

Enterprise Data Archiving



Report to the Joint Legislative Oversight Committee on Information Technology

Chris Estes

State Chief Information Officer

Office of Information Technology

December 2013



This page left blank intentionally



Contents

- Legislative Request1
- Executive Summary1
- Introduction2
- Policy Considerations.....3
- Proposed Data Archiving System.....4
- Scope of Data Archiving System4
- Enterprise Features5
- Roadmap.....5
- Issues & Cost.....8
- Next Steps.....8
- Appendix.....9



This page left blank intentionally



Legislative Request

Section 7.11 of Session Law 2013-360 directed the State Chief Information Officer (State CIO) to investigate the feasibility of creating an enterprise data archiving system for state agencies that will:

- (i) Allow for the effective management of data from multiple sources;
- (ii) Provide for efficient, timely responses to discovery requests and investigations; and
- (iii) Ensure real-time state agency access to and use of archived files.

The provision in the 2013-2015 budget bill also required a report to the Joint Legislative Oversight Committee on Information Technology by December 1, 2013, and authorized the State CIO to implement the system upon approval of a business case by the Office of State Budget and Management.

The full text of the legislation can be found in Appendix A.

Executive Summary

State government generates thousands of digital records and information in its daily operations and in providing services to citizens. Records and publications created by state government are public although some may be confidential or have confidential and non-confidential information co-mingled.

The volume of electronically stored records is growing rapidly along with as the need to store it. Permanently retaining all data would be prohibitively expensive, even with the rapidly falling cost of data storage. At the same time, state law and rules require retention of some data for varying amounts of time. Some must be retained permanently.

The discipline of records management, sometimes called information life cycle management, provides some answers to this conundrum. It focuses on records life cycle from creation to disposition—either held permanently, transferred to the State Archives or destroyed. It also includes policies, processes, technologies and tools to support these life cycles.

In preparing this report, the State Chief Information Officer's office reviewed current records management capabilities at the Department of Cultural Resources and various in-progress initiatives, and consulted market research analysts. Staff also reviewed relevant laws and policies. Based on that analysis, with the ultimate goal to reduce cost and risk and still meet regulatory compliance, the State CIO recommends the following approach.

- 1) For archiving, discovery and investigation of permanent legal and/or historical (non-active) records, publications, social media postings and the state's website, the state should continue to use statewide solutions managed by the Department of Cultural Resources.
- 2) For archiving, discovery and investigation of active records, the state should adopt a roadmap with two-phase approach; with phase one delivering results within 18 months and covering a



majority of the archiving needs. The initiative under phase one is already in progress, as a part of larger program to transition state government to a cloud-based communication and collaboration platform. The State CIO's Office is working with the Department of Cultural Resources to ensure this initiative conforms to statewide records management guidelines.

Phase one addresses an overwhelming majority (80% to 90%) of archiving needs. After phase one, the state should re-evaluate cost benefits to address remaining e-discovery needs and weigh the benefit against other business priorities. Based on the cost-benefit analysis, the state should make a decision about proceeding with phase two.

One note: e-discovery and archiving are often used interchangeably, but they are not the same. Archiving is a more general term for a searchable system of stored records. E-discovery is much more specific, referring specifically to electronic data that is located, secured and searched as potential evidence in a civil or criminal court. The two systems have different rules and requirements. For this report, we use the more general term of archiving.

Introduction

In the past decade or more, North Carolina government automated major functions and processes to deliver greater value. As a result, larger numbers of business communications and records have gone digital. In addition, the volume, variety and speed of digital information being created has increased, resulting in huge volumes of data. Because many of these records are the property of the people, the state has an obligation to maintain and make them available on request. Laws and an Executive Order governing public records include:

- a) G.S. § 132 establishes a definition of public record and entitles any citizen to examine public records. The statute also directs the Department of Cultural Resources to provide and administer records management services for state agencies and local government.
- b) Executive Order 12, signed by Governor Pat McCrory, directed agencies to archive email communications for five years and set other rules for email retention.

Federal statutes and rules also come into play.

