Washington State

# Digital Archives Project Feasibility Study

August 2003

Washington Secretary of State
SAM REED

# TABLE OF CONTENTS

# 1. Executive Summary

**Purpose**

The Office of the Secretary of State, Division of Archives and Records Management, is mandated by statute to insure the proper management and safeguarding of public records and facilitating citizen and government access to those records. This mandate encompasses a wide range of responsibilities including centralizing the archives of the state of Washington, developing retention schedules and insuring the maintenance and security of all state public records regardless of format.

Technology and the "electronic revolution" are having a substantial impact on the way governments conduct business and present challenges for capturing, preserving, managing, storing and making accessible electronic records. Significant amounts of critical electronic data have already been lost. The primary purpose of the Digital Archives is to preserve and provide access to records of enduring legal and historical significance. As government records are increasingly generated and stored in computer-based information systems, the state faces the challenge of managing and preserving these digital documents. Many are critical to the survival of Washington's history and culture, captured in the day-to-day business of government.

**Project History**

The agency began strategic planning for the Digital Archives in March of 2000, when the project first appeared in the agency's Information Technology Portfolio. Planning for the physical design and technical infrastructure of the facility occurred during CY 2002. The state's 2001-2003 Capital Budget (SSB 6155) authorized the Secretary of State to enter into a financing contract for the construction. Early site work for the facility in Cheney, Washington began July 2002 with construction beginning January 2003. This facility will serve as the physical "hub" for the Digital Archives and is scheduled for completion in May 2004. This two-story facility will house both the Eastern Washington Regional Archives (traditional paper archives) as well as the Digital Archives serving both state and local government agencies.

Concurrently, research began on the programmatic and technological aspects of the Digital Archives. Site visits by project team members were made to the National Archives and the Library of Congress. A strategic plan was developed that included extensive involvement of staff, executive management and external stakeholders. The State's 2001-2003 Capital Budget also authorized some financing authority for the Office of the Secretary of State for the purchase of technology equipment and software for the Digital Archives. That authority was contingent on completion of a feasibility study for the project's technology and subsequent approval by the Information Services Board.

As part of the feasibility analysis, alternatives to development of a centralized Digital Archives were discussed and subsequently rejected. No other alternative would meet existing legal mandates, adequately protect against the current loss of electronic archival records, or be as cost effective. The decisions to be made were more a matter of scale – how much technological capacity was required and at what point in time would that capacity be needed. In March 2003, GlassHouse Technologies, Inc., a vendor-neutral technology firm with specific expertise in mass

storage architecture, was hired to assist the agency with assessing the technical feasibility of the project, proof-of-concept testing, determining system requirements, designing the system architecture and working with the agency to develop cost estimates. That work was completed in June 2003.

## The Solution

The project was determined to be both exciting and technologically very feasible. The preferred solution will be one that integrates practical work in electronic records management with network-based system development methodologies and business improvement practices.

It will result in tools and models that create:

- Backup and security of essential legal and historical data.
- Seamless e-workflow processes recognizing that client agencies aren't archivists.
- Easily searchable, accessible, viewable, and printable data.
- Legal compliance for retention of e-records.
- Reduced cost over handling paper.
- The ability to assist geographically dispersed agencies to better serve their customers

The Digital Archives technology will require a mass storage environment that can provide ease of scalability as the demand on the Digital Archives grows, rapid access to the information to serve the customer needs, and a stable, redundant environment to ensure maximum "up time" of the system. The system architecture will be comprised of:

- LAN Server Platforms
- Network switches and on-line disk storage
- Near-line tape drives, tape library and associated backup software
- Index and file databases to store electronic archival records

Web servers will deliver electronic archival records to customers, providing remote access. A content management system will provide intelligent, context based searching capability of metadata and file attributes.

The technology solutions identified are all off-the-shelf components requiring little customization and represent proven stable technologies, mitigating the project risk. Models exist that utilize SAN technology for very large-scale data archiving, web spidering and email archiving. Management and records search capability will be provided by a proven content management solution. A combination of service bureaus and in-house solutions will be utilized for legacy hardware and software conversion.

## Project Implementation

The Digital Archives technology implementation is planned in four staged phases, each phase building on the knowledge and experience of the previous one, while increasing the robustness and capacity of the system architecture over time:

- **Phase I**: Initial rollout of the SAN architecture and content management system at the Digital Archives. One records series at the local government level will be identified and accessioned into the Digital Archives from all 39 counties. Additionally, the Office of

the Secretary of State will serve as the pilot agency to test the remote accession capability of the content management system.

- **Phase II**: Four to six state and local government agencies of varying technological abilities and presence will enter into a Memo of Understanding with the Digital Archives. By working with these partner agencies, the Digital Archives will test the various technologies and accession methods (content management, FTP, manual entry). Technical assistance and training for these initial partner agencies will be provided.
- **Phase III**: Additional state agencies will be targeted which generate records of known archival value (such as the Governor's Office, Legislature, etc.). Again, agencies will be asked to enter into a Memo of Understanding with the Digital Archives; technical assistance and training for these agencies will be provided.
- **Phase IV**: All other remaining state and local government agencies will be able to sign a Memo of Understanding with the Digital Archives and begin transmitting data. This phase will also continue to focus on capturing and archiving email from policy makers in all the major agencies at the state level.

## Ensuring Success

A Quality Assurance Plan is also required for this project. Several strategies have, and will continue to be employed to ensure project success. These include, but are not limited to:

- Use of an external Quality Assurance contractor
- Expert outside consulting services with industry expertise, as needed
- Strong executive sponsorship and involvement
- Regular weekly meetings of the project team to review progress, establish milestones and deadlines
- Phased implementation, incrementally developing and expanding the system rather than trying to do too much initially
- Targeted training for agency staff to increase knowledge and skill levels
- Hiring new staff with required skill levels
- Use of competitive procurement processes, where identified, to obtain the required functionality at competitive prices
- Careful contract negotiation and development to ensure that requirements are met
- Continued involvement by the agency's Department of Information Services OITO liaison and conformity with state policies and standards
- Revision to the Washington Administrative Code (WAC) addressing the archiving of electronic records using a process that involves all key stakeholder groups.

# 2. Background and Needs Assessment

## 2.1 Business Objectives

The Secretary of State's Division of Archives and Records Management is mandated by statute to insure the proper management and safeguarding of public records and facilitating citizen and government accessibility (RCW 40.14.020). This mandate encompasses a wide range of responsibilities including centralizing the archives of the state of Washington, developing retention schedules and insuring the maintenance and security of all state public records regardless of format.

In June 2001, the Office of the Secretary of State completed a strategic planning process culminating in the agency's published 2001-2007 Strategic Plan. That plan recognizes the need to establish a digital archives program for storage of electronic records in order to prevent losses to the state's documentary heritage.

Retention of electronic records has reached crisis proportions. In an attempt to comply with general obligations regarding digital records, many agencies are either retaining almost every electronic file or they are deleting records from servers without sorting short from long term items. The collective record of state and local government is at risk of significant loss.

Technology is one solution to improving government business processes. Government agencies are applying new technologies to conduct an increasing amount of their business electronically over networks. Older records captured through proprietary systems or on platforms no longer supported, the rapid evolution of technology solutions both hardware and software, and the volume of electronic data that is increasing exponentially are some of the challenges facing government agencies.

The technological challenges are having a substantial impact on agencies' ability to create, manage, and use electronic records to support their legal responsibilities and business needs.

The specific objectives for the Digital Archives are solutions that provide:

- Simple, reliable, persistent methods to capture, identify, index, store and retrieve digital records for their statutory retention periods or permanently in the case of archival material.
- Cost-effective means to retain and maintain, through migration processes, the readability and accessibility of the historical record of government in the state.
- Public access to the collection(s) so that citizens, including students, have the ability to search and retrieve information and historical objects, such as photos and maps, to explain the role of government in Washington State, optimally via remote access.

## 2.2 Business Environment

Challenges in Managing and Preserving Electronic Records

Electronic records, unlike paper records, are susceptible to undetectable changes in content and format unless they are held securely and under defined and auditable procedures. This is essential if electronic records are to be acceptable as evidence in legal proceedings. These procedures must ensure that electronic records are an authentic and accurate representation of the transaction and have been kept safe from alteration.

The fragile nature of the electronic medium, and the dynamic way in which information technology is deployed, threaten the reliability and authenticity of the record if appropriate information management disciplines are not applied.

Perhaps the greatest challenge to electronic record keeping is the evolution of technology itself. New hardware and software are replacing the products and methods used to record, store and retrieve digital information on cycles of two to five years. No system is currently capable of more than thirty years of retention and access.

Content migration presents it own set of challenges to preserve the integrity of the original electronic record. Migration includes refreshing as a means of digital preservation. However, it is not always possible to make an exact digital copy or replica of a database or other information object (due to hardware and software changes) and still maintain the compatibility of the digital data with the new generation of technology. Some content, functionality, or structure may be lost during migration. Dependence on migration as a strategy for long term retention of electronic records is a form of risk management. The organization must weigh the benefits of maintaining a record's full information content against the cost of migration. Many smaller state agencies do not have the financial resources to maintain a file's readability or function. The great danger is that the agency will commit to electronic records, only to discover that they may lose the information eventually. Clearly a more certain and enduring solution is needed.

Finally, in many cases the only metadata for a large and complex record collection exists as a poor fifth-generation paper copy of a printout, or worse only as a set of handwritten notes. The focus of past custodians was often on data preservation rather than record preservation. As a result we may have all the numbers, but little indication of what those numbers are trying to tell us or why.

One of the primary challenges for the Digital Archives is the sheer number of agencies within state and local government that will be submitting data in a wide variety of formats. Currently the archives has retention schedules for over 30,000 divisions of state agencies, and there are over 2,000 local government agencies. Tasks such as managing permissions become more complex when driven by this type of volume. There are no current comprehensive models for this type of

archiving which coordinates numerous agencies and data types. There are, however, models that are applicable for most parts of our required solution. For example, there are examples of very large-scale data archiving. Models also exist for web spidering and e-mail archiving.

In addition to the technical complexity of managing the electronic data, there is further complexity in addressing the business needs of the stakeholder groups. Electronic data requires a new way of looking at archiving and records retention policies. The volume of electronic data is staggering, and precludes the ability to visually inspect the data first to determine its archival value -- as has traditionally be done with paper records. In addition, many agencies want to keep and display their own data because it is a source of revenue resulting in a reluctance to transmit the records to the archives, or to allow the archives to web enable access to the records. These issues will need to be addressed agency by agency, with interagency agreements developed between them and the Division of Archives and Records Management.

Project History

Records are specific pieces of information produced or received in the creation, conduct or completion of an institutional or individual activity and set aside as evidence of that activity. The use of electronic information technologies is transforming the conduct of government business. The down side of this electronic revolution is the staggering accumulation of electronic and paper records that chronicle government decisions and actions.

In March of 2000, a strategy was developed by the Secretary of State's Division of Archives and Records Management, to deal with the identification of technology, policy, and management factors that would ensure that electronic records are created, maintained, and stored in a manner that facilitates access for agency operations, secondary uses, evidentiary needs, and archival purposes. That strategy included a vision of a central repository and it's incorporation into the Eastern Regional Archives facility currently being constructed in Cheney, WA.

The project was included the 2001-2011 Capital Plan and the passage of HB 1926 created a surcharge of $1-per-page, collected by county auditors on recorded documents, to help pay for the facility and on-going operations.

Strategic planning sessions were begun to address the myriad of issues involved with electronic records created in over 3,000 local and state government agencies with varying retention requirements.

A trip to the National Archives, in College Park, Maryland and the Library of Congress in Washington, DC in May of 2002 was undertaken by the Digital Archives project team. Valuable information was gathered and a strategic plan was developed that included extensive involvement of staff, executive

management and external consulting to begin the process of creating the first digital archives incorporating state and local government records.

Planning the physical design and technical infrastructure of the facility occurred during CY 2002. The services of Sparling, a technology consulting firm, were employed to help design the technology specific spaces.

Digital Archives Technologies/Best Practices

The Washington State Digital Archives, operated under statutory authority of the Washington State Division of Archives and Records Management, will be a central repository for electronic "essential" records and archival/historical records, which require long-term retention. Typical holdings would include e-mail folders, directories, databases, documents and web pages.

Principal functions are:

- Retention and preservation of electronic records: Uses state-of-the-art technology and special environmental controls.
- Centralized Access: Assures citizens convenient access to public electronic records. Uses Internet access and dynamic electronic index.
- Migration of Records: Uses generic platforms and applications to migrate entire collections.
- Security: Stringent procedures: use of digital signatures and keys.
- Records integrity and audit controls: Use of digital signatures, metadata, and audit trails to guarantee integrity.
- Disaster Recovery Assistance: Secure storage of "essential" records to assist with recovery efforts.
- Research & Teaching Aid: A multi-media classroom/conference room offering opportunities for high technology students, teachers and industry experts to study storage area networking, mass storage architectures, open indexing standards, and advanced access and retrieval systems. Ultimately this space will be equipped to offer remote learning opportunities.
- E-Commerce Transactions: Long-term storage of selected items with historical or enduring value.

Best Practices will include the following:

- Proper up-front planning of design of file plan and hierarchy.
- Careful planning of metadata requirements and flexibility.
- Establishing and implementing clear record-keeping rules and policies.
- Participation from business drivers, executives, IT and the user community.
- Training in relevant technology areas.
- Minimizing the amount of involvement to capture a record in to the system – if there are a lot of steps or process, user acceptance will degrade.

## 2.3 Business Need

Purpose of a Washington State Digital Archives

Significant amounts of critical electronic data have already been lost, including:

- Governor Gardner's administrative files – originally on a Wang system which was erased when Gardner left office.
- Governor Spellman's reports and letters on "Mag-Cards" and obsolete floppies.
- Reports, correspondence, etc. "preserved" on floppies of various formats (some as large as 12") and most using obsolete software and operating systems (i.e. CP/M, EBSDC, etc.) from most state agencies.
- Many state agency data bases "preserved" on 9-trak tape, created with obsolete hardware and software.
- Electronic images "preserved" on old hard-drives and other old storage systems.
- E-mail from all state agencies - either they have been deleted or everything has been retained at such high volumes that individual records are virtually impossible to locate.
- State agencies such as the Department of Ecology and the State Parks and Recreation Commission have GIS systems that have been updated on a regular basis, with no "snapshots" saved from earlier days. Failure to retain earlier versions of dynamic data results in loss of historical information such as trends or development transitions.
- Voter Registration database and other similar databases which have been updated through the deletion of data as it becomes obsolete and the addition of new data, without preserving "snapshots" along the way.

The purpose of the Digital Archives is to preserve and provide access to records such as those mentioned above – records of enduring legal and historical significance. The Digital Archives will provide a controlled environment with temperature and humidity controls and other facility features to maximize retention of information in electronic format. The facility will provide a central repository for these electronic records.

Public access to collections will be enhanced through indexes and/or cataloging of records. State and local researchers and the general public will benefit from a centrally accessible point to search for government records. Easy-to-use Internet access will encourage the public to use the Digital Archives. All records retained will carry metadata adequate to establish any restriction from public disclosure. The principal product will be a dynamic electronic inventory of related information for use by government offices and citizens. Information will be available seven days a week, twenty-four hours a day.

Market Analysis

Many state and local agencies including towns and districts do not have the resources to retain electronic records for their full retention period. The state is

directed by statute to operate records centers for inactive and essential state records and essential local records. Without such support for electronic records, the record of government activities in the state will be incomplete and significantly jeopardized. Some examples of essential and/or archival electronic records include:

- Land records – the majority of County Auditors have electronic databases for land records including digital images of the documents. Paper documents are not retained by the Auditors offices. Pilot projects have been conducted for e-transactions – soon many property transactions will occur with no paper document.
- Court records – County Clerks have also moved ahead with digital technology and the majority of counties have electronic databases. As with land records, there is a strong move toward electronic transactions.
- Maps – The Department of Natural Resources maintains surveyed documents in digital format as do most County Auditors. In addition, there are a myriad of departments that maintain GIS systems and the Geology Division maintains digital images of geological maps and reports.
- Vital Records – The Department of Health maintains a database of vital records including birth, death and marriage records. In addition, most counties have moved toward imaging marriage records.
- Retirement Systems – The Department of Retirement Systems is in the middle of a major scanning project to image all retirement documents.
- Corrections – The Department of Corrections maintains a database of all inmates.
- Codes, Ordinances, Statutes – These are maintained in electronic format by numerous state and local agencies.
- Public Disclosure Commission – The PDC maintains records of political campaign and candidate disclosure reports. Reports can be submitted by candidates electronically.
- Legislature – Both the State House of Representatives and the State Senate maintain digital images of members.

While electronic record-keeping has greatly improved and eased our ability to handle large quantities of electronic records, making many formerly tedious office tasks very convenient, this capability is a mixed blessing. The rapid accumulation of state and local government records and information in all formats, from paper to microfilm and electronic media has dramatically accelerated the need to manage this “information behemoth”. Government offices produce ever-increasing volumes of electronic records in the form of email, e-documents, e-forms, e-transactions, databases, web pages and other records. Web usage is increasingly sophisticated and e-mail is utilized extensively. This information is fragile and perishable. For long-term retention, an authentic copy should be obtained as soon as the legal purpose or business need is met (for example, as soon as an e-mail has been opened and read) and then transmitted to an archival facility.

Nationally, the typical office is creating and needing to store and manage double the volume of information every three years. Without the application of records management principles and practices, offices are failing to reap the rewards of an effectively managed operation. Records are lost through misfiles. Saving (archiving) all electronic records as if they had the same value and retention makes retrieval chaotic and inefficient. Filing of electronic records is seldom systematic or coordinated with legal retention requirements. Storage of obsolete records means that excessive storage space is used on hard drives and other devices that could be freed for current needs driving up costs of the system. Identifying "essential" records that would be crucial to an office's ability to recover operations in the event of a disaster are limited at best. Identifying archival legal and historical electronic records of permanent and enduring value by the Washington State Division of Archives and Records Management is complicated by the lack of a storage facility to house them.

The most cost effective structure for the Digital Archives will be to share facilities with the Eastern Regional Archives. The Regional Archives system has long term community and legislative support, with three of the five regions housing their collections in their own facilities (Northwest Branch Archives in Bellingham at Western Washington University, Puget Sound Branch Archives at Bellevue Community College, and Central Branch at Central Washington University).

## 2.4 Statutory and Legal Requirements

There are several statutory and legal requirements that provide significant context for the Digital Archives:

- State laws outlining the statutory responsibilities and authorities of the Archives and Records Management Division for the preservation and accessibility of public records
- State laws regarding electronic access to public records
- The central role of records in litigation and discovery
- State laws related to public disclosure

Public Records Retention and Archiving

RCW 40.14 defines public records and assigns responsibilities to the Office of the Secretary of State's Archives and Records Management Division for the preservation, protection and access of public records. Assigned responsibilities include the power and authority to operate a central state archives, establish record retention and transfer schedules for all government agencies, establish safeguards against unauthorized removal or destruction of records, establish and operate state records centers for the purpose of preserving and protecting state public records that must be preserved either temporarily or permanently.

State agencies are required to retain records until one of two events occurs. Either the record reaches the end of an established retention period as defined in the state and local government records retention schedules, or the State Archivist has

determined the record to be historically valuable and permanently accessions it into the central archives. Records should be retained, regardless of media, for the period required by the agency's records retention program for any legal, user, historical or other purpose.

The Archives and Records Management Division currently maintains three types of facilities for the storage of records:

(1) The Division operates a state records center for the storage of records that cannot yet be destroyed but which are not currently needed in the agency that created or received them. The records center operates essentially as a warehouse service for other agencies. If the retention period for a particular record has not yet expired (and so it cannot yet be destroyed), but the record is not immediately needed in the agency, it can be stored in the records center. Records stored in the records center remain the property of the originating agency, and remain in that agency's legal custody.

(2) The centralized state archives contains state government records of long term legal and historical significance (40.14.020(2)). Once records are transferred to the central archives, they become the property of the Archives and Records Management Division. Such records are maintained permanently for purposes of reference and scholarship.

(3) Local government records of long term legal and historical significance are held in five regional facilities: Puget Sound (Bellevue), Northwest (Bellingham), Central (Ellensburg), Southwest (Olympia) and Eastern (Cheney). As with the state records, once records are transferred to the archives, they become the property of the Archives and Records Management Division. Such records are maintained permanently for purposes of reference and scholarship.

In addition, the state archives manages a security microfilm program centralized in Olympia that provides secure, environmentally controlled storage for original microfilm of archival records.

As stated above, "public records" include "any…machine readable material, compact disc…or other document, regardless of physical form or characteristics" received by the state of Washington in the course of business. RCW 40.14.020 also authorizes the State Archivist to establish rules to "…facilitate access to, photographic, optical, electronic, or other images used as public records;..." A January 10, 2001 legal analysis by the Attorney General of Washington states: "A reasonable interpretation of this legislative direction would therefore require that the archivist should make electronic records easy to locate or reach in the archives…"

Electronic Access to Public Records

Washington State is one of only about ten states that have enacted statutes (RCW 43.105.270) that require government agencies to consider public access when

planning or purchasing computer systems, and one of only about four states that require remote or online access of public records.[1] Although the emphasis is on the provision of current (not archival) records and information, the statute also allows that archival material should be "made available digitally as resources allow or as a need arises."

Use of Records in Litigation and Discovery

State and local government agencies involved in litigation – an increasing phenomenon – must respond to requests to produce records in the discovery process. The written discovery process involves a series of requests for records that the other side must turn over. Agencies are expected to be able to find the record, retrieve it, and provide the requester with either a copy or the opportunity to view it. Failure to properly provide the requested records could cost substantial amounts of money. It is therefore vital that state agencies not only properly retain their records, but also keep them in an organized fashion.

Records retention requirements make no distinction between formats, requirements are media neutral. E-mail records pose particular challenges for a number of reasons. In recent years, the use of e-mail has proliferated – an estimated forty messages per day per worker are created. The volume alone makes these records difficult to manage. In addition, as a communication tool, e-mail is treated very informally by most users, communications are not private, and there is a loss of control over distribution. Due to these factors, e-mail is frequently a source of liability in cases of fraud, defamation, harassment, etc.

Recent high-profile cases (e.g. the Tobacco Litigation, Enron, Arthur Anderson, Microsoft, the findings of the Committee on Government Reform regarding White House e-mail failures) underscore the importance of record retention and demonstrate the high stakes. As the scope and stakes of litigation keep increasing, so will the demands on government to provide more and more records in litigation discovery. Litigators depend on locating and obtaining records and documents to support their contentions and prove their case. Failure to properly provide requested records could cost the state substantial amounts of money in the form of sanctions. Records that are unorganized and difficult to find can result in time and money spent searching for them. The inability to find records, inconsistent record keeping practices or failure to follow established processes can undermine the government agency's credibility and ultimately lose a legal case, again costing the agency substantial amounts of money. Systematic record keeping can save a lot of time, effort and ultimately money to government agencies.

The increasing use of computers over the last two decades means that there are more records that may be requested and must be produced in forms other than traditional paper-based formats. In addition to requirements to locate and produce electronic

[1] Access Denied, by Charles N. Davis and Sigman L. Splichal, 2000, pgs. 43 and 50

records, government agencies must be able to guarantee that the record is authentic and has not been altered. RCW 40.14 sets out provisions for the State Archives to provide certified copies of records in its custody. That certification states, in part, that the copy "is a full, true and correct copy of the original record in the official custody of the state archivist of the State of Washington." There is a presumption that public records reflect accurate information produced by trustworthy procedures. Further, regardless of the retention period, records must continue to exist when litigation, government investigation or audit is pending, or imminent.

The information maintained on the media and the ability of the system to produce records from the information must achieve the required retention period. This means that for some technologies it may be necessary to periodically convert, regenerate, copy or transfer the information from one medium or technology to another to preserve the information for the required period.

Records used in litigation and discovery must be proven to be authentic with the content in tact. Electronic records, unlike paper records, are susceptible to undetectable changes in content and format unless they are held securely and under defined and auditable procedures. This is essential if electronic records are to be acceptable as evidence in legal proceedings. The agency must be able to show that the process or system used to store and reproduce a public record is trustworthy in terms of producing an accurate result. This requires consistent and reliable policy application, establishing a rules-based retention program, developing functional and performance criteria for the system, and documenting chain of custody for the record.

Public Disclosure Act

Public records play a key role within our overall system of government. Open government, where decisions are made publicly and records are available for full public review, is fundamental to a free and democratic society. If the public is to have the ability to monitor government and be involved in its decision making, access to records and knowledge of decision making are essential. [2] Recognizing the critical importance of public access to government records, the Washington State Legislature initially passed RCW 42.17 in 1972. The Declaration of Policy stated in 42.17.010, subsection (11) states: "That, mindful of the right of individuals to privacy and of the desirability of the efficient administration of government, full access to information concerning the conduct of government on every level must be assured as a fundamental and necessary precondition to the sound governance of a free society." The chapter goes on to say: "The provisions of this chapter shall be liberally construed to promote complete disclosure of all information respecting the financing of political campaigns and lobbying, and the financial affairs of elected officials and candidates, *and full access to public records so as to assure continuing public confidence of fairness of elections and governmental processes, and so as to assure that the public interest will be fully protected."* (italics added)

[2] Citizen's Guide to Open Government in Washington State, May 1999

Under this act, members of the public have broad rights to review state and local government agency records and copy them if desired. Agencies are specifically required to make public records available for public inspection and copying, unless specifically exempted (42.17.260). Washington's law has an expansive definition of public records, defined as "any writing which contains information relating to the conduct of government or the performance of any governmental or proprietary function." In addition, the law defines "writing" as virtually all means of recording any form of communication regardless of physical form or characteristics. Records, therefore, can include not only documents and papers, but maps, photographs, films, sound recordings, electronic records and information on compact disks, manuals, budgets, etc. Records are increasingly maintained on electronic systems; many are never converted to print. This use of technology in the past twenty years is posing unique challenges in the retention of and access to public records.

Digital WAC

Additions to the Washington Administrative Code (WAC) are needed to address the unique elements associated with management of electronic records. Following the publication of federal Department of Defense regulations in 1995 related to federal electronic records, and the State of Florida's subsequent adoption of those same standards, the State of Washington made an initial attempt to adopt similar rules in 2001. Although the draft WAC went through three revisions and included stakeholder involvement, the new WAC was never adopted. In retrospect, it is clear that some key factors contributed to the lack of success:

- Insufficient communication with stakeholders regarding the policy basis for the new WAC
- Failure to address key stakeholder comments in subsequent revisions to the original draft; stakeholders felt they had not been heard.
- The makeup of the stakeholder group was inadequately balanced between state agencies, local agencies and higher education. In addition, the state agency group representation was too heavily focused on information technology managers rather than at the policy level.
- The Washington State draft WAC was not sufficiently modified to reflect technological changes that had occurred since the 1995 Department of Defense rules were adopted.
- Stakeholder meetings and communications were sometimes disorganized and did not provide participants with adequate notice of upcoming meetings.

A renewed effort to develop and adopt WAC rules will begin in 2003 and be completed by 2004, establishing a process that corrects the key deficiencies that led to the earlier failure.

## 2.5 Business Service Goals and Opportunities

Service Goals

A growing public expectation combined with state and local government agencies electronic records explosion is a driving force for Washington State to improve the way it conducts business and provides services to citizens.

The Washington State Digital Archives will provide a proper facility for long term storage of electronic records created by state and local agencies including towns and districts. The Washington State Digital Archives will supply a dramatic means to enhance community links between government, business and education.

The repository will provide relief to state and local jurisdictions. Access, using search engines, will bring a wealth of government information together and orders it for more meaningful presentation and study. Centralization will provide an important degree of control and standardization to records of historic and enduring value.

Business Opportunities

The opportunity exists for the facility to provide significant research, training and teaching opportunities relating to information technology, preservation of digital records, and other electronic records management issues and policies. The archive will be a "working laboratory" where pure research, applied research and practical technical training may be pursued in the discipline of long term mass storage and retrieval of archival data.

Co-location on the campus of Eastern Washington University will make available technical exposure for the host academic institution and it is expected that the facility and its functions will draw worldwide attention, just as the San Diego Supercomputer Center did for the recent National Archives and Records Administration project.

