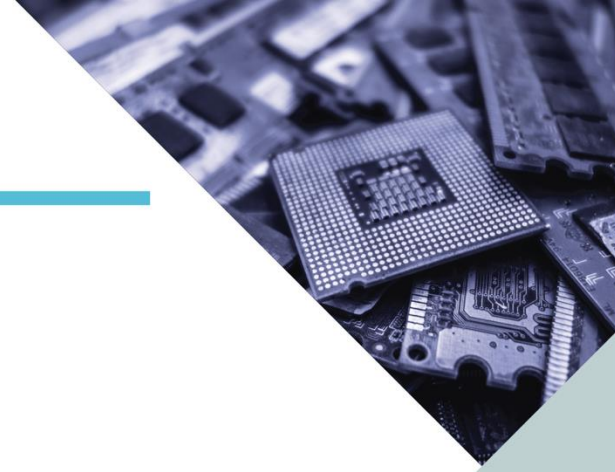


TEXAS DIGITAL STORAGE STUDY

2018

TEXAS DEPARTMENT OF INFORMATION RESOURCES
IN COLLABORATION WITH
TEXAS STATE LIBRARY AND ARCHIVES COMMISSION

Texas Digital Storage Study

Note: In this report "state agency" means a board, commission, office, department, council, authority, or other agency in the executive or judicial branch of state government that is created by the Texas Constitution or by statute. The term does not include institutions of higher education.

EXECUTIVE SUMMARY

Information is a state asset like property and buildings are assets. Loss of information is a liability that can be disastrous to the health of the state. As governments and organizations contend with the mass proliferation of information in both paper and electronic formats, they must examine all the components needed to manage the intellectual contents of those records, provide for the digital storage and transmission of those contents, and ensure access to those records over the short and long-term. This work involves establishing policies and practices to deploy and manage an effective information technology (IT) and infrastructure for those records, implementing policies and standards to appropriately categorize and retain records, classifying records to ensure the contents of the records are described and ultimately accessible, establishing protocols for access and use of those records, and developing and implementing practices for the disposition, retention or archival preservation of records.

All of these tasks rely on the availability of a robust mechanical or IT infrastructure of resources, policies, and standards as well as of a knowledge management framework to organize, classify, retrieve, and preserve information. The Texas Department of Information Resources (DIR) works on providing a secure and robust IT infrastructure for the state, while the Texas State Library and Archives Commission (TSLAC) provides the means to create and manage the knowledge architecture for the state's record resources. Together, these related but separate domains of action provide the broad policy apparatus to manage state informational resources and infrastructure.

A growing area of concern for both DIR and TSLAC is that of electronic records storage and management. Data is the multitude of smaller discrete bits of information that make up our electronic information resources. The volume, malleability, and vulnerability of data is changing the way organizations operate. The proliferation of data has been attributed to software, the Internet, mobile devices, cloud services, social media, and the "Internet of things" where everyday devices are collecting and distributing data. It is no surprise then that the state of Texas stores petabytes of data costing millions of dollars each year.

As such, the Texas Legislature, through House Bill 8 (85R), 2017, required DIR and TSLAC to conduct a study on state agency digital data storage and records management practices and the associated costs to this state. This study includes:

- current digital data storage practices of state agencies in this state;
- costs associated with those digital data storage practices;
- digital records management and data classification policies of state agencies and whether the state agencies are consistently complying with the established policies;
- whether the state agencies are storing digital data that exceeds established retention requirements and the cost of that unnecessary storage;
- the adequacy of storage systems used by state agencies to securely maintain confidential digital records;
- possible solutions and improvements recommended by the state agencies for reducing state costs and increasing security for digital data storage and records management; and
- the security level and possible benefits of and the cost savings from using cloud computing services for agency data storage, data classification, and records management.

In a decentralized IT environment, agencies classify, clean, store, and archive their data based on agency needs, regulatory requirements, and data quality criteria. Records management and

data classification policies in Texas agencies exist; however, some agencies are not always judicious about deleting records that are no longer needed. Effective records management practices can reduce costs associated with unnecessarily stored data. A better understanding of the types of data agencies store, how often and for what purpose, better inform how we can use data to make government more efficient. The costs per gigabyte varies based on the technology solution each agency uses to store its data and the ways in which the agency is retrieving and accessing the data on an ongoing basis.

