently evidenced by the 2013 National Digital Stewardship Alliance (NDSA) meeting *Preserving.exe: Toward a National Strategy for Preserving Software.* The Library has not to date considered its legacy software from a curatorial perspective, wherein software can be seen as cultural artifacts in their own right. This project aligns well with the 2015 NDSA National Agenda for Digital Stewardship which recognizes software preservation as a key priority for community attention. The project also addresses the key topics of the Digital Preservation Outreach and Education (DPOE) curriculum by identifying NLM software products; determining what to preserve; how products will be stored, protected, and managed; and how NLM will provide access to the software products.

Required Knowledge and Skills for Residents

Graduate degree in Library and Information Science, or equivalent.
Additionally, the successful candidate will have the following:
- General knowledge of archiving and collecting practices, digital preservation concepts and technologies
- Ability to communicate clearly and effectively in writing, meetings and interviews

Preferred Knowledge or Experience

Conducting oral histories, writing narratives to provide historical context

---

[1] Knecht, L. & Shooshan, S. (2001). Internet Grateful Med to be Retired; Reminder of NLM Gateway Availability. NLM Technical Bulletin. (318; e3). Retrieved from: http://www.nlm.nih.gov/pubs/techbull/jf01/jf01_igm_phaseout.html

[2] Collins, M. (2007). An Overview of the DOCLINE System, its Functions, Purposes and Descriptions of Participating Libraries. Journal of Interlibrary Loan, Document Delivery & Electronic Reserve. (17, no. 3: 15-28). DOI: 10.1300/J474v17n03_05

[3] U.S. National Library of Medicine. (2014). MeSH Browser. Retrieved from: http://www.nlm.nih.gov/mesh/mbinfo.html