In addition, business opportunities include:

- Cost efficiencies
- Less travel as records would be transferred over high speed data lines
- Backup/security storage
- Potential for a new revenue stream
- Labor efficiencies
- Digitizing once – capitalizing on IT investments

Relationship to Agency Strategic Plan

The Washington State Digital Archives supports goals of consistency of service and improvement of data collection and information use statewide. The digital project is both a new way to capture and organize information and a catalyst for organizational change. The planning and design phases generated many kinds of

organizational learning. The testing and implementation steps will surely do the same.

The project strategic plan relationship includes the following activities:

- Develop and evaluate generic system-level functional requirements for electronic recordkeeping.
- Review current system of development methods in networked computing environments.
- Review business process improvement and business process reengineering methodologies.
- Develop framework or set of questions to elicit a complementary suite of technology, management, and policy mechanisms to support records management to be used as part of the digital archiving process.
- Evaluate the effectiveness of the prototype in meeting business objectives, including records management requirements.
- Produce a wide ranging method for use in other organizations.

# 3. Project Impacts

## 3.1 Impact on State and Local Government

The Division of Archives and Records Management has the responsibility for providing secure long-term storage and access to archival records of legal and historical value for all agencies within state and local government. As such, all agencies stand to benefit from the establishment of centralized archiving of electronic records.

The project integrates practical and theoretical work in electronic records management with network-based system development methodologies and business improvement practices.

It will result in tools and models that create:

- Backup and security of essential legal and historical data.
- Seamless e-workflow processes recognizing that client agencies aren't archivists.
- Easily searchable, accessible, viewable, and printable data.
- Legal compliance for retention of e-records.
- Reduced cost over handling paper.
- The ability to assist geographically dispersed agencies to better serve their customers.

Client agencies will be asked to enter into a Memorandum of Understanding that identifies which agency electronic records are of archival value and the method of transmitting those records to the digital archives. It is anticipated that some electronic records not currently being saved would be identified in this process.

## 3.2 Impact on Constituent Groups and Other Agency Customers

The archives provides access to records in its custody not only to other government agencies, but to businesses and individuals as well. A number of benefits will accrue to anyone who requires access to the records:

- Centralizing the data means users have one point of contact rather than being required to contact individual agencies to access information.
- Geographically dispersed users will have remote access to data.
- With web technology, users will have access to web-enabled data twenty-four hours a day, seven days a week.
- Data which is not currently being retained, or is not readily searchable will become available to users.

# 4. Major Alternatives Considered

This section presents the major alternatives considered for the Washington State Digital Archives. The alternatives are:

4.1 Pursue a "Business-as-Usual" path
4.2 Require conversion of electronic files to paper
4.3 Require state and local governments to archive their own electronic files and make them accessible
4.4 Acquire an existing Digital Archiving system
4.5 Develop Digital Archiving system – Full system implementation
4.6 Develop Digital Archiving system – Phased system implementation

Information about each alternative is presented in the following format:

- Description
- Major assumptions
- Risk factors
- Cost/benefit summary

## 4.1 Pursue a "Business-as-Usual" Path

Description

This alternative would not entail any major changes in how state and local governments currently preserve and retain their electronic records, despite the fact that paper records are increasingly being replaced by digital records. The Washington State Records Center and Archives Division would continue to work with state agencies -- within currently available resources -- to retain and preserve paper records. State and local government agencies would continue to retain or destroy electronic records in whatever manner that they currently do. Within currently available staffing levels, assistance and consultation would be available to state and local government agencies on the preservation and retention of electronic records.

Major Assumptions

- Some state and local government agencies would preserve and retain important digital records and documents; others would delete or write over them.
- Agencies will be responsible for retention of their own electronic records.
- Some agencies would print out electronic records on paper and continue current paper retention practices.

Risk factors

- Low level of compliance with retention and preservation requirements for public records.
- High level of loss for historical records, possibly increasing as paper records are replaced by digital records.

- Older versions of dynamic data, with historical or archival significance, would continue to be deleted and/or written over.
- Low level of access for the public, legal and research needs due to lack of centralized point of access.
- Higher cost of retrieval of existing records.
- The Digital Archives portion of new facility, currently under construction and scheduled for completion in 2004, would be inactive.

Cost/benefit summary
This alternative has a high cost to benefit ratio.

Costs:
- The state would continue to pay for the Digital Archives facility currently being constructed (which is a portion of the two-story Eastern Washington Regional Archives/Digital Archives facility in Cheney, WA), although no equipment would be purchased for the Digital Archiving function. At the same time, the cost of storing paper records would continue to increase.
- Without a centralized or coordinated way to retain and preserve archival/historical and essential digital records, preservation and retention efforts would be inconsistent at best, and the cost of developing methodologies would be duplicated to some extent throughout state and local governments.

Benefits:
- None.

### 4.2 Convert All Electronic Files to Paper

Description
State and local government agencies would be required to convert electronic records to paper format in order to meet retention and archival requirements for long-term retention and storage. The paper records would then be transmitted to the Records Center and Washington State Archives as is currently done with other paper records.

Major assumptions
- Some state and local government agencies would perform the conversion to paper; others would not. Rather, they would delete or write over them.
- Current WAC rules would require updating to make this a requirement.
- The Division of Archives and Records Management would hire and retain one or two of the Digital Archives staff authorized in the operating budget to provide limited assistance and consultation to state and local government agencies on the preservation and retention of electronic records.

Risk factors

- Increased state and local government agency costs for: staff to select, index and print pertinent records, inventory and prepare transmittals, printing, storage space and supplies to store paper records, and delivery of paper records to the records center.
- Increased difficulty in accessing and searching for needed records. Does not take advantage of easier computer mechanisms for searching for and retrieving information. Records would not be centrally and electronically available via the web; users would need to access paper records in fixed locations, requiring more time on the part of staff and users.
- Inadequate space for storing greater volume of paper records. Facilities are already at maximum capacity.
- Violation of federal law regarding "authentic" records (see Chapter 2.4, Statutory and Legal Requirements).
- The Digital Archives portion of the Eastern Washington Archives/Digital Archives facility, currently under construction and scheduled for completion in 2004, would be inactive.
- Higher cost of retrieval for existing records.

Cost/Benefit summary

This alternative has a high cost to benefit ratio.

Costs:

- The state would continue to pay for the Digital Archives facility currently being constructed (which is a portion of the two-story Eastern Washington Archives/Digital Archives facility in Cheney, WA), although no equipment would be purchased for the Digital Archiving function.
- The need for facilities in which to store paper records would increase exponentially.
- Additional staff time in state and local government agencies to select, print, box and index massive amounts of records.
- Higher cost of retrieval for existing records.

Benefits:

- State and local governments could follow existing and familiar rules and processes

### 4.3 Require State and Local Governments to Archive their Electronic Files and Make Them Accessible

Description

Each state and local government agency would need to develop more sophisticated systems and procedures for selecting, preserving, storing and ensuring accessibility to their own electronic records. Paper records, including electronic records converted to paper form, could continue to be shipped to and retained at the Records Center and the Washington State Archives.

Major Assumptions

- WAC would be updated to require preservation of archival/historic/ essential records by each agency, but would not assume any centralized system, mechanism or support from the Washington State Division of Archives and Records Management. Each agency would need to meet the requirements in whatever way they could.
- Agency would hire Digital Archivist and possibly one information technology specialist to provide limited assistance and consultation to state and local government agencies.

Risk Factors

- Each state and local government agency would need to develop sophisticated systems and procedures for selecting, preserving (including conversion of materials to ensure future accessibility and readability), storing and making accessible (via the web or other means) their own electronic records.
- The result would be inconsistent systems, practices and policies throughout state and local governments.
- Without a centralized system, there would be duplication of effort among state and local government agencies.
- There would be limited and difficult access for the public and researchers without a centralized repository, search capability or web access.
- There would likely be a low level of compliance among state and local government agencies, as this would be seen as another unfunded mandate competing with other agency priorities.

Cost/Benefit Summary

This alternative has a high cost to benefit ratio.

Costs:

- Staff time in each agency.
- Purchase of hardware and software, including that needed to convert from legacy systems, would be required by each agency, with duplication of costs and effort.
- Staff cost for Digital Archivist and information technology FTEs at the Division of Archives and Records Management.

Benefits:

- None

## 4.4 Acquire an Existing Digital Archiving System

Description

This alternative explored the potential of importing and using another state or federal government digital archiving system. Through contacts in various

professional associations including the National Archives and Records Management Association (NARMA), National Archives and Records Management Association (NARA), General Accounting Office (GAO) and others, it was determined that although both the federal and other state governments continue to explore the issue, none have developed a comprehensive system incorporating local and state archival records.

Major Assumptions

- All levels of government face the same issues and have similar needs and requirements for the selection, preservation and retention of electronic records.
- Other governmental agencies are exploring systems and processes for digital archiving.

Risk factors

- Had a functioning and successful system been identified, this option would have minimized the time and cost for development, testing and implementation.
- Existing systems – had they been available – may have required significant modifications in order to meet Washington State's needs, increasing the risk of failure.

Cost/benefit summary

No cost/benefit summary is possible since another system does not exist.

### 4.5 Develop Digital Archiving System – Full system implementation

Description

This alternative assumes full scale start-up based on the architecture plan created by the consultant. Following system development and testing, update of the WAC, and completion of the physical facility and equipment purchase/set-up in Cheney, all local and state government agencies would be notified to begin electronically transmitting their archival, historic and essential electronic records to the new Digital Archives facility in Cheney. Staff at the Digital Archives would then convert records to a platform-neutral format to ensure future readability and access, web-enable those records that are of most interest and value to researchers and the public, and store the records for long-term. A cost flow analysis is provided for this option in Section 9.1.

Major Assumptions

- WAC update in place requiring preservation of archival/historic/essential records by each agency.
- State and local government agencies understand the process and the requirements.
- All necessary legacy hardware for conversion would be in place.

- The Washington State Digital Archives is fully staffed. Staff members are adequately prepared to handle the anticipated volume, and the volume can be accurately estimated.

Risk Factors

- The capture and transmittal of electronic records to the Washington State Digital Archives would occur before adequate preparation and testing.
- Would require more staff, funding and resources than are currently authorized and available.
- Attempting to convert all records that are currently in some electronic/digital format would require a higher volume and wider array of legacy equipment than might be available initially.
- Would likely overwhelm existing staff and systems and create unrealistic expectations from customer agencies. Subsequent failure to meet those expectations would result in a loss of credibility for the program/agency, and possibly diminished future levels of compliance.

Cost/Benefit Summary

This alternative has a relatively high cost to benefit ratio in the development stages of the project.

Costs:

- Digital Archives operating expenses, including costs of full staffing, facility related costs, equipment, hardware/software, etc.
- Customer agency costs and staff time for preparing and transmitting electronic records.
- Wide array of legacy equipment.
- Costly adjustments to the system, policies and procedures if needed due to inadequate testing.

Benefits:

- Brings state and local government agencies into compliance sooner.
- Centralized access to the wider range data by the public/researchers would potentially be available sooner.

### 4.6 Develop Digital Archiving System – Phased system implementation (Preferred Alternative)

Description

The selected approach is development of a digital archiving system with implementation phased in over time. This alternative is also based on the architecture recommended by the consultants, but assumes a slower, phased-in implementation plan. This option is described in Section 5, Proposed Solution.

Initial architecture will begin with minimum requirements for scalable enterprise-level infrastructure expandable as additional agencies are connected to the Digital

Archives. Each of the four phases would be designed to build on the knowledge and experience of the previous phase. A cost flow analysis for this alternative is provided in Section 9.2 and is the preferred alternative.

Major Assumptions

- WAC update in place requiring preservation of archival/historic/essential records by each agency.
- The Washington State Division of Archives and Records Management staff would work with state and local agencies prior to initiating capture and transmittal of electronic documents to ensure understanding of the process, requirements and expectations.
- Adjustments to the system, policies and procedures would be made as needed based on early testing and initial implementation phases, prior to full implementation.
- The Washington State Digital Archives is fully staffed.

Risk Factors

- Extends the time during which state and local government agencies may not be adequately retaining and preserving electronic archival records.
- Only limited data would be available to the public and researchers initially. Centralized access to the wider range data would be delayed, as customer agencies will be brought on in phases.

Cost/Benefit Summary

This alternative has the lowest relative cost to benefit ratio.

Costs:

- The Washington State Digital Archives ongoing operating expenses, including costs of full staffing, facility related costs, equipment purchase/financing payments for hardware/software, etc.
- Customer agency costs and staff time for preparing and transmitting electronic records.

Benefits:

- Orderly process.
- Mitigates cost risk.
- Easier to adapt new technology.
- Initial demonstrated success and meeting more limited initial expectations will increase buy-in from state and local government agencies and boost future compliance.
- Allows for development of models.

# 5. Proposed Solution

The focus of the Digital Archives is to maintain electronic records of historical, legal or fiscal significance for long term accessibility. To that end, the Digital Archives must be able to capture records in a variety of formats and bring them into a centralized repository. In order to make the records as accessible as possible, they must be searchable through a web interface with a consistent presentation. The best method to facilitate the searching of the records will be to "wrap" all electronic records into an XML format, which converts the records into an international standard for marked-up text, and applying metadata 'tags' to the documents. These metadata tags will contain structured information about the source of the information, such as date, author, agency and subject. By searching on the metadata tags, more complex, comprehensive searches can be conducted that will locate related documents which may not have been sent to the archives at the same time.

A graphic overview of the electronic archiving system is shown in Exhibit A on page 31.

Each participating agency at the state or local level generating records of archival nature will enter into a Memo of Understanding (MOU) with the Digital Archives. This MOU will detail the responsibilities and deliverables of both the Digital Archives and the partner agency. Once the appropriate paperwork has been completed and an agreement has been reached, a team of experts from the Digital Archives will work with agency representatives to identify those electronic records series of long-term, archival value that are to be transferred to the Digital Archives.

Once the Digital Archives is operational, the process for accessioning, or accepting records into the Archives repository, will consist of a three tiered approach to address the various levels of technology that is available within participating target agencies:

At the *top tier* (participating agency has high level of technology capability), a content management software package installed at the partner agency will automatically transfer identified records to the Digital Archives content management database upon creation. The partner agency content management software will also add the appropriate metadata tags to the records and wrap them into an XML format prior to transfer.

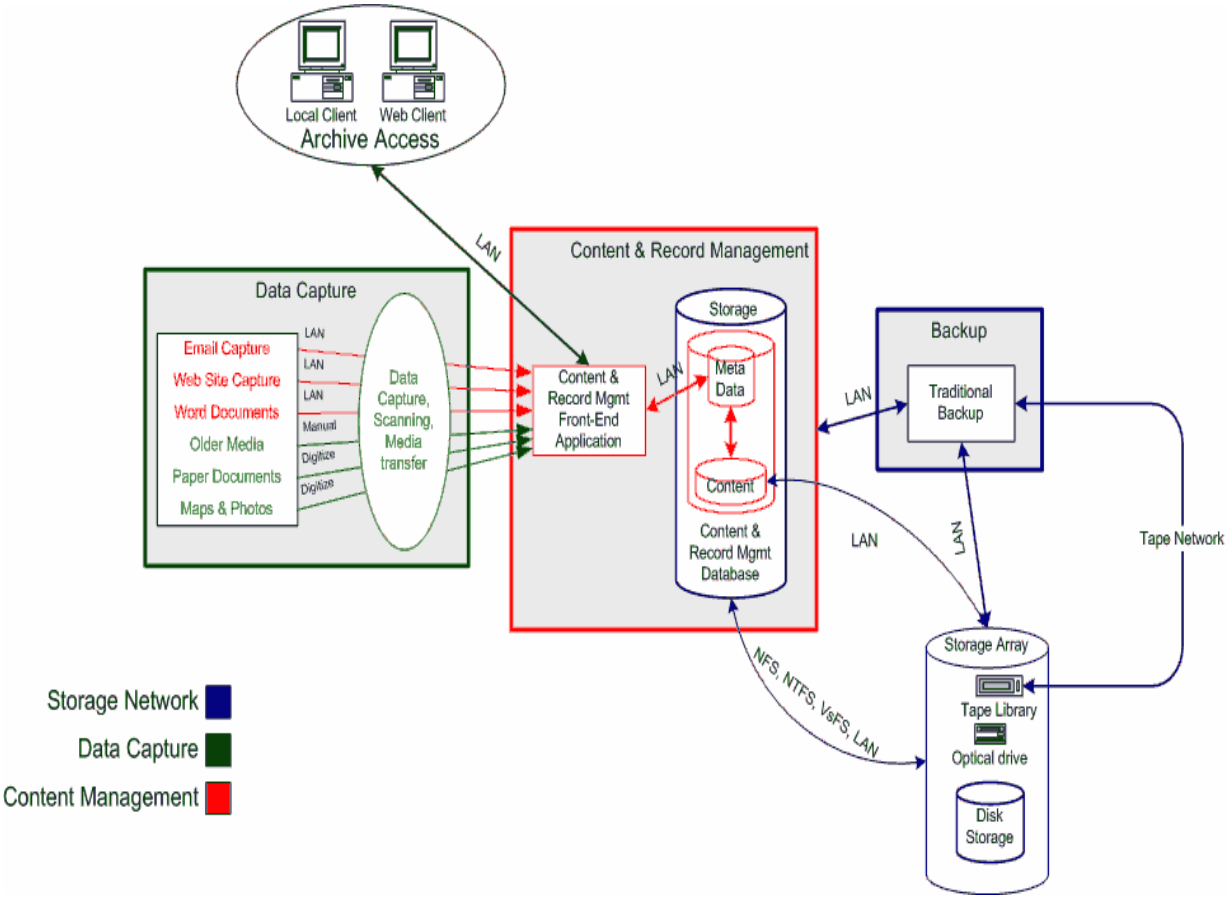
The *middle tier* will consist of designated personnel at the partner agency using an off the shelf applet to create a metadata text file containing the required information about a record series. Another applet will be used to convert the records into an XML format. Then both the metadata and the XML records will be FTP'd into a specific folder on the FTP server at the Digital Archives. The Archives staff will then take the FTP'd records, check for accuracy and import them into their content management database.

The *bottom tier* option (participating agency has little technology capability) will entail designated personnel at the partner agency filling out a text form file to capture the appropriate metadata, transferring the text files and records to a CD, tape or diskette and then sending this media to the Digital Archives via USPS, UPS, campus mail, etc. The Archives staff members

will then check the transfer for accuracy, convert the records to XML with the appropriate metadata tags and import the records into their content management database.

## Exhibit A

*Electronic Archiving System Overview*



## 5.1 System Architecture

Overview

The Digital Archives will require a mass storage environment that can provide ease of scalability as the demands on the Digital Archives grows, rapid access to the information to serve the customer needs, and a stable, redundant environment to ensure maximum "up time" of the system. As shown in Exhibit B, page 39, the network architecture will be comprised of:

- LAN Server Platforms (Either UNIX or Windows).
- Network switches and On-line disk storage.
- Near-line tape drives, tape library and associated backup software.
- Index and file database to store electronic archival records.

As there are several distinctly different technologies available that meet the requirements of the Digital Archives, extensive research into the various options was conducted. The following summarizes these architecture options and is the foundation of the design and purchasing phases of this project.

LAN Server Platform

UNIX or Windows-based servers are the two viable options. UNIX is typically found in server farm environments where large, enterprise scale operations need to be run. Windows has a larger install base, large support group and a much wider availability of compatible software. As the exact performance metrics (i.e. software overhead, user volumes, etc.) will not fully be understood until implementation, the final server selection should focus on flexibility and scalability (i.e. room for growth) to allow the Digital Archives to quickly accommodate any additional processing requirements without changing the server model/platform. The platform, once selected, will become the "standard" for future purchases to maximize economies such as staff skill-set and vendor maintenance. Finally, technologies such as blade servers and clustering could be considered during the later phases of this project, but are not required at initial implementation.

***Proposed solution: A windows-based server will be chosen due to the greater variety of compatible products, large user base, and in-house expertise with Windows systems.***

Network Switches and Appropriate Storage Network Hardware

- *Storage Network Switched Environment*. There are three identified architectures for mass storage: File Servers, Network Attached Storage and Storage Area Networks. File servers have the storage drives attached directly to a server that controls access and requests to the drives. Of the three solutions, it is the least expensive; but it also has the lowest level of performance, scalability and reliability. Network Attached Storage (NAS) attaches a JBOD (just a bunch of discs) to the LAN. This allows the storage

array to be seen by all the servers on the network, but data exchange happens on the same LAN network as the user requests. The addition of the storage traffic to the LAN degrades performance, particularly when backing up data onto tape. Storage Area Networks create a network behind the servers specifically for moving the stored files, either for user requests, tape back up or storage management needs. By containing the storage array to its own network, LAN traffic is not affected, even during large scale tape back ups. While SANs are more expensive initially than file servers or NAS, it maximizes scalability, performance and reliability.

***Proposed solution*: *An enterprise-class storage array connected to a SAN. A RAID-5 implementation offers the most cost-effective data protection schema, and should be implemented on the array. The SAN fabric should be based upon switches with dual-fabric architecture to provide an additional level of redundancy in the environment.***

- *Storage Infrastructure*. For external storage connectivity, there are two primary options for the archives infrastructure: Fibre Channel and IP Storage-based protocols such as iSCSI. Fibre Channel protocol has been an open standard since 1991 which has been demonstrating increased performance and flexibility over the past few years as technology develops. IP storage protocols are still new to the marketplace with a limited installed base and at this time is it not known how this new technology will weather the market.

***Proposed Solution: The SAN network will use the Fibre Channel Protocol. IP storage may be an option during subsequent phases of this project to facilitate specific additional requirements such as long-distance storage replication over a wide area network, depending on industry adoption of iSCSI or like protocol.***

Near-line Storage

- *HSM functionality*. Hierarchical Storage Management is generically defined as a data storage system that automatically moves data between high-cost and low-cost storage media. HSM systems exist because high-speed storage devices, such as storage arrays, are more expensive per byte stored than slower devices, such as magnetic tape drives. Infrequently accessed files can be moved to lower-cost storage as long as the delay in retrieving a file (i.e. when the file is read) is acceptable. Options exist for the type of near-line storage, including but not limited to magnetic media. Optical (DVD) is also an option for near-line HSM storage. The primary benefit of optical near-line storage is access time – sub-second versus tape's best 11-second access time. However, all other features have significant challenges when compared to today's tape technologies; these include overall cost, media capacity, total footprint per terabyte, and throughput (read/write speed). In addition to HSM, several other specialized storage architectures were analyzed– such as cheap ATA (Advanced Technology Attachment) disk and Content Addressed

Storage. These alternative technologies will be revisited when making this decision.

***Proposed Solution: As HSM adds a level of complexity and adds expense to the project it is not recommended at the initial roll-out phase, but will be re-evaluated in subsequent phases as storage requirements continue to grow and data access patterns are better understood. "Cheap Disc" ATA storage will be utilized at the start-up of the Digital Archives. As the initial storage capacity is exceed, HSM –whether utilizing tape or optical drives -- will be evaluated for expansion of the Digital Archives.***

- *Tape Library*. Tape remains a viable technology and continues to evolve in the areas of reliability, capacity, performance, and cost-effectiveness. There are a handful of proven and reliable tape options, all which meet the business backup/restore requirements for both speed and capacity. Hierarchical Storage Management (HSM) and other data-recall tape features (i.e. time to access data located anywhere on the tape) will also be considered in the future. Additionally, the tape format roadmap, library support, "write once" feature-set, and backwards-read compatibility will also factor into the final decision. The Archives will use a tape library as the foundation of the automation behind the backup/restore process, as well as the optional HSM functionality in the future. The use of robotic tape libraries provides the automation required to expedite and simplify access to archived data in an efficient manner. There are three industry-leading tape library vendors; each can satisfy the requirements of the project. Important decision criteria include integration of the selected tape format, support by the selected backup (and optional HSM) software, and scalability and manageability of the selected library (including ease of ejecting tapes for off-site rotation) to meet both initial and future near-line data storage requirements. The backup/restore segment of storage software is a very crowded category with literally hundreds of products to choose from. It is often difficult to differentiate among features and benefits of each product. The Digital Archive's storage requirements mandate an enterprise-class backup software product that provides ease of use as well as industry-proven reliability and scalability (in combination with the selected tape library and drives).

***Proposed Solution: Magnetic tape is recommended as the near-line format to backup and protect the Digital Archive's data. A multi-bay tape library based on a well established industry standard will be used along with enterprise-class backup software. Both the tape library and the backup software will be compatible with HSM, as well as scalable to meet future needs.***

- *Storage Management Tools*. There are many approaches to storage management; some solutions are proprietary and tightly integrated into the selected hardware or software, while other solutions are stand-alone, from vendors independent of the hardware or software selection. Industry averages

vary depending upon the environment and staff skill-set, but it is generally accepted that a storage administrator can manage between two and ten terabytes without special third-party storage management tools.

***Proposed Solution: As the initial deployment will have less than 10 TB of storage, the Digital Archives will use native, integrated management tools during the initial phases of this project. As the storage capacity increases, the Digital Archives will investigate stand-alone storage management tools to more effectively manage the growing and more complex environment, which may or may not include HSM.***

Index and file database

- *Archiving Content*. Record and Content Management Systems provide the ability to centralize the storage of records into a repository where common services such as *capture*, *secure*, *maintain* and *workflow* are applied consistently. They also provide flexibility in decentralizing components to support scalability and changes in the environment. A management system is the preferred method of maintaining records for preservation purposes. There are open source frameworks from which a custom management system can be developed that offer lower upfront costs; however, the long term development costs and functional design specifications and development cycles may be more costly. Records should be stored in a way that minimizes dependence on specific hardware or software resources. Electronic archiving standards exist, such as (Department of Defense) DoD 5015.2, which present a uniform requirement for any records to be accessioned into the database.

  Records are often maintained in their original format. However, it is common practice to generate renditions in alternate formats to facilitate search and retrieval. A preferred long-term format for storing records is XML (Extensible Markup Language), where data is "wrapped" with descriptive tags which can be used to manage the retention, security, accessibility, and retrieval of records. XML is a viable format for digital preservation projects because it is "self-describing" and flexible. Additionally, XML is non-proprietary and is supported by all major software vendors. A large volume of data in the Digital Archives will consist of unstructured content, including text and image files created by a variety of applications. The ability to "tag" these files with appropriate metadata that accurately describes the contents and allows them to be searched, as well as the ability to render and display these documents, is a critical requirement.

***Proposed Solution: The Digital Archives will deploy at the Archives facility, a content management system with a web interface to store the electronic records and allow these to be searched. Electronic records brought into the Digital Archives will be converted into an XML format with defined metadata fields that will facilitate comprehensive searches. Individual agencies will have the option of incorporating a compliant content management system into their***

***agency or using XML wrapper applets. Best practices will be established for the use of metadata, with fields and accepted abbreviations defined. Partner agencies will be expected to follow the metadata formats specified to facilitate accurate searching capabilities.***

*E-mail Records*. In addition to non-structured content, the Digital Archives will archive e-mail records. There are two implementation decisions to make: stand-alone versus integrated tools, and server versus client side. An integrated management system with e-mail archiving built-in provides a consistent management interface for capturing and managing all records. Several management systems provide e-mail archiving capabilities in their base product offerings. While stand-alone email archive solutions provide considerable benefits to the IT community and appear to also provide a greater feature set than integrated offerings, they do not natively offer a mechanism to extract a record from its repository into a records management system. Additionally, e-mail archiving can be implemented either on the server-side or the client-side. The server-side management system has to process all mail to determine what is, or is not, a record, but does not require user interaction. Client-side systems capture only those e-mail messages and attachments that are determined to have archival significance, but require user involvement.

***Proposed Solution: The Digital Archives will implement a server side capture of identified individual's email, as defined in the MOU with participating agencies. Whether the product used will be stand-alone or integrated will depend largely on the content management software package chosen. The Digital Archives team will continue to monitor and contribute to the DIS EARS project, as it may reflect an opportunity for extracting archival email from a centralized DIS source in an automated and cost effective manner.***

- *Web Content and Websites*. As the value of data in web sites and the significance of the content and contextual representation of that data are recognized, it is desirable to be able to retain it for preservation purposes. Options for preserving web content are: saving web files off onto an archive share/folder or using a 'spider' utility to extract the files through an external web interface. Saving files to a share folder requires manual intervention and can be done every time the web page is modified, whereas spiders are more of an automated process that is run externally.

