

## Exploring the Concept of Temporal Interoperability as a Framework for Digital Preservation\*

Margaret Hedstrom, University of Michigan, Ann Arbor, MI USA

**Abstract:** This paper explores a new way of thinking about digital preservation and introduces a new requirement for interoperability that I refer to as temporal interoperability. The concept of temporal interoperability concerns the interoperability of systems and the access to heterogeneous collections over time. The paper discusses recent developments in digital preservation that begin to approach preservation as a problem of interoperability. It also suggests areas where methods to enhance interoperability in real time can support continuing access and long-term preservation.

\*\*\*

As organizations of all sorts build rich digital resources there is increasing concern over the ability to preserve those resources and provide access to them over time. Typically, this issue is defined as the problem digital preservation or archiving where organizations develop collection- or institution-specific methods for preserving digital information in light of regular change in the underlying storage, data management, presentation, and delivery technologies. In this paper I propose a new way of thinking about digital preservation and introduce a new requirement for interoperability that I refer to as *temporal interoperability*. By temporal interoperability, I mean the ability of current systems or legacy systems to interoperate with future systems that may use new formats, data models, languages, communication protocols, and hardware. Although the concept of temporal interoperability is in a very formative stage, recent developments in interoperability and in digital preservation could be leveraged against each other to provide a framework for more systematic and scalable approaches to digital preservation.

Interoperability is a complex problem for distributed systems generally and for digital libraries in particular. As more content becomes available in libraries, archives, museums and galleries through the conversion of conventional materials to digital form and through the acquisition of born-digital content, demand for access grows from users scattered around the world using different networks and client systems. The goal of interoperability is to allow components of a system to evolve independently without sacrificing the ability of those components to communicate with each other. Interoperability involves work on technical issues (hardware, data formats, applications, and communication protocols); representation issues (language, metadata, visualization and other aspects of semantics); and social issues (institutional cooperation, legal rights and responsibilities, and cultural sensitivities) [2].

Temporal interoperability is not simply a new term of digital preservation. By thinking about digital preservation as an interoperability requirement, many of the measures proposed for interoperability in real time can be leveraged to support interoperability over time. It can refocus preservation research and development toward standards, languages, and technologies that have a better chance of scaling than many of the approaches

currently proposed for long-term preservation. Temporal interoperability also promises to make the digital archives of the future as interoperable as the digital libraries of today.

The greatest technical impediment to long-term preservation is incompatibility between the hardware and software used originally to create, store, manage, and disseminate digital information and current or future computing platforms. The lack of interoperability between older and newer generations of computing platforms is evident whenever established storage media, data formats, applications, programming languages, representation schema, or rendering tools are replaced by new methods without backwards compatibility. One particular technical challenge for preserving complex digital objects is that different components of systems evolve at different rates. Rather than encountering an abrupt point where all of the components of a system become incompatible with each other, storage media, operating systems, applications, and data formats tend to evolve independently over the course of many years. Even if backward compatibility or translation services exist between different versions of each component, eventually some link in the chain of dependencies breaks [8].

The digital preservation strategies that are currently in use rely primarily on minimizing the extent to which continuing access to any preserved digital resource is dependent on a particular computing platform, software application, or data format. Standardization is one method used to reduce the number of dependencies because compliance with widely adopted standards for storage media, data formats, and metadata reduces the heterogeneity in a digital archive. If the collections in a repository conform to a minimum number of standard formats, repository managers can monitor the evolution and fate of a small set of standards and develop transition plans from standards that are becoming obsolete to the newer standards that replace them. Standardization works best in environments where the creators, producers, and users of resources use common tools, data formats, and terminology and where they already comply with standards to support their work, such as sharing information and comparing results. Archives, libraries, and data repositories have been less successful at convincing producers to adopt standards simply because they facilitate long-term preservation. Moreover, the standards adopted by one community of data producers may not be compatible with standards used by other producers.