FINDINGS

- Texas agencies estimate having a total volume of 15 petabytes of data at a cost of approximately $42 million in FY 2017.
- 81% of agencies state their storage solutions allow the agency to meet the service delivery and security requirements of their data.
- Agencies indicate few have experienced security or cost benefits for cloud computing in data storage, data classification, or records management.
- Almost half (45%) of agencies state they store digital data beyond its established retention requirements, but only 22% track the volume of digital records disposed.
- Cost, competing priorities, and agency culture were cited by agencies as some of the largest barriers regarding electronic records and digital data storage.

RECOMMENDATIONS

1. Implement a statewide information governance coordinator position within TSLAC's State and Local Records Division to prioritize information governance and increase resources and outreach for state agency records management programs.
2. Require implementation of data classification, data security, and data retention requirements at project initiation.
3. Establish a periodic training schedule requirement for all agency staff (new hire and ongoing) regarding the agency-specific records retention schedule and records management program. Agencies may use training materials provided or made available by TSLAC.
4. Require DIR, in collaboration with TSLAC, to develop a strategy for agencies to transfer appropriate eligible archival electronic state records to the Texas Digital Archive and to leverage the Statewide Technology Center's long-term non-archival storage options, in coordination with records management and archival practices.

METHODOLOGY

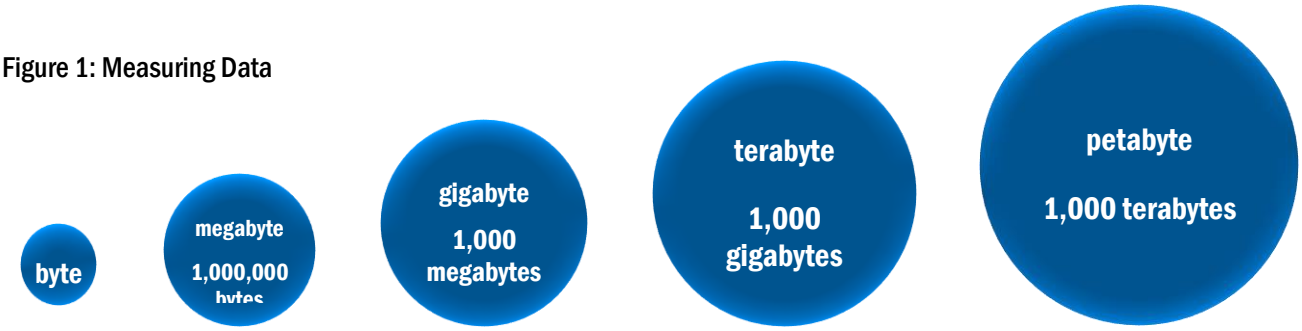
DIR developed this report in collaboration with TSLAC. Information about current agency practices and initiatives were collected via the 2018 Information Resources Deployment Review (IRDR), a self-assessment of the information resources environment and compliance with state IT rules and statutes of 78 agencies. DIR and TSLAC also hosted a facilitated session and continuous engagement with agency volunteers who provided insight on their successes, challenges, and recommendations regarding digital data storage and records management practices. To meet the requirements of HB8, the scope of this report excluded paper records and was limited to digital data and electronic records.

OVERVIEW

The implications of the vast amounts of data are important for state agency business needs and governance processes. As data is created and used for analytics and information, a more nuanced approach is necessary to determine what information is relevant and useful. A better understanding of the types of data and information agencies are storing, how often and for what purpose, will better inform how we can use data to make government more efficient.

Texas agencies store approximately 15 petabytes of data. Table 1 shows the size volume difference with the more commonly quoted storage sizes. To further put it in perspective, one petabyte equals 20 million four-drawer filing cabinets filled with text.

Figure 1: Measuring Data



Source: NIST and IEC International Standard.

DIGITAL STORAGE PRACTICES

Current law requires agencies to establish and maintain a records management program for the creation, use, maintenance, retention, preservation, and destruction of state records, including electronic records. The records management program in each agency should include:

- the development of records retention schedules
- the management of filing and information retrieval systems in any media
- the adequate protection of state records that are vital, archival, or confidential according to accepted archival and records management practices
- the economical and effective storage of inactive records
- control over the creation and distribution of forms, reports, and correspondence
- maintenance of public information in a manner to facilitate access by the public

TSLAC sets forth rules for the creation, protection, maintenance, and storage of electronic state records. While TSLAC establishes the minimum standards and procedures for electronic state records practices, each agency is required to establish and implement their own electronic state records management policies and procedures. Records management programs vary agency by agency and agencies employ a variety of storage solutions for the management of their data.