Under state law, public records generally include, but are not limited to all documents, papers, letters, maps, books, photographs, films, sound recordings, magnetic or other tapes, electronic data-processing records, emails, instant messaging, and social media postings. For the purpose of this report, the records are broadly categorized as active records and non-active records. Active records refer to current records, while non-active records refer to permanent records of historical importance or legal permanent value. The Department of Cultural Resources manages permanent records and publications for all state government entities, including counties and cities. Most archiving requests and investigations pertaining to non-active records are currently satisfied by the Department of Cultural Resources, using their current archival system.



Active records remain with the source or originating agency or department; once they become historical records or publications, they are turned over to the Department of Cultural Resources. Most public record requests are for active records. There is an opportunity for a statewide solution for archiving of active records that could reduce the time spent complying with public records and e-discovery requests and share resources. This report focuses on archiving, discovery and investigation of active records.

State CIO staff informally surveyed several agencies and learned that many public record requests are for email communication and documents related to a project or a topic. As a result, we suggest an initial focus on email, instant messages, and project-related documents.

Work is already underway on a project that would begin addressing this need. The Office of Information Technology Services (ITS) currently provides email services to about 60,000 users from various state agencies, boards and commissions. This is an on premise service, with the infrastructure residing in ITS data centers. As a part of the state's effort to fix and modernize IT, ITS is planning to move from on premise email solution to a cloud-based communication and collaboration platform that will provide

- Email,
- Instant messaging (chat),
- Web conferencing,
- Enterprise files repository for storing all types of documents and files,
- E-discovery capabilities for email, instant messaging and enterprise file repository.

The platform will provide archiving and e-discovery capabilities as well. The initiative called "MS Office," was funded by the General Assembly in the IT Reserve Fund.

Policy Considerations

The State CIO is actively working with various agencies on this initiative and has made significant progress. However, there are major challenges and policy considerations, which are expected in any enterprise-wide program.

Technology implementation is often easier than making sure people, processes, and technology are aligned with the business needs to deliver desired results.

Governance is one of the main challenges. Information technology in state government remains federated, with each agency making an independent decision with regards to process, licensing model, procurement and how it addresses e-discovery needs. Adoption of uniform process across agencies often requires agencies to give up agency specific tools and practices and agree on common process, common licensing model, and ability to share unused licenses across agencies.



A fair cost allocation and charge back model is also the key to enterprise-wide adoption. Since each agency is at a different level of maturity in terms of archiving, records management and e-discovery process, adoption of common records management process requires education and training. A strong governance model with agency representation is required to address these enterprise-wide decisions.

Proposed Data Archiving System

The proposed archiving and solution will be delivered as a cloud-based communication and collaboration platform. ITS has not made a decision about the tools or the vendor for the MS Office initiative. ITS will leverage the Innovation Center's "try before you buy" approach to try the product and understand if it meets the business needs. This will give the state an opportunity to see the benefits before making any investments.

Scope of Data Archiving System

Once fully implemented, the cloud-based communication and collaboration initiative will provide archiving and e-discovery for email, instant messaging, web conferencing and file share for over 60,000 users from 26 state government agencies, boards and commissions listed below:

- 1) Department of Administration
- 2) Department of Commerce
- 3) Department of Cultural Resources
- 4) Department of Environment and Natural Resources
- 5) Department of Health and Human Services
- 6) Department of Public Safety with the exception of Law Enforcement Division
- 7) Department of Revenue
- 8) Department of Transportation
- 9) Office of the State Controller
- 10) Department of Agriculture
- 11) Department of Insurance
- 12) Office of the Lieutenant Governor
- 13) Department of Labor
- 14) Information Technology Services
- 15) Office of State Budget and Management
- 16) Office of Human Resources
- 17) Office of Administrative Hearings
- 18) Alcohol Beverage Control Commission
- 19) Banking Commission
- 20) Credit Union Division
- 21) Industrial Commission
- 22) North Carolina Cemetery Commission
- 23) North Carolina State Board of Opticians
- 24) Public Staff Utilities Commission
- 25) State Board of Elections
- 26) Wildlife Resources Commission



State agencies that will not be included initially are the Secretary of State, the State Auditor, the State Treasurer, the Administrative Office of the Courts, the university system, the Department of Justice and the General Assembly. The courts, university system and General Assembly are not required to use ITS services, but may do so if they choose. The Department of Public Instruction and the community college system already use a cloud-based communication and collaboration offering designed for education. Agencies involved in this initiative comprise a majority of the Executive Branch and Council of State.