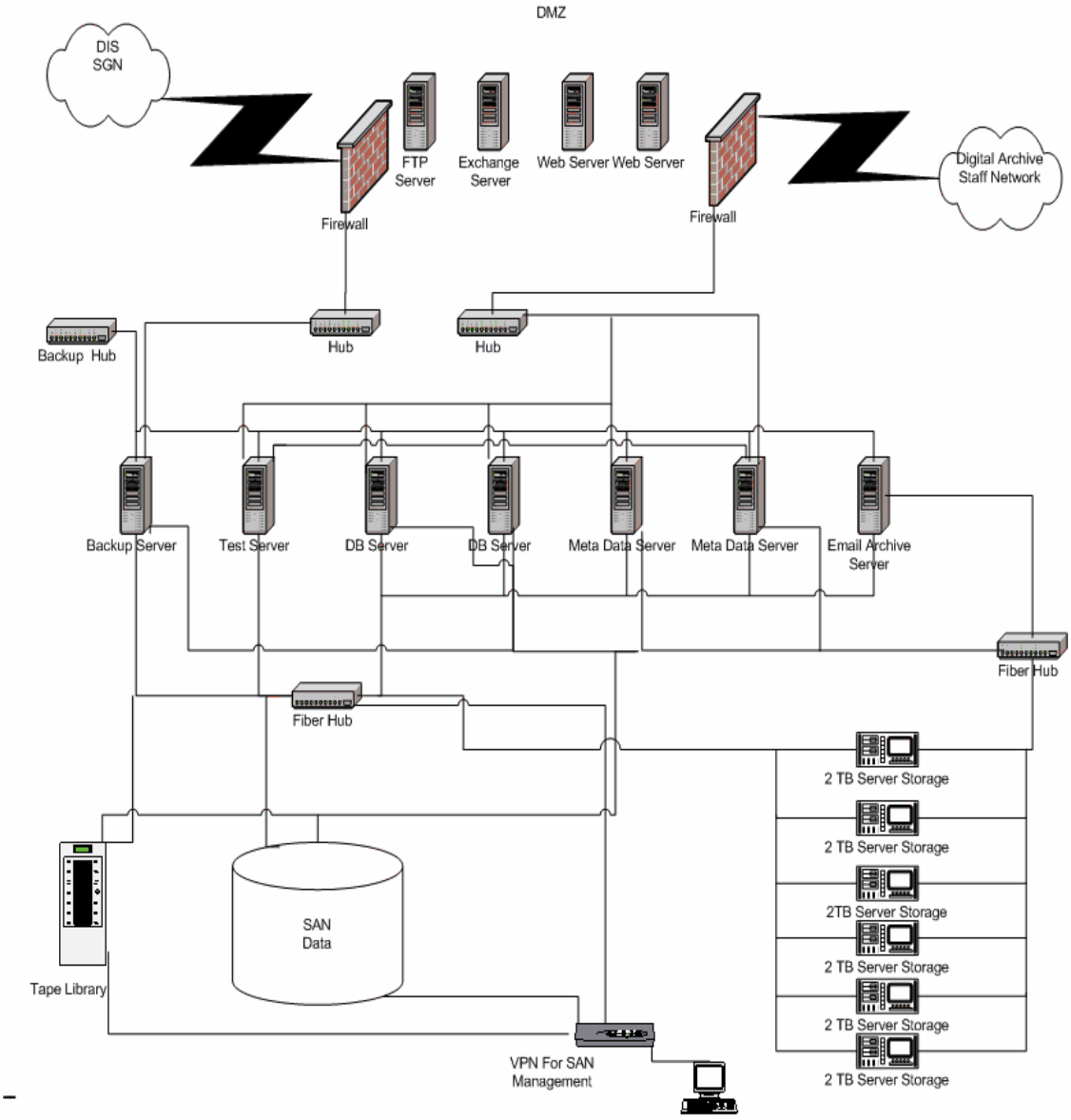
***Proposed Solution: The Digital Archives will deploy web spiders to capture agency web sites at a specified frequency, as defined in the MOU with participating agencies. Whether the spider used will be stand-along or an integrated package will depend on the content management software selected. Best practices and procedures will be created that establish how deep to "spider", how to handle links to external sites, and make recommendations as to object types that can be successfully captured. Partner agencies will be expected to follow these practices in order to facilitate efficient web spidering.***

- *Databases and Corresponding Records*. Databases are unique to the archiving paradigm in that they are containers within themselves and do not facilitate migrating content with contextual information into another repository. One approach to archiving databases is to extract database information into a format such as XML that adequately gathers metadata as well, then saving data to a medium that is transferable to the Digital Archives. Another option is to extract data via an ODBC (Open Database Connectivity) interface into a content management system. As a method of last resort, databases can be converted into ASCII format, although this method loses much of the native database functionality.

***Proposed Solution: Whenever possible, the database will be converted into an ODBC format with an XML version saved, with metadata, into the content management database. For those databases of historical importance that cannot be converted, third party vendors can be contracted to create 'translators' which will allow legacy databases to be converted into a more usable format.***

## Exhibit B

*Proposed System Architecture*



- *Establish Policies and Best Practices*. Electronic records serve a variety of purposes. It would be optimum to preserve both the "look and feel" as well as the meaning of the contents. Elements of preservation need to be articulated as do processes and methodologies for capturing, storing and retaining records. A consistent approach to applying metadata, establishing a file plan and hierarchy, implementing security, and workflow processes regardless of record type will need to be created. Experiences of several other organizations that have undertaken similar projects provide common themes. The majority of the challenges these organizations faced were related to proper design of the file plan and hierarchy, establishing policies, and acceptance by and training of the user community. Throughout the entire process, document integrity with chain of custody will be documented and secured to prove authenticity of the electronic record. This will allow the Digital Archives to certify that the records in their custody are true and original copies.

***Proposed Solution: The Digital Archives is drafting a WAC, with strong stakeholder input, to ensure that electronic records are treated with the same care and preservation that paper records now have. Furthermore, as the project develops a 'best practice' manual will be developed as a guidebook for archiving electronic records, and will encompass desktop documents, email, databases, and web pages.***

## 5.2 Converting from Legacy Hardware and Software:

Overview

The most cost-effective solution for long-term media conversion for the Digital Archives will be a combination of establishing a hardware lab for ongoing conversion work, supplemented by periodic outsourcing of media conversion and translation. Outsourcing will be considered when the cost of hardware acquisition, software acquisition, and/or software development exceeds the cost of outsourcing to an external vendor.

- The hardware lab for long-term conversion work will include hardware and software for the most common, widely available, and still-used storage formats (magnetic drives and tape, digital tape).
- The data conversion portion of the Digital Archives project is very feasible, despite the relative immaturity of standards for data conversion. The Digital Archives will begin by capturing and converting data into a neutral, archival digital form, using the formats discussed and recommended in the following sections.
- Moreover, while there are a number of standards and formats available, the Digital Archives will periodically revisit these standards and formats, and consider proactively converting to better formats as they emerge.

Conversion Hardware, Software, and Service Bureaus

The market for media conversion can be divided into three major components:

- Software – Currently available to allow Windows-based systems to access data stored on tapes, cartridges and other media in heterogeneous media system arrays. Data can be saved to a network, hard drive or burned to a CD or other media for later conversion to the appropriate archival format.
- Hardware – Also widely available and offered by a large number of vendors – each with a wide array of new and refurbished devices for reading magnetic media. While many vendors specialize in specific media (such as nine-track systems), there are several vendors that maintain a varied stock of new and refurbished drives for one-stop shopping.
- Service Bureaus - Offered by a plethora of companies providing services to move media off-site, or to transfer it to hard drives (or the media of your choice). They range in size from one-person shops, to medium-size businesses with banks of tape arrays, to agencies and corporations with off-shore processing facilities. These services can also include "computer forensics" experts who will assess the quality of archaic media, and can also retrieve data from damaged media.

There is not a single, monolithic approach to conversion. Researchers and archivists have a number of options for content conversion methods and target formats. In some cases, document material is captured as page images only, sometimes as page images with full text captured for search and retrieval, and sometimes in a neutral format such as XML for later reprocessing. PDF-A is an emerging potential standard, but is in the earliest planning stages and will not be available as a draft standard until the summer of 2004.

***Proposed Solution: The Digital Archives will look at a combination of these approaches both initially and over time. Appropriate hardware will be acquired for those media types that are in abundance. In terms of software solutions, the Digital Archives will focus on general purpose data conversion tools that convert multiple proprietary document and data formats into a neutral format, such as XML. The Digital Archives will rely on several different methods for converting various legacy materials—such as, using page image scans plus encoded text for land records and cultural collections, and full text for legislative, birth, and death records. In terms of service bureaus, the Digital Archives will competitively bid potential projects to companies that have established track records in government archival work.***

Proof of Concept Testing

The Digital Archives has conducted some initial proof of concept testing for the data conversion processes, e-mail archiving and web spidering.

- *Legacy Conversion.* For proof of concept testing of legacy files, the Digital Archives team concentrated on files created on the Intel/PC platform. A sampling of about 100 files was taken from the Office of the

Secretary of State, the State Library Networks, and files turned in to the state archives by state agencies. Samples selected included files from the Governor Lowery's Office, Senate Floor debates, and the Insurance Commission. These selected data files represent a mix of simple to complex formats dated from 1991 to 2003. These files were also given to several file conversion vendors by Glasshouse, our consultant on the Digital Archive Project.

For in-house testing, the following vendor products were used:

-- Stellent Outside IN
-- XML Export Outside In

The Stellent product supports over 225 input file formats and is used in many e-mail systems such as Microsoft Exchange and Novell's GroupWise product.

XML Export leverages the power of XML and Outside In Technology to deliver all information about a document's contents, presentation information, and meta-data to an application developer. XML Export Outside In normalizes all of the information to an XML schema provided in the form of a DTD. The application can either directly consume the XML or further transform it to a schema specified by the application developer.

> 97% of all documents converted from native formats to XML had no change in format, feel or function.

Several different Legacy Hardware configurations were used in order to recover a file off of an obsolete media format, depending on the format. This allowed the Digital Archives team to move the file to a current IT standard hard-drive, CD-ROM, or by a server system for processing.

Approximately 97% of the files tested could be converted with equivalent look and feel of the original document. On a few documents containing internal links to a support file, if the file was not found, parts of the original document would be missing pieces. Additionally, some files would have minor changes in formatting. The formatting changes were due to the differences caused by how various software packages handle software-formatting codes. This can be corrected manually or by having a custom macro modify the file back to the original style.

Research conducted by the Secretary of State's Office and Glasshouse, an independent technology consulting firm, revealed that most of the files to be archived can be converted in-house or sent to a file conversion service provider at

reasonable cost. As a result of this research it was determined that the plan to take legacy files in to the Digital Archives is feasible.

- *Email Archiving*. Tests were conducted for two different methods of e-mail archiving of the Secretary of State's Novell GroupWise and Microsoft Exchange system to an offsite Microsoft Exchange 2000 system.

  *Client side*. Client side is based on using a forward, or delegation rule, set on a individual's e-mail account such as Outlook or GroupWise desktop client. This set of rules can be set by an individual or the e-mail administrator for both sending and receiving of e-mail. The rule function setup works by matching a set of conditions then forwarding the e-mail to the Digital Archive's e-mail server for processing.

  *Server side*. Server side is based on using the journaling feature on the Microsoft Exchange server. The journaling feature works by capturing copies of user's messages within the Exchange system. Journaling lets an administrator capture all messages to another recipient (i.e., mailbox, custom recipient, or public folder) as soon as anyone submits or receives the message. The daily journal is then removed from the exchange server on a regular basis for processing at the Digital Archive. This function is not available on the GroupWise system.

  Both the client side and the server side proof of concept tests function effectively and accurately. All targeted emails were successfully moved to the remote server, where they could then be searched, printed or converted into another format.

- *Web Spidering*. Several off-the-shelf web spidering utilities were tested by mirroring the OSOS external website onto a CD. Several of the utilities have advanced option capabilities allowing the selection of depth of spidering, servers to explore, frequency of spidering, and types of files to download. Based on the testing conducted, the Digital Archives team is very confident that spidering a web site using stand alone software that captures content, links, metadata and files is possible. An integrated product that performs as a function of the chosen content management system will expand the capabilities of the Digital Archives to preserve web content.

Based on the testing conducted, the Digital Archives Team is very confident that spidering a web site using stand alone software that captures content, links, metadata and files is possible. An integrated product that performs as a function of the chosen content management system will expand the capabilities of the Digital Archives to preserve web content.

> The Office of the Secretary of State's external web site — consisting of over 1200 files and 100 MB of data — was successfully spidered using an off-the-shelf commercial web spidering utility, in less than 3 minutes.

## 5.3 Delivering Digital Content Via The Web

Web Servers

The Digital Archives will contain one or more web servers running Windows 2000 and/or Windows 2003 server with IIS. Because the storage of media will be handled by the SAN system, they will handle HTTP requests and not be used solely for direct storage purposes. The system will have sufficient bandwidth available for the transfer of archival data. This same connection will be able to accommodate all web site visitors with no problems in network traffic. If increased server power is needed, additional web servers can be purchased and load balanced in order to handle the increased load. The servers will be monitored remotely by designated Secretary of State web staff, with additional hardware support by on-site IT administrators.

Enterprise-Level Database

The system for delivering media and the subsequent database structure will be created and run on Windows 2000 SQL Server according to currently implemented OSOS standards. The database servers will be kept locally at the Digital Archives and maintained by designated IT staff, with additional hardware support by on-site IT administrators. Supplemental graphics and images will be stored on the SAN system (archival media) and/or the web servers (supported images/media for web sites, etc.).

Programming

The web site used to deliver archival media will be created and maintained by designated OSOS web staff. Microsoft .NET technology will be used to deliver this content which may include a variety of tools including XML, Web Services and SOAP to deliver archival media and metadata associated with that media based on specific customer needs as they are identified.

Other Digital Content Areas

Because of the unique nature of the Digital Archives, with a large and growing collection of data types and files required to be accessible to customers, the web site will need to remain flexible to successfully deliver the dynamic content. Therefore, various technologies may need to be used to address various situations as they arise.

## 5. 4 Phased Implementation

The Digital Archives project has been broken into four staged phases; each phase building on the knowledge and experience of the previous one, while increasing the robustness and capacity of the system architecture.

**Phase One** will consist of initial rollout of the SAN Architecture and content management system at the Digital Archives. During the initial phase, one records series at the local level will be identified and accessioned into the Digital Archives from all 39 counties. Depending on the IT capabilities of the local agency, the records series will be transmitted via direct network connection through the firewall, FTP or through disc/tape. Additionally, the Office of the Secretary of State (OSOS) will be the pilot agency to test the remote accession capability of the content management system. As the OSOS will be connected to the Digital Archives content management application, many of the records processes will be automated. Appropriate policy makers in the agency will be identified and their emails will be archived at the server side and automatically sent to the Digital Archives. The agency web page will be spidered by the Digital Archives on a periodic basis. Based on the results of this three pronged approach (content management, email archive and web spidering) the best practices manual will be developed and expanded to serve as the electronic archiving guide for future partner agencies.

**Phase Two** will begin with MOUs being signed by 2-3 state agencies and 2-3 local agencies. The agencies will be selected carefully among those that already have a strong technological presence in their agency as well as one or two that have very little technological ability. At least one of these partner agencies should have a content management system in place (preferably different than the Digital Archives content management system) that will allow for automated, remote transmission of electronic records to the Digital Archives. Two to three of the partner agencies will utilize an XML applet to wrap their archival records with metadata and transmit the packages to the Digital Archives by FTP. Lastly, one partner agency will create text files with the appropriate metadata tags and send these records to the Digital Archives by way of CD, Tape or diskette. Digital Archives staff will be working closely with partner agencies to provide technical support, as necessary. By careful selection of the phase two agencies, multiple access points to the Digital Archives will be tested and processes refined. During phase two, internet access to the records stored at the Digital Archives will be made available to the public in a highly structured, searchable format. The best practices manual will continue to be refined based on findings from Phase Two.

By the beginning of **Phase Three**, multiple accession methods will have been tested (content management, FTP, manual entry) so the focus will be on targeting specific agencies which generate records of known archival value (such as Governor's office, Legislators, Judiciary, etc). MOUs will be reached with each of the target agencies and Digital Archives staff will train the agency staff in the

processes required to archive records, providing any technical assistance that may be required.

**Phase Four** will focus on signing MOUs with all other interested state and local agencies, and shifting more of the less technical agencies to a more automated process. Phase Four will also continue to focus on capturing and archiving email from policy makers in all the major agencies at the state level.

## 5.5 Disaster Prevention & Recovery

Compliance with Policies and Standards

The Washington State Digital Archive Project has been designed to be in compliance with the ISB and DIS Security Policy and Guidelines. Also incorporated in the project are many of security features from the RFC 2196 Site Security Document http://www.faqs.org/rfcs/rfc2196.html, and the accepted National Institute of Standards and Technology (NIST) documentation (see Engineering Principles for IT Security (EP-ITS) (http://csrc.ncsl.nist.gov/)

Fire Suppression(s)

Plans include the installation of an early warning fire detection system which monitors both temperature and air quality to warn of a potential fire before it turns into a real flame. If, however, a true fire were to erupt within the Digital Archive's Data Center, the FM-200 Fire Suppression System includes tanks of FM-200, an environmentally safe, gas-based fire suppression agent. Unlike other fire suppression alternatives, the FM-200 fire suppression agent does not leave behind any oily residues, particulates, water, or corrosive materials and would not cause any collateral damage in the event it was activated. The FM-200 system monitors on a zoned basis and has dozens of sensors and suppression heads located both above and below the raised tile floor. Fire resistant walls that are true floor-to-ceiling, passing through the false ceiling tiles, further protect the Digital Archive's Data Center in the event of a fire. Windows that allow a view from the Data Center to the offices are being installed with fire safety glass to allow over an hour of burn time before a fire could cause any serious collateral damage. Pre-action sprinklers, required by code, are also included. As part of the pre-action system, Louvers in the air handling system have been designed to preclude smoke or flame from reaching the Data Center zone through the air duct system.

Security Systems

The Digital Archives Data Center will be housed within a semi-public building and security clearance is required for admittance to all non-public areas. Access to the data center will require clearance through the door security system by using a proximity identity card. In addition to the building security, many security features have been implemented within the building. Current security features include proximity card access, motion detectors, perimeter and door sensors, and 24/7/365 remote video monitoring of all access points with the additional benefit of live video feeds that are viewable and recorded from remote locations.

Data security starts with firewall protection inside the DIS firewall system. The Digital Archives has further increased this protection by incorporation of additional firewall protection zones. In addition, best practices are being enforced on system policy and file access management.

Data Rebuild/Recovery

The Digital Archives will be built with redundant hardware and network systems to prevent common failures from occurring. System architecture also includes test servers to allow for testing and recovery of the tape backup process.

If any type of file or system restoration is required, all archival data will be immediately retrievable from the tapes within the local tape library on a day-to-day basis. Should a disaster or compromising event occur, the data will be available from tape(s) stored at an off-site location, with a restoration process as well as restoration location to be determined at that time. It is important to recognize that this disaster restoration process, including recovery location, disaster equipment (servers, storage, etc.), rotation vendor, policy and procedures, will be more thoroughly investigated and defined by Washington State Archives as the project progresses through the RFP process.

Back-up & Off-Site Storage Options

The Digital Archives project has unique storage attributes. As all archive data being stored is static data, it will not and can not be changed or modified after being initially added into the archival system. Therefore, an overly complex backup schema is not required at initial implementation and, an overly complex Disaster Recovery schema is not mandated. There are two primary areas to be backed up locally and protected remotely for Disaster Recovery purposes.

- The operating system, application software, backup software, and other data residing on the server's internal disks.
- The newly archived data such as e-mail, databases, web-sites, etc., that is added each day (or other selected frequency as determined by the archives manager).

The recommended backup frequency is to back up all new data at the end of each day on two separate backup tapes and to rotate one of those tapes off-site on a nightly basis for Disaster Recovery purposes.

The Digital Archive's Data Center back-up tape storage area will be completely enclosed within a FireLock vault with a FM-200 fire protection system.

UPS & Back-up Power Generation

For added lightning protection, power will be fed to the Washington State Digital Archive's Data Center through underground wiring. Once the power reaches the Digital Archives building, it will pass through a transfer switch and into dual redundant 75kVA UPS systems where it will be cleansed and then sent to the Data Center. In the event that something were to happen that prevented the utility from providing power to the Digital Data Center, a full generator backup and power cleansing system will be installed.

The generator will be capable of powering the Data Center for up to 15 hours straight without refueling. However, the generator will have the ability to run indefinitely by being refueled while in operation.

Redundant Fiber/IP Access & Back-up

The Digital Archive's Data Center will be equipped with two redundant fiber-optic lines through Century Telephone. To prevent loss of network connection, the fiber-optic lines will be routed to the Digital Archives building from two different locations connected through an automatic fail-over router system. These fiber connections will be directly connected, and not shared, to the Department of Information SGN network.

## 5.6 Eastern Washington Digital Archives Facility

Facility Overview

The facility is being constructed on a 2-acre site in the Spokane-Cheney area, on the basalt flats of Eastern Washington, providing a significant level of geologic stability. The building is a two-story design, with a building footprint of 24,000 square feet. The competed square footage for both floors including common space is 46,900 GSF. The first floor will house the traditional paper archives for the 11 eastern Washington counties, with the second floor devoted to the digital archives functions, multi-media presentation classroom, public research area and two classrooms available for lease/use by Eastern Washington University.

Consultant Team

The Eastern Washington Archives Facility has been in the planning stages since the late 1990's. Many people were involved at the conceptual beginning which included both Washington state and national archivists, Assistant Secretary of State and Washington state IT Professionals. The current facility consultant team is comprised of the following:

**Integrus Architecture, P.S**. brings more than 45 years of civic architectural programming, design and management experience. The firm has a special design interest in libraries and the integration of new information technologies, and recently won an award in an international design competition for a new library for the information age. Integrus Architecture is a 60-person Architecture and Engineering firm with offices in Seattle and in Spokane.

**Sparling Technology Consultants**, with more than 130 professionals, is the largest specialty electrical consulting firm in the nation providing integrated design services for electrical engineering, telecommunications, audio/video and broadcast systems, and architectural lighting. The firm's diverse expertise also includes high-reliability power systems for today's e-commerce businesses, data network design, and systems that support video conferencing, multi-media presentation, and distance learning systems.

The Washington State Digital Archives Project Team, is comprised of archives and IT professionals from the Secretary of State office, and is supplemented by vendor neutral technical consultants (See Section 7, Project Management).

Building Construction

**Paras**, a construction firm from Spokane, Washington, completed early site work in January 2003. Their work included completion of the utility tunnel and preparation of the construction site pad.

**Panco**, a construction firm from Spokane, Washington, won the contract for construction of the facility which began January 2003. Completion is scheduled for spring 2004.

Programming Assumptions

The facility will provide a controlled environment with temperature and humidity controls and other facility features to maximize retention of information in electronic format. The program plan assumes that the new facility will make dramatic improvements in our delivery of services to citizens and client agencies. The ability to store large volumes of records in less space than traditional paper archives will be achieved, with the capacity to provide client agencies and citizens, with faster information delivery.

Programming assumptions provide for:

- Five technical staff and a Digital Archivist
- Secure storage of public records in electronic format
- Records integrity and audit controls
- Records migration
- Temperature and humidity controls to insure long term preservation of electronic records in various media formats
- Disaster recovery assistance
- Electronic access via Internet to archival public documents
- Multi-media presentation and digital classrooms
- Public research

Data Processing Center

The data processing center is designed with “delta” features to include sufficient space and infrastructure to support the operation. The “delta” space will allow additional equipment installation without disruption or downtime of current operations.

The space is designed to be modular, floor to ceiling for ease of space re-allocation. The Digital Archives, Digital Archives Operations, and Control Rooms contain:

- Suspended ceiling
- Low profile raised floor

- Pre-action sprinklers
- Floor drains
- FM-200 fire suppression system
- Proximity-type access control
- Temperature range 64-75 deg F
- Relative Humidity Range: 30% - 55% non-condensing

Tape Library

- 480 square feet
- Viewable from the Control room
- Access and contiguous to Nearline Storage
- Access and contiguous to servers, HD array and communications
- Temperature controlled to 70-75 deg F and 45 – 65% relative humidity

Digital capacity projections are:

| Media | 1st year | 3rd | 5th | 10th | 20th |
|---|---|---|---|---|---|
| Tape* | 4 TB | 10 TB | 45 TB | 350 TB | 800 TB |
| Disk** | 3 TB | 19.2 TB | 70 TB | 275 TB | 595 TB |

- TB = Tera Bytes (1 TB = 1,000,000,000,000 bytes)
- * = High Density Tape Cartridges averaging 100 – 200 Gb
- ** = High Capacity Large Disk Array

Monitor and Control Room

- Approximately 300 square feet
- Low profile raised floor
- Three workstations
- 700 watt workstation for terminals, etc., and 2000 watt for A/V equipment
- View (glass) to corridor, tape library HD array, servers, and communications
- Overhead flat-screen monitors
- Enough space for a group of people behind the operators (either visitors or staff involved in resolution of a system problem.) Half-height wall separation.

Storage Vault

This modular 21’ x 34’ x 12’ vault is capable of withstanding temperatures well above 2000° F for five hours of test. Both the outer vault door and inner vault door will meet the required media fire rating. The system is an above-grade (second floor) vault and is therefore a six-sided panel system. The vault includes:

- Water shield roof deck
- Vapor barrier
- Door stanchions and cable tray assembly into the vault chamber
- Spun ceramic material

Legacy Hardware & Software Laboratory

With the constant upgrade of operating systems, applications, and storage technology, obsolescence of access will overtake any durable assortment of hardware and software. The Legacy Hardware and Software Laboratory space will house equipment and software needed to deal with obsolescence issues with records created in multiple formats by state and local government agencies. Space includes tech benches, shelving and space for Legacy hardware.

Located in front of the legacy workshop space will be a "Legacy Technology Museum" showcasing legacy equipment displays for the public to enjoy.

Digital Reading Room

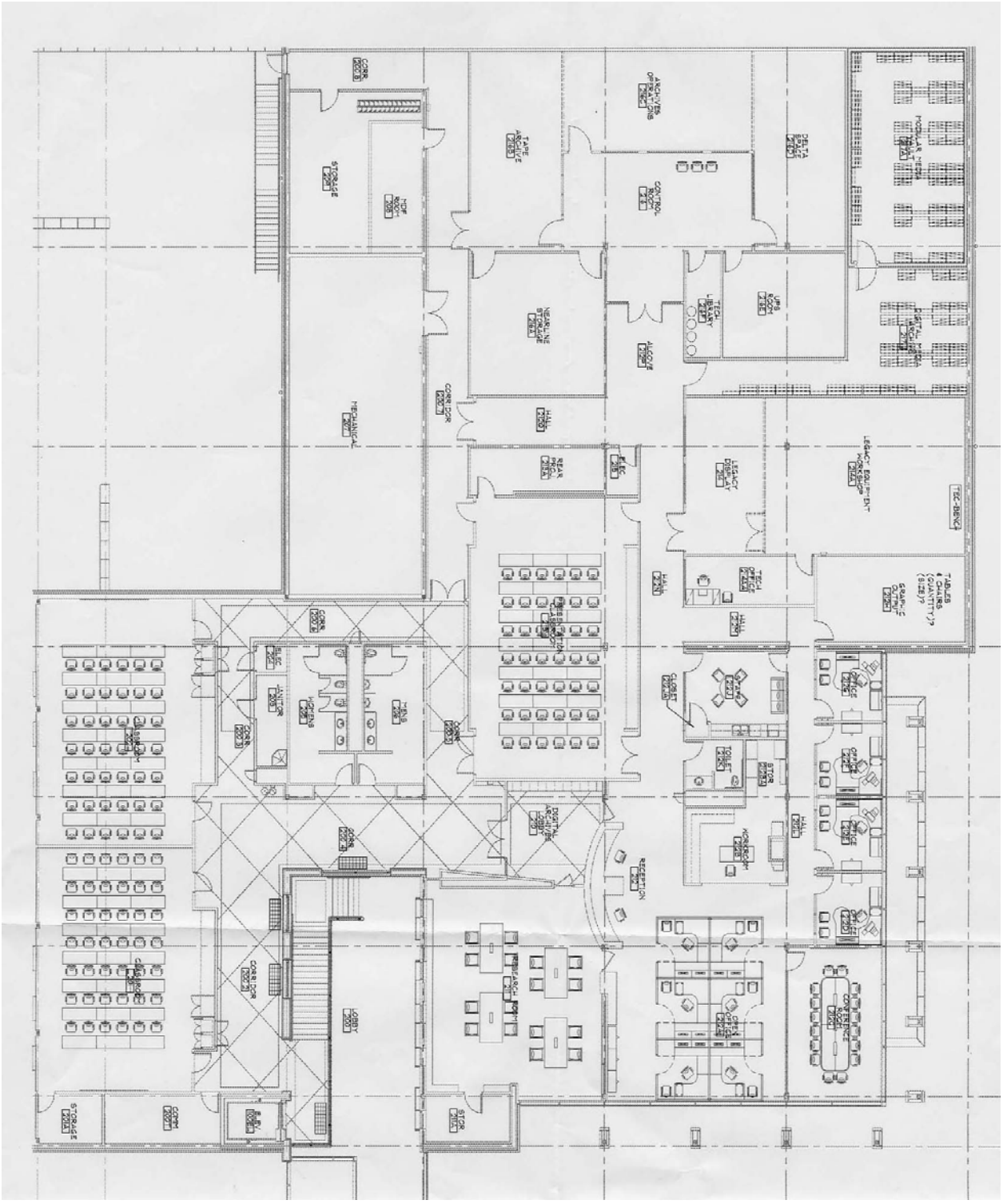
The digital research room located adjacent to the main reception area will accommodate up to 16 researchers within approximately 1,300 square feet. Digital Archives holdings will be available for research and review. Modeled after the National Archive's reading rooms that are equipped with computer workstations, for laptop connections, this facility will allow citizens convenient centralized access to public electronic records using Internet access and dynamic electronic indexes. Archives holdings not currently web enabled will be compiled and made available to researchers by Archives technicians.

