The ECHO DEPository is a 3-year NDIIPP-partner research and development project at the University of Illinois at Urbana-Champaign in partnership with OCLC and a consortium of content provider partners. This project supports the Library of Congress in its development of a technical foundation for digital preservation by conducting research and building tools to help us GET, SAVE and KEEP our digital cultural heritage. Specifically, we are building web archiving tools, evaluating existing repository software, developing an architecture to enhance existing repositories’ interoperability and preservation features, and modeling next-generation repositories for supporting long-term preservation.

**GET: We must identify and collect at-risk Web content now. How do we begin to do this?**

Manual item-level selection fails due to the enormous number of resources on the Web. Fully automated Web capture approaches result in substantive materials being buried under a mountain of irrelevant information. To address this, OCLC is building a suite of open-source Web archiving tools based on archival practice that bridges the gap between manual selection and automated capture. The Web Archives Workbench (WAW) comprises four tools that identify, describe, select and harvest Web-based content for storage in any repository.

Though still in development, the suite supports real-world archiving projects today. A 'lite' version of the full suite has been in testing since December 2005, allowing beta-tester institutions to discover and describe Web domains of interest, to describe and harvest content from these sites, and to deposit and manage the harvested content in the OCLC Digital Archive. To date, our testing organizations of five state libraries and archives and 20 non-partner beta testers have run over 1500 spiders and completed over 500 ingest of content to the Digital Archive. Current development of the suite focuses on building powerful features for targeting relevant content within Web sites.

**SAVE: Once we have obtained digital materials to be preserved, we need to store and maintain them. How well do existing repository systems support long-term preservation? What can we do now to enhance the interoperability and support for long-term preservation of our existing repository systems?**

To help ensure that we store content in reliable systems that will stand the test of time, UIUC is conducting an evaluation of four existing open-source repository systems (DSpace, ePrints, Fedora and Greenstone), and developing a “hub-and-spoke” proof-of-concept suite of scripts and tools for enhancing the interoperability and preservation features of those systems. The evaluation work includes the ingestion and manipulation of half a terabyte of heterogenous content in each repository system, as well as the development of a preservation-focused repository evaluation framework based on current thinking about what systems, resources and support must be in place for an institution to be a trusted long-term curator of digital resources.

The “hub and spoke” interoperability model addresses real-world library problems for managing multiple repository systems and ensuring valuable preservation data is not lost today. This work includes the development of a common “vanilla” standards-based method for packaging content that supports the collecting of technical and provenance information crucial for long-term preservation, and allows digital objects to be moved in and out of more repositories more easily. This model has potential wide applicability and is already being implemented in real-world settings.

**KEEP: What must repository systems do tomorrow to support long-term preservation?**

Current first generation repository systems preserve the structure of information, not its meaning or semantics. When we move content from one system to another, this structure may be subtly or unsubtly transformed. To meaningfully preserve our digital content over time, we therefore have to infer meaning or semantics from structures that change over time. Because of volume, we need to be able to do this with automated tools.

The UIUC Graduate School of Library and Information Science (GSLIS) is working with NCSA to contribute to the development of next-generation archives with semantic analysis capabilities to reduce long-term preservation risks. Using a new repository system developed at NSCA and automated reasoning tools developed at GSLIS, we are building a proof-of-concept semantic archive. This work includes developing a more formally structured understanding of how descriptive information about archived digital resources is structured by analyzing real-world examples from content exchange experiments in the repository evaluation activities. This work will model how semantic inference capability may help next-generation archives head off long-term preservation risks.