The State CIO's Office will work with the Department of Cultural Resources (DCR) to manage active records and assign records retention policies to records/information; so the records that become permanent can be easily transferred to DCR.

Enterprise Features

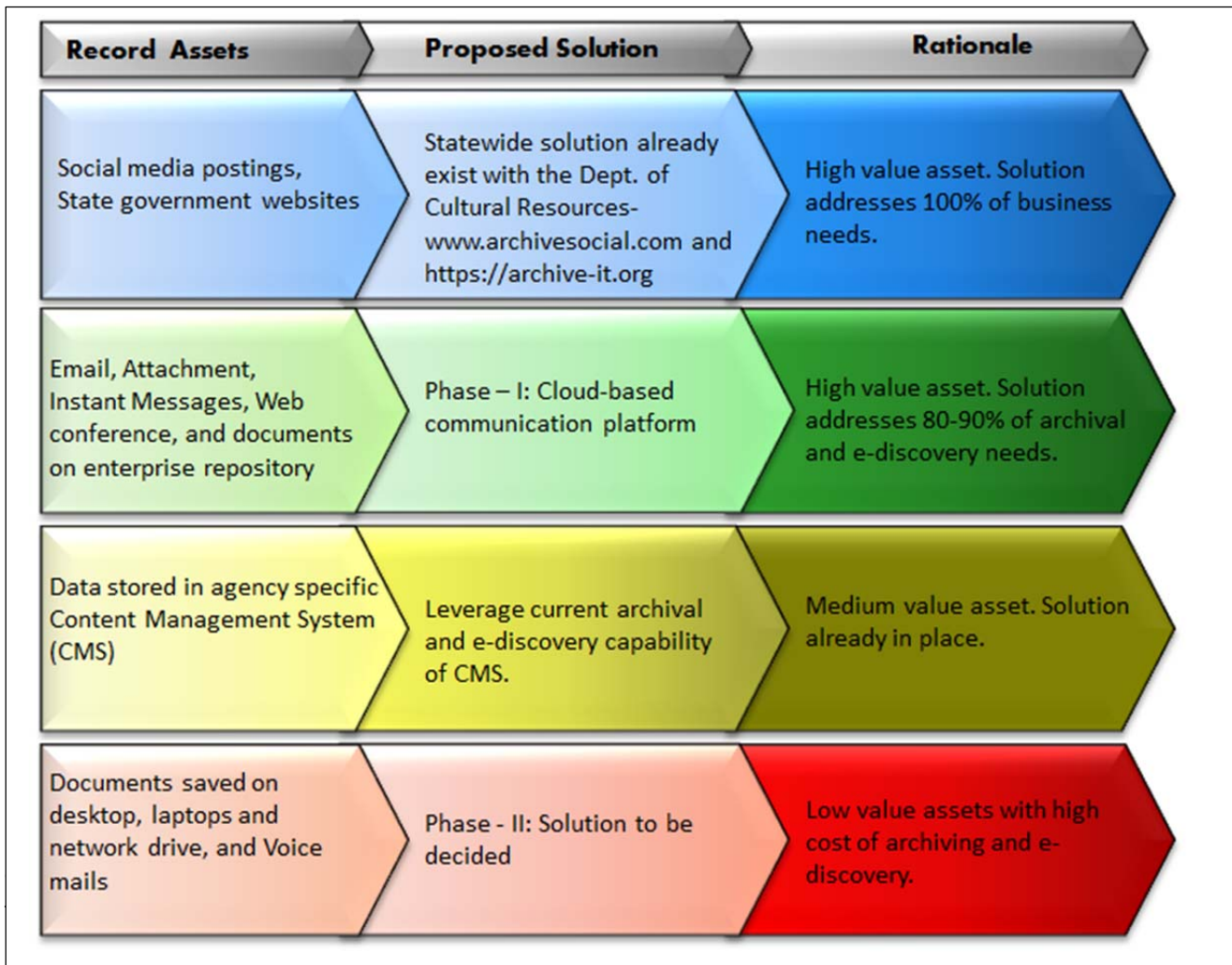
The statewide system will be hosted in the cloud and used as Software as a Service (SaaS) solution. It will have multi-tenancy, with each agency's data segregated from other agency. Each agency will have access to its own data and the ability to perform its own archival and e-discovery. The solution will have scalability, redundancy, high availability and disaster recovery built into it. After successful "try before buy" pilot phase, the state will negotiate with the vendor to meet its needs.