Multimedia Presentation Room

With space for up to 42 individuals, the state-of-the-art multimedia presentation room will house an advanced video and teleconferencing system. Space is planned with infrastructure to ultimately support:

- Rear screen projection system
- Room speaker and sound system
- High speed telecommunication connections that support Video Teleconferencing
- Telephone connections to support Audio Teleconferencing
- Video switching system to receive and project video from various sources (video teleconferencing system video, control room video cameras, along with internet and network video feeds)

Washington State Digital Archives – 2nd floor



# 6. Conformity with Agency Information Technology Portfolio

## 6.1 Strategic Focus (Business and IT Goals)

The business strategies and goals that this project serves are described in Section 2 – Background and Needs Assessment. This project has been assigned a Rating Level 2 on the Risk/Severity Matrix; however, the legislature required that this project be closely followed and documented in the Agency IT Portfolio.

The Digital Archives will fulfill the following IT strategic initiatives:

- Continuing to provide a consistent means via the Internet for the public to access Washington State Historical Records.
- Using the state-of-the-art technology and information systems that support advanced computing, enhance customer service, and improve the processes of the Archives Division.
- Being the forerunner of Digital Archiving in enhancing the template to obtain, index, store, maintain, and retrieve digital archival objects.

## 6.2 Effect on Existing Technology Infrastructure

The Office of the Secretary of State's Digital Archives project is in full compliance with the Standards and Protocol Directions set forth by the Department of Information Services (DIS), Information Services Board (ISB), implemented on April 10, 2002.

The existing standards, both hardware and software, of the Office of the Secretary of State were considered in the first stages of this project. Hardware was purchased, and will be purchased, based on these standards and there should not be an impact to the existing IT structures in place.

There are no planned changes to any internal network and desktop support services technologies (including printers, workstations, laptops, peripherals, etc) due to this project. Planned upgrades to existing workstations have been budgeted and do not impact this project's budget.

The two major impacts will be bandwidth and storage space. Increased traffic via the Internet is expected and is accounted for in the networking plan. The performance of the existing network has been excellent, although it will be essential that technical support staff continue to be trained on the digital storage technology, as well as effective network tools and practice in order to eliminate any downtime to this mission critical application.

The storage of digital objects does, of course, require a large amount of storage space. Previously, Storage Area Network (SAN) technology was not used in this agency. Additional storage media are included in the budget and will not negatively impact existing agency storage requirements.

# 7. Project Management and Organizational Effects

## 7.1 Project Management

Roles, Responsibilities, and Management Qualifications

OSOS is committed to providing the most qualified staff available to assure project success. Staffing required includes *executive leadership*, *technical staff* and *administrative and business support.* The management of this project is strengthened by utilization of an interdisciplinary team within the agency that meets together weekly. Utilizing the variety of expertise that is available from within the agency minimizes the degree to which external consultants are needed.

Listed below are the staff positions associated with each of these three areas, their roles and responsibilities, and the qualifications of incumbent staff. A project team organization chart is shown in Exhibit C.

Executive Leadership

The success of the initiative depends on strong executive support and sponsorship. For the Washington State Digital Archives Initiative, both the Assistant Secretary of State and the Deputy Secretary of State provide active support and oversight, as well as continuous, weekly involvement in the project.

*Steven Excell, Assistant Secretary of State*

Mr. Excell serves as the executive sponsor for the State Digital Archive Initiative. He presides over weekly status meetings; establishes overall policy and direction; reviews and approves contracts related to the project and has final approval on project deliverables.

Steve brings extensive experience from both the private and public sector. He has substantial experience as a construction project manager and real estate development Vice President managing complex projects ranging in size from $3.5 million to $100 million. Since 1987 Mr. Excell served as the principal partner of a Seattle-based communications firm, which since 1995 has specialized in web-based communications and applications. In the public sector, Mr. Excell served as Legislative Director for Congressman Joel Pritchard and as Chief of Staff for Governor John Spellman. As the Governor's Chief of Staff, he participated in the planning and development of information and accounting systems for state government. Mr. Excell currently serves as Assistant Secretary of State (Chief of Staff) for Secretary of State Sam Reed. He has a Bachelor of Arts degree from the University of Washington and a Juris Doctor (law) degree from Georgetown University.

*Dan Speigle, Deputy Secretary of State*
Mr. Speigle serves as the executive sponsor of Network and Desktop Support and Applications Development groups. He assists in budget and costing decisions related to the project.

Dan has a Bachelor of Science in Accounting with a minor in Data Processing. He is a Certified Public Accountant (CPA) and a Certified Information Systems Auditor (CISA). While at the State Auditor's Office, Dan conducted general IT controls and application controls reviews of clients (both state agency and local government) computing environments and systems, as well as provided technical support to field auditors in computer assisted audit techniques. While at the Department of Labor and Industries Dan was an IT auditor, responsible for providing IT audit support to other audit staff, review of agency applications, and assisted in project development.

Technical Staff
The technical expertise required by the project includes, but is not limited to:

- Expertise in the field of archives and records management.
- Strong technical leadership to implement the nation's first state-wide Digital Archives.
- Ability to provide required technical support to state and local client agencies.
- Expertise in converting data from legacy hardware and software.
- Expertise to prepare data from client agencies.
- Expertise to determine the archival value of digital data.
- Ability to ensure continuous operation and data availability 24 hrs x 7 days x 52 weeks.

The required expertise will be provided by a combination of existing agency staff who contribute time and expertise to the Washington State Digital Archives Initiative, staff hired specifically to support the Washington State Digital Archives on a full-time basis, and outside technical consultants, as appropriate.

Funding is currently approved that will support the following two and a half positions specific to the Washington State Digital Archives:

- Full-time Digital Archivist
- Full-time network/SAN specialist
- Part-time Archives Assistant

Additional funding will be required to support three and a half additional positions (for a total of six):

- Programmer
- Electronic records manager
- Web master
- Additional hours to make Archives Assistant full time

These additional positions will be required in order to make the system fully operational, and provide adequate assistance and services to client agencies. Funding to support the additional positions was requested as part of the agency's 2003-2005 biennial budget request, from a combination of state revolving fund charges and the State Archives existing local revenue stream.

*Jerry Handfield, State Archivist*
Mr. Handfield attends weekly status meetings; provides oversight and direction regarding archive laws, policies and procedures; ensures integration of digital and paper archives policies; serves as a liaison with stakeholder groups representing state and local government agencies.

Appointed Washington State Archivist by Secretary of State Sam Reed in 2001, Jerry was recruited from Indiana where he served as State Archivist for 14 years. His career includes many years of leadership at the local state and national levels on public records and historic preservation issues. He has a Masters degree and doctoral program work in American History, and many years of teaching college level history. He served as the Director of the Indiana Commission on Public Records prior to coming to Washington State.

*Diana Rae Bradrick, Deputy State Archivist*
Ms. Bradrick serves as project manager for planning and construction of the Eastern Washington Regional/ Washington State Digital Archives facility in Cheney, WA; attends weekly status meetings; provides day-to-day supervision of Washington State Digital Archives initiative and staff members; provides oversight and management assistance on development and implementation of state archive laws, policies and procedures.

Diana Rae has over 12 years of project management experience including project planning, contract negotiation and management, developing scope of work and requirements documents, procuring hardware and software and serving as point of contact with the vendors. For 11 years, Diana Rae managed implementation of two major imaging systems for records of long-term archival value. She has a Masters of Business Administration from Portland State University, and is an NACRC Certified Public Official in Records Management.

*Adam Jansen, Digital Archivist*
Mr. Jansen serves as full-time project manager for the Washington State Digital Archives initiative. Adam attends weekly status meetings; serves as on-site manager in Cheney, Washington for the Washington State Digital Archives; conducts strategic planning; establishes, monitors and evaluates project goals, objectives and milestones; serves as primary liaison with state and local government customers and stakeholders; and coordinates and manages Washington State Digital Archives technology issues.

Adam is formally trained in both Archives and Records Management as well as information technology and IT management, bringing to the project a rare and ideal combination of skills and experience. He has a Master of Information Technology certification from the Association of Information and Image Managers, and certification as a Document Imaging Architect by Comp/TIA. Adam is a Microsoft Certified networking professional, a Certified Records Manager and has done Masters degree work in Archives and Records Management. He is a published author and speaker in the field of electronic records, and is trained in the care and preservation of electronic media.

*Network/SAN Specialist (recruitment in process)*
This position provides on-site, full time technical support for the Washington State Digital Archives Initiative. He/she will attend weekly status meetings; assists in developing the RFP for acquisition of technical solutions and bid review process; serve as liaison with stakeholder/customer technical representatives; maintains Storage Area Network and tape library. The position will also be responsible for network operations and security; expansion and coordinating IT issues with other agency IT staff assigned to the Washington State Digital Archives Initiative.

*Archives Assistant (vacant)*
Serves as primary customer interface; assist clients with access and on-site research; provides administrative support to Washington State Digital Archives staff.

*Electronic Records Manager (pending funding)*
This position will be assigned full time to the Washington State Digital Archives Initiative and will be responsible for development and management of record retention schedules, disposition of records, and ensuring the authenticity of electronic records.

*Digital Archives Webmaster (pending funding)*
This position will be assigned full-time to the Washington State Digital Archives Initiative and will be responsible for Web enabling digital records; coordinating the indexing, retrieval and web-arraying of archived electronic records of state and local government agencies; assisting with SAN maintenance.

*Programmer (pending funding)*
This position will be assigned full-time to the Washington State Digital Archives Initiative and will serve as the principle programmer for the Washington State Digital Archives in Cheney, WA. The position will be responsible for development of applications for connecting with remote locations and transmitting their electronic records to the Washington State Digital Archives; indexing and storing records; providing web-based retrieval.

*Patti Prouty, Agency IT Applications Manager*
Ms. Prouty has agency-wide responsibility for the agency IT portfolio. Patti is responsible for ensuring that the Washington State Digital Archives Initiative conforms to the portfolio and agency standards. She will oversee the development of any in-house applications, working with other agency applications staff as needed. She will coordinate with the Digital Archivist on long-term IT goals and special projects. Patti will continue to serve as a primary liaison with DIS and their Customer Advisory Board, stakeholder groups, on project-related issues. Finally, she will work with the agency and project-specific web masters on providing web-based content, specifically database security and database standards.

Patti brings over fifteen years of experience in the field of computer applications. For the past eight years, she has been responsible for managing the agency's IT Portfolio; maintaining and overseeing development of agency software applications; developing and implementing software application standards; and working in conjunction with webmaster(s) to ensure seamless deployment across the Internet. Prior to working for OSOS, Patti worked in both the private and public sector on database development, management and administration, network management and special applications. Patti has a Bachelor of Science in Computer Science with a minor in Computer Engineering from Santa Clara University. She also is a Microsoft Certified Systems Engineer.

*Matthew Edwards, Agency Webmaster*
As the overall agency webmaster, Mr. Edwards will be creating and/or coordinating the creation of web sites and applications for the delivery of Archived media. This may include graphic design, user interface design, web programming, database design as well as web server administration and maintenance. Work will be accomplished in conjunctions with the Digital Archives IT Manager or other designated OSOS staff.

Matthew has over seven years of web design/programming experience, with work on dozens of web sites and applications, including those for the movie and video game industries, multi-national corporations, nonprofit organizations, small and large businesses and state government. He has experience building web applications for a transactions, demographic and data-mining tools, multimedia and multi-language web sites, among others.

*Mike Huntley, Network and Desktop Services Manager*
Mr. Huntley serves as the consultant to the State Digital Archives team regarding overall network administration and operations, and provides system trouble-shooting to both the agency and vendors to ensure seamless network operations.

Michael brings over nine years of experience in network operations and desktop support. For the past four years, he has been responsible for managing the agency's internal and external network infrastructure, hardware and software

equipment acquisition, network security, virus protection, upgrade and maintenance of all desktop, as well as server hardware and software. Michael is also charged with ensuring the agency receives the highest standard of customer service set forth by the Deputy Secretary of State.

Prior to working for OSOS, Michael worked for public schools at the district level, and the Department of Defense. Mike is a Microsoft Certified Systems Engineer and Trainer and has technical degrees in telecommunications and electronics.

Mr. Huntley is supported by **Paul Longwell** and **Larry Gratton**, both of whom are responsible for agency-wide network administration. Each has been extensively involved in the research and design for the State Digital Archives, including SAN configuration and conversion software research and testing.

*Paul Longwell, Information Technology Systems Specialist*
Mr. Longwell is responsible for network administration in the agency and has been extensively involved in the research and design for the Washington State Digital Archives, including SAN configuration and conversion software research and testing.

Paul has a Bachelor of Science degree in Electronic Engineering Technology along with several operating system certifications: MCP, CNA, A+, Network+, CET. Paul also has over 20 years' experience in the computer field dealing with networking, system administration, and purchasing of computer network systems.

*Outside technical consultants*
To date, technical assistance and consulting services have been acquired primarily from two sources:

Sparling & Associates**,** under contract with Integrus Architecture, provided consulting assistance on the Eastern Washington Regional/Washington State Digital Archives facility planning. They were responsible for planning and designing voice/data connection, security, and audio/visual planning along with programming the physical space (legacy equipment room, operations area, tech rooms, etc.)

GlassHouse Technologies, Inc**.** was engaged to assist the agency with some of the more technical aspects of the Washington State Digital Archives Feasibility Study, including SAN requirements and configuration, conversion of data from legacy hardware and software, and digital archiving technologies. GlassHouse Technologies, Inc., headquartered in Framingham, MA, is a pioneer in the storage services sector, providing a full range of vendor-neutral solutions from strategic development to operations and implementation. They are recognized as industry experts and frequently requested presenters at key industry events and conferences. They have completed over 50 storage-related consulting projects for

a wide variety of clients, including engagements focused on email, web-site and database applications and product evaluation and assessment. Qualifications for the primary consultants utilized by OSOS are attached as Appendix 11.1.

These contractors, and others, will continue to be utilized as needed. To date, GlassHouse Technologies, Inc. has been engaged to assist the agency with the feasibility analysis, requirements analysis, development of reference architecture and costs estimates.

*Stakeholders and Advisors*
A stakeholder group, consisting of representatives of both state and local government client agencies, will be utilized to provide advice and assistance to the project management team. Initially, the group will be convened to develop additions to the Washington Administrative Code (WAC) needed to address the unique elements associated with management of electronic records. The new WAC revisions are targeted for completion in 2004. Following that, stakeholder groups will continue to be utilized to assist with project implementation and evaluation.

Business and Planning Support
Project success also depends on administrative support and infrastructure. The following resources are available to support the project.

*Cathy Turk, Financial Services Manager*
Ms. Turk manages the agency's Financial and Support Services Unit. She and the staff in her unit provide the accounting, budget, contracting and purchasing support for the agency, including the Digital Archives Initiative. Her specific responsibilities to the project include serving as project manager for development of the Washington State Digital Archives Initiative Feasibility Study, Washington State Digital Archives Project Management Plan and the Washington State Digital Archives Investment Plan. Cathy also assist in providing budget monitoring for the Eastern Washington Regional/Washington State Digital Archives facility construction project.

Cathy brings over 25 years of management and administration experience, including eight at the executive management level in Washington State government. She brings particular expertise in the areas of budgeting and contracting, strategic planning and goal setting. Cathy completed all coursework for a Masters in Public Administration at Evergreen State College.

*Betty Craig, Administrative Support*
Ms. Craig provides administrative and logistical support for project team meetings. She is responsible for formatting and editing solicitation documents (e.g. RFQQ, RFP), contracts, required plans and associated documents. This is a temporary appointment through July 2003 and is specific to the Washington State Digital Archives Initiative.

In addition to a teaching background in business education, Betty brings over 15 years experience in both the private and public sectors providing administrative support in document creation, editing and meeting support. She has a Bachelors degree in Education/Business Education – Economics and teaching certification from Western Washington University.

Decision Making Process

*Strategic and policy decisions.* Strategic and policy decisions are discussed by the project team which meets every week. Strategic and policy decisions include decisions on funding requests in support of the project. The Assistant Secretary of State, executive sponsor for the project, is an active member of the project team and ultimately makes these decisions, with input from team members.

These decisions and the thinking behind them are, and will continue to be, documented in written analyses and plans following extensive planning processes. Policies governing retention and archiving of electronic records for state and local governments will be reflected in revisions to the existing Washington Administrative Code (WAC).

*Technology decisions.* The Washington State Digital Archives project will primarily utilize commercially available hardware and software solutions, with limited initial requirements for custom development. As a result, most remaining decisions regarding the technology solution, involve selecting hardware and software from vendor options available that meet the business requirements of the project. Selection will be based on additional proof of concept testing and the Request for Proposal (RFP). The RFP will include the specifications required, based on the completed Feasibility Study analysis. Vendor and product selection following the RFP process will be based on the bid criteria defined in the RFP; those specifications will be developed by the Project Manager with assistance from the agency IT staff and external consultation and review from outside consultants as needed.

*Day-to-day operational decisions.* Project team staff members have developed and continually update budget tracking documents related to construction of the new Eastern Washington Regional/Washington State Digital Archives facility, the operational and capital equipment budgets. Project schedules and timelines are also developed and continually updated. Updates are shared with project team members and assist the project team in making decisions and adjustments as needed.

Following completion of the physical facility and installation and testing of the Washington State Digital Archives technology, day-to-day operational decisions on site at the facility in Cheney, WA will generally be left to the Digital Archivist who serves as the project manager. Weekly meetings of the project team will continue, with Cheney staff members participating generally by telephone

conference. This will be especially important in the early phases of the project. As implementation proceeds, issues will undoubtedly arise that will require refinement and adjustment of processes.

Quality Assurance Strategies

A Quality Assurance Plan is required for this project by legislation, with the intent of ensuring that the project is successfully completed, with success factors including accomplishment of the functionality, meeting the anticipated schedule and budget parameters.

*Feasibility Study*

The agency has engaged in extensive investigation and analysis in preparing the Washington State Digital Archives Feasibility Study and the Washington State Digital Archives Investment Plan. All efforts were geared toward presenting a high quality and reliable business analysis of alternatives for meeting the objectives of the project.

*Acquisition of expert consulting services.*

Procured through an RFQQ process, GlassHouse Technologies, out of Framingham, Massachusetts, was selected to assist the agency with the following:

- analyzing the technical feasibility of the project;
- identifying and describing the range of technical options available, the advantages and disadvantages of each, and the key decision points for narrowing the options (see Appendix 1);
- assisting the agency in conducting a requirements analysis to narrow the options to those that will best suit the business needs of the agency;
- preliminary identification and evaluation of key vendors, manufacturers and products that may meet agency specifications;
- design of a reference architecture and cost estimates for the developmental phase and beyond.

*Regular progress reporting with our Office of Information Technology Oversight (OITO) consultant.* In addition, agency staff have attended DIS briefings and presentations on the DIS Email Archiving and Retention System (EARS) project, and initiated a joint meeting to discuss strategies for integration and coordination.

*Regular, weekly meetings of project team*.

Weekly meetings are used to review progress, establish milestones and deadlines. The Assistant Secretary of State attends all weekly meetings and serves as executive sponsor for the project.

*Targeted training for key agency staff*. Agency IT staff members continue to actively acquire additional knowledge and training that will be required in order to plan and implement the project. As an example, IT staff members

have, and continue to acquire, SAN knowledge through formal training classes, attendance at conferences, on-site visits with both public and private organizations where SAN solutions have successfully been implemented. Examples include:

- Microsoft Exchange 2000 training to increase knowledge in email archiving;
- Cisco and IBM SAN Solution workshop;
- On-site visits to the National Archives, Weyerhauser and state agencies using SAN technology;
- On-site visits to vendors such as ADIC and EMC to increase knowledge of different SAN solution product lines;
- Brocade Fiber channel workshop to increase knowledge base on fiber channel products and their configuration;
- Managing Electronic Records training;
- SQL Server Administration training;
- Windows 2000 Server training;
- Training on Cisco router configuration

Agency staff will participate in training in SAN technology, content management, electronic records management, XML and metadata, and other appropriate subjects. Initial intensive training will occur July 2003 – January 2004 with ongoing training as technology evolves.

*Hiring of staff with specific knowledge and technical skills.*
The legislature authorized some funding for hiring additional key staff to help ensure successful implementation of this project. One of the new positions, the state's Digital Archivist position, is on board and has been able to work with the project team and the consultants on development of the feasibility analysis. Additional technical staff hires are in process.

*RFP Process*
The Request for Proposals (RFP) process will provide another opportunity to refine and/or clarify stated requirements. The RFP process will be initiated following ISB approval of the project and will be used to acquire the hardware and software required by the project (except where a Master Contract may be utilized). Specific steps that will help ensure a successful outcome for the RFP process include:

*Requirements analysis facilitated by GlassHouse Technologies.*
The requirements analysis provided a targeted requirements definition based on project service objectives and technical direction. In addition, a reference architecture has been developed based on the gathered requirements. This reference architecture, based upon specific requirements, will be reviewed prior to the RFP process, further refined as needed, and incorporated into the

RFP process to ensure that bids received conform to specific agency requirements for functionality.

*Pricing estimates are based on the requirements analysis and developed with assistance by industry experts.*
Price estimates are highly accurate for the initial implementation phase. Accurate price estimates from the outset will help ensure that the project stays within the project budget and avoid surprises.

*Use of successful RFP examples in state government*.
The agency will work closely with our OITO consultant, other units of DIS and other agencies to develop a successful RFP that will provide the desired results. Consulting assistance from public or private sources with a track record of developing successful RFP documents may also be utilized.

*Use of pre-bid proposal conference(s).*
A pre-bid proposal conference will be held to assure clarity of requirements and common understanding among potential bidders.

*Bid review and contract negotiation*

*A thorough bid review process.*
Individuals with technical expertise, business/management background and program specific knowledge will be part of the bid review team. Review team members will include agency staff and expertise from outside the agency (such as staff from DIS). Reference checks and bidder interviews will augment the review of the written bids.

*Contract negotiation and development will ensure that all requirements are retained.* This will help ensure that there is no opportunity for compromise in the quality of the product/solution, timeline or budget.

*Project Execution*

*The agency will retain a quality assurance contractor.*
The contractor will be obtained through a competitive procurement process, and may utilize the state's master contract process, supplemented by state agency references and evaluations of contractors. A quality assurance plan will be a deliverable from the contractor during the initial phase of the project and must be approved by the project team.

*Continued training of key staff and hiring staff with needed expertise and knowledge.*
Specialized training will help ensure that the agency has the skill sets to initially implement and expand the project through full implementation.

*Phased technical implementation*.
Implementation of the technology in Phases, as described in Section 5.4, will help ensure project success by not trying to do everything at once, and implementing the project in manageable stages such that each successive phase will be implemented only after successful implementation of the previous one. This will also allow the agency to manage expectations from client agencies.

*Continued involvement of OITO consultant.*
Continued consultation with the agency's OITO consultant will be important during project implementation. The agency looks forward to an ongoing relationship geared towards ensuring project success.

## 7.2 Organizational Effect of Completed Project

Impact on work processes
The Digital Archives project will have two primary effects on agency work processes:

*Assure archiving of digital records.*
Currently, most electronic data of archival value is not transmitted to the archives. The electronic records are retained within individual agencies, or is deleted or over-written. This project will provide a central repository for archival digital data. All archives staff will have access to this data in order to respond to research requests, and be able to refer customers to on-line data so they may conduct their own research. The volume of available archival data will be increased, but manual processes for research (mail, phone or fax requests for records for example) will be reduced since geographically dispersed customers will be able to conduct research on-line.

*Automated transfer of records*.
The project will make it easier to transfer electronic records from client agencies, with fewer steps. The capture and transfer process will be much more automated, and in real time. Currently, the capture and transfer of electronic records to the Archives is a very manual process. Electronic records stored on a tape or disk are sent to the Archives via mail or "sneaker net". Those records are then labeled and boxed. The Digital Archives project will allow for server to server automatic transfer of records. Automatic transfer of records as they are created will reduce the possibility of accidental loss of data. A revision to the Washington Administrative Code will set out the rules and procedures for state and local government agencies to follow for participation in the management of electronic records (also see Chapter 2, Section 2.4, Statutory and Legal Requirements).

*Regional Archives Branches.*
Regional Archives branches (currently in Ellensburg, Cheney, Bellevue, Bellingham and Olympia) may, in the future, serve as backup storage sites for

electronic records from client agencies within their region. Electronic records may first be sent to the regional branch, where regional staff would document the transfer and prepare or add information to the records as needed, before electronically transferring it on to the Cheney facility. Regional archives staff are typically more familiar with the client agencies in their regions, and the nature of the records that will be sent for archiving. This process is not much different than what the regional staff members currently do for paper records; however the format of the records will be different.

Training Needs

The Digital Archives project includes additional staffing, located at the Digital Archives facility in Cheney, WA. New staff will be hired with many of the skills and knowledge base required, although some additional training may still be needed. Existing agency staff, including both Olympia area Information Technology staff and the regional archives branch staff, will require an upgrade in their technical knowledge. One of the primary areas where training will be required is in Storage Area Networking. This is an area that is new for existing staff. Regional staff will need to be trained in the handling and management of electronic records, since their focus has primarily been paper-based archiving systems.