LOCALLY MANAGED STORAGE

Locally managed storage, also referred to as on-premise storage, involves data stored on servers, computers or other devices within and managed by the agency. Because the data is locally stored and managed, agency staff have both physical and logical access to the data and have direct and exclusive control of the acquisition, configuration, management, and security of the computing infrastructure and data. Locally managed storage is used because:

- it allows for physical control over data backup
- by storing critical data in-house, no third party has access to that data
- there is no dependency on an Internet connection for access to data
- it can be more cost-effective for small to mid-sized data storage requirements

However, locally-managed storage requires agency staff to ensure the security, backup, and retention of digital records, which is often a cost not calculated when assessing the total cost of ownership. Moreover, the content being locally managed is often on legacy systems, portable devices, desktop documents, e-mail, social media, text messages, or repurposed content.

TEXAS DATA CENTER SERVICES

The Texas Data Center Services Program (DCS), a private cloud dedicated to state agencies and other governmental entities as defined by Government Code, Chapter 2054, Subchapter L, provides a managed data center services solution. This program provides infrastructure, including various data storage options, and support services for agencies to implement their specific solution. Because DCS is a fully managed services solution, state agencies are able to be more agile in responding to the exponential growth of their storage needs. DCS makes the necessary investments in hardware, software, and IT staff to meet customer needs, so that customers can focus on their core missions. DCS offers:

- two hardened data centers designed to meet the highest standards of security and disaster recovery
- multiple-vendor model to provide the most efficient, cost-effective data management solution possible
- tiered service levels to accommodate the differing needs and resources of its customers
- active, back-up and redundant services for agency data

In March 2017, Hybrid Cloud Services were introduced to the DCS program to provide customers expanded public cloud storage options, while meeting the business, security, and regulatory requirements of Texas state government. Building upon the existing DCS private cloud, the DCS hybrid cloud enables data and compute residing in the state's consolidated data centers to connect directly with data and compute residing in multiple commercial or public clouds.

PUBLIC CLOUD STORAGE

Public cloud services provide elastic, low cost computing resources (networks, servers, storage, and application services) where users pay for access on-demand and can often times be acquired and implemented faster than the state's private cloud resources. In the public cloud, data is held and managed by a third party through applications available over the Internet. Types of public cloud data storage models include pure public cloud storage where an organization's data is wholly managed by a third-party public cloud provider or hybrid cloud storage which is a combination of the public and private cloud storage.

Cloud storage features include:

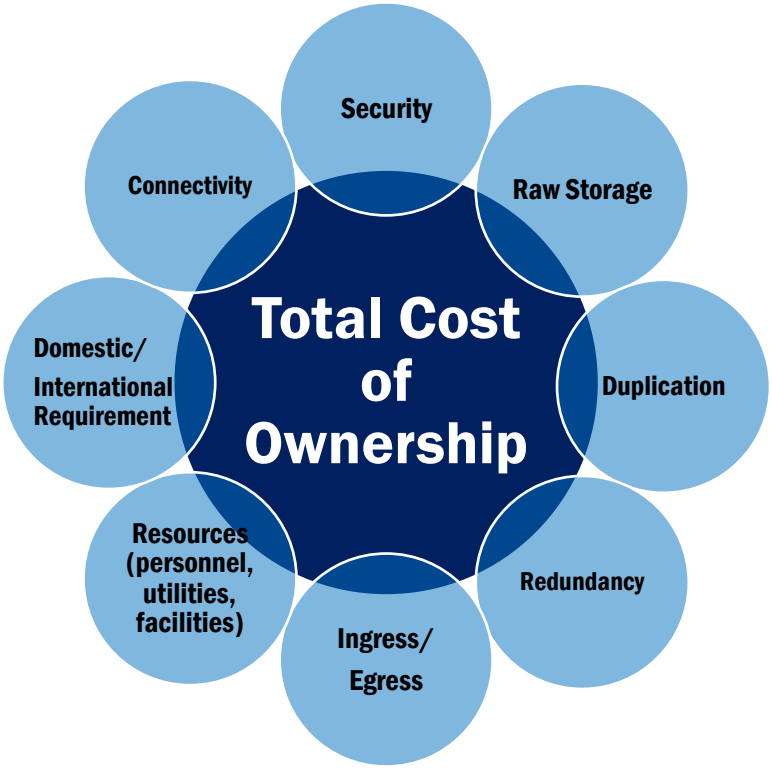
- Accessibility: users can access data stored on the cloud from anywhere with Internet access
- Data recovery: storing data offsite helps ensure business continuity
- Cost: organizations can reduce spending on their own storage equipment by using remote storage that is maintained by a cloud provider
- Storage becomes an operational expense versus a capital investment