Roadmap

To develop a roadmap, we listed elements that constitute a record, and applied the Pareto principle, also known as 80/20 rule, to address a majority of archiving and e-discovery needs. The figure 1 below shows list of major record assets and associated solutions.

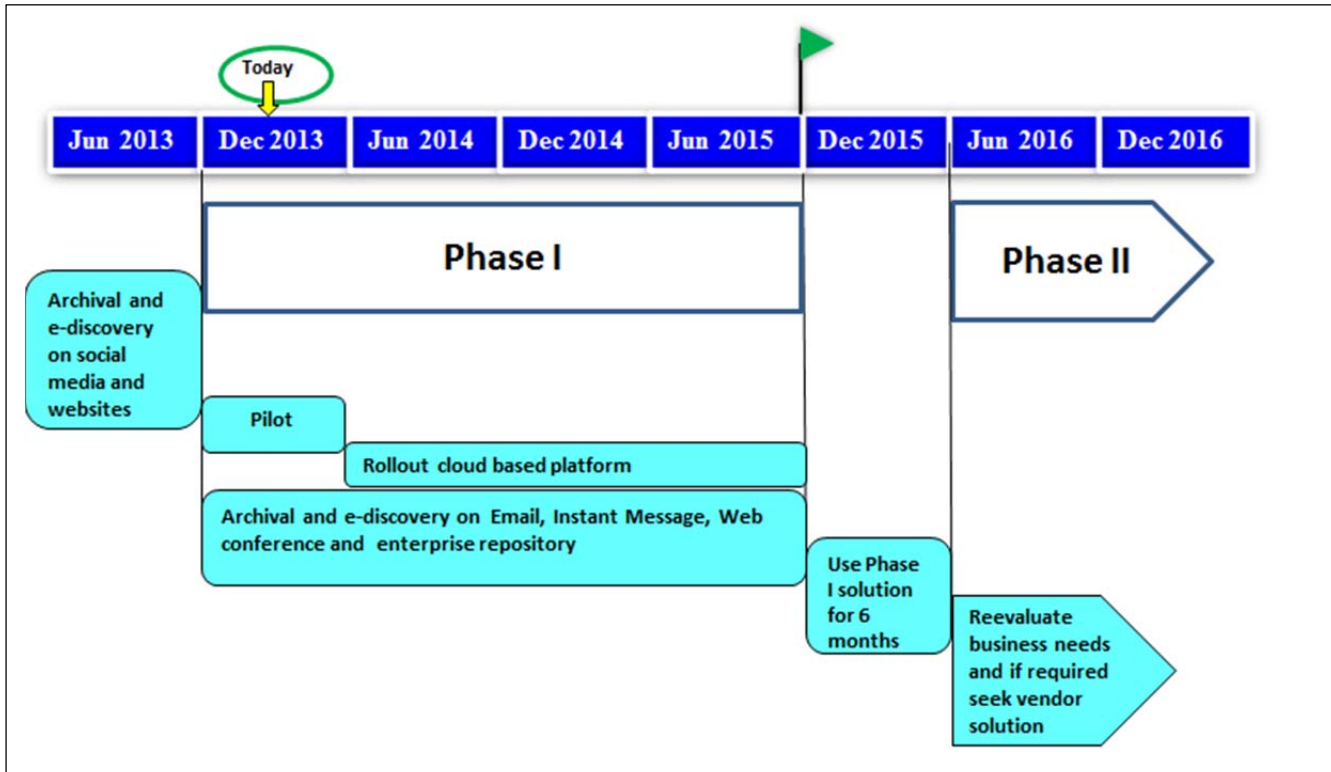


Figure 1: Records Assets and Proposed Solutions



As shown in Figure 1 above, social media postings and state government websites from agencies are currently archived upon request. The solution is managed by the Department of Cultural Resources using ArchiveSocial.com and Archive-It.org. The state should continue to use these solutions for website and social media archiving needs.

Figure 2: E-discovery Roadmap



Phase one of the roadmap includes a pilot with cloud-based communication platform. The pilot is currently in progress with about 200 users. The intent of the pilot is to test the product and make sure it will work before the state invests money to buy the product. The pilot is scheduled to be completed by end of December 2013. After a successful pilot the next step is the enterprise wide rollout of the solution. The rollout process is expected to be complete by end of fiscal year 2015. The solution should deliver archival and e-discovery capability for majority of the records artifacts.

After completing phase one, the state should use the solution for six months, determine whether there are unmet business needs and decide if the remaining records artifacts should be addressed in phase two. Phase two would consist of evaluating a vendor provided solution to archive and run e-discovery on records that are not easy to track and maintain. If the cost of phase two justifies the value delivered, then the state should proceed with phase two.

Another statewide initiative already in-progress and related to data archiving is GDAC, Government Data Analysis Center. It is a program and consists of two major projects.

- NCFACS - North Carolina Financial Accountability and Compliance Technology System
- CJLEADS - Criminal Justice Law Enforcement Automated Data Services

GDAC consists of statewide data integration and business intelligence initiatives. It provides a central repository to store, manage and analyze a broad array of information from all state government agencies. The central repository reduces duplicative data integration design and development work,



data storage and repetitive maintenance and support. The repository can serve as a source for data archiving and can provide timely responses to discovery requests and investigations.

