

Preservation Environments that Manage Technology Evolution

Reagan Moore

Richard Marciano

San Diego Supercomputer Center

*International Research on Permanent Authentic Records in Electronic Systems
(InterPARES) Presentation*

Preservation can be thought of as the extraction of digital records from the environment in which they were created, and their import into a preservation environment. The preservation environment, in turn, is the set of software, management policies, and infrastructure that shields the records from changes in technology. In effect, the archives and their authenticity and integrity properties are maintained independently of the choice of technology used for storage, access, and management.

The ability to manage technology evolution depends upon the ability to characterize the properties of the preservation environment. InterPARES has identified the fundamental preservation requirements as ensuring authenticity and integrity of the digital records. Authenticity corresponds to assertions made by the creator of the record about the governing institution, the reason the record was created, the name of the creator, the date the record was submitted into the preservation environment, that the chain of custody has been tracked, among other parameters. Integrity corresponds to assertions made by the archivist that the record remains uncorrupted, that the authenticity metadata remains linked to the record, that an audit trail records all operations performed on the record, that access controls have ensured no unauthorized changes, that sufficient replicas exist to mitigate risk of data loss, among other parameters.

The concept of infrastructure independence is added to the preservation environment to characterize the ability to migrate the archives onto new technology without loss of authenticity or integrity. In practice this means not only that the choice of storage technology introduces no proprietary format or naming conventions that make the use of new technology difficult, but also that the preservation environment simplifies the use of new technology, making it easier to employ new technology that is more cost effective and more robust.

As part of InterPARES 2, data grid technology has been evaluated for its utilization as the preservation environment. Data grids provide the interoperability mechanisms that make it possible to simultaneously use both old and new technologies. At the point in time when the archives is migrated onto new technology, the old technology is also present. Thus an interoperability mechanism facilitates the use of new technology. Data grids also provide persistent naming to all points of interaction between the preservation environment and the external world. The persistent naming means that no matter which technology is used, the archivist retains control over the names used to identify archivists, to identify digital records, to identify authenticity and integrity attributes, to arrange the digital records, and to manage access controls. This paper describes multiple use cases in which new technology is inserted into a preservation environment, and discusses how data grid technology simplifies the task.

An earlier case study was presented in Focus 2, called “Persistent Archives Based on Data Grids”, where the minimal capabilities needed within grid technology for preservation of governmental records were examined. This case study led to an article in the *Journal of American Archivist* (to be published in August 2006). However, this case study and resulting paper are really two years old. What is proposed in this book chapter is different and will include development over the last two years as well as examples of the technology. This IP2 book chapter will also be distinct from the VanMAP case study, which discusses the use of data grids for recordkeeping and preservation purposes but focuses primarily on GIS records.