Job Content

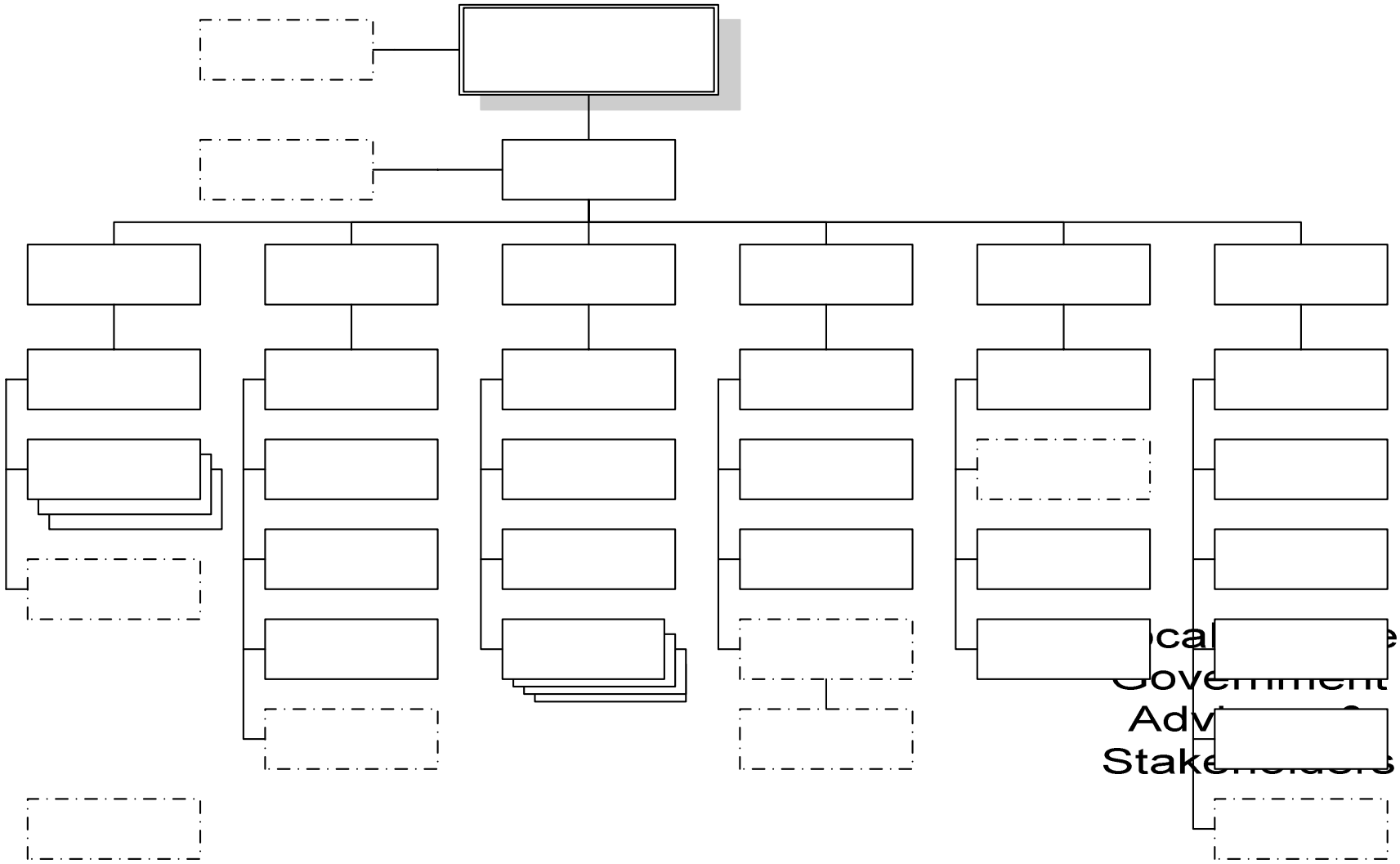
The essential nature of existing positions in the Archives and Records Management Division will not change. The Digital Archives project reflects the addition of archival records in a different format only. Although the format will be different, the basic processes will not be.

Organizational Structure

See Exhibit C on page 71.

**Exhibit C**

**PROJECT TEAM ORGANIZATION CHART**



# 8. Estimated Timeframe and Workplan

## 8.1 Construction, Procurement and Operation Schedule

The Digital Archives will be an ongoing program of the Division of Archives and Records Management. The projected timeline for the project is shown in Exhibit D.

## 8.2 Legislative Appropriations, Partnerships and Other Funding by Project Phase

Legislative Appropriations

The 2001-2003 Capital Budget (SSB 6155, Section 907) authorized the Office of the Secretary of State to "enter into a financing contract in the amount of $653,800 plus financing expenses and required reserves pursuant to chapter 39.94 RCW 10 to purchase technology equipment and software for an electronic data archive, provided that authority to expend funding for acquisition of technology equipment and software associated with the electronic data archive is conditioned on compliance with section 902 of the 2001-2003 operating budget bill (information services projects)."

The $653,000 in financing authority was not utilized during the 2001-2003 biennium, as the Feasibility Study and Investment Plan were not yet completed. Based on the now completed Feasibility Study, with assistance from the independent and vendor-neutral GlassHouse Technologies, Inc. of Framingham, MA, the agency has more accurately identified agency requirements and needs (both for the initial development phase and future implementation). In addition GlassHouse Technologies, Inc. has updated estimates based on currently available technologies, and has conducted more thorough research of costs based on the consultant's own industry knowledge as well as multiple vendor contacts. We have identified, as a result of this investigation, that costs will be higher than initially estimated.

Based on the revised cost estimates included in Section 9, Cost Flow and Benefit Analysis, the agency has determined that local fee revenue is adequate to support the initial technology investment and the local fee portion of ongoing operations. The agency will submit a 2004 Supplemental Budget request for the current biennium and subsequent decision packages in future biennia to meet the required state revolving fund portion of the total costs.

Financial Partnerships

There are at least three partnerships for this project that will or could result in leveraging other financial resources:

**Eastern Washington University**

The Eastern Washington Archives/Washington State Digital Archives facility itself is being constructed on a 2-acre site owned by Eastern Washington University in the Spokane-Cheney area under a 99-year lease agreement at virtually no cost ($1.00) to the Office of the Secretary of State. Under the agreement, the Archives will have classroom and conference space available for joint use by the University, provide work-study and internship opportunities for the University's computer science and history students, and provide mass-storage and supercomputing opportunities for computer science faculty and students. The Archives will be responsible for utilities and services, and any taxes, levies, and assessments assessed or levied against the property and which are caused by the Archives activities on the leased premises, or which will directly benefit the Archives. Under separate agreement, the Archives and EWU have an agreement regarding maintenance, repair and upkeep of the premises utilizing existing university staff and shops.

**Department of Information Services, E-Mail Archiving System**

The Department of Information Services is pursuing development of an E-mail archiving and retention system (EARS). DIS' original plan was to provide EARS as an optional, fee-for-service product for its state agency customers. If this was a required format for all state agencies, it may tie in very well with the Digital Archives' collection of selected emails. In this scenario, DIS would be responsible for providing a standard enterprise solution for near-line e-mail storage and retrieval to state agencies whether by hosting, implementation or purchase of a system. Local governments would be given the option of using the statement of requirements to develop a solution that will meet their needs as well. State and local government e-mail administrators will be required to be trained to implement and maintain a new standardized system in order to provide a cost effective system for storing , managing and retrieving email (with associated attachments) for their agency. Although only a fraction of all state government e-mail will be appropriate for capture in the Digital Archives, the EARS project may reflect an opportunity for extracting the archival email from a centralized DIS source in an automated and cost effective manner. The Digital Archives project team has met with the EARS Project Team and has provided them feedback on the requirements analysis with the goal of ensuring compatibility between the two projects.

**Local government**

RCW 36.22.175 established a $1.00 surcharge "…per instrument for every document recorded after January 1, 2002. Revenue generated through this surcharge shall be transmitted to the state treasurer monthly for deposit in the archives and records management account to be used exclusively for the construction and improvement of a specialized regional facility located in Eastern

Washington designed to serve the archives, records management and digital data management needs of local government." Use of this dedicated revenue stream is monitored by the Archives Oversight Committee, appointed by the Secretary of State and representing local government agencies, the Archives Division and Washington Association of County Officials (WACO). The funds are held in an account subject to legislative appropriation, that will provide approximately 60% of the cost of the Digital Archives operating budget, including the ongoing technology required.

# 9. Cost Flow and Benefit Analysis

Cost information is provided for two alternatives considered. The first alternative is based on the system architecture and cost estimates developed by GlassHouse Technologies. The second alternative is similar but assumes a slower phased implementation and assumes that future storage costs will decrease. Both of these alternatives are described in Section 4 (4.5 and 4.6).

For each of the two alternatives the following information is provided:

- Cost Flow Analysis
- Hardware Cost Estimate Detail
- Software Cost Estimate Detail
- Software Cost Estimate Analysis
- Legacy System Conversion Cost Detail
- Rationale for Cost Estimates
- Discussion of Benefits

## 9.1 Develop a Digital Archiving System – Full System Implementation

See the following tables.

## COST FLOW ANALYSIS
## FULL IMPLEMENTATION OPTION

| Budget Category | Object | Development Phase 2003-2005 | Development Phase FY 2006 | FY 2007 | FY 2008 | FY2009 | FY2010 | FY2011 | Total |
|---|---|---|---|---|---|---|---|---|---|
| Personal Service Contracts | C | | | | | | | | |
| - Quality Assurance | | $150,000 | $0 | $0 | $0 | $0 | $0 | $0 | $150,000 |
| - Software integration, conversion & automation | | $136,000 | $50,000 | $50,000 | $50,000 | $50,000 | $56,000 | $50,000 | $442,000 |
| - HW configuration, integration, optimization, RFP development | | $280,500 | $178,500 | $123,250 | $114,750 | $42,500 | $42,500 | $42,500 | $824,500 |
| Hardware Maintenance | EE | $33,204 | $18,197 | $30,880 | $167,514 | $230,244 | $237,610 | $376,508 | $1,094,157 |
| Software Maintenance/Upgrade | EE | $268,179 | $213,930 | $238,636 | $321,948 | $344,790 | $586,238 | $638,803 | $2,612,524 |
| Goods & Services | | | | | | | | | |
| -- Software training | EG | $46,250 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $66,250 |
| -- Purchased Services | ER | $67,000 | $134,000 | $134,000 | $134,000 | $67,000 | $33,500 | $23,000 | $592,500 |
| Hardware Purchase | | | | | | | | | |
| - Capitalized | JC | $1,224,059 | $407,176 | $427,532 | $847,705 | $1,430,815 | $1,393,150 | $1,335,492 | $7,065,929 |
| Software Purchase | | | | | | | | | |
| - Capitalized | JC | $1,094,282 | $776,724 | $26,913 | $785,993 | $47,240 | $1,820,000 | $1,417,245 | $5,968,397 |
| - Non-capitalized | JA | $5,000 | $5,000 | $5,000 | $5,000 | $2,000 | $0 | $0 | $22,000 |
| TOTAL | | $3,304,474 | $1,783,527 | $1,036,211 | $2,436,910 | $2,214,589 | $4,168,998 | $3,893,548 | $18,838,257 |

## Hardware Related Cost Estimate Detail

| *Costs by Component* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| **Enterprise Disk** | $781,121 | $254,869 | $254,869 | $724,481 | $883,069 | $1,030,861 | $1,140,030 |
| **ATA Disk Subsystem** | $266,198 | $62,746 | $85,228 | $224,677 | $267,799 | $510,983 | $426,736 |
| **SAN Fabric** | $87,130 | $0 | $5,064 | $140,064 | $92,194 | $5,064 | $43,878 |
| **Tape Library & Drives** | $198,329 | $32,423 | $12,713 | $12,713 | $78,313 | $15,818 | $78,313 |
| **Backup Solution** | $117,963 | $0 | $3,245 | $111,974 | $26,219 | $1,325 | $27,544 |
| **Consulting Services** | $280,500 | $178,500 | $123,250 | $114,750 | $42,500 | $42,500 | $42,500 |
| **Host Systems** | $267,204 | $59,850 | $111,637 | $36,000 | $348,981 | $59,850 | $59,850 |
| **TOTAL** | **$1,998,445** | **$588,388** | **$596,006** | **$1,364,659** | **$1,739,075** | **$1,666,401** | **$1,818,851** |

| *Costs by Budget Category* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| **Hardware Maintenance** | $29,520 | $12,713 | $24,497 | $160,230 | $222,060 | $229,426 | $368,324 |
| **Software Maintenance** | $169,965 | $0 | $23,806 | $3,974 | $26,456 | $1,325 | $50,290 |
| **Personal Services Contracts** | $280,500 | $178,500 | $123,251 | $114,750 | $42,500 | $42,500 | $42,500 |
| **Hardware Purchase** | $1,203,594 | $397,175 | $422,532 | $842,705 | $1,425,815 | $1,393,150 | $1,335,492 |
| **Software Purchase** | $314,866 | $0 | $1,920 | $243,000 | $22,244 | $0 | $22,245 |
| **TOTAL** | **$1,998,445** | **$588,389** | **$596,006** | **$1,364,659** | **$1,739,075** | **$1,666,401** | **$1,818,851** |

| *Component Specifics* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| **Enterprise Disk (TB)** | 3.00 | 2 | 2 | 9 | 9 | 14.25 | 14.25 |
| **ATA Disk Subsystem (TB)** | 3.50 | 3.50 | 8.50 | 13.50 | 30.75 | 48.00 | 76.50 |
| **SAN Fabric - ports** | 64 | 64 | 64 | 64 | 128 | 128 | 128 |
| **SAN Fabric - switches** | 2 | 2 | 2 | 2 | 4 | 4 | 4 |
| **Tape Library (TB cap)** | 120 | 120 | 120 | 120 | 150 | 150 | 180 |
| **Tape Library (slots)** | 400 | 400 | 400 | 400 | 500 | 500 | 600 |
| **Tape Media (carts)** | 450 | 600 | 600 | 600 | 750 | 750 | 900 |
| **Drives in Tape Library** | 4 | 4 | 4 | 4 | 6 | 6 | 8 |
| **Host Systems** | 9 | 9 | 9 | 9 | (Replace 7) 16 | 16 | 16 |

| *Total Terabytes* | Year 1 | Year 2 (est) | Year 3 | Year 4 (est) | Year 5 | Year 6 (est) | Year 7 |
|---|---|---|---|---|---|---|---|
| **Tape Library** | 10 | 28 | 45 | 103 | 160 | 255 | 350 |
| **Storage Array Disk** | 6.5 | 8.4 | 15.25 | 29 | 54.75 | 86.25 | 129 |

## Software Cost Estimate Detail

| *Costs by Budget Category* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| Hardware Maintenance | $0 | $0 | $0 | $0 | $0 | $0 | $0 |
| Software Maintenance & upgrade | $97,314 | $212,130 | $212,130 | $314,374 | $314,374 | $580,953 | $584,553 |
| Personal Services Contracts | | | | | | | |
| Professional Services | $136,000 | $50,000 | $50,000 | $50,000 | $50,000 | $56,000 | $50,000 |
| Training | $46,250 | $0 | $0 | $10,000 | $0 | $0 | $10,000 |
| Hardware Purchase | $0 | $0 | $0 | $0 | $0 | $0 | $0 |
| Software Purchase | $779,415 | $776,724 | $24,993 | $542,993 | $24,995 | $1,820,000 | $1,395,000 |
| Total | **$1,058,979** | **$1,038,854** | **$287,123** | **$917,367** | **$389,369** | **$2,456,953** | **$2,039,553** |

| *Component Specifics* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| Content Management | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Email Archive - # mailboxes | 501 | 1001 | 2001 | 3001 | 5001 | 5001 | 6001 |
| Content Contributors (concurrent) | 100 | 600 | 600 | 1100 | 1100 | unlimited | unlimited |
| Content Viewers (concurrent) | 1001 | unlimited | unlimited | unlimited | unlimited | unlimited | unlimited |
| Web Spidering was included in 2 products | | | | | | | |
| XML Conversion - all<br>Content Categorization – all | | | | | | | |

## Software Cost Estimate Analysis

| Year | 1st Year | 2nd Year | 3rd Year | 4th Year | 5th Year | 6th Year | 7th Year | Total Software |
|---|---|---|---|---|---|---|---|---|
| **Software Costs** | | | | | | | | |
| Software Product A | $496,000 | $750,000 | $0 | $450,000 | $0 | $866,000 | $0 | $2,562,000 |
| Software Product B | $364,750 | $112,500 | $0 | $225,000 | $0 | $1,467,250 | $0 | $2,169,500 |
| Software Product C | $779,415 | $776,724 | $24,993 | $542,993 | $24,995 | $1,820,000 | $1,395,000 | $5,364,120 |
| Software Product D | $1,118,250 | $4,387,500 | $150,000 | $4,462,500 | $314,750 | $8,775,000 | $150,000 | $19,358,000 |
| Software Product E | $836,500 | $926,500 | $709,000 | $1,535,000 | $1,060,000 | $3,195,000 | $1,795,000 | $10,057,000 |
| | | | | | | | | |
| mean (average) | $718,983 | $1,390,645 | $176,799 | $1,443,099 | $279,949 | $3,224,650 | $668,000 | $7,902,125 |
| mean(without Product D) | $619,166 | $641,431 | $183,498 | $688,248 | $271,249 | $1,837,063 | $797,500 | $5,038,155 |
| median (middle) | $779,415 | $776,724 | $24,993 | $542,993 | $24,995 | $1,820,000 | $1,395,000 | $5,364,120 |
| | | | | | | | | |
| **Software Support and Maintenance** | | | | | | | | |
| Software Product A | $99,200 | $249,200 | $249,200 | $339,200 | $339,200 | $512,400 | $512,400 | $2,300,800 |
| Software Product B | $72,950 | $95,450 | $95,450 | $140,450 | $140,450 | $433,900 | $433,900 | $1,412,550 |
| Software Product C | $97,314 | $212,130 | $212,130 | $314,374 | $314,374 | $580,953 | $584,553 | $2,315,828 |
| Software Product D | $201,285 | $991,035 | $1,018,035 | $1,821,285 | $1,877,940 | $3,457,440 | $3,484,440 | $12,851,460 |
| Software Product E | $59,400 | $180,270 | $282,240 | $409,860 | $553,860 | $795,960 | $1,025,460 | $3,307,050 |
| | | | | | | | | |
| mean (average) | $106,030 | $345,617 | $371,411 | $605,034 | $645,165 | $1,156,131 | $1,208,151 | $4,437,539 |
| mean(without Product D) | $82,216 | $184,263 | $209,755 | $300,971 | $336,971 | $580,803 | $639,078 | $2,334,057 |
| median (middle) | $97,314 | $212,130 | $212,130 | $314,374 | $314,374 | $580,953 | $584,553 | $2,315,828 |
| | | | | | | | | |
| **Personal Services** | | | | | | | | |
| Software Product A | $159,540 | $50,000 | $50,000 | $50,000 | $50,000 | $59,060 | $50,000 | $468,600 |
| Software Product B | $179,875 | 50000 | $50,000 | $50,000 | $50,000 | $65,000 | $65,000 | $509,875 |
| Software Product C | $136,000 | $50,000 | $50,000 | $50,000 | $50,000 | $56,000 | $50,000 | $442,000 |
| Software Product D | $182,200 | $50,000 | $50,000 | $50,000 | $182,200 | $50,000 | $50,000 | $614,400 |
| Software Product E | $117,200 | $10,000 | $40,000 | $40,000 | $40,000 | $49,000 | $40,000 | $336,200 |
| | | | | | | | | |
| mean (average) | $154,963 | $42,000 | $48,000 | $48,000 | $74,440 | $55,812 | $51,000 | $474,215 |
| mean(without Product D) | $148,154 | $40,000 | $47,500 | $47,500 | $47,500 | $57,265 | $51,250 | $439,169 |
| median (middle) | $136,000 | $50,000 | $50,000 | $50,000 | $50,000 | $56,000 | $50,000 | $442,000 |

| | 1st Year | 2nd Year | 3rd Year | 4th Year | 5th Year | 6th Year | 7th Year | Total Software |
|---|---|---|---|---|---|---|---|---|
| **Training** | | | | | | | | |
| Software Product A | $47,560 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $67,560 |
| Software Product B | $46,250 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $66,250 |
| Software Product C | $28,000 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $48,000 |
| Software Product D | $74,800 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $94,800 |
| Software Product E | $20,000 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $40,000 |
| | | | | | | | | |
| mean (average) | $43,322 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $63,322 |
| mean(without Product D) | $35,453 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $55,453 |
| median (middle) | $46,250 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $66,250 |

## Legacy System Conversion Cost Estimate Detail

| *Costs by Component* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| **Media Server** | $8,890 | $0 | $0 | $0 | $0 | $0 | $0 |
| **Additional Drives** | $11,575 | $10,000 | $5,000 | $5,000 | $5,000 | $0 | $0 |
| **Extraction Software** | $0 | $2,500 | $2,500 | $2,500 | $1,000 | $0 | $0 |
| **Conversion Software** | $5,000 | $2,500 | $2,500 | $2,500 | $1,000 | $0 | $0 |
| **Image Conversion Services** | $42,000 | $84,000 | $84,000 | $84,000 | $42,000 | $21,000 | $10,500 |
| **Document Conversion Services** | $20,000 | $40,000 | $40,000 | $40,000 | $20,000 | $10,000 | $10,000 |
| **Miscellaneous Conversion Services** | $5,000 | $10,000 | $10,000 | $10,000 | $5,000 | $2,500 | $2,500 |
| **Maintenance on WH and SW** | $4,584 | $7,284 | $9,084 | $10,884 | $12,144 | $12,144 | $12,144 |
| **Total** | **$97,049** | **$156,284** | **$153,084** | **$154,884** | **$86,144** | **$45,644** | **$35,144** |

| *Costs by Budget Category* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| **Hardware Maintenance** | $3,684 | $5,484 | $6,384 | $7,284 | $8,184 | $8,184 | $8,184 |
| **Software Maintenance** | $900 | $1,800 | $2,700 | $3,600 | $3,960 | $3,960 | $3,960 |
| **Purchased Services** | $67,000 | $134,000 | $134,000 | $134,000 | $67,000 | $33,500 | $23,000 |
| **Hardware Purchase** | $20,465 | $10,000 | $5,000 | $5,000 | $5,000 | $0 | $0 |
| **Software Purchase** | $5,000 | $5,000 | $5,000 | $5,000 | $2,000 | $0 | $0 |
| **Total** | **$97,049** | **$156,284** | **$153,084** | **$154,884** | **$86,144** | **$45,644** | **$35,144** |

Notes:
Image conversion costs are based on an average of $0.21 per converted image. Thus, Year 1 assumes 200,000 images converted at a cost of $42,000, Year two 400,000 images at a cost of $84,000, etc.

Assumes image conversion to TIFF.

Document conversion costs are based on an average of $2.00 per converted page. Thus, Year 1 assumes 10,000 pages converted at a cost of $20,000, Year 2 20,000 pages at a cost of $40,000, etc.

Assumes document conversion to XML-encoded text files.

**Rationale for Cost Estimates: Full Implementation**

Hardware Related Costs

The hardware architecture on the full scale start-up option is based directly on the architecture plan created by the consultant. It was designed with the theory of acquiring a large complex enterprise system and then allowing the Digital Archives program to grow into it. The initial architecture will be comprised of: dual load balanced database servers, dual load balanced content management servers, dual load balanced internet servers, an email archiving server, a backup server and a test server. First year storage capacity is planned at 3TB of enterprise-level storage and 3.5TB ATA disc storage. SAN connectivity will be provided through dual redundant fibre channel switches and backup will be done using an automated tape library. The growth rate of the architecture is based on an aggressive accumulation of data and exponential growth in the number of customers accessing the information over the web. To reflect this growth, additional CPUs and RAM are added to all servers on a yearly basis and additional servers are added every few years to help load balance the demands. Prices for the equipment were determined by the consultant through price quotes from the various vendors or resellers.

Year two would add:
- 2TB of enterprise storage
- 5TB of ATA storage
- one CPU and additional 1GB RAM to all nine servers to handle increasing demands

Year three would add:
- 2TB of enterprise storage
- 5TB of ATA storage
- two additional servers to load balance the internet connectivity
- one server for management of the SAN
- one CPU and additional 1GB RAM to all servers

Year four would add:
- 9TB of enterprise storage
- 17.5TB of ATA storage
- SAN management software

Year five would add:
- 9TB of enterprise storage
- 17.5TB of ATA storage
- two additional fibre channel switches for increased SAN connectivity
- two additional drives for the tape library to decrease backup time
- technology refresh, replacing the original servers with updated models

Year six would add:
- 14.25TB of enterprise storage
- 28.5TB of ATA storage
- one CPU and 1GB of RAM to all servers

Year seven would add:
- 14.25TB of enterprise storage
- 28.5TB of ATA storage
- two additional drives to the tape library
- one CPU and 1GB of RAM to all servers

Software and Related Costs

The content management software implementation on the full scale start up, as recommended by the consultant, was designed with the purpose of allowing multiple users from each agency access at the system simultaneously throughout the day. Additionally, multiple content management servers are configured in the initial startup for load balancing and a test server is configured in the system. All servers have XML conversion capability and the metadata indexing is automated to the greatest extent possible. E-mail archiving is configured from the start, with additional key policy makers identified yearly and their emails brought into the Digital Archives. Year one initial configuration has 100 concurrent contributor seats, a minimum of 1000 web viewer seats (used for price comparison on those software packages which charge for web viewers) and 500 email accounts identified for archival preservation. The cost figures are based on the average of five major content management software vendors.

Year two will add:
- 500 concurrent contributors to handle increased number of contributors from new agencies/departments
- 4000 concurrent web viewers will allow more public/external read-only access to the electronic records
- 500 email accounts will allow for the archiving of additional, identified policy makers from state and local agencies

Year three will add:
- 1000 email accounts

Year four will add:
- 500 concurrent contributors
- 5000 concurrent web viewers
- 1000 email accounts

Year five will add:
- 1000 email accounts

Year six will add:
- two additional content management servers for handle additional demands on system resources
- one additional test server to more closely mirror the production environment
- 1000 email accounts
- unlimited concurrent contributors

Year seven will add:
- 1000 email accounts

Legacy Conversion Costs

The pricing for the legacy electronic records conversion is based on the current cost of the MediaServer by Shaffstall. As this is a very elite, specific market, no other vendors have currently been identified which produce a similar product. The MediaServer will allow up to 28 drives of different media types to be installed, with appropriate conversion software to operate the hardware and decode, then translate the electronic records. As obsolete media is identified that contains records of archival nature, additional drives and appropriate software will be purchase to allow the migration of the electronic records to the Digital Archives. Prices are based on a current quote received by the consultant.

Quality Assurance

This estimate for an outside quality assurance (QA) contractor assumes the need for QA through Phase II of the project, with an estimated duration of 18 months. The estimate is based on two external QA contracts entered into by other state agencies, one costing $90,000 for 15 months, the other $96,000 for 12 months.

**Discussion of Benefits**

The cost benefit analysis forms do not reflect any dollar benefits for the Digital Archives project. Their exclusion does not mean that there aren't any financial benefits. Rather, their exclusion reflects the fact that many of the benefits cannot be accurately projected and would potentially be misleading, if a projection was attempted. Other benefits are intangible and not quantifiable in terms of dollars.

The agency evaluated potential benefits in four categories, as shown below. Following the table, a description of each category is provided.

## TABLE OF BENEFITS

| Cost Savings | Cost Avoidance | Cost Recovery | Intangible Benefits |
|---|---|---|---|
| | Legal fines and sanctions | Copies of certified records | Improved public access |
| | Growth in storage facilities for paper records | State recoveries from lawsuits and settlements (e.g. Tobacco settlement) | Legal compliance |
| | | | Public trust in government |
| | | | Preservation of state history |
| | | | Staff efficiency |
| | | | Improved record security |

Cost Savings

Cost savings was defined as funds *currently* being expended by the Secretary of State that will no longer need to be expended once the project is up and running. No cost savings are projected.

Cost Avoidance

*Legal Fines and Sanctions*

The state can face legal fines and sanctions if the event of:

- failure to retain records that are required to be retained in accordance with the state records retention requirements, or
- failure to produce public documents requested in the course of a public disclosure request or in litigation.

This issue is discussed in more detail in Section 2.4, Statutory and Legal Requirements. The absence of an organized system of records management – including electronic records – can cost agencies substantial amounts of money. If an agency cannot find their records when they need them, the time and money spent looking for them is a cost. If the state does not provide a record to a requester because it cannot be found, the agency could face monetary sanctions under the Public Disclosure Act or in litigation.

For the purposes of state public records laws, it doesn't matter whether records are in an electronic form or not. The Digital Archives will serve the purpose of providing state agencies with an organized way of storing electronic records that otherwise run the risk of being lost in various ways, including technological migration, inconsistent systems of organization and retention, and others.

*Growth in Storage Facilities for Paper Records*

Governments are far from the paperless operations predicted in earlier years, even in the face of technological developments. Despite continuing exponential growth in the creation of *electronic* records, there will be a continuing need to properly store and retain paper records as well. In fact, there is some evidence to support the notion that with the technologies available, agencies are creating more paper than ever before! However, it is hoped that over time the required growth in storage facilities for paper records will begin to diminish, with a corresponding increase in electronic storage. The Washington State Digital Archives should facilitate such a trend for Washington State.

Between 1987 and 2001, the Division of Archives and Records Management has constructed new or expanded facilities to house paper records at an average cost for any 10-year period of $17 million:

- The new Eastern Regional Archives and Washington State Digital Archives facility is currently being built for approximately $13 million and is expected to open in the spring of 2004 (the first floor devoted to the regional "paper" archives).
- The state Records Center is currently being expanded.
- The agency's ten-year capital plan also addresses the need to possibly move the Southwest Washington "branch" out of the central Archives building in downtown Olympia to a separately constructed facility somewhere in southwest Washington.
- The central Archives is currently full and an expansion or new facility combining the Archives and the State Library is in the early planning stages.

As a result, the agency is not projecting any short term cost avoidance for physical storage facilities; a slowing in the need to construct new facilities for paper record storage, however, might be anticipated in the future.

