The International Journal of Digital Curation
Issue 1, Volume 3 | 2008

Towards a Theory of Digital Preservation

Reagan Moore,
San Diego Supercomputer Center

June 2008

Abstract

A preservation environment manages communication from the past while communicating with the future. Information generated in the past is sent into the future by the current preservation environment. The proof that the preservation environment preserves authenticity and integrity while performing the communication constitutes a theory of digital preservation. We examine the representation information that is needed about the preservation environment for a theory of digital preservation. The representation information includes descriptions of the preservation management policies, the preservation processes, and the state information that is needed to verify the correct working behavior of the system. We demonstrate rule-based data grids that can verify that prior policies correctly enforced preservation properties, while sending into the future descriptions of the current preservation management policies.[1]

[1] This work was supported by the National Archives and Records Administration under NSF cooperative agreement 0523307 through a supplement to SCI 0438741, "Cyberinfrastructure: From Vision to Reality" and by the National Science Foundation grant ITR 0427196, "Constraint-based Knowledge Systems for Grids, Digital Libraries, and Persistent Archive". The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied, of the National Science Foundation, the National Archives and Records Administration, or the U.S. government.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. ISSN: 1746-8256 The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre.

Introduction

The concept of preservation can be characterized as communication with the future. We know that in the future new technology will be used that is more cost effective and more sophisticated than current technology. Communication with the future then corresponds to moving records onto new choices of technology. The preservation environment will need to incorporate new types of storage systems, new protocols for accessing data, new data-encoding formats, and new standards for characterizing provenance. Thus a major challenge that confronts preservation is how to incorporate new technology effectively, while conserving preservation properties such as authenticity, integrity, and chain of custody. Data grid technology provides the required ability to incorporate new technology, without compromising the preservation environment properties, through the concept of infrastructure independence (Moore, 2006). We will explore this concept under "Infrastructure Independence".

We can also view preservation as the validation of communication from the past. An archivist makes assertions about the application of prior preservation processes and their controlling preservation policies. Any claims about the current state of authenticity and integrity rely upon a complete description of prior actions. This is the second major challenge for preservation environments, the ability to characterize how prior preservation processes have been controlled by preservation management policies. Rule-oriented data grids support explicit characterization of both policies and processes (Rajasekar, Wan, Moore & Schroeder, 2006). We will explore this concept under "Integrated Rule-Oriented Data System".

Given a characterization of the preservation processes and management policies, it is then possible to verify properties of the preservation environment. In particular, it is possible to verify trustworthiness as defined by the RLG/NARA Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC) (2007). This is the third major challenge for preservation environments, the ability to verify that they are working correctly. We explore this requirement under "Assessment Criteria".

A theory of digital preservation then becomes possible. It is based on the definition of the minimal set of preservation processes that are needed to implement management policies, and the minimal set of preservation metadata (persistent state information) needed to validate assessment criteria. The goal is to show that the preservation environment maintains the desired properties as the underlying technology evolves. When we examine the components that are required to build a theory of digital preservation, we note that representation information about the preservation environment itself is required. Preservation environments require not only the ability to protect records from changes that occur in the preservation infrastructure, but also the ability to characterize changes that occur in preservation policies. We explore how rule-oriented data grids provide these capabilities under "Theory of Preservation", and close with a description of the components of a theory of digital preservation.

Infrastructure Independence

A preservation environment is the software middleware that shields records from the rapid evolution of technology. The preservation environment provides standard operations on persistent name spaces that can be used to implement preservation processes. The standard operations are ported to each new version of the infrastructure. Thus from the perspective of the archivist, the names that are used to identify records, archivists and storage systems should not change, and the operations that are performed upon the records should remain consistent over time. The preservation environment effectively shields the records from any dependencies upon the current choice of infrastructure.

Data grids are software middleware that enable the formation of shared collections from distributed data. The data may reside in different types of storage systems, located in different institutions that are linked by the Internet. The data grid maintains the information needed to identify where a file is actually stored, manages descriptive metadata about each file, manages replicas of files, and manages communication over wide area networks. Persistent name spaces are maintained that include the names used to identify archivists, the names used to identify records, the names used to identify descriptive attributes, and the names used to identify storage resources. Data grids implement trust virtualization through authentication of each user and management of explicit access controls on the files, metadata, and storage systems. Data grids also provide a standard set of operations for accessing the remote files. In effect, data grids manage all properties of the records independently of the choice of storage and database technology. This capability is the basis for the concept of infrastructure independence.

