

Chronopolis Project Launched Under Library of Congress Partnership to Preserve At-Risk Digital Information

SDSC, UCSDL, NCAR and UMIACS Focus on Cross-Domain Sharing for Vital Collections

April 14, 2008

Jan Zverina

The Chronopolis Digital Preservation Demonstration Project, one of the Library of Congress' latest efforts to collect and preserve at-risk digital information, has been officially launched as a multi-member partnership to meet the archival needs of a wide range of cultural and social domains.

Chronopolis is a digital preservation data grid framework being developed by the San Diego Supercomputer Center (SDSC) at UC San Diego, the UC San Diego Libraries (UCSDL), and their partners at the National Center for Atmospheric Research (NCAR) in Colorado and the University of Maryland's Institute for Advanced Computer Studies (UMIACS).

A key goal of the Chronopolis project is to provide cross-domain collection sharing for long-term preservation. Using existing high-speed educational and research networks and mass-scale storage infrastructure investments, the partnership is designed to leverage the data storage capabilities at SDSC, NCAR, and UMIACS to provide a preservation data grid that emphasizes heterogeneous and highly redundant data storage systems.

"Chronopolis is part of a new breed of distributed digital preservation programs," said Brian E.C. Schottlaender, a Principal Investigator on the project and UCSD University Librarian. "We are using a virtual organizational structure in order to assemble the best expertise and framework to provide data longevity, durability and access well into the next century."

"The Chronopolis team leverages broad experience in the use and access of research data from the science and engineering community with deep experience from the library and archival communities on the preservation of cultural assets," said Francine Berman, director of the San Diego Supercomputer Center at UCSD, and also a Principal Investigator for Chronopolis. "The project allows innovation in multiple dimensions and will give us experience with a scalable framework for developing preservation grids."

Specifically, the partnership calls for each Chronopolis member to operate a grid node containing at least 50 TB of storage capacity for digital collections related to the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP). For reference, just one terabyte of information would use up all the paper made from about 50,000 trees.

The Chronopolis methodology employs a minimum of three geographically distributed copies of the data collections, while enabling curatorial audit reporting and access for preservation clients. The partnership will also develop best practices for the NDIIPP community for data packaging and transmission among heterogeneous digital archive systems.

With its launch, two collections from within the NDIIPP community will be incorporated into the Chronopolis preservation grid. The Inter-university Consortium for Political and Social Research (ICPSR), based at the

University of Michigan, will provide up to 12 TB of data from its world-renowned archive of social science and political research data sets, marking the first time that the collection is completely stored outside the state of Michigan.

In addition, the California Digital Library (CDL) will provide up to 25 TB of content from its "Web-at-Risk" collections, which were first selected under the auspices of the original NDIIPP preservation partnerships to preserve political campaign websites in 2004.

Chronopolis partners will generate permanent information for each collection using the Audit Control Environment (ACE), developed by UMIACS, and the Storage Resource Broker (SRB) developed by the Data-Intensive Computing Environments (DICE) group at SDSC. Each collection from ICPSR and CDL will have a community-targeted Web interface to check on the status and authenticity of the collections while they are being managed within the Chronopolis data grid.

Media Contacts: Jan Zverina, SDSC Communications, 858-534-5111 Warren R. Froelich, SDSC Communications, 858-822-3622