Cost Recovery

*Copies of Certified Records*

The archives has a fee structure that allows recovery of costs to produce copies of records in our collection. An initial estimate for the digital archives, based on historical trends from this cost recovery stream among all branches of the division, is $12,000 per year. It is anticipated that this may increase over time as more data becomes available electronically. Records series such as maps are in high demand, but their large size and the fragile condition of the paper originals precludes the provision of copies. Once these maps are scanned in and become available in digital format, demand for full-size copies will likely increase the cost recovery revenue.

*State Recoveries from Lawsuits and Settlements*
Government agencies also stand to gain from lawsuits and settlements. Starting in 1996, Washington joined several other states in a suit filed against the tobacco companies, alleging illegal targeting and marketing to minors and violating Washington's consumer protection and antitrust laws. This suit resulted in a settlement for Washington State of approximately $4.5 billion through 2025 to help rectify the harm caused by tobacco. The total tobacco settlement is the largest financial recovery in legal history.

This financial recovery, benefiting the citizens of Washington State, was won in part through the ability to recover and produce records. As described in a white paper by Jeffrey Bean from the Washington State Attorney General's office (June 8, 1999) many of the documents required by the Attorney General's office were "aged and ripened well beyond their retention period – many from the 1950s, some from the 1940s, and some even earlier." Much of the documentation was obtained through the Division of Archives and Records Management. A well organized records management system, for records in *any* format, improves the ability to produce documentation and evidence that can ultimately result in a legal settlement.

Intangible Benefits

Most of the immediate benefits resulting from the Digital Archives fall into this category, but are no less important.

*Improved Public Access*
One of the functions of the Digital Archives will be to serve as a centralized repository for the state's archival and historical records. Government agencies, researchers and other citizens will no longer be required to contact individual agencies to access information. State and local researchers and the general public will benefit from a centrally accessible point to search for government records, and geographically dispersed users will have remote access to the data seven days a week, twenty four hours a day. This will reduce the public's cost in time, travel, parking, etc. The inventory will be searchable through use of metadata associated with the records, facilitating ease of searching by source, date, author, agency, subject, and other key descriptors.

*Legal Compliance*
Section 2.4, Statutory and Legal Requirements, describes the legal issues surrounding the retention of electronic records. The Secretary of State's Division of Archives and Records Management is mandated by statute to insure the proper management and safeguarding of public records and facilitating citizen and government accessibility. This program will allow the archives to fulfill this mandate by providing the means to safeguard, and provide access to, electronic records – a function we are currently unable to perform. Further, all state and local government agencies have a legal obligation to establish and follow records

retention schedules for all records regardless of format. This program will help ensure that agencies meet that obligation by providing reliable, persistent methods to capture, identify, index, store and retrieve digital records for their statutory retention periods.

*Public Trust in Government*
When public documents and records are difficult to access, the public's trust in government diminishes. Conversely, greater accessibility and availability of government records and documents can increase that trust.

*Preservation of State History*
As described in Section 2.3, Business Need, significant amounts of critical electronic data have already been lost. The primary purpose of the Digital Archives is to preserve and provide access to records of enduring legal and historical significance. As government records are increasingly generated and stored in computer-based information systems, the state faces the challenge of managing and preserving these digital documents. Many are critical to the survival of Washington's history and culture, captured in the day-to-day business of government.

*Improved Record Security*
In a distributed environment such as we have today, agencies are often unable to identify archival data due to the large volume of electronic records. As a result, records of enduring value are often deleted, overlooked, or cannot be located. The digital archives will provide the ability to clearly identify and accumulate in one centralized location records of long-term archival value and the ability to universally apply mandated security to those records.

*Staff Efficiency*
Because of the lack of reliable methods for capturing, maintaining, and providing long-term access to archival data in electronic form, staff resources are often utilized to manage both a printed "hard copy" of the record along with the electronic data. While this process helps ensure the availability of critical records, it drives up the cost of storing managing, researching and retrieving the documents. The centralized nature of the Digital Archives along with the ability to clearly identify archival records will result in increased efficiency in staff resource utilization for management of that data.

### 9.2 Develop a Digital Archiving System - Phased Implementation

See the following tables.

## COST FLOW ANALYSIS
## PREFERRED ALTERNATIVE: PHASED IMPLEMENTATION

| Budget Category | Object | Development Phase 2003-2005 | Development Phase FY 2006 | FY 2007 | FY 2008 | FY2009 | FY2010 | FY2011 | Total |
|---|---|---|---|---|---|---|---|---|---|
| Personal Service Contracts | C | | | | | | | | |
| - Quality Assurance | | $150,000 | $0 | $0 | $0 | $0 | $0 | $0 | $150,000 |
| - Software integration, conversion, and automation | | $136,000 | $50,000 | $50,000 | $50,000 | $50,000 | $50,000 | $50,000 | $436,000 |
| - HW configuration, integration, optimization | | $46,750 | $0 | $0 | $0 | $0 | $0 | $0 | $46,750 |
| Hardware Maintenance | EE | $13,044 | $18,197 | $24,161 | $192,603 | $178,846 | $269,466 | $287,039 | $983,356 |
| Software Maintenance/Upgrade | EE | $247,889 | $135,302 | $135,328 | $182,769 | $233,457 | $258,289 | $362,822 | $1,555,856 |
| Goods & Services | | | | | | | | | |
| - Software training | EG | $8,000 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $28,000 |
| - Purchased Services | ER | $67,000 | $134,000 | $134,000 | $134,000 | $67,000 | $33,500 | $23,000 | $592,500 |
| Hardware Purchase | | | | | | | | | |
| - Capitalized | JC | $1,238,472 | $491,822 | $347,616 | $303,067 | $773,829 | $483,673 | $835,760 | $4,474,239 |
| Software Purchase | | | | | | | | | |
| - Capitalized | JC | $665,413 | $201,560 | $101,410 | $350,050 | $386,295 | $25,000 | $148,881 | $1,878,609 |
| - Non-capitalized | JA | $5,000 | $5,000 | $6,920 | $5,000 | $2,000 | $0 | $0 | $23,920 |
| TOTAL | | $2,577,568 | $1,035,881 | $799,435 | $1,227,489 | $1,691,427 | $1,119,928 | $1,717,502 | $10,169,230 |
| 60% local fee revenue | | | $621,529 | $479,661 | $736,493 | $1,014,856 | $671,957 | $1,030,501 | $6,101,538 |
| (Biennial Total) | | | $1,101,190 | | $1,751,350 | | $1,702,458 | | |
| 40% state revolving fund revenue | | | $414,352 | $319,774 | $490,996 | $676,571 | $447,971 | $687,001 | $4,067,692 |
| (Biennial Total) | | | $734,126 | | $1,167,566 | | $1,134,972 | | |

## Hardware Related Cost Estimate Detail

| *Costs by Component* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| **Enterprise Disk** | $781,121 | $254,869 | $254,869 | $402,864 | $446,192 | $592,242 | $680,039 |
| **ATA Disk Subsystem** | $0 | $266,198 | $62,746 | $85,228 | $121,962 | $138,312 | $313,135 |
| **SAN Fabric** | $118,330 | $0 | $5,064 | $140,064 | $83,024 | $5,064 | $43,878 |
| **Tape Library & Drives** | $198,329 | $32,423 | $12,713 | $12,713 | $78,313 | $15,818 | $78,313 |
| **Backup Solution** | $117,963 | $0 | $3,245 | $111,974 | $26,219 | $1,325 | $27,544 |
| **Consulting Services** | $46,750 | $0 | $0 | $0 | $0 | $0 | $0 |
| **Host Systems** | $76,500 | $8,000 | $25,000 | $0 | $210,000 | $16,000 | $33,000 |
| **Network Connectivity** | $350,000 | $0 | $0 | $0 | $0 | $0 | $0 |
| **TOTAL** | **$1,688,994** | **$561,490** | **$363,637** | **$752,842** | **$965,711** | **$768,760** | **$1,175,910** |

| *Costs by Budget Category* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| **Hardware Maintenance** | $9,360 | $12,713 | $17,777 | $185,319 | $170,662 | $261,282 | $278,855 |
| **Software Maint./Upg.** | $147,484 | $22,481 | $1,325 | $26,456 | $3,974 | $23,806 | $39,049 |
| **Personal Srv Contracts** | $46,750 | | | | | | |
| **Hardware Purchase** | $868,007 | $481,822 | $342,616 | $298,067 | $768,829 | $483,673 | $835,760 |
| **Software Purchase** | $267,393 | $44,475 | $1,920 | $243,000 | $22,245 | | $22,246 |
| **Other** | $350,000 | | | | | | |
| **TOTAL** | **$1,688,994** | **$561,490** | **$363,637** | **$752,842** | **$965,711** | **$768,760** | **$1,175,910** |

| *Component Specifics* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| **Enterprise Disk (TB)** | 3.00 | 2.00 | 2.00 | 3.88 | 3.88 | 5.75 | 7.67 |
| **ATA Disk Subsystem (TB)** | 0 | 3.60 | 8.60 | 13.60 | 22.10 | 30.60 | 45.60 |
| **SAN Fabric - ports** | 64 | 64 | 64 | 64 | 128 | 128 | 128 |
| **SAN Fabric - switches** | 2 | 2 | 2 | 2 | 4 | 4 | 4 |
| **Tape Library (TB cap)** | 120 | 120 | 120 | 120 | 150 | 150 | 180 |
| **Tape Library (slots)** | 400 | 400 | 400 | 400 | 500 | 500 | 600 |
| **Tape Media (carts)** | 450 | 600 | 600 | 600 | 750 | 750 | 900 |
| **Drives in Tape Library** | 4 | 4 | 4 | 4 | 6 | 6 | 8 |
| **Host Systems** | 9 | 9 | 9 | 9 | (Replace 7) 16 | 16 | 16 |

| *Total Terabytes* | Year 1 | Year 2 (est) | Year 3 | Year 4 (est) | Year 5 | Year 6 (est) | Year 7 |
|---|---|---|---|---|---|---|---|
| **Tape Library** | 10 | 28 | 45 | 103 | 160 | 255 | 350 |
| **Storage Array Disk** | 3 | 8.6 | 15.6 | 24.6 | 37.1 | 51.6 | 74.6 |

## Software Cost Estimate Detail

| *Costs by Budget Category* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| Hardware Maintenance | $0 | $0 | $0 | $0 | $0 | $0 | $0 |
| Software Maintenance & Upgrade | $99,505 | $111,021 | $131,303 | $152,713 | $225,523 | $230,523 | $319,813 |
| Personal Services Contracts | $144,000 | $50,000 | $50,000 | $60,000 | $50,000 | $50,000 | $60,000 |
| Professional Services | $136,000 | $50,000 | $50,000 | $50,000 | $50,000 | $50,000 | $50,000 |
| Training | $8,000 | $0 | $0 | $10,000 | $0 | $0 | $10,000 |
| Hardware Purchase | $0 | $0 | $0 | $0 | $0 | $0 | $0 |
| Software Purchase | $398,020 | $157,085 | $101,410 | $107,050 | $364,050 | $25,000 | $126,635 |
| Other | | | | | | | |
| Total | **$641,525** | **$318,106** | **$282,713** | **$319,763** | **$639,573** | **$305,523** | **$506,448** |

| *Component Specifics* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| Content Management | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Email Archive - # mailboxes | 501 | 1001 | 2001 | 3001 | 5001 | 5001 | 6001 |
| Content Contributors | 100 | 600 | 600 | 1100 | 1100 | unlimited | unlimited |
| Content Viewers | 1001 | unlimited | unlimited | unlimited | unlimited | unlimited | unlimited |
| Web Spidering was included in 2 products | | | | | | | |

## Software Cost Estimate Analysis

| Year | 1st Year | 2nd Year | 3rd Year | 4th Year | 5th Year | 6th Year | 7th Year | Total Software |
|---|---|---|---|---|---|---|---|---|
| **Software Costs** | | | | | | | | |
| Software Product A | $155,000 | $101,500 | $131,250 | $0 | $487,500 | $0 | $0 | $875,250 |
| Software Product B | $154,750 | $37,125 | $55,125 | $0 | $78,750 | $0 | $179,750 | $505,500 |
| Software Product C | $398,020 | $157,085 | $101,410 | $107,050 | $364,050 | $25,000 | $126,635 | $1,279,250 |
| Software Product D | $292,250 | $547,500 | $675,000 | $150,000 | $1,725,000 | $150,000 | $150,000 | $3,689,750 |
| Software Product E | $501,000 | $277,000 | $449,000 | $420,000 | $515,000 | $325,000 | $331,000 | $2,818,000 |
| | | | | | | | | |
| mean (average) | $300,204 | $224,042 | $282,357 | $135,410 | $634,060 | $100,000 | $157,477 | $1,833,550 |
| mean(without Product D) | $302,193 | $143,178 | $184,196 | $131,763 | $361,325 | $87,500 | $159,346 | $1,369,500 |
| **median (middle)** | **$398,020** | **$157,085** | **$101,410** | **$107,050** | **$364,050** | **$25,000** | **$126,635** | **$1,279,250** |
| | | | | | | | | |
| **Software Support and Maintenance** | | | | | | | | |
| Software Product A | $31,000 | $51,300 | $77,550 | $77,550 | $175,050 | $175,050 | $175,050 | $762,550 |
| Software Product B | $30,950 | $38,375 | $49,400 | $49,400 | $65,150 | $65,150 | $101,100 | $399,525 |
| Software Product C | $99,505 | $111,021 | $131,303 | $152,713 | $225,523 | $230,523 | $319,813 | $1,270,401 |
| Software Product D | $58,450 | $167,950 | $302,950 | $332,950 | $345,000 | $677,950 | $737,950 | $2,623,200 |
| Software Product E | $100,200 | $155,600 | $145,200 | $173,800 | $187,000 | $168,000 | $131,200 | $1,061,000 |
| | | | | | | | | |
| mean (average) | $64,021 | $104,849 | $141,281 | $157,283 | $199,545 | $263,335 | $293,023 | $1,223,335 |
| mean(without Product D) | $65,414 | $89,074 | $100,863 | $113,366 | $163,181 | $159,681 | $181,791 | $873,369 |
| **median (middle)** | **$99,505** | **$111,021** | **$131,303** | **$152,713** | **$225,523** | **$230,523** | **$319,813** | **$1,270,401** |
| | | | | | | | | |
| **Personal Services** | | | | | | | | |
| Software Product A | $159,540 | $50,000 | $50,000 | $50,000 | $50,000 | $59,060 | $50,000 | $468,600 |
| Software Product B | $179,875 | $50,000 | $50,000 | $50,000 | $50,000 | $65,000 | $65,000 | $509,875 |
| Software Product C | $136,000 | $50,000 | $50,000 | $50,000 | $50,000 | $50,000 | $50,000 | $436,000 |
| Software Product D | $182,200 | $50,000 | $50,000 | $50,000 | $50,000 | $50,000 | $50,000 | $482,200 |
| Software Product E | $117,200 | $10,000 | $40,000 | $40,000 | $40,000 | $49,000 | $40,000 | $336,200 |
| | | | | | | | | |
| mean (average) | $154,963 | $42,000 | $48,000 | $48,000 | $48,000 | $54,612 | $51,000 | $446,575 |
| mean(without Product D) | $148,154 | $40,000 | $47,500 | $47,500 | $47,500 | $55,765 | $51,250 | $437,669 |
| **median (middle)** | **$136,000** | **$50,000** | **$50,000** | **$50,000** | **$50,000** | **$50,000** | **$50,000** | **$436,000** |

| Year | 1st Year | 2nd Year | 3rd Year | 4th Year | 5th Year | 6th Year | 7th Year | Total Software |
|---|---|---|---|---|---|---|---|---|
| **Training** | | | | | | | | |
| Software Product A | $12,000 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $32,000 |
| Software Product B | $8,000 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $28,000 |
| Software Product C | $12,000 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $32,000 |
| Software Product D | $8,000 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $28,000 |
| Software Product E | $8,000 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $28,000 |
| | | | | | | | | |
| mean (average) | $9,600 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $29,600 |
| mean(without Product D) | $10,000 | $0 | $0 | $10,000 | $0 | $0 | $10,000 | $30,000 |
| **median (middle)** | **8,000** | **0** | **0** | **10,000** | **0** | **0** | **10,000** | **$28,000** |

## Legacy System Conversion Cost Estimate Detail

| *Costs by Component* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| **Media Server** | $8,890 | $0 | $0 | $0 | $0 | $0 | $0 |
| **Additional Drives** | $11,575 | $10,000 | $5,000 | $5,000 | $5,000 | $0 | $0 |
| **Extraction Software** | $0 | $2,500 | $2,500 | $2,500 | $1,000 | $0 | $0 |
| **Conversion Software** | $5,000 | $2,500 | $2,500 | $2,500 | $1,000 | $0 | $0 |
| **Image Conversion Services** | $42,000 | $84,000 | $84,000 | $84,000 | $42,000 | $21,000 | $10,500 |
| **Document Conversion Services** | $20,000 | $40,000 | $40,000 | $40,000 | $20,000 | $10,000 | $10,000 |
| **Miscellaneous Conversion Services** | $5,000 | $10,000 | $10,000 | $10,000 | $5,000 | $2,500 | $2,500 |
| **Maintenance on WH and SW** | $4,584 | $7,284 | $9,084 | $10,884 | $12,144 | $12,144 | $12,144 |
| **Total** | **$97,049** | **$156,284** | **$153,084** | **$154,884** | **$86,144** | **$45,644** | **$35,144** |

| *Costs by Budget Category* | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 | Year 6 | Year 7 |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| **Hardware Maintenance** | $3,684 | $5,484 | $6,384 | $7,284 | $8,184 | $8,184 | $8,184 |
| **Software Maintenance** | $900 | $1,800 | $2,700 | $3,600 | $3,960 | $3,960 | $3,960 |
| **Purchased Services** | $67,000 | $134,000 | $134,000 | $134,000 | $67,000 | $33,500 | $23,000 |
| **Hardware Purchase** | $20,465 | $10,000 | $5,000 | $5,000 | $5,000 | $0 | $0 |
| **Software Purchase** | $5,000 | $5,000 | $5,000 | $5,000 | $2,000 | $0 | $0 |
| **Total** | **$97,049** | **$156,284** | **$153,084** | **$154,884** | **$86,144** | **$45,644** | **$35,144** |

Notes:
Image conversion costs are based on an average of $ 0.21 per converted image. Thus, Year 1 assumes 200,000 images converted at a cost of $42,000, Year 2 400,000 images at a cost of $84,000, etc.

Assumes image conversion to TIFF.

Document conversion costs are based on an average of $2.00 per converted page. Thus, Year 1 assumes 10,000 pages converted at a cost of $20,000, Year 2 20,000 pages at a cost of $40,000, etc.

Assumes document conversion to XML-encoded text files.

**Rationale for Cost Estimates: Phased Implementation**

Hardware and Related Costs

The phased hardware acquisition architecture is based on recommendations from the consultants on the minimum requirements for a scalable enterprise-level infrastructure that will maintain expandability throughout the entire growth process of the Digital Archives. Through careful monitoring of the system resources, hardware will be added once demand on the system passes a certain threshold. Additional capacity will be added in direct proportion to agencies connected to the Digital Archives and as funding for expansion is available. To that end, the initial architecture will be comprised of: a 3 TB enterprise-level storage array; dual redundant fibre channel switches; a automated tape library and software for data backup; servers for the content management, database, Internet, email archiving and backup. Each proceeding year, storage availability and computing power and capacity requirements will be increased based on the anticipated agencies planned to be connected to the Digital Archives in the coming year. Prices for the equipment are according to those quoted under the state master contract, when available. For those items not available under master contract, prices used are those developed by the consultant for FY 2003-2008. Storage hardware costs were reduced 5% from current prices in FY 2009, 7% from current prices in FY 2010, and 10% from current prices FY 2011 to reflect an assumption of future, cheaper storage prices.

Year two will add:

- 2TB to the enterprise storage array for added capacity as additional agencies are brought on board
- 3.6TB to a secondary ATA disc array for less expensive storage of infrequently accessed records
- one additional CPU and 1GB RAM to each existing servers to better handle increasing resource utilization

Year three will add:

- 2TB of enterprise storage
- 5TB of ATA storage
- An additional load balancing server for the database
- A test server to replicate the content management environment for continued development of the system outside of the production environment
- One CPU and 1GB RAM for existing servers.

Year four will add:

- 4TB of enterprise storage
- 5TB of ATA storage
- SAN Management software to optimize the storage environment
- Server for SAN management utilities

Year five will add:

- 4TB of enterprise storage
- 8.5TB of ATA storage
- technology refresh on the dual fibre channel switches
- technology refresh on content management, database, email archiving, and backup servers
- additional tape drives for automated tape backup system for decreased backup time on larger storage environment

Year six will add:

- 6TB of enterprise storage
- 8.5TB of ATA storage
- additional fibre channel switch to add capacity for additional storage
- one CPU and 1GB RAM for existing servers

Year seven will add:

- 8TB of enterprise storage
- 15TB of ATA storage
- additional tape drives for automated tape backup system
- additional servers for load balancing

Software and Related Costs

The content management software implementation using a phased approach operates under the philosophy of purchasing the minimum number of concurrent licenses required. In order to function effectively with a smaller number of licenses, concurrent contributor usage of the system will be actively monitored and each contributor will have a carefully planned 'window' of time usage assigned to them. Data dumps will be automated throughout the night to the extent that it is possible. All servers have XML conversion capability and the metadata indexing will be automated to the greatest extent possible. Email archiving will be configured from the start, with additional key policy makers identified yearly and their emails brought into the Digital Archives. Year one initial configuration has 10 concurrent contributor seats, a minimum of 1000 web viewer seats (used for price comparison on those software packages which charge for web viewers) and 500 email accounts identified for archival preservation. The cost figures are based on the average of five major content management software vendors.

Year two will add:

- 40 concurrent contributors to handle increased number of contributors from new agencies/departments
- 1000 concurrent web viewers will allow more public/external read-only access to the electronic records

- 500 email accounts will allow for the archiving of additional, identified policy makers from state and local agencies

Year three will add:
- 100 concurrent contributors
- 1000 concurrent web viewers
- 1000 email accounts
- one load balancing server for the content management software to better handle increased demand on system capacity

Year four will add:
- 1000 concurrent web viewers
- 1000 email accounts

Year five will add:
- 350 concurrent contributor
- 1000 email accounts
- 5000 concurrent web viewers

Year six will add:
- 1000 email accounts
- Unlimited concurrent web viewers, if possible

Year seven will add:
- 1000 email accounts
- two additional load balancing content management servers
- one additional test server

Legacy Conversion Costs

The pricing for the legacy electronic records conversion is based on the current cost of the MediaServer by Shaffstall. As this is a very elite, specific market, no other vendors have currently been identified which produce a similar product. The MediaServer will allow up to 28 drives of different media types to be installed, with appropriate conversion software to operate the hardware and decode, then translate the electronic records. As obsolete media that contains records of archival nature, additional drives and appropriate software will be purchased to allow the migration of the electronic records to the Digital Archives. Prices are based on a current quote received by the consultant.

Quality Assurance

This estimate for an outside quality assurance (QA) contractor assumes the need for QA through Phase II of the project, with an estimated duration of 18 months. The estimate is based on the rough proration of two external QA contracts entered into by other state agencies, one costing $90,000 for 15 months, the other $96,000 for 12 months.

**Discussion of Benefits**

The benefits under this option are the same as those described in the full implementation option.

# 10. Risk Management

## 10.1 Severity and Risk Matrix

The Severity and Risk Matrix (Exhibit E) was reviewed by the agency with the agency's OITO technology consultant. The resulting overall ranking of project indicated medium risk (Level 2) using the criteria in the matrix. Comments below each Dimension evaluated provide additional explanation on the risk level ranking.

Overall, although the project will be deployed across all state and local government agencies, this project represents primarily a difference in the *format* of the records being archived; the requirements and processes for archiving government records (primarily paper records) has been in place for a long time. The project will employ technologies that will not require large initial investments by client agencies and will be designed to work in the background in an automated manner without manual intervention on the part of client agencies. The project will utilize primarily off-the-shelf technology and applications, with limited customization. Such technology did not exist a few years ago but is increasingly available to meet the needs of this project.

## 10.2 **Risk Management Matrix** (Exhibit F)

The Risk Management Matrix includes agency-identified potential risks, the impact of those risks (should they occur), the likelihood of occurrence and the steps the agency will take to mitigate against those risks.

Quality Assurance

The project will utilize a primary outside quality assurance contractor as required by legislation. See Quality Assurance Strategies and Responsibilities, in Section 7.1 Project Management for further information on steps the agency will take to mitigate risk and ensure a successful project.