The standard operations provided by data grids for manipulating data have been designed to support the processes required for data sharing, data publication, and data preservation. The operations include byte-level manipulation (open, close, read, write, seek, stat, etc.), file level manipulation (replicate, create a version, backup, aggregate in containers, parse metadata), and collection level manipulation (recursive operations on directories, bulk access, bulk data movement). Currently, the Storage Resource Broker (SRB) data grid provides over 80 operations in support of remote data access and manipulation[2].

We observe that the storage protocols supported by a particular type of storage system differ in either semantics (intent of the operation) or manipulation (different result generated) across vendor products. The SRB technology implements standard operations that are storage system-independent by creating storage system-specific drivers to map from the standard operations to each vendor's storage system protocol. The Storage Resource Broker data grid is an example of a system that provides infrastructure independence for preservation environments.

SRB data grids are in production use around the world, varying in extent from local environments that manage data on local disk and tape archives, to international data grids that move hundreds of terabytes of data between North America and Europe. At the San Diego Supercomputer Center, SRB data grids manage a Petabyte of data

[2] Storage Resource Broker data grid http://www.sdsc.edu/srb/index.php/

comprised from more than 200 million files. Production SRB data grids support digital libraries, preservation environments, real-time sensor data streams, and data analysis platforms.

The ability of data grids to incorporate transparently new technology forms the basis for their selection as the infrastructure on which preservation environments can be built. At the point in time when a new storage technology becomes available, a new data grid storage resource driver can be implemented that maps the standard operations to the new storage access protocol. Records that have been stored on the old technology can be migrated to the new storage system without the names of the records changing, without the access controls on the records changing, and without the links between the descriptive metadata and the record being broken. The data grid transparently manages all required administrative metadata to ensure that authenticity and chain of custody are preserved while new technology is brought into the preservation environment.

Data grids also manage data integrity. Checksums are maintained of each record, along with multiple replicas. Administrative commands can be issued that verify the integrity (checksum correctness), verify the existence of replicas (synchronization), and verify that for each archived record appropriate descriptive metadata is present. Based on these capabilities, multiple preservation environments have been created. Notable examples include the NARA Transcontinental Persistent Archive Prototype (TPAP)[3], the National Science Digital Library persistent archive[4], and the California Digital Library Digital Preservation Repository[5]. Given this success, the question is what else is needed for a preservation environment to be feasible?

We observe that the labor required to manage large 100-terabyte-sized collections that have tens of millions of files can become onerous. It is not sufficient to provide administrative functions (preservation processes) that can be executed to recover from detected errors. Instead the data grid must automate as much as possible administrative functions, including verification of integrity and authenticity, for large preservation systems to be viable.

Integrated Rule-Oriented Data System

The automation of management policies is feasible if policies can be expressed as rules that are enforced on each data access or manipulation. The iRODS integrated Rule-Oriented Data System provides a mapping from assessment criteria, to preservation policies that enforce the assessment criteria, to the preservation processes that manipulate the records. Data grid middleware software is used to map from the desired management policies to commercially available storage systems and databases. Based on experience with the Storage Resource Broker data grid (Baru, Moore, Rajasekar, & Wan, 1998), we recognize that the operations supported by the remote storage systems do not correspond directly to preservation capabilities. We use the standard operations supported by the SRB data grid to define micro-services that are executed directly on the remote storage location. Thus each preservation capability is an aggregation of multiple micro-services, and each micro-service is an aggregation of multiple operations at the remote storage system.

[3] NARA Transcontinental Persistent Archive Prototype http://www.sdsc.edu/NARA
[4] National Science Digital Library http://nsdl.org/
[5] California Digital Library Digital Preservation Repository
http://www.cdlib.org/inside/projects/preservation/dpr/

Preservation Environment	*Conserved Properties*	*Control Mechanisms*	*Capabilities*
Management Functions	Assessment Criteria	Preservation Policies	Preservation Processes
	Data grid – Management virtualization		
Data Management Infrastructure	Persistent State	Rules	Micro-services
	Data grid – Data and trust virtualization		
Physical Infrastructure	Database	Rule Engine	Storage System

Table 1. Characterization of preservation management policies as rules.