Migration, or the conversion of data from older to newer formats, is another strategy commonly used for used long-term preservation. Repositories often convert data on intake from the media supplied by the producer to the media currently used by the repository and from non-conforming formats and applications to those that the repository supports. For example, a repository might acquire numerous data sets on CD-ROMs with their own database management systems and directory structures. For long-term preservation, the repository would copy the data to its own storage system and convert the data into a database management system that the repository supports. Over time, however, the storage technologies, file formats, database systems, and directory methods that a repository uses will become obsolete. Digital archives are then faced with the need to convert their holdings to current media and current formats in order to continue to manage and provide access to their holdings. These problems are not unique to digital

libraries and archives. Any organization that needs to access digital information beyond the useful life of its original computing platform faces the problem of inter-generational incompatibilities among the components of their systems.

There are many limitations to digital preservation strategies that rely on standards and migration. From an interoperability perspective, standards and migration are problematic because they focus on making preserved resources compatible with current technologies rather than developing ways for older technologies to interoperate with newer ones. Typically each collection has to be analyzed in detail and a specific method of conversion has to be developed for each resource. The history of changes in storage media, data formats, programming languages, and software applications suggests that such transformations have to occur at a minimum of every ten to fifteen years, adding to the expense of digital preservation and the risk of introducing errors into the data. From a user's perspective, each conversion of a resource from an obsolete to a current format creates a new version which, over time, deviates more and more from the original.

The Open Archival Information System Reference Model (OAIS) is a significant development that begins to provide a more systematic framework for digital preservation. The Consultative Committee on Space Data Systems developed the OAIS Reference Model out of concern for long-term preservation of data from space missions and scientific instruments [1]. Involvement of interested parties from the library and archival communities, however, helped produce a model that is broadly applicable to organizations that create information which may warrant long-term preservation. The OAIS Reference Model provides a high-level definition of the environment of an archival information system, defines the functions of a digital archive, and presents a logical model for the information stored in an archival repository. Although the OAIS Reference Model does not specify any particular design or implementation for a digital archive, its data management recommendations favor compliance with standards and periodic migration of storage media and data formats.

Researchers are also investigating emulation as a method for long-term preservation. The basic idea behind emulation is to develop programs (emulators) that can be run on current computing platforms to emulate the functions and behavior of obsolete technologies. Using the emulation approach, a repository would preserve an exact copy of a digital object on current storage media along with a copy of the original applications needed to access and render the original object. Access to the original object is achieved via its native application running under emulation on a current computing platform. Emulation as a digital preservation strategy still is largely experimental and several different technical approaches that are being investigated [4, 5, 9]. Nevertheless, emulation is currently being used in several areas to provide interoperability between obsolete and current systems or between two currently incompatible systems. Hardware manufacturers frequently provide emulation modes for processors and peripherals when they ship new hardware, such as the DOS mode that is included with most Windows platforms. Several products, such as SoftWindows, Virtual PC, and Wine, emulate the Wintel environment on non-Wintel machines. Computer hobbyists and game enthusiasts

are using emulation to preserve hundreds of computer games, some of which have not been marketed or supported for more than fifteen years [4].

From the perspective of temporal interoperability, emulation has a number of distinct advantages over other digital preservation methods. Rather than dealing with incompatibilities at the level of documents, objects, and data formats, proponents of emulation contend that a relatively small number of emulators will be needed for hardware components, application program interfaces (APIs), and operating systems. Emulation also allows users to interact with an exact copy of an original digital object rather than a reformatted version or a surrogate. Of all of the technical approaches being developed to support long-term preservation, emulation comes closest to the concept of temporal interoperability. Rather than converting older digital collections to current formats so that they can be disseminated and used with current technologies, emulation is oriented toward interoperability between components of older systems and current computing platforms.

Regardless of the specific digital preservation method used, there are many potential points of synergy between interoperability in real time and temporal interoperability. On a technical level, interoperability and digital preservation share the goals of information access and sharing that are independent from particular computing platforms, data formats, document models, and languages [6]. Many of the technical approaches that are being developed for contemporaneous interoperability also are potentially useful in more refined strategies for digital preservation. Widespread conformance with strong standards reduces the number of data formats, formal models and languages with which a digital repository must contend. Persistent naming conventions for digital objects, components, and versions simplifies long-term management of heterogeneous collections that a repository may acquire from many different sources. External wrappers, gateways, and schema translation can provide frameworks for transforming exogenous resources into data and object models that a digital repository can handle. Specification-based interaction and efforts towards software reuse are particularly compatible with efforts to develop emulation as a digital preservation strategy [6, 7, 10].