## Exhibit E

**Severity Level Criteria: Overall Ranking of 2.5**

The severity matrix assesses the proposed project's impact on citizens and state operations, its visibility to stakeholders, and the consequences of project failure.

| | Categories | | | |
|---|---|---|---|---|
| **Levels** | **Impact on Clients** | **Visibility** | **Impact on State Operations** | **Failure or Nil Consequences** |
| **High (3)** | • Direct contact with citizens, political subdivisions, and service providers - including benefits payments and transactions. | • Highly visible to public, trading partners, political subdivisions and Legislature.<br>• Likely subject to hearings.<br>• System processes sensitive / confidential data (e.g. medical, SSN, credit card #'s). | • Statewide or multiple agency involvement / impact.<br>• Initial mainframe acquisitions or network acquisitions. | • Inability to meet legislative mandate or agency mission.<br>• Loss of significant federal funding. |
| **Medium (2)** | • Indirect impacts on citizens through management systems that support decisions that are viewed as important by the public.<br>• Access by citizens for information and research purposes. | • Some visibility to the Legislature, trading partners, or public the system / program supports.<br>• May be subject to legislative hearing. | • Multiple divisions or programs within agency. | • Potential failure of aging systems. |
| **Low (1)** | • Agency operations only. | • Internal agency only. | • Single division.<br>• Improve or expand existing networks or mainframes with similar technology. | • Loss of opportunity for improved service delivery or efficiency.<br>• Failure to resolve customer service complaints or requests. |

| **Ranking** | **2.0** | **2.5** | **3.0** | **2.5** |
|---|---|---|---|---|
| **Comments** | | • Project is intended to be an ongoing effort & work seamlessly background without burdening state/ local government agency staff.<br>• Project will likely be of interest nationally to other state & local governments. | • It is not anticipated that state and local governments will have to purchase hardware.<br>• The technology is not high risk.<br>• The 3.0 ranking is solely because the initiative will be deployed statewide. | • Failure to implement would increasingly result in an inability to meet the agency's legal mandate for records management including digital assets.<br>• However local govern- ments will retain records needed for daily governance and use.<br>• Daily government operations will be able to continue to function if system is down.<br>• Records could mostly be downloaded to paper in the short term if needed for access. |

## Risk Level Criteria: Overall Ranking of 2.0

The risk matrix measures the impact of the project on the organization, the effort needed to complete the project, the stability of the proposed technology, and agency preparedness.

| | Categories | | | |
|---|---|---|---|---|
| **Levels** | **Functional Impact on Business Processes or Rules** | **Development Effort & Resources** | **Technology** | **Capability & Management** |
| **High (3)** | • Significant change to business rules.<br>• Replacement of a mission critical system.<br>• Multiple organizations involved.<br>• Requires extensive and substantial job training for work groups. | • Over $5 million.<br>• Development and implementation exceeds 24 months.*<br>• Requires a second decision package.<br><br>*Clock starts after feasibility study or project approval and release of funding. | • Emerging<br>• Unproven<br>• Two or more of the following are new for agency technology staff or integrator, or are new to the agency architecture: programming language; operating systems; database products; development tools; data communications technology.<br>• Requires PKI certificate.<br>• Complex architecture-greater than 2 tier | • Minimal executive sponsorship.<br>• Agency uses ad-hoc processes.<br>• Agency and/or vendor track record suggests inability to mitigate risk on project requiring a given level of development effort. |
| **Medium (2)** | • Moderate change to business rules.<br>• Major enhancement or moderate change of mission critical system.<br>• Medium complexity business process(es).<br>• Requires moderate job training. | • Under $5 million but over agency delegated authority.<br>• 12 to 24 months for development and implementation. *<br><br>*Clock starts after feasibility study or project approval and release of funding. | • New in agency with 3rd party expertise and knowledge transfer.<br>• One of the technologies listed above is new for agency development staff. | • Executive sponsor knowledgeable but not actively engaged.<br>• System integrator under contract with agency technical participation.<br>• Agency and/or vendor record indicates good level of success but without the structure for repeatability. |

| | Categories | | | |
|---|---|---|---|---|
| **Levels** | **Functional Impact on Business Processes or Rules** | **Development Effort & Resources** | **Technology** | **Capability & Management** |
| **Low**<br><br>**(1)** | • Insignificant or no change to business rules.<br>• Low complexity business process(es).<br>• Some job training could be required | • Within agency delegated authority<br>• Under 12 months for development and implementation.*<br><br>*Clock starts after feasibility study or project approval and release of funding. | • Standard, proven agency technology. | • Strong executive sponsorship.<br>• Agency and vendor have strong ability to mitigate risk on a development project.<br>• Project staff uses documented and repeatable processes for tracking status, problems, and change.<br>• Agency or vendor is CMM Level 3 equivalent or above. |
| **Ranking** | **2.0** | **2.0** | **2.0** | **2.0** |
| Comments | • Archiving principles & work processes will not significantly change: the primary change will be the format of the records (electronic vs. paper).<br>• System will be designed to work seamlessly in the background without burdening state/ local governments' agency staff. | | • Some technology (e.g. SAN operation) will be new to agency staff. Currently & increasingly, these are off-the-shelf tools & technology solutions that did not exist a few years ago. | • Phased implementation of the project will help ensure success. Agency has recent history of successful implementation of complex technology projects.<br>• Expect technical consultation is being utilized to augment agency IT staff knowledge & experience. |

**Exhibit F**

# RISK MANAGEMENT MATRIX

| Risk | Impact | Likelihood of Occurring | Mitigation Plan |
|---|---|---|---|
| Local and state government client agencies may demonstrate some resistance to archiving of electronic records | High | Medium | The WAC 434 governing Archives will be revised with participation from all stakeholder groups to ensure their understanding of requirements for archiving electronic records and that the requirements are doable. With client agency involvement in development of the rules, we anticipate a sense of ownership on their part and an additional desire to participate. The Secretary of State will meet with management of state and local agencies, working with them to determine which records are of archival value and the method by which they will be transmitted to the Digital Archives. Processes will be established to allow for transmittal of records with a minimum of manual intervention, and instead focus on automated processes that will require initial set up only. A Memo of Understanding will be written, at least initially, with participating agencies. Digital Archives staff will be available for assistance. The project will be phased in, beginning with only a few agencies to allow for successful piloting and provide other agencies with proven practices and methods prior to full implementation. |
| Inability to import and use legacy data due to outdated technology | Low to Medium | High | At a minimum, the Digital Archives will be able to convert data to the lowest common denominator (ASCII). The Digital Archives plans also include a legacy equipment lab that can be used for conversion and/or use of outside commercial conversion services if necessary. The percentage of unreadable data will decrease over time. |

| Risk | Impact | Likelihood of Occurring | Mitigation Plan |
|---|---|---|---|
| State and local government client agencies may need to utilize/learn some new technologies. | Low | Medium (High for medium to small size agencies) | The Digital Archives will strive to capture data in current systems, rather than force state and local governments to convert to a new technology. Digital Archives staff will provide technical assistance, on-site when required. Training documentation that is easy to read and use will be provided, and some of the local government archives training dollars can be used to assist with local government training. If customer agency training needs exceed current budget, we will consider seeking additional appropriation authority from dedicated revenue streams in the future to provide such training. |
| Archival data will be centrally stored in a mass storage system (SAN), potentially subject to loss of power and data corruption. | Medium | Low | SAN will be vendor installed and staff will be fully trained. The SAN will be protected by redundant power sources (generator, UPS), RAID 5, a backup tape library, including off-site storage. Data will be able to be fully restored if necessary. If the system were temporarily down, there would be no cost consequences to client agencies and staff would still be able to respond to customers to fill any emergent needs. |
| There is no universal content management system used by state and local government client agencies. Lack of a universal system necessitates more manual and less automated systems for capturing, managing and preparing data for inclusion in the archives so that the data can be easily searched and retrieved. | Medium | Medium | The agency will not immediately impose universal use of a content management system. The Digital Archives will utilize a combination of off-the-shelf applications, sound data structure and processes, use of content management software at the Digital Archives and manual processes to manage and prepare the data. The agency will continue to work with client agencies in how to manage their content. |

# 11. Appendices

Appendix 11.1: GlassHouse Technologies, Inc.

Appendix 11.2: Digital Archiving and Mass Storage Glossary

Appendix 11.3: Office of the Secretary of State 2002 IT Portfolio

## Appendix 11.1

## GlassHouse Overview

Company Information
GlassHouse Technologies, Inc.
200 Crossing Boulevard
Framingham, MA 01702
(508) 879-5729 (phone)
(508) 879-7319 (fax)
info@glasshousetech.com

A privately owned company, GlassHouse Technologies is recognized as a pioneer in the storage services sector. GlassHouse Technologies provides a full range of vendor-neutral solutions from strategic development to operations and implementation. From strategic assessments to solution maintenance, the company's goal is to help organizations realize the business benefit from storage and backup technologies, enhancing client environments and meeting their business needs.

Glasshouse was founded in August 2001 by a group of IT industry veterans who recognized a growing need to transform data storage from an IT initiative to a business imperative. The executive team brings experience in management, storage technology and consulting for both private and public organizations. They are frequently the featured speakers at key industry events, including Storage Decision, NEDRIX (New England Disaster Recovery xChange) and Comdex. GlassHouse leaders and industry experts have written and published many articles in publications such as STORAGE Magazine, Intelligent Enterprise and InfoStor, and are often quoted in publications ncluding CIO Magazine, Byte and Switch, and Mass High Tech.

As an independent, vendor-neutral provider of storage services and consulting, they have no formal affiliations with any hardware or software manufacturers or vendors and do not serve as a seller or reseller of any vendor products.

Listed below are the members of the GlassHouse consulting team that provided technical consulting to the Office of the Secretary of State in development of the Washington State Digital Archives Feasibility Study.

Consultants
*Jay Ramsperger, Project Manager*
Jay Ramsperger is a Senior Consultant at GlassHouse Technologies, Inc. He has 15 years experience delivering solutions as a program manager, business analyst and interface architect. Mr. Ramsperger has expertise in strategic planning with decisive ability to identify and define processes that produce a positive impact on the bottom line. He has held management positions including Engagement Manager, Director, Principal Consultant, and Project Lead with industry leading consulting and technology companies.

Mr. Ramsperger has leveraged extensive experience to lead development and implementation of business strategies to increase market share for a leading storage vendor; has been responsible for managing and leading client engineering engagements within the Storage Professional Services area; and has managed certification and interoperability lab engagements, assessment/requirements gathering and analysis, infrastructure design and proof of concept testing.

Mr. Ramsperger's related experience includes:

- Identified, designed, implemented and supported operational processes including delivery process, SAN Lab management and customer management processes.
- Led integration of business strategies that address solution delivery methodology, storage architecture, services portfolio, organizational structure, roles and responsibilities and supporting business systems.
- Coordinated efforts for several initiatives including evaluation of Professional Services Automation (PSA) tools, implementation of standards and practices for project management, development of new service offerings, and implementation of supporting materials.
- Responsible for all aspects of client interaction including Proposal/Statement of Work authoring, coordinating all aspects of equipment testing and deliverable creation with heavy emphasis and extensive experience in Technical Writing.
- Provided leadership, developed and delivered training to numerous functional teams.

*William (Bill) R. Peldzus, Section Leader: Storage Area Network*
Bill Peldzus is a Senior Storage Manager at GlassHouse Technologies, Inc. where he is responsible for key client engagements. Prior to GlassHouse, Bill was a Storage Consulting Marketing Manager and Sr. Storage Architect in Imation Corporation's Storage Professional Services. Peldzus joined Imation in September of 1999 from StorageTek, where he most recently was Storage Marketing Consultant for the wordwide StorageTek SAN Operations Business Group.

Bill often serves as a content expert in storage networking presentations, including speaking engagements at the Fibre Channel Technical Conference, PC Expo's SAN Summit, Comdex, IDC's StorageVision, SAN East/West, Storage Networking World and Networld+Interop. He is frequently interviewed and quoted by leading industry trade publications such as *InformationWeek, eWeek, InfoStor, Washington Technology, Storage,* and *Network World.* In addition, he contributed to a regular column titled, "From the Lab" for the publication *Computer Technology Review,* and has led webinars on storage networking issues for both SeachStorage and eWeek.

Mr. Peldzus holds a Bachelor of Science degree in computer science with an emphasis in Mangement Information Systems (MIS) from the University of St. Francis in Joliet, Ill.

*Bill Trippe: Section Leader: Conversion from Legacy Hardware/Software*
Bill Trippe is President of New Millennium Publishing, a Boston-based consulting practice formed in 1997. Bill has more than 20 years of technical and management experience in content management, XML, and related technologies. He specializes in large-scale applications of

content management technology, working with publishers who are typically converting expensive legacy databases and systems into more contemporary, open technology. He brings a unique blend of strategic and hands-on knowledge of the products and trends that are shaping the publishing and content technology marketplace.

Clients include major companies where publishing is the core business (Houghton Mifflin Company, Pearson Education, Reed Business Information, Thomson Learning, Random House) and where publishing is a demanding "second business" (General Motors, United Technologies). Bill advises clients through all phases of the technology lifecycle, from the business case through requirements analysis, vendor selection, integration, and deployment, and he typically plays an active part in determining how XML will be used as a central source driving output through print, the Web, and other online channels.

In addition to his role at New Millennium, Bill is associate editor of The Gilbane Report, where he covers trends and technologies in the content management industry and develops conferences and tutorials on XML and content management. He is a sought-after speaker and author in the XML field, and is the XML columnist for Transform, a Consulting Associate with the consulting and market research firm CAP Ventures, and a regular contributor to the magazine, EContent. He recently co-authored Digital Rights Management: Business and Technology (John Wiley & Sons) and is co-author of a new book, SVG for Designers: Using Scalable Vector Graphics in Next-Generation Web Sites (Osborne/McGraw-Hill).

*Jenney Fields: Section Leader: Digital Archiving*
Ms. Fields has over ten years experience implementing solutions based on Microsoft and other technologies. She has expertise in the design and implementation of highly available solutions for Microsoft technology areas, and recently published a book on optimizing SQL Server 2000.

Ms. Fields' related experience includes combining design and implementation services with training development and delivery and has provided technical expertise and direction in the management and implementation of enterprise computing solutions. She has provided expertise in tactical analysis associated with risk, scope and requirements management for all projects and services, and has discussed, defined, developed and delivered on high-end custom training on a variety of enterprise topics.

Ms. Fields' technical skills include:

- Software: SQL Server 2000, V7 and V6.5, Exchange V5.5, Microsoft Cluster Server, Groupwise, TGV Multinet/Multiware, Process TCPware, Saleslogix, NetApp
- Network and OS: Windows 2000, Windows NT, Windows 95/98. OpenVMS, OS/2, DOS/Windows, Macintosh OS; Windows 2000, Windows NT, Novell NetWare, DEC PATHWORKS, and Appleshare, NetApp Data ONTAP and WAFL
- Certifications: Microsoft Certified Database Administrator (MCDBA), Microsoft Certified Systems Engineer Windows 2000 and Windows NT (MCSE), Microsoft Certified Trainer (MCT), Novell Certified NetWare Engineer V3.11 (CNE)

- Publications: "Microsoft SQL Server 2000 Optimization Guide"; published by Prentice Hall, November 2000. “Networking Windows NT and Windows 95”; distributed by Sutton Technology Group, 1998/1999, “Integrating PATHWORKS with Windows NT”; Pathworks Manager, January 1993 “Manageworks”

## Appendix 11.2

## Digital Archiving & Mass Storage Glossary

**3GIO**
The third-generation I/O architecture (code-named 3GIO) designed as a highly flexible, reliable, serial I/O architecture that will scale to the theoretical limits of copper. It will serve multiple markets segments such as desktop, mobile, server, communications and embedded applications and remain compatible with the existing PCI programming model. Products based on third-generation I/O architecture are expected to begin emerging in the marketplace in the second half of 2003. Companies working on this standard are known as the "Arapahoe Work Group."

**8-mm tape**
A format (named for the width of the tape) first introduced by Exabyte Corp. in 1987, using a helical scan technique that put 2.5GB of storage on one cartridge. Since then, Exabyte has expanded the line to include cartridges and drives with capacities up to 40GB.

**AIT**
Advanced Intelligent Tape™ - is a Sony brand of 8-mm magnetic tape cartridge with built-in memory chips to speed the process of data retrieval. With a capacity of up to 35GB and a data transfer rate of up to 4MB/sec., Sony's AIT-1 drive is the entry-point to AIT. Starting in the third quarter of 2001, AIT-3 is the new internal and external Sony tape drive and has a 3.5-inch form factor and features 100 gigabytes (GB) of native capacity (260GB with 2.6:1 compression) and a 12MB/sec sustained native transfer rate (31MB/sec with 2.6:1 compression). Like the previous generations, AIT-3 drives are fully read- and write-backward compatible with AIT-1 and AIT-2 media. Sony predicts it will offer an 800GB per cartridge before the end of the decade.

**ALA**
The American Library Association is the professional association for America's librarians with over 63,000 members. The ALA is working on digital archiving and indexing through its Networked Resources and Metadata Committee.

**ADC**
ATA Disk on Chip is a chip which includes a solid state disk and an industry standard PC AT (ATAPI) bus interface.

**Array**
A collection of SCSI hard disks connected to a single controller which, depending upon RAID settings, can act as independent disks, mirrored disks or a single virtual harddrive

**Archiving**
The act of storing records to ensure permanence due to their intrinsic historical, legal, or continuing value.

**ASCII**
American Standard Code for Information Interchange is the dominant character set encoding used by present-day computers (this may slowly change.) Current ASCII uses 7 bits of data for each character, allowing for 128 distinct character code points. ASCCII code is probably the lowest common denominator for exporting and importing data between diverse applications.

**ATA**
Acronym for Serial ATA, a reference to the bus architecture used to connect the hard drive to the CPU. ATA defines the physical, electrical, transport, and command protocols for the internal attachment of storage devices.

**ATM**
Asynchronous Transfer Mode is a high-speed networking technology designed to deliver transmissions consisting of many different types of information including text, voice, audio and video. ATM uses small, fixed-length packets that it multiplexes onto the carrier. ATM can achieve speeds up to 622 Mbps over fibre optic cable. The basic unit of an ATM transmission is a Cell, a packet consisting of 5 bytes routing information and a 48-byte payload of data.

**Authenticity**
The authenticity of a digital object refers to the degree of confidence a user can have that the object is the same as that expected based on a prior reference or that it is what it purports to be (See The CEDARS Glossary of Commonly Used Terms and BIBLINK Authentication of Electronic Publications). The digital environment poses particular challenges for establishing authenticity. This is due to the ease with which digital material may be altered and copied, resulting in the possibility of a multiplicity of versions of a particular document. Methods used in converting, storing, transmitting or rendering digital objects may result in distortions and therefore need to be documented. The process of migrating information from one system or format to another may result in changes which also need to be recorded. Aspects such as a document's functionality, its dependence on particular software and its relationship to other documents are all features which need to be considered in the establishment of its authenticity.

**Best practices**
An assessment recommending the most appropriate way of handing a certain type of task, based on an observation of the way that several organizations handle that task.

**Bit**
Bit is the smallest amount of information in a binary digital system. It can be used to represent two states of information, such as: YES or NO; SOMETHING or NOTHING.

**Block**
(1) The unit in which data is stored and retrieved on disk and tape devices. Blocks are the atomic unit of data recognition (through a preamble and block header) and protection (through a CRC or ECC). (2) A unit of application data from a single information category that is transferred within a single sequence.

**Bridge Controller**
A target controller that uses SCSI for the connection between the initiator and some other bus (or SCSI) to connect to the peripheral device. Bridge controllers are commonly used in RAID arrays or in optical disk or tape libraries.

**Byte**
Abbreviation for Binary Term, a unit of storage capable of holding a single character. On almost all modern computers, a byte is equal to 8 bits. A byte can represent 256 different states, for example numbers, processor instructions, or a combination of letters and numbers as in the ASCII code. Large amounts of memory are indicated in terms of kilobytes (1,024 bytes), megabytes (1,048,576 bytes), gigabytes (1,073,741,824 bytes), and terabytes (1,099,511,627,776). A removable disk that can hold 1.44 megabytes, for example, is capable of storing approximately 1.4 million characters, or about 3,000 pages of information.

**CD-ROM**
High capacity, read-only storage media in the form of an optically read 4 ¾ “ disc.

**Clones**
A clone is a physical copy of a volume. The clone is created as a "third mirror" of the volume. Once the clone is made, it acts as a mirror until the clone is broken off. Typically, the application will be paused before the clone is broken off to ensure point in time data integrity. If the application is not paused, the clone will contain "crash consistent" data. Once broken off, the clone can be mounted on another host connected to the HSG80 controller. The Clone operation can be applied to RAID 0, RAID 1 or RAID 0+1 volumes. See snapshot. Clones and snapshots can be used in similar ways. Both can be used for offline backup. In offline backup, the backup is done by a dedicated backup server using the clone or snapshot. This allows the application to continue to run without interruption. In traditional backup, the application is typically unavailable during the backup process. Even with applications supporting "Hot Backup" there is typically performance degradation during the backup process. Clones are most useful for operations that may change the data, such as production testing or data warehousing. Snapshots are most useful for operations that require a quick copy of the data. One case is where snapshots are made on an hourly basis to ensure a quick restore of files that are accidentally deleted or destroyed. Snapshot is required if the volume to be copied is parity RAID (RAID 3/5).

**Clustering**
The grouping of multiple servers in a way that allows them to appear to be a single unit to client computers. Clustering is a means of increasing network capacity and improving data security.

**Content Management (CM)**
(1) Generally, the ability to manage content pulled from any part of an extended organization or network of organizations, including employees, partners, customers and others to be used and reused in any language by any number of users. CM allows access to both structured and unstructured data. Unstructured data typically consists of business documents, emails, rich media and other data residing on individual PCs and servers spread throughout an organizations. (2) Software that helps developers track the location of information (especially important in medium and large systems, where the location of content can be easily forgotten) and relationships among the different pieces of information.

**Controller**
The hardware and software that provides the control logic in a storage subsystem that performs command transformation and routing, aggregation (RAID, mirroring, striping, or other), high-level error recovery, and performance optimization for multiple storage devices. Control logic also exists on a disk or tape that performs command decoding and execution, host data transfer, serialization and deserialization of data, error detection and correction, and overall management of device operations.

**DAS**
Directly Attached Storage. At one time, all mass storage devices such as disk and tape drives were directly attached to the nearest computer (or were located inside the same box) so there was no need for a term describing this. But since the technologies which enable storage networking have become more prevalent, the term DAS has been used to describe those parts of a wider storage network in which this local connection is still used.

**DAT**
Digital audio tape; a tape technology designed for very high-quality audio recording and data backup. With their 12GB native capacity, DAT data cartridges are most often used for backing up small servers. DAT uses a scheme called *helical scan* to record data. A DAT cartridge is slightly larger than a credit card in width and height and contains a magnetic tape that can hold from 2 to 24 gigabytes of data. It can support data transfer rates of about 2 MBps. Like other types of tapes, DATs are sequential-access media. The most common format for DAT cartridges is DDS (Digital Data Storage).

**Data Mining**
The process of using software tools to review information in a collection of databases and making new connections among otherwise seemingly unrelated information.

**DDP**
Disk-based Data Protection, in which a disk or RAID system is used as a backup system instead of a tape drive or tape library.

**DDS**
Digital Data Storage

**Digital Archiving**
The process of transferring an electronic record (i.e., data) from an author's or originating entity's online composition system to an off-line storage medium in a manner that allows for ready access and retrieval of such record with a high degree of authenticity. Archival formats vary widely and common ones include SGML (text only), ASCII (text only), PostScript or Acrobat PDF (pages, which may include text and/or images), and TIFF or EPS (graphical images of text and/or images).

**Disaster**
Any event—whether natural, accidental or engineered—that exceeds an organization's maximum allowable tolerance of going without information.

**Disaster Recovery Planning**
Advanced planning to minimize data loss and downtime in the event of a disaster.

**DLT**
Digital Linear Tape is a magnetic tape technology originally developed by Digital Equipment Corp. and now sold by Quantum Corp. DLT cartridges provide capacities from 10GB to 70GB. DLTtape technology has been the de facto standard for backup and archiving. The first of the next generation of Super DLTtape drives is the SDLT 220 drive. SuperDLT operates at provides 110 GB (native) capacity per cartridge and 39.6 GB/hour (native) throughput per drive, nearly tripling the data storage capacity and doubling the throughput of previous generation DLTtape technology -- and provides backward-read compatibility with previous generations of DLTtape IV media cartridges recorded on DLT 4000, DLT 7000, DLT 8000, and DLT1 drives. Future generations of Super DLT drives will store more than one terabyte of capacity per data cartridge. Like previous generations of DLTtape drives, Super DLTtape drives generally support most major systems and platforms. The installed base is estimated to be 1.7 million DLTtape drives in use and more than 70 million DLTtape media cartridges storing data. Tens of thousands of Super DLT drives have recently shipped.

**DOD 5012.2**
A set of standards published in 1997 to establish the core requirements for electronic records management programs that vendors dealing with the ODD must follow. It sets forth mandatory baseline functional requirements for Records Management Application (RMA) software used by the DOD Components in the implementation of their records management programs; defines required system interfaces and search criteria to be supported by the RMAs, and describes the minimum records management requirements that must be met, based on current National Archives and Records Administration (NARA) regulations.

**DTD**
Document Type Definition is a formal SGML specification for a document. A DTD defines the structural elements and combinatorial rules that can be used to create instances of documents. The HyperText Markup Language is defined using one of these formal definitions.

**Dublin Core**
Dublin Core Metadata Element Set consists of 15 descriptive data elements relating to content, intellectual property and instantiation. The elements are title, creator, publisher, subject, description, source, language, relation, coverage, date, type, format, identifier, contributor and rights. They are to be supplied by the producer of the resource. The Warwick Framework set out a conceptual approach to implementing the Dublin Core, one of which is embedding the data in an HTML document using the META tag. DC is being widely discussed and there is a growing corpus of implementation projects in over 10 countries. There is a Dublin Core-USMARC mapping. See: http://dublincore.org/documents/dces/

**DVD**
Digital Versatile Disc, an ultra-high capacity, read only storage media in the form of an optically read 4 ¾“disc encompassing audio, video, and computer data.

**Emulation**
Emulation refers to the process of mimicking, in software, a piece of hardware or software so that other processes think the original equipment/function is still available in its original form. Emulation is essentially a way of preserving the functionality of and access to digital information that might otherwise be lost due to technological obsolescence.

**Encapsulation**
Encapsulation is the preserving digital materials by a technique of grouping together a digital object and anything else necessary to provide access to that object. This technique aims to overcome the problems of the technological obsolescence of file formats because the details of how to interpret the digital bits in the object can be part of the encapsulated information. Encapsulation can be achieved by using physical or logical structures called "containers" or "wrappers" to provide a relationship between all information components, such as the digital object and other supporting information such as a persistent identifier, metadata, software specifications for emulation. The encapsulation may be composed of analog and digital components. An example of an analog component would be human readable instructions, such as writing on the outer case of a physical format carrier, to describe how to use the carrier and interpret the outer most layer of the digital component, most likely the wrapper, which will in turn provide the information required to use the rest of the digital information contained. The analogue component of the encapsulated object may change as the carrier needs to be refreshed but ideally there should be no change required of the digital component when it is stored this way. The types of supporting information that should be included in an encapsulation, apart from the digital object itself, are described by the Reference Model for an Open Archival Information System (OAIS) which are based on the report "Preserving Digital Information" from the Task Force on Archiving of Digital Information.

**Encoded Archival Description (EAD)**
The EAD Document Type Definition (DTD) is a standard for encoding archival finding aids using the Standard Generalized Markup Language (SGML). The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress (LC) in partnership with the Society of American Archivists.

**Exabyte**
2 to the 60th power (1,152,921,504,606,846,976) bytes. An exabyte is equal to 1,024 petabytes. This is not yet a commercially-available, practical unit of storage, but does exists in some government and research university R&D labs.

**Fabric**
In the context of a storage area network, the fabric, refers to the detailed make up of the network such as cards and attached devices.

**Fibre Channel**
Fibre-Channel is an interface standard for connecting computers to mass storage devices such as disk drives and tape libraries. It was developed more than a decade after SCSI, which it was intended to replace for high performance applications. Fibre-Channel standard was specified around faster data throughput speeds, and longer distances because of its use of fiber-optic cable. In the meantime, new versions of SCSI have caught up in speed terms, so the performance differences are now blurred. The Fibre Channel transmission scheme specifies signaling and data-handling techniques for a variety of connection media, including coaxial and fiber-optic cable, at speeds ranging from 266 megabits/second to over 4 gigabits/second (Gbps). Speeds up to 1 gigabit per second (1 Gbps) are commercially available. Fibre Channel can support for distances up to 10 km. The standard provides for highly reliable connections and assured delivery of data. Fibre Channel connections are very adaptable and can carry other connection schemes such as SCSI, Internet Protocol (IP), IPI, HIPPI-FP, and audio/video. In a Fibre Channel network, legacy storage systems are interfaced using a Fibre Channel to SCSI bridge. IP is used for server to server and client/server communications. Also known as the Amercan National Standards Institute, National Committee for Information Technology Standards, T11 I/O interface. Note: The difference in spelling of “Fibre Channel” from “Fiberoptic” was done deliberately by the standard creators.

**FCIP**
Fibre Channel over IP. See “Fibre Channel.”

**Fiberoptic Cable**
Network cabling that transmits signals optically with light (rather than electrically as do coaxial, Cat-5 and twisted-pair cables). The light-conducting glass or plastic fibre is called the core and is surrounded by a refractive layer called the cladding that traps light and keeps it bouncing along the central fibre. Outside, both the core and cladding are covered by a coat or jacket of plastic. Fibreoptic cable can transmit clean signals at speeds up to 2 Gbps. Because fibre transmits light, not electricity, it is highly secure from evesdropping. Note: The difference in spelling of “Fiberoptic” from “Fibre Channel.”

**Firewall**
A protective mechanism, usually a combination of hardware and software, designed to protect an internal network from unauthorized access via an external network such as the Internet. Firewalls forbid access to anyone but individuals authorized to remotely access the internal network.

**FTP**
File Transfer Protocol used for downloading files on the Internet.

**GBIC**
GigaBit Interface Converter is the interface module which converts the light stream from a fibre channel cable into electronic signals for use by the network interface card

**Gigabyte**
2 to the 30th power (1,073,741,824) bytes. One gigabyte is equal to 1,024 megabytes. Gigabyte is often abbreviated as *G* or *GB*. This is about one-and-a-half times the capacity of a standard music CD-ROM, or about a million times the capacity of a standard 3.5 inch floppy disk.

**GILS**
Government Information Locator Service was developed in the U.S. and now being adopted in other countries. GILS is a decentralized collection of systems containing databases of GILS records describing location and access information for publicly-available government information resources. Z39.30 is the access mechanism that has been specified for searching these systems, but they can also be searched through the Web. There is a GILS-USMARC mapping, and an SGML profile has been developed for GILS records.

**HBA**
See Host Bus Adaptor.

**Helical Scan**
A tape-recording method that uses a spinning read/write head and diagonal tracks. Proponents claim higher capacity and better data integrity than with traditional recording methods.

**High Availability**
In information technology, high availability refers to a system or component that is continuously operational for a desirably long length of time. Availability can be measured relative to “100% operational” or “never failing.” A widely-held but difficult-to-achieve standard of availability for a system or product is known as "five 9s" availability. For storage, a redundant array of independent disks is one approach. A more recent approach is the storage area network.