As shown in Table 1, we express a preservation process as a set of micro-services. We express the management policies as sets of rules that control the execution of each micro-service. We evaluate the assessment criteria through queries on persistent state information that is generated by application of the rules. The rules are stored in a rule engine that is installed at each remote storage location, and the persistent state information is stored in a central metadata catalog.

Note that even if the micro-services remain unchanged over time, we still need a standard set of operations that is applied at the remote storage systems that also remain unchanged. The two levels of data virtualization shown in Figure 1 make it possible to build a preservation environment that executes standard micro-services (for expressing preservation processes) that are composed from standard operations (for interacting with storage systems). Thus preservation processes that are expressed as sets of standard micro-services will work across both old and new storage systems through the standard operations.

The iRODS micro-services can be thought of as defining the minimal set of preservation functions that need to be carried forward in time. Similarly, the iRODS rules can be thought of as defining the minimal set of management policies that are needed to enforce trustworthiness. Given a minimal, complete description of the components of a preservation environment, one can then create a rule-based system that is internally self-consistent. The required operations generate the persistent state information needed to validate assertions about trustworthiness.

Besides the ability to characterize the preservation environment, a second implementation requirement is the automation of the execution of management policies. We observe that as collections increase in size, the amount of labor required to recover from problems that occur in distributed environments becomes onerous (Moore, Wan, & Rajasekar, 2005). The iRODS system automates the application of the rules as either atomic operations, executed on each preservation process invocation, or as deferred operations, executed when resources become available, or as periodic operations. The support for deferred operations is needed to handle situations where the remote storage system is temporarily inaccessible. Periodic operations are needed to manage validation of integrity and authenticity. A rule can be written that validates checksums, synchronizes replicas, and corrects copies that have been corrupted. Similarly a rule can be written that compares provenance metadata to the required metadata for a record series, and then either identifies missing authenticity information, or extracts the required information from submission agreements.

A related issue when implementing a scalable preservation environment is the management of a large number of records. This is addressed by creating operations that act on sets of records. In practice, the set of SRB operations that are performed at each storage system or database depends upon the level of aggregation that is imposed on each of the logical name spaces used to identify records, archivists, and storage systems:

- Operations on the user logical name space {single user, user group, data grid federation}
- Operations on the storage resource logical name space {single storage system, compound system with a data cache, cluster}
- Operations on the file logical name space {single file, container aggregating multiple files, hierarchical directory of files}
- Operations on user-defined metadata {single attribute, hierarchical table, collection}

These multiple levels of aggregation are required for scalability of the preservation environment. A similar set of aggregation levels is being implemented in the iRODS system through NSF funding for the development of generic rule-based data management systems.

A final design criterion for a preservation environment is support for the evolution of the preservation environment itself. This implies that the preservation management policies and processes can evolve to handle new types of records, new types of provenance metadata, and new legal requirements. To allow the rules and micro-services to evolve, we support three additional logical names spaces in the iRODS environment that identify micro-services, rules, and persistent state information. We also add similar levels of aggregation within each name space:

- Operations on the micro-service logical name space {atomic, deferred, periodic}
- Operations on the rule logical name space {single micro-service, set of micro-services, recursive rule hierarchy}
- Operations on the persistent state information logical name space {single attribute, audit trails on events, parsing of audit trails for compliance with assessment criteria}

These three name spaces enable management virtualization. One can add new management policies, new preservation capabilities, and new persistent state information without affecting the ability to execute previous management policies and processes. With these three additional logical name spaces, the preservation environment can also evolve. Also, the ability to characterize management policies as rules implies that management policies can be automated, minimizing the labor requirements needed for large collections.

By defining rules that automate the validation of assessment criteria, a preservation environment can be defined that validates its own trustworthiness. It also becomes possible to migrate the records, rules, micro-services, and persistent state information to another independent preservation environment. The management policies that were being applied in the first preservation environment can continue to be applied in the second preservation environment. This means that assertions about authenticity and integrity can continue to be validated as records are moved between different implementations of preservation environments.