- The technology for backup and security undergoes regular refresh as the technology evolves automatically by the cloud provider

COST AND BUDGET

Agencies make IT purchasing decisions according to their unique business needs, resources, and budgeting practices. When comparing different options, agencies should consider the total cost of ownership, which includes direct costs like infrastructure, connectivity, and security, as well as indirect costs, including personnel and utility services. Costs also vary greatly depending on the length of time the asset will be deployed and the extent of usage during its deployment. Figure 2 shows some of the contributing factors to total cost of ownership for digital storage.

Figure 2: Factors for Total Cost of Ownership for Digital Storage



TOTAL COST OF OWNERSHIP

When agencies procure storage solutions, they first consider their requirements for accessing the data, with higher transaction volumes requiring higher-cost, higher-performance solutions, and lower transaction volumes requiring lower-cost, lower-performance solutions. Storage solutions vary in cost based on storage hardware performance capacity, which refers to the rate at which transactions between data and end-user applications can take place. Data accessed in high volumes and at high frequencies, often used by business-critical database applications, will require a storage solution of a higher-performance tier than data constituting less active stored data. In addition, while there is usually not a charge for "ingress" or the movement of data into a public cloud storage solution, there is almost always a charge for "egress," or removal of data from a public cloud solution, which includes instances of applications accessing the stored data.

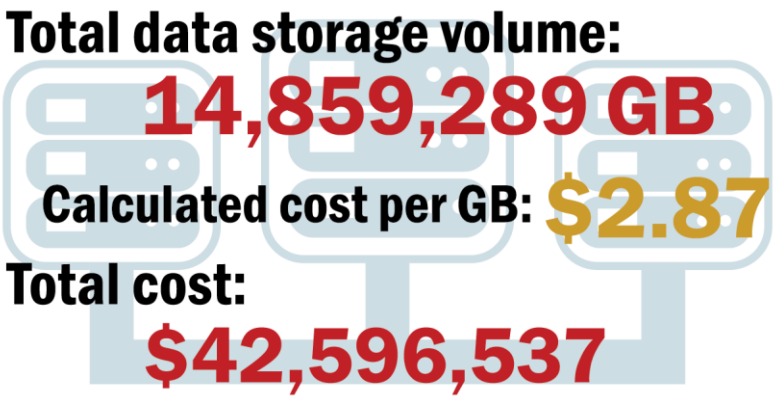
Because storage cost depends so heavily upon data processing requirements, agencies should devote resources to properly classify data and choose appropriate storage solutions.

In addition to hardware, network connectivity – the method by which an agency accesses stored data – can significantly affect the cost of any storage solution. For example, data with fewer security requirements can be transferred via a virtual private network (VPN) connection to a data center, while data requiring the highest level of security may be restricted to transfer through a direct cable connection between agency servers and data centers, an option chosen by many state agencies using Data Center Services. While the cost for these requirements may be mitigated by encryption of data in transit and at rest, encryption technologies can impact the overall performance of the processing of the data. For data with high security requirement but low transactional volume, encryption may be an appropriate mitigating strategy; however, data that has high security requirements and high transaction rates, the DCS private cloud would be a more viable solution. Furthermore, for enhanced security and continuity of operations, Data Center Services also provides geographic redundancies, which allows agency data to be stored in two geographically distant facilities. This is often a critical component not considered in the total cost of ownership of locally managed storage.

Agency responses to the 2018 Information Resources Deployment Review show inconsistency in the way agencies measure, budget for, and report digital storage. Total cost of ownership for each storage solution and the variation of reporting methodologies should be considered when discussing costs for digital storage.