Issues & Cost

As outlined in the roadmap, this effort is broken into very distinct initiatives. Some of the initiatives are already in place and the state has embraced them. There is no additional cost associated with them. The Department of Cultural Resources is already working on improving its records management system and seeking commercial off-the-shelf (COTS) solution with enhanced e-discovery capability. In addition, the department is working with ArchiveSocial and providing input to influence next release of the product.

The cloud-based communication platform initiative is already funded by the General Assembly in the IT Reserve Fund.

Several agencies have agency specific Content Management Systems (CMS) in place; and there is no further cost associated with them. Based on product maturity, CMS usually have archive and e-discovery capabilities.

Phase two is a proposed phase and its business case and funding should be dealt with only after phase one is fully implemented and the state feels there is unmet business need.

Next Steps

The state should focus on phase one for delivering a cloud-based communication platform on a statewide basis for the 26 agencies involved.

After phase one implementation, the state should use the solution for at least six months, before reevaluating additional archival and e-discovery needs.



Appendix

Here is the text of the data archiving provision in Session Law 2013-360.

STATE INFORMATION TECHNOLOGY DATA ARCHIVING

SECTION 7.11.(a) The State Chief Information Officer (CIO) shall investigate the feasibility of creating an enterprise data archiving system for State agencies that will (i) allow for the effective management of data from multiple sources; (ii) provide for efficient, timely responses to discovery requests and investigations; and (iii) ensure real-time State agency access to and use of archived files. The system shall be financed only by savings accrued as a result of the project.

SECTION 7.11.(b) By December 1, 2013, the State CIO shall report to the Joint Legislative Oversight Committee on Information Technology and the Fiscal Research Division on the results of the feasibility assessment.

SECTION 7.11.(c) Subsequent to making the report required by this section, and only if the State CIO has developed a business case that is validated by the Office of State Budget and Management, then the State CIO may initiate the development of an enterprise data archiving system.