Assessment Criteria

The preservation community is developing assessment criteria for validating the trustworthiness of a digital repository. A system that is able to validate the assessment criteria can be considered trustworthy, and thus would be a reasonable environment for the preservation of data for the long term. An initial set of assessment criteria has been proposed by the Research Libraries Group and the National Archives and Records Administration (RLG & NARA, 2007). Analyses of the assessment criteria and mappings of the criteria to management policies have been published (Moore & Smith, 2006; Smith & Moore, 2006). The expectation is that one can define management policies that ensure trustworthiness, define rules that apply the policies, define capabilities that implement the required preservation functions, and define preservation metadata that capture information about the application of the preservation functions. One can then query the preservation metadata to assess whether the assessment criteria have been satisfied. An approach based on assessment criteria attempts to define all of the management policies that are needed to prove that a preservation environment will successfully preserve records.

The assessment criteria are based on traditional preservation principles:

- Authenticity: assertions about the provenance of the records
- Respect des fonds: assertions about the original organization of the records
- Chain of custody: assertions about the ownership of the records
- Integrity: assertions about the management of the records

Each of these preservation principles defines properties that the preservation environment should maintain. At a minimum, the preservation environment needs persistent names for identifying the records, the archivists, and the storage repositories. Assertions about the management of the records can then be based on attributes associated with the persistent name spaces. Examples include the name spaces that are used by the preservation environment to track provenance (descriptive metadata), integrity (rules, preservation processes, system state information), chain of custody (archivist names, storage resource names), and respect des fonds (record names and descriptive metadata). These name spaces need to remain invariant over time, ensuring that an operation performed by a prior archivist can be correctly interpreted. The assessment criteria also require that the functions performed by preservation processes remain consistent over time. This means a preservation environment should be quantified in terms of the actual preservation operations that are performed and the management policies that control the execution of the preservation processes. We thus have two separate contexts that require description; the properties of the records, and the preservation environment policies and processes.

Rule-oriented data systems define the minimal set of preservation processes and management policies on which a preservation environment can be based. Rule-oriented systems also track changes to the preservation environment. We need to know how the preservation processes we are applying today are related to preservation processes that were applied in the past, in order to make assertions about integrity and authenticity.

The characterization of the minimal set of preservation management policies is difficult. We need to map from assessment criteria, to the management policies that enforce the preservation assertions, to the preservation capabilities that implement preservation processes. The iRODS data management system prototype is designed to

implement a provably complete preservation environment. Current development efforts include:

- Implementation of the Electronic Records Archives capabilities as iRODS rules, micro-services, and state information (National Archives and Records Administration [NARA], 2003). A major challenge is the design of the correct level of aggregation of remote storage operations into micro-services. The micro-services need to be simple enough that all preservation capabilities can be implemented from standard micro-services. But if they are too low-level (byte-oriented operations) they become difficult to apply. The current assessment has defined 174 rules and a smaller number of micro-services that need to be implemented. The current iRODS environment provides 73 micro-services. The supposition is that if the correct level of aggregation of the standard operations performed at remote storage systems is defined for each micro-service, a minimal number of micro-services will be created, capable of expressing every required preservation capability. The initial set of iRODS' micro-services is based upon the data-handling operations supported by the Storage Resource Broker data grid.
- Implementation of the RLG/NARA assessment criteria for trusted digital repositories. We have defined 105 management policies that are needed to express the assessment criteria in terms of iRODS rules and micro-services.
- Comparison of preservation metadata-driven approaches with rule-driven approaches. One can define the required preservation metadata, and then identify the set of preservation rules needed to manage the metadata. We have taken the converse approach. We define the required preservation management policies, and then identify the preservation metadata that result from application of the policies. This approach provides a more comprehensive set of preservation metadata about not only the records, but also the preservation environment.

The iRODS rule-oriented data system is available as open source software from a wiki[6]. The first production release of the software was made on January 23, 2008.

Theory of Preservation

Preservation can be thought of as communication with the future. Information that is understood today is transmitted to an unknown system in the future where it will be interpreted and displayed. The future system may have not only different hardware and software, but also different standards for encoding information. Effectively, we need to send into the future not only the information (records), but also a description of the environment that is being used to manage and read those records.

The future preservation system may be linked over time to the original preservation system. To maintain the ability to interpret and display the records, the preservation environment must characterize its own evolution, and the impact that preservation environment evolution has on record management. A theory of preservation makes assertions about the ability to maintain the information context, arrangement, and management of records, as well as the information context

[6] iRODS http://irods.sdsc.edu/

(management policies and preservation procedures) of the preservation environment.

An example of the management of contextual information is defined in the OAIS standard (Consultative Committee for Space Data Systems [CCSDS], 2002). This focuses on the ability to access and interpret records through the creation of representation information. The representation information defines the structures present within a record and their semantic labels. A designated community is defined that maintains the ability to interpret the semantic labels. However, the OAIS standard does not provide representation information about the preservation environment itself.

The concept of infrastructure independence quantifies assertions about the preservation environment, including not only the name spaces, but also the preservation processes, and the preservation policies. By demonstrating that the preservation environment controls the information context needed to preserve the ability to apply preservation procedures, we can create a theory of preservation in which the information content of the records and the information context of the preservation environment are communicated into the future.

Since no system can be completely self-describing, the theory of preservation needs to define the minimal set of assumptions on which preservation environments are based, and then show how these assumptions are conserved as the preservation environment evolves. To do this, we need a few more concepts for a theory of preservation. We need a fundamental unit of "function" on which preservation processes can be based. The candidates for a unit of preservation function are the minimal set of management policies and the minimal set of micro-services required to express and implement required preservation capabilities. We want to characterize the impact of applying a preservation process as a change of state information, and a transformation function that is applied to the record. For viable preservation processes, we need reversible transformations: the ability to transform back to the original record. This involves characterizing records as follows:

- Every record is a sequence of bits
- Information content is described by defining the structures present in the bit sequence, and then naming the structures. The structure names represent the semantic terms used to define the meaning of the record.
- Knowledge content is defined as relationships between and on the structures. Examples include:
    - Logical relationships. The semantic term can be mapped into an ontology, and reasoning done on inferred attributes (semantic grid).
    - Temporal relationships. The structure may represent a time stamp that may be used to enforce causality.
    - Spatial relationships. The structure may be mapped to a coordinate system that can be mapped in turn to a geometry and displayed in a GIS system.
    - Procedural relationships. The structure may represent the outcome of a process in a workflow such as the application of a preservation process.
    - Functional relationships. The structure may require the application of a transformation algorithm for creating derived data products.

These characterizations of records support the concept of persistent objects

(Moore, 2003). It is possible to create a persistent object that can be displayed in the future using technology as yet unknown, even though the object's internal structures and relationships are based on present-day technology. To enable display of persistent objects, we need one more concept, namely that future manipulations of records can be expressed in terms of the manipulation of the structures and relationships that have been described for each record. We base the ability to display records through the following levels of indirection:

- Characterize the standard structures and relationships present within the record
- Characterize the standard operations that can be applied on each type of relationship for each type of structure.
- Characterize the manipulations performed by a display application in terms of standard operations on standard relationships. In effect, the display application does not manipulate records. Instead it executes standard operations on standard relationships on standard structures. One can map from the actions of the display application to the standard operations. Given this mapping, any display application can manipulate any record, or at least the structures within the record for which the required relationships are defined. The software that implements the standard operations can be turned into a micro-service that is managed by the preservation environment.

The preservation process then consists of the manipulation of structures in records, or the assignment of properties to sets of records, or the establishment of relationships between records. If we have a defined set of fundamental reversible preservation processes, we can assert that any future preservation environment can transform all records back to their original form. The future preservation environment can correctly interpret the preservation information context from the past and apply the same preservation policies. The concept of "persistent objects" effectively combines the concepts of migration (transformation of the format of a record to a new standard), and emulation (migrate the capabilities of the original creation application onto new storage technology).

A theory of preservation needs to demonstrate internal consistency between the assessment criteria, the rules that control execution of preservation processes, and the persistent state information generated by application of the preservation processes. The system is internally consistent if all preservation attributes needed to quantify the preservation principles (authenticity, integrity, chain of custody, respect des fonds) can be generated or validated through the application of management processes; and if the persistent state information generated by the application of management policies are retained as preservation attributes that can be queried to verify preservation properties. We cannot have a situation in which a preservation attribute that is needed for assessing preservation principles cannot be controlled or verified by one of the preservation rules. Nor can we have a situation in which persistent state information generated by application of the rules is not included in the representation of the preservation environment context that is migrated forward into the future.

The acid test of a preservation environment is whether it describes the entire preservation information context sufficiently well that the records can be migrated into an independent preservation environment without loss of authenticity or integrity. This requires migrating not only the records, but also the characterization of the

preservation environment context. The new preservation environment would have to apply the same management policies, the same preservation processes, use the same logical name spaces, and manage the same persistent state information. If all of these context components can be expressed and migrated to a new preservation environment, then the preservation context is correctly described.

The expectation is that we can develop a theory of preservation. Its components are:

Characterization:

- Definition of the persistent name spaces
- Definition of the standard operations that are performed upon the persistent name spaces
- Characterization of the changes to the persistent state information (associated with each persistent name space) that occur for each standard operation
- Characterization of the transformations that are made to the records on each standard operation

Completeness:

- Demonstration that the set of preservation processes is complete, enabling the decomposition of every preservation process onto sets of micro-services that execute standard operations
- Demonstration that the preservation management policies are complete, enabling the control of all preservation processes
- Demonstration that the persistent state information is complete, enabling the validation of all assessment criteria

Assertion:

- If the operations are reversible, then a future preservation environment can recreate a record in its original form, maintain authenticity and integrity, support access, and display the record
- A corollary is that such a system would allow records to be migrated between independent implementations of preservation environments, while maintaining authenticity and integrity

The iRODS rule-based environment is a first step towards the creation of a trustworthy digital preservation repository. The definition of standard rules, standard micro-services, standard operations, and standard persistent state information provided by iRODS can lead to a theory of digital preservation.

Finally, we observe that digital libraries (Moore, Rajasekar, & Wan, 2005) and shared collections can be implemented from the same capabilities that are used to support a preservation environment, provided their management policies are enforced. The iRODS data grid is a generic data management infrastructure that can be tuned to support data preservation, data publication, data sharing, or data analysis through specification of appropriate data management policies.

Acknowledgements

This work was supported by the National Archives and Records Administration under NSF cooperative agreement 0523307 through a supplement to SCI 0438741, "Cyberinfrastructure: From Vision to Reality" and by the National Science Foundation grant ITR 0427196, "Constraint-based Knowledge Systems for Grids, Digital Libraries, and Persistent Archive".

References

Baru, C., Moore, R., Rajasekar, A., & Wan, M. (1998). The SDSC storage resource broker. In Proceedings of *CASCON'98 Conference, Toronto, Canada*.

Consultative Committee for Space Data Systems (CCSDS). (2002). Reference model for an open archival information system (OAIS). Retrieved June 10, 2008, from http://public.ccsds.org/publications/RefModel.aspx

Moore, R. (2003). The San Diego Project: Persistent objects. *Archivi & Computer, Automazione e Beni Culturali*. l'Archivio Storico Comunale di San Miniato: Pisa, Italy.

Moore, R. (2006). Building preservation environments with data grid technology. *American Archivist 69,* 139-158.

Moore, R., Rajasekar, A., & Wan, M. (2005). Data grids, digital libraries and persistent archives: An integrated approach to publishing, sharing and archiving data. Special Issue of the Proceedings of *the IEEE on Grid Computing, vol. 93, no. 3*, pp. 578-588. IEEE Computer Society: Piscataway, New Jersey.

Moore, R., & Smith, M. (2006). Assessment of RLG trusted digital repository requirements. In *Joint Conference on Digital Libraries workshop on Digital Curation & Trusted Repositories*: *Seeking Success*. Chapel Hill, North Carolina.

Moore, R., Wan, M., & Rajasekar, A. (2005). Storage resource broker: Generic software infrastructure for managing globally distributed data. In Proceedings of *IEEE Conference on Globally Distributed Data*, pp. 65-69. IEEE Computer Society: Piscataway, New Jersey.

National Archives and Records Administration. (2003). Electronic records archive capabilities list. Retrieved June 10, 2008, from http://www.archives.gov/era/pdf/requirements-amend0001.pdf

Rajasekar, A., Wan, M., Moore, R., & Schroeder, W. (2006). A prototype rule-based distributed data management system. In *High Performance Distributed Computing workshop on Next Generation Distributed Data Management*. Paris, France.

Research Libraries Group & National Archives and Records Administration. (2007). Trustworthy repositories audit & certification: criteria and checklist. Retrieved June 10, 2008, from http://wiki.digitalrepositoryauditandcertification.org/pub/Main/ReferenceInputDocuments/trac.pdf

Smith, M., & Moore, R. (2006). Digital archive policies and trusted digital repositories. In Proceedings of *The 2nd International Digital Curation Conference:Digital Data Curation in Practice*. Glasgow, UK.