**Host Bus Adaptor**
Host Bus Adapter is an interface card that plugs into the computer's bus and connects it to the network. An HBA is an I/O adapter that sits between the host computer's bus and the Fibre Channel loop and manages the transfer of information between the two channels. In order to minimize the impact on host processor performance, the host bus adapter performs many low-level interface functions automatically or with minimal processor involvement. The Host Bus Adapter permits the generic attachment of any particular device to a server running a proprietary operating system for a particular hardware platform. Absent a Host Bus Adaptor, a given operating system may prevent the addition of file redirectors (installable file system) to accommodate various SAN functions. With the Host Bus Adaptor, the operating systems should handle fabric management identically with common discovery via agreed upon standards. The storage/SAN management and other functionalities should be transparent to the native operating system. The ideal Host Bus Adapter would be generic allowing the generic attachment of any particular device. Unique function built into the HBA, by definition, will make it proprietary. Any function required in the HBA for SAN management should be downloadable into the HBA and applicable to all HBAs through a common interface. NOTE: The name adapter (as opposed to controller) is due to the fact that while the SCSI card or circuitry is in the host computer, all of the intelligence is located in the target controller on the peripheral device

**Hot Swappable**
Originally, it was the ability to remove and replace SCSI devices on the bus. There are four "levels" or "cases" of hot-swapping. Case 4 is true hot-swapping as it requires that the bus remain running during the plugging action. Now the term is commonly used for any network or computer component (such as controllers, hard disks, and power supplies) that can be swapped without powering down and without loss of data. Also known as hot plugging.

**HSM**
Hierarchical Storage Management is a software and hardware system that automatically moves files from hard disks to slower, less expensive storage media such as optical or tape. The idea is to go from fast but relatively expensive storage to slower, more economical storage as the need for access to the files diminishes. HSM optimizes the cost and efficiency of various storage media – since storage technologies have different characteristics in terms of access speed, density and cost. It is not economic to use just one type of storage for all data needs. Some data is needed more frequently than other data, so it is possible to manage the data which is stored in a heirarchy which matches the organization's applications and economics, and moves data up or down from RAM, local disk, network attached RAID, optical jukebox, tape library, and web storage as needed.

**HUB**
A hardware device to which nodes connect on star-wired networks. Hubs can be passive, active, or intelligent. A passive hub simply acts as a connection point. An active hub acts as a connection point, but has the ability to regenerate signals. An intelligent hub is one with added capabilities such as the ability to configure the network. A switch can typically perform the same functions more quickly than a hub.

**IDE**
Intelligent Drive Electronics is an interface standard which was developed as a low cost method of connecting a native PC bus to a mass storage device such as a disk drive. For high-availability computing applications and mass storage, it has been replaced by the RAID-controlled SCSI interface which is capable of generating redundant back-ups on separate disk drives.

**InfiniBand**
An interconnect or I/O architecture that connects servers with remote storage and networking devices, and other servers. It can also be used inside servers for inter-processor communication. InfiniBand is a channel-based, switched fabric, point-to-point interconnect, which provides scalability and performance for a wide range of platforms. InfiniBand provides a scalable performance range of 500 MB/s to 6 GB/s per link (entry level to high-end enterprise systems).

**I/O Adapter**
An adapter that converts between the timing and protocol requirements of an intelligent device's memory bus and those of an I/O bus or network.

**IP**
The Internet Protocol or TCP/IP, is responsible for routing packets of data. IP runs at the internetwork layer in the TCP/IP model – equivalent to the network layer in the ISO/OSI Reference Model.

**iSCSI**
A software package which emulates SCSI protocols, but the connection method is via an IP network instead of a direct SCSI compatible cable. Early implementations of SCSI used ribbon cable and industry standard logic levels which limited users to using a SCSI disk installed in their PC or a single array of multiple disks in proximity to the server. SCSI is an intelligent protocol which enables data blocks to be read from or sent at high speed to a storage device such as a disk or tape drive and ISCSI emulates that protocol without using a ribbon cable and industry standard logic levels. Storage management software which was originally written for the well established SCSI standard, can now be used to make a remote disk or tape drive on a network operate just like a local disk. The network can be a local area network such as Ethernet, or even the Internet. The benefit is that users can connect to remote storage devices to replicate data without having to develop customized software.

**JBOD**
Slang for Just a Box Of Disks, a term used for a storage enclosure which is supplied with disk drives pre-integrated. The systems integrator can incorporate their own choice of RAID controller or just use the JBOD as an economic way to add more disk storage.

**Kilobyte**
In decimal systems, *kilo* stands for 1,000, but in binary systems, a *kilo* is 1,024 (2 to the 10th power). Technically, therefore, a kilobyte is 1,024 bytes, but it is often used loosely as a synonym for 1,000 bytes. For example, a computer that has 256K main memory can store approximately 256,000 bytes (or characters) in memory at one time.

**Knowledge Management (KM)**
Capturing, storing, transforming, and disseminating information within an organization, with the goal of promoting efficiency at the least and innovation and competitive advantage at the most.

**LAN**
Local area network. A group of computers and other devices spread over a relatively limited area. They are connected by a communications link that enables any device to interact with any other device on the network and establishes security access to those resources.

**Library**
A multi-tape storage device that loads and unloads the tape drive. The capacity varies by model and some units can contain hundreds of tape slots and one or more tape drives.

**LTO**
Linear Tape-Open (LTO) technology was developed jointly by HP, IBM and Seagate as a Tape Library solution for handling a wide range of backup, archive, and disaster recovery data storage needs. LTO technology is an "open format" technology which is available to other vendors. The Accelis format is the "fast access" implementation of LTO technology. The Ultrium format is the "high capacity" implementation of LTO technology. LTO competes with Super-DLT, DLT, 8mm, 4mm, and 1/4-inch tape drives for streaming data applications such as backup.

**LUN**
Logical Unit Number is a method of expanding the number of multiple devices that can share one SCSI domain ID. Logical Unit Numbers address up to seven devices at each SCSI ID on an 8-bit bus or up to fifteen devices at each ID on a 16-bit bus.

**MARC**
The MARC formats are standards for the representation and communication of bibliographic and related information in machine-readable form as developed by the Library of Congress Network Development and MARC Standards Office. See: http://lcweb.loc.gov/marc/

**Mass Storage**
Refers to various techniques and devices for storing large amounts of data. The earliest storage devices were punched paper cards, which were used as early as 1804 to control silk-weaving looms. Modern mass storage devices include all types of disk drives and tape drives. Mass storage is distinct from memory, which refers to temporary storage areas within the computer. Unlike main memory, mass storage devices retain data even when the computer is turned off.

**Megabyte**
When used to describe data storage, 1,048,576 (2 to the 20th power) bytes. *Megabyte* is frequently abbreviated as M or MB.

**Metadata**
Metadata is structured information that describes resources. It is more commonly thought of as cataloging information either embedded into an electronic file (e.g., embedded in an HTML header), or placed onto a wrapped file (OAIS information package), or as a separate file (e.g., MARC Record) to aid in the search and retrieval of the original file. Typical metadata includes keywords, description, creation date, author and similar information to assist future researchers in finding and retrieving the content being described. There are a wide variety of metadata models and operations supported (search and retrieve; disclosure and distribution; hierarchal storage management). Metadata may be created by the author or by catalogers, curators, archivists, librarians, web site developers, database administrators, volunteers, authors, editors, and other interested persons.

**Migration**
The migration of digital information refers to the "periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation." This term is sometimes used to refer to the transfer of information to non-digital media such as paper.

**Mirroring**
A means of protecting data on a network by duplicating it, in its entirety, on another disk or disks. Mirroring is one of many back-up strategies that can be implemented with RAID security.

**MTTR**
When a computer network or SAN incurs a component failure, Mean Time To Recovery is the time measured in seconds to detect a fault, recover data, and bring applications back online for users.

**NANO CUBIC technology**
NANO CUBIC technology is an ultra-thin layer coating that results in higher resolution for recording digital data, ultra-low noise and high signal-to-noise ratios that are ideal for magneto-resistive (MR) heads. It is capable of catapulting data cartridge and digital videotape to one-terabyte native (uncompressed) capacities and floppy disk capacities to three gigabytes. To help visualize the potential, 1TB can store up to 200 two-hour movies. What sets the NANO CUBIC coating technology apart is its ability to be applied cost-effectively to a mass production manufacturing process while requiring only small modifications to current coating equipment for application. Fujifilm is beginning to work with drive manufacturers to develop new, high capacity magnetic storage products using NANO CUBIC technology.

**NAS**
Network Attached Storage are stand alone storage devices connected directly to a local area network. These devices typically don't contain the full network operating system found in a file server but use a microkernel OS designed to handle specific data read and write functions for a variety of operating systems.

**NIC**
Network Interface Card. This term is mainly used in the context of switch manufacturers. However the same card when plugged into a server may be called an HBA.

**NIST**
The National Institute of Standards and Technology was founded in 1901, and is a non-regulatory federal agency within the U.S. Commerce Department's Technology Administration. NIST's mission is to develop and promote measurement, standards, and technology that enhances productivity, facilitates trade, and improves the quality of life.

**ODBC**
Open Database Connectivity, a standard method of sharing data between databases and programs.

**Open Archives Initiative**
The Open Archives Initiative develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The Open Archives Initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication. Continued support of this work remains a cornerstone of the Open Archives program. The fundamental technological framework and standards that are developing to support this work are, however, independent of the both the type of content offered and the economic mechanisms surrounding that content, and promise to have much broader relevance in opening up access to a range of digital materials. Support for Open Archives Initiative activities comes from the Digital Library Federation, the Coalition for Networked Information, and from National Science Foundation Grant No. IIS-9817416. See: http://www.openarchives.org/

**Open SANs**
The term "open", when used in the IT community, has the perceived meaning of interoperability between heterogeneous servers and operating systems. For example, an open software product in theory can be installed and run independent of the server platform or operating system. Storage Area Networks extend the concept of "openness" by allowing heterogeneous storage and server attachment to the SAN. The "open" elements of a SAN can be summarized into the following: Server operating systems and host bus adapters; Fibre channel interconnect components (bridges, hubs, switches); Storage subsystems and/or devices; Storage abstraction (management appliance or software); SAN management software (network and storage management); and Software to exploit SAN capabilities (backup, copy, etc.).

**Optical disc**
A storage media that is read by means of reflecting a laser beam off of the surface of the disc. Typical optical drives are CD drives and videodisc drives. However, there are many types of optical drives including those that can be written on and re-written on much like floppy discs or computer tape.

**Periodicity**
How often a particular backup is scheduled (Daily, weekly, etc).

**Persistent Identifier**
An unique identifier not dependent upon the location of the resource being sought on the World Wide Web. The current method of discovering and locating resources on the World Wide Web relies on allocating an identifier to all resources. At present these identifiers are Uniform Resource Locators or URLs, and are allocated according to the location of the resource. They can identify the network zone, machine, disk, directory(ies) and the actual name of the file within which the resource resides. Although URLs have been serving the combined purpose of identifying a resource and describing its location for some time now, they are not a satisfactory means of persistently identifying a digital resource. The URL simply points to the current location of the resource. If a resource is moved to a new location, the previous URL is no longer useful, and links to the resource which are embedded in other documents or databases will be redundant. A persistent and unique identifier would preserve access to that resource regardless of its location, as long as it still existed on the Internet.

**Petabyte**
2 to the 50th power (1,125,899,906,842,624) bytes. A petabyte is equal to 1,024 terabytes.

**QIC**
Quarter-Inch Cartridge, pronounced “quick,” a standard for magnetic tape drives. QIC now comes in different cartridge sizes and with various combinations of wider and longer tapes. QIC tapes are among the most popular tapes used for backing up personal computers due its low cost. QIC tapes are divided into two general classes: full-size (also called *data-cartridge*) and minicartridge. The QIC-40 and QIC-80 standards are sometimes referred to as *floppy tape* standards because they are designed to use a personal computer's existing floppy disk drive controller instead of requiring a customized controller. The newest set of QIC standards are based on the Travan technology developed by 3M. The various QIC standards are controlled by a consortium of manufacturers called the *Quarter-Inch Cartridge Drive Standards, Inc.* The term *QIC*, therefore, is used to refer both to the type of tape and to the standards-producing organization.

**RAID**
Redundant Array of Independent Disks (originally, Redundant Array of Inexpensive Disks). RAID hard disk systems use software or firmware to split data across several drives to increase performance and data redundancy. Different RAID levels organize the data across the drives in different ways. An array of many disks can look electronically to the operating system just like a single, bigger disk (called a virtual disk) by having the disks work in parallel by being connected to a RAID controller interface. The combined system can be programmed to provide desirable characteristics such as faster data throughput (for example a 4 disk wide system could have a data throughput capability 4 times faster than a single disk). RAID can also provide fault tolerance, because redundant disks can be added into the array and the data split up in such a way that there is no loss of performance or data availability if any single disk fails. However, RAID cannot speed up access time, which is why some applications use solid state disks.

**RAID Advisory Board**
An industry, open-membership organization, formed in July of 1992, to standardize and classify RAID levels and to assist end-users in making more informed storage procurement decisions. The *RAB* has developed Functional and Performance specifications and established the *RAB* RAID Level Conformance Program and the *RAB* Disk System and Array Controller Classification Program. *RAB* has published many books and articles and sponsored numerous conferences and seminars. Prominent among its publications are the *RAID*book and the *Storage System Enclosure Handbook* . In the area of standards, the *RAB*, working closely with the American National Standards Institute (ANSI), has developed the *SCSI Controller Commands* and *SCSI-3 Enclosure Services Command Set* standards. Over 20 *RAB* members are currently licensed to display the *RAB* Logo and legends indicating that the products identified by the logos have met certain criteria established by the *RAB*. See: http://www.raid-advisory.com/

**Robot**
(1) The "arm" mechanism that moves tapes between the drive and the library slots. (2) The software program a search engine uses to “crawl” across the World Wide Web automatically to collect information such as web page URLs, metatag data and other information for a search engine’s database. Also known as a bot, crawler or spider.

**SAA**
Founded in 1936, the Society of American Archivists is North America's oldest and largest national archival professional association. SAA's mission is to serve the educational and informational needs of more than 3,400 individual and institutional members and to provide leadership to ensure the identification, preservation, and use of records of historical value. See: http://www.archivists.org/

**SAN**
A Storage Area Network is a dedicated, centrally managed, secure information infrastructure, which enables any-to-any interconnection of servers and storage systems. A SAN usually includes multiple servers working off a centralized data store made up of highly reliable and redundant hardware, including RAID. The object is to provide a single point of storage with sophisticated management. Fibre Channel is an important part of the SAN concept, because it works with existing data storage technologies, including SCSI and network connections.

**SAS**
SAN Attached Storage is a term used to refer to storage elements that connect directly to a storage area network and provide file, database, block, or other types of data access services to computer systems. Abbreviated SAS. SAS elements that provide file access services are commonly called Network Attached Storage, or NAS devices.

**SANmark**
The SANmark Qualified Program is sponsored, operated, and promoted as an open industry, conformance test suite development program, within the non-profit Fibre Channel Industry Association (FCIA). This program addresses product interoperability and device level compatibility for companies participating in the program. The program is open to members and non-members of the FCIA organization, and to the entire storage networking industry.

**SCSI**
Abbreviation for Small Computer System Interface, a hardware peripheral interface used most often with storage devices such as disk, CD-ROM, and tape drives. SCSI comes in several incarnations, which can move data at rates ranging from 5 to 40 MBps. More than one SCSI controller can be connected to a SCSI cable for increased reliability. Originally SCSI used logic level (TTL) signals over ribbon cable to daisy chain up to 8 devices, this standard has been enhanced over the years to include newer voltage levels, higher speeds and many more devices. The standard enables large amounts of data to be requested using a small number of intelligent commands. Due to the popularity of this standard, some manufacturers have developed extenders which enable connection over a long distance, converters such as IDE to SCSI, and routers which translate the SCSI interface and commands to fibre-channel compatible signals.

**SCSI Cable Lengths**
The SCSI specifications provide maximum recommended cable lengths for various implementations of SCSI. It is further recognized that in "engineered" installations these cable lengths may be exceeded. Maximum recommended SCSI cable lengths:

| SCSI TYPE – SPEED | MAX CABLE LENGTH |
|---|---|
| SINGLE ENDED - SLOW (FAST-5) | 6 Meters / 19.7 Feet |
| SINGLE ENDED - FAST (FAST-10) | 3 Meters / 9.8 Feet |
| SINGLE ENDED - ULTRA (FAST-20) | 1.5 Meters* / 4.9 Feet |
| DIFFERENTIAL (HVD) - (ANY SPEED) | 25 Meters / 82 Feet |
| LVD - ULTRA2, ULTRA3 or ULTRA4 | 12 Meters / 39.4 Feet** |

* Maximum cable length with eight addresses. May be up to 3 meters with four devices.
**This may be increased to 25 meters (82 Feet) in point to point applications.

**SFP**
Small Form-factor Pluggable (GBIC) provides the same functionality as a regular GBIC but in a smaller and denser physical size. It is used in network switches for fibre-channel, Gigabit Ethernet and Infinband. The newer generation of SFPs operates at faster data transfer rates of 2.5 Gbps and above.

**SGML**
Standard Generalized Markup Language is a formatting language that allows the transfer of information between applications. This is an ISO standard [8879:1986] document definition, specification, and creation mechanism that makes platform and display differences across multiple computers irrelevant to the delivery and rendering of documents. It is an “international” standard for the publication and delivery of electronic information. Commands are embedded within a document so that an SGML-compatible application can interpret them and format the document accordingly.

**SLR**
Trademark of Tandberg Data. It relates to a family of compatible tape drives and media. SLR tape drives offer capacities from 525 MB to 100GB and up to 4TB with SLR automation In March, 2001 - Tandberg Data announced that it had shipped its third million SLR tape drive to the market.

**Snapshot**
(1) A snapshot is a virtual copy of a volume. The snapshot is created instantly. Like the clone, the application should be paused before the snapshot is taken to ensure data integrity. The snapshot operation can be applied to RAID 0, RAID 1, RAID 0+1 or RAID 3/5 volumes. It creates a fully usable copy of a defined collection of data that contains an image of the data as it appeared at the point in time at which the copy was initiated. A snapshot may be either a duplicate or a replicate of the data it represents. (2) The CIM snapshot class. An optional construct that can be used to represent a storage extent that contains either a full copy of another storage extent or the changes to that extent (in the case of a delta before or delta after copy). A CIM snapshot is not equivalent to a volume or file-based snapshot, or a point in time copy. It represents storage used to hold a copied image of an extent, or to hold changes to an extent. See Clones. Clones and snapshots can be used in similar ways. Both can be used for offline backup. In offline backup, the backup is done by a dedicated backup server using the clone or snapshot. This allows the application to continue to run without interruption. In traditional backup, the application is typically unavailable during the backup process. Even with applications supporting "Hot Backup" there is typically performance degradation during the backup process. Clones are most useful for operations that may change the data, such as production testing or data warehousing. Snapshots are most useful for operations that require a quick copy of the data. One case is where snapshots are made on an hourly basis to ensure a quick restore of files that are accidentally deleted or destroyed. Snapshot is required if the volume to be copied is parity RAID (RAID 3/5).

**SNIA**
The Storage Networking Industry Association was incorporated in 1997 as an umbrella organization to serve as the point of cohesion for developers of storage, storage networking, system integrators, application vendors, service providers, and IT professionals. It encompasses all storage technologies and seeks to deliver architectures, education and services that will provide interoperable storage solutions. See http://www.snia.org/about/

**SNM**
Storage Network Management is a set of tools that creates a map of all devices on the storage network and monitors for errors, such as a network or server failure. SNM automates what has been a manual process for some time.

**SOAP**
Simple Object Access Protocol is an RPC mechanism based on XML that sends commands and receives values over HTTP.

**SPC**
The Storage Performance Council is an industry standards organization that defines and promotes storage benchmarks as well as disseminates objective, verifiable performance data and related test tools to the computer industry and its customers.

**Spider**
The software program a search engine uses to "crawl" across the World Wide Web automatically to collect information such as web page URLs, metatag data and other information for a search engine's database. Also known as a robot, bot or crawler.

**SSD**
A Solid State Disk is a semiconductor system which functions just like a disk drive, but which contains no moving parts. The advantages are faster access time and better resilience to vibration and temperature. The disadvantages are higher cost (more expensive per megabyte than disk drives) and larger size (an individual SSD can be mechanically compatible with a winchester disk drive, but because it has lower capacity, more drives will be needed to implement the same storage capacity.)

**SSF**
Formed under the auspices of the Storage Networking Industry Association (SNIA) and the Technical Support Alliance Network (TSANet), the SNIA Supported Solutions Forum (SSF) provides a process for coordinating rapid response technical support for multi-vendor storage networks that use interoperable products from member companies. The organization certifies multi-vendor solutions as interoperable, and it provides a mechanism for technical support professionals from all of the participating companies to take rapid, coordinated action to support joint customers. The program includes 7x24 support and extends to sites around the world.

**SSP**
Storage Service Providers are typically companies which sell storage via the web, but the terms may include other firms providing content delivery with agreed service levels, storage security monitoring, web backup, or on-site storage management.

**Storage Virtualization**
This software manages a variety of proprietary devices so they think alike, thus increasing interoperability among storage devices. The technology helps storage administrators ignore the complexities of networks in favor of logical, data-centric views.

**Striping**
A means of protecting data on a network by spreading it across multiple hard disks. Striping is combined with parity (error-correcting information) to ensure that if any data is lost, it can be reconstructed. Performance is also better than a single drive since the workload is balanced between the array members. This array type is for high performance systems. Identical drives are recommended for performance as well as data storage efficiency. The disk array data capacity is typically limited to the number of drive members times the smallest member capacity. For example, one 1GB and three 1.2GB drives will form a 4GB (4 x 1GB) disk array. The stripe size is a value can be set from 1KB to 1024KB sector size. The size can directly affect performance. In the FastBuild BIOS, the "Desktop" default is 8KB while "Server" and "A/V Editing" size is 64KB.

**SuperDLT**
See DLT.

**Switch**
A network infrastructure component to which multiple nodes attach. Unlike hubs, switches typically have internal bandwidth that is a multiple of link bandwidth, and the ability to rapidly switch node connections from one to another. A typical switch can accommodate several simultaneous full link bandwidth transmissions between different pairs of nodes.

**Tape Drive**
A device, like a tape recorder, that reads data from and writes it onto a tape. Tape drives have data capacities of anywhere from a few hundred kilobytes to several gigabytes. Their transfer speeds also vary considerably. Fast tape drives can transfer as much as 20MB (megabytes) per second. The disadvantage of tape drives is that they are *sequential-access* devices, which means that to read any particular block of data, you need to read all the preceding blocks. This makes them much too slow for general-purpose storage operations. However, they are the least expensive media for making backups.

**Terabyte**
2 to the 40th power (1,099,511,627,776) bytes. This is approximately 1 trillion bytes, or about 50 times the size of a typical PC disk drive. This unit of storage is most commonly used for backup devices such as tape drives and tape libraries connected to a SAN

**Technological Obsolescence**
Technological obsolescence is the result of the evolution of technology: as newer technologies appear, older ones cease to be used. For example, new media for storing digital information rapidly replace older media and reading devices for these older media become no longer available. Newer versions of software constantly render older versions obsolete and the hardware required by this software also changes over time. Consequently, information which relies on obsolete technologies becomes inaccessible. Currently, it seems that the lifetime of digital storage media generally exceeds the life of the technology that supports it. Strategies for dealing with technological obsolescence include: migration of digital information to technologies from which they are accessible, the emulation of obsolete systems, and the preservation of obsolete technologies.

**TIFF**
Tagged Interchange File Format; a format used for storing graphics. A TIFF file essentially contains high-resolution bitmapped information, but since compression factors are used, the files are smaller than EPS versions. Commonly used for imaging paper records so they can be indexed, retrieved and re-printed using a computer.

**Travan**
A tape technology that evolved out of the industry-standard QIC. Travan delivers more capacity than previous QIC formats. With capacities ranging from 400MB to 20GB, Travan formats are best for backing up individual machines or small servers. ADD: A magnetic tape technology developed by 3M Corporation that allows for higher data densities. Travan has been standardized by the QIC consortium, and is backward compatible with older QIC standards. This means that Travan tape drives can read and write older QIC tapes a well as the newer high-capacity Travan tapes. The following table shows tape capacities (uncompressed) and QIC compatibilities for the different Travan levels, 1 to 4.

| Travan Level | Uncompressed Storage | QIC compatibility |
|---|---|---|
| TR-1 | 400 MB | QIC-80-MC |
| TR-2 | 800 MB | QIC-3010-MC |
| TR-3 | 1.6 GB | QIC-3020-MC |
| TR-4 | 4 GB | QIC-3095-MC |

**Ultra SCSI**
Also called "Fast-20". An enhancement of SCSI that results in doubling the FAST SCSI data throughput speeds to 20 MBytes/sec for 8-bit and 40 MBytes/sec for 16-bit. Reduces maximum allowable single-ended SCSI cable length to 1.5 meters (4.9 feet) for five to eight addresses and 3 meters (9.8 feet) for four or fewer addresses. Maximum allowable differential (HVD) SCSI cable length is 25 meters (82 feet). Defined in the SPI document of the SCSI-3 specification.

**Ultra2 SCSI**
Also called "Fast-40". An enhancement of SCSI that results in doubling the Ultra SCSI data throughput speeds to 40 Megatransfers/sec (40 MBytes/sec for 8-bit and 80 MBytes/sec for 16-bit). The SCSI specification recognizes only LVD Ultra2 SCSI and single-ended is not defined at this speed. Maximum allowable Ultra2 cable length is 12 meters (39.4 feet) with more than two active SCSI IDs or 25 meters point to point. Backward compatible through the single-ended interface using Ultra or slower speeds and cable lengths.

**Ultra3 SCSI**
Also called "Fast-80" or "Ultra160". Described in SPI-3, Ultra160 again doubles the data throughput speeds to 80 Megatransfers/sec (160 MBytes/sec for 16-bit). Single-ended not defined at this speed. Requires LVD signalling. SPI-3 also obsoletes HVD, SCAM and the 8-bit and 32-bit wide bus. U160/m is a sub-set of Ultra160 SCSI that includes double transition clocking, CRC and domain validation. Maximum allowable Ultra3 cable length is 25 meters (82 feet) point to point or 12 meters (40 feet) with more than two active SCSI IDs. Backwards compatible through the single-ended interface using Ultra or slower speeds and cable lengths.

**Ultra4 SCSI**
Also called "Ultra 320" or "Fast-160". Again doubles the data throughput to 320 MBytes/sec. Defined in the SPI-4 document of the SCSI-3 specifications, Ultra 320 requires LVD signaling. To reduce overhead and increase data throughput Ultra 320 will include QAS and Information Units (packetizing). Maximum allowable Ultra4 cable length is 12 meters (39.4 feet) with more than two active SCSI IDs and up to 25 meters (82 feet) point to point. Ultra320 is backward compatible through the single-ended interface using Ultra or slower speeds and cable lengths. As with Ultra3 SCSI, single-ended is not defined at Ultra4 throughputs and HVD is obsolete.

**Virtual Disk or Virtual Hard Drive**
A virtual disk is an array of disks that appears as a single physical disk from the operating system's viewpoint. A set of disk blocks presented to the operating system as a range of consecutively numbered logical blocks with disk-like storage and I/O semantics. Sometimes referred to as a Logical Disk.

**VXA**
The trademarked name for a tape technology developed by Ecrix.

**WORM**
When used in all capital letters, *WORM* is an acronym for *write once, read many,* an optical disk technology that allows you to write data onto a disk just once. After that, the data is permanent and can be read any number of times. Unlike CD-ROMs, there is no single standard for WORM disks, which means that they can only be read by the same type of drive that wrote them. This has hampered their acceptance, although they have found a niche market as an archival media. WORM is also called CD-R.

**Wrapper**
(1) In data transmission, a wrapper is the data that is put in front of or around a transmission that provides information about it and may also encapsulate it from view to anyone other than the intended recipient. A wrapper often consists of a *header* that precedes the encapsulated data and the *trailer* that follows it. (2) In database technology, a wrapper can be used to determine who has access to look at or change the data that is wrapped. It is useful in digital archiving since a wrapper may be used to prohibit alterations to an electronic record thus preserving its “authenticity.”

**XML**

Extensible Markup Language is a simple, reduced subset of SGML designed (in 1996) for ease of implementation and interoperability with both full SGML and HTML. Currently a draft meta-language application profile, it is simpler than SGML (reducing a 500-page reference to 26 pages). Unlike HTML, XML supports (optionally) user-defined tags and attributes, allows nesting within documents to any degree of complexity, and can contain an optional description of its grammar for use by applications that need to perform structural validation. Every valid XML document will be a conformant SGML document. Not backward compatible with HTML documents, although those conforming to HTML 3.2 can easily be converted. Not intended to supplant HTML but to complement it. The XML character set is Unicode. XML is being widely discussed currently and future releases of MS Internet Explorer and Netscape browsers may be XML-enabled. See: http://www.w3.org/XML