Semantic interoperability, or the ability to transfer meaningful information from senders to recipients across space and over time, is a very difficult problem for both synchronous and deeply asynchronous exchanges [3]. Some of the models for descriptive information developed to support long-term digital preservation may be useful for real time interoperability. The OAIS reference model, for example, distinguishes between the content of digital objects and their representation information. Representation information includes both structural and semantic information about digital objects and their context, provenance, and relationships to other digital objects. Typically, the descriptive models for preserved digital objects go beyond common bibliographic models and metadata schema by encompassing taxonomies, reference systems, registration systems, and a variety of other descriptors organized into complex representation networks based on assumptions about the knowledge base of a designated user community [1, pp. 4-19 – 4-30]. Given that semantic operability becomes immensely

complicated over time as natural languages change and use environments evolve, temporal interoperability will require constant attention to changing semantics.

Expressing digital preservation as an interoperability problem offers a number of benefits to research and development in both real time interoperability of digital libraries and interoperability over time. Temporal interoperability has the potential to provide much more scalable for continuing access to digital collections because it focuses on bridging incompatibilities between older and new technologies rather than on constantly transforming a much larger number of older collections to the requirements of new computing platforms. Addressing current and temporal interoperability issues jointly also has the potential to place digital preservation issues in the mainstream of digital library research rather than treating preservation as a separate issue or an afterthought. Viewing digital preservation as an issue of interoperability recognizes that communities that are building interoperable digital libraries face many of the same issues as those who are concerned primarily with long-term preservation, such as the need for greater independence from particular computing platforms. Finally, common approaches to interoperability for both current and long-term use of digital collections will enable the integration of current and archival collections into a seamless set of resources that is not divided arbitrarily between "current" and "non-current" or between libraries and archives. Most of all, users may come to expect that systems which interoperate across organizations, nations, languages, and cultures should also be able to interoperate between the past and the future.

## REFERENCES

\*Research for this paper was supported in part by the National Science Foundation, Award #9905935, Digital Library Initiative - International, Emulation Options for Digital Preservation.

1. Consultative Committee for Space Data Systems (2001). *Reference Model for an Open Archival Information System (OAIS)*. Red Book, CCSDS 650.0-R-1.2. Mountain View, CA (June 21).
2. Fox, E. and Marchionini, G. (1998). Toward a Worldwide Digital Library. *Communications of the ACM* 41:4, 29-32.
3. Heiler, S. (1995). Semantic Interoperability. *ACM Computing Surveys* 27:2, 271-73.
4. Holdsworth, D. and Wheatley, P. (2001). Emulation, Preservation and Abstraction. *RLG DigiNews* 5:4 (August 15). Available: <http://www.rlg.org/preserv/diginews/diginews5-4.html#feature2>
5. Lorie, R.A. (2001). Long-term Archiving of Digital Information, *Proceedings of the First Joint Conference on Digital Libraries*, Roanoke, VA, June 23-26, 2001.

6. Paepcke, A., Chang, C. C. K., Garcia-Molina, H., and Winograd, T. (1998). Interoperability for Digital Libraries Worldwide. *Communications of the ACM* 41:4, 33-42.
7. Payette, S. Blanche, C., Lagoze, C., and Overly, E.A. (1999). Interoperability for Digital Objects and Repositories: The Cornell/CNRI Experiments. *D-Lib Magazine* 5:5 (May 1999). Available: <http://www.dlib.org/dlib/may99/payette/05payette.html>.
8. *Preserving Digital Information: Report of the Task Force on Archiving Digital Information* (1996). Washington, DC: Commission on Preservation and Access and the Research Libraries Group.
9. Rothenberg, J. (1999). *Avoiding Technology Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. Washington, DC: Council on Library and Information Resources.
10. Scherlis, W. L. (1996). Repository Interoperability Workshop. Towards a Repository Reference Model. *D-Lib Magazine* (October). Available: <http://www.dlib.org/dlib/october96/workshop/10scherlis.html>