According to Gartner, Inc. the 2017 average annual storage cost per *raw* 1 GB of capacity was $1.65. In response to the 2018 IRDR, agencies provided total cost of ownership estimates for various data storage solutions. The total cost estimate is based on direct and indirect costs for FY 2017, as seen in Figure 3.

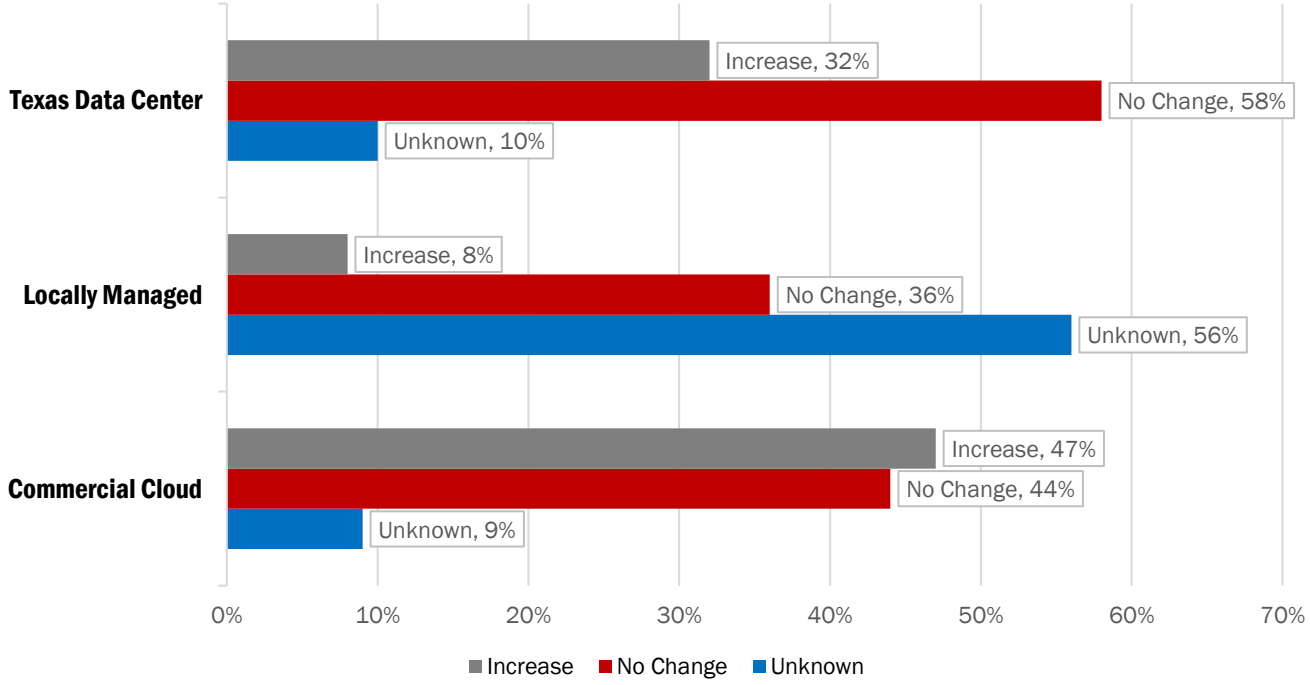
Figure 3: Agency-Assessed Data Storage Volume and Total Estimated Costs, FY 17



Source: 2018 Information Resources Deployment Review. Includes agency self-assessed estimates of total cost of storage, including direct and indirect costs. These figures include only completed answers provided by agencies with both estimates for volume and storage.

Agencies were asked whether they anticipate an increase, decrease, or no change in their spending on digital storage within the next two years. No agency stated they anticipate a decreased use of each of the storage types, as shown in Figure 4.

Figure 4: Agency Anticipated Usage by Storage Type: No Decrease in Use Expected



Source: 2018 Information Resources Deployment Review

Twenty-two agencies (28%) state that they make specific budget allocations for storage. The total budgeted for storage among those 22 agencies was reported to be $15,960,218, approximately 38% of statewide storage expenditures.

POLICIES AND COMPLIANCE

RECORDS MANAGEMENT

Each agency is required to establish and implement policies and procedures for their electronic records management practices. In 2018, DIR asked agencies to identify their policies relating to digital data and records management practices. Most agency policies address records disposition based on retention schedules (81%), email (73%), data classification (67%), and employee separation (54%). Fewer agencies have policies that address issues of duplicate file management, hierarchical storage, mobile storage and backup and social media.

TSLAC has developed the Texas State Records Retention Schedule which establishes minimum retention periods for records commonly found in state agencies. State agencies are required by law to prepare and submit agency specific records retention schedules to TSLAC and by administrative rule are required to recertify them for approval every five years. Some agencies require employees to retain data, including electronic information, in accordance with agency policy and the approved record retention schedule.

DIR asked agencies if they comply with the State Electronic Records Rules in Texas Administrative Code (13 TAC 1) and if archival electronic state records are properly preserved by the agency. State agency responses regarding compliance are encouraging in that between 65% and 77% said they comply with the requirements. This presents an opportunity for training and coordination between information technology and records management departments to

inventory their electronic records and systems to identify records and their retention requirements. Agency responses are listed in Table 2.

Table 1:Agency Compliance with State Electronic Records Rules

Rule/Statute	Each Agency Must:	Yes	No	N/A	Unknown
13 TAC §6.93	Meet the minimum requirements for the policies and procedures required for the management of all electronic state records	65%	18%	3%	14%
13 TAC §6.94	Meet the minimum requirements for the management of all electronic state records	73%	12%	3%	13%
13 TAC §6.95	Meet the additional record requirements for archival, permanent, and vital electronic state records	68%	13%	5%	14%
13 TAC §6.96	Stay up-to-date on Texas State Library and Archives Commission resources for electronic state records	77%	8%	3%	13%
13 TAC §6.97	Meet the minimum requirements for the final disposition of all electronic state records	77%	8%	3%	13%
13 TAC §6.98.	Meet the minimum requirements for the management of all electronic transactions and signed records	68%	8%	10%	14%
Government Code, Sections 441.186 and 441.180(2)	Ensure that electronic records in its custody that are archival state records or that need archival review are properly preserved	69%	12%	6%	13%

Source: 2018 Information Resources Deploymnet Review. Percentages may not total 100% due to rounding

DATA CLASSIFICATION

Data classification is the identification of security and privacy requirements for appropriate handling and protection of records. Data classification can also be the process of identifying groups of records, or records series, that are often filed together and have the same retention period.

In a decentralized IT environment, agencies classify, clean, and store their data based on agency needs and data quality criteria. When surveyed, agencies reported varying levels of compliance with data classification policies:

- 31% of agencies state data classification policies and processes are defined and repeatable across their organization. This means only one-third have a common understanding of what the organization's most important and sensitive information is and data owners have been identified for most of that information.
- 28% state their organization's data-classification policies are aligned with applicable regulations and the organization's own risk assessments. The organization takes enforcement actions such as spot checks, audits, process controls, awareness communications, and data-leak prevention controls that reinforce these classifications.

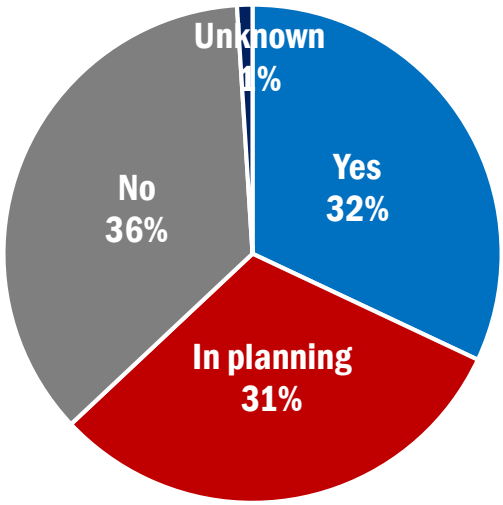
- 23% state data classification policies exist but classifications are inconsistent, unreliable, and inaccurate
- 6% state data classifications and procedures do not exist in their agency.

DATA MANAGEMENT

Data can be one of an organization's most valuable assets and the exponential increase in data has created both challenges and opportunities for organizations. Although some agencies have implemented fundamental data management, governance, policies, and best practices, they indicate slow adoption of a master data management plan that governs the collection, classification, use, and disposal of agency data.

While they are related, it is important to distinguish between data management and data governance. Data management is the organization, administration and standards for defining data and document management. Data governance addresses the decision rights and accountabilities for information related processes derived from the data. Implementing a data governance model allows more efficient data management, improved data quality and security, and better decision-making.

Figure 5: Percent of Agencies with a Master Data Management Plan



Source: 2018 Information Resources Deployment Review

CHALLENGES

As volumes of digital data increase, so do the challenges that agencies face when managing the data. Data exists within various programs and departments of an agency and can be costly without clear data retention and storage practices. A successful records management program requires support from employees and executive leadership, and changes in records and retention policies often require a shift in agency culture.

State agencies expressed that implementing a data governance program has been challenging. Lack of dedicated personnel, competing priorities, and lack of qualified staff are some of the largest barriers that are hindering agency efforts to implement a data governance program.

TANGIBLE

The storage costs of data can vary based on the technology solution each agency uses to store its data and in the ways in which the agency is retrieving and accessing the data on an ongoing basis. Costs remain the number one cited obstacle to adequately address records management and digital storage. Texas agencies face competing priorities and budget constraints that do not always favor investment in technology that may help address the increasing influx of data. Smaller agencies remain highly paper-based, making records retention difficult and new digitization initiatives cost-prohibitive.

LOGISTICS

Organizations often do not have a complete picture of their existing data, making it difficult to develop and adhere to established plans and procedures. For example, a large agency with thousands of employees and many locations, by necessity, has a decentralized system with multiple records coordinators. They have noted it is difficult to identify every electronic record and to determine the appropriate retention within their organization. Adding to this complexity, agencies sometimes lack adequate staffing resources to administer management of electronic records and digital data storage.

AGENCY CULTURE

One of the biggest challenges for agencies in implementing a successful records management program is agency culture. Agencies have cited employee resistance, difficulty in applying retention schedules, and lack of policy and enforcement as some of the barriers in their management of electronic records and digital storage.

In 2018, agencies identified the following as the largest barriers regarding electronic records and digital data storage and implementing a data management and governance program.

Table 2: Barriers Identified by Agencies

Barriers regarding Electronic Records and Digital Data Storage		**Barriers for Implementing a Data Management and Governance Program**	
Cost	56%	Lack of dedicated personnel	69%
Competing priorities/initiatives	47%	Competing priorities	60%
Difficulty applying retention schedules	40%	Lack of qualified staff	24%
Underdeveloped data management practices	31%	Resistance from data owners	9%
Unclear understanding of data	21%	Lack of perceived interest	8%
Lack of policy and enforcement	14%	Poor data quality/integrity	5%
Lack of executive engagement	0%	Other	6%
Other	5%		

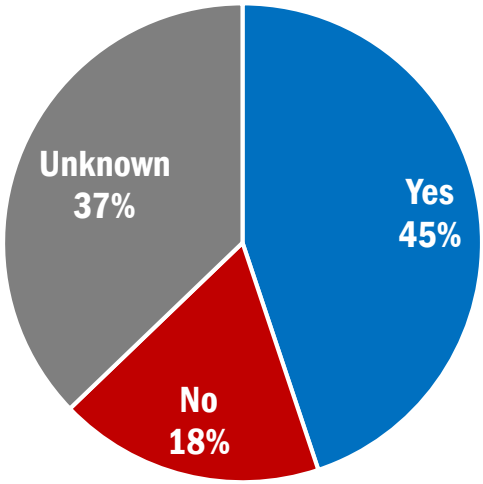
Source: 2018 Information Resources Deployment Review. Percentages reflect multiple choices allowed.

RETENTION

Retention is governed by law and best practices. State agencies must keep records according to legal retention requirements and business practices. How long to keep records is documented in each agency's certified retention schedule. Many agencies do not regularly review their digital data in terms of complying with their retention requirements: 48%, of agencies indicate they only look at retention intermittently, as needed, and 11% say they don't assess for retention compliance. Only 24% indicate any routine assessments.

Almost half (45%) of agencies state they store digital data beyond established retention requirements and 18% state they do not store data beyond those requirements.

Figure 6: Percent of Agencies that Store Digital Data that Exceeds its Established Retention Requirements



Source: 2018 Information Resources Deployment Review

The top four reasons agencies report for retaining digital data past retention are:

- organizational practices/culture
- potential for reuse/reference/research
- administrative, audit, or legal holds
- competing priorities

REDUNDANT, OUTDATED, TRIVIAL (ROT)

ROT is the concept that a high percentage of the data storage space may be occupied by data that is **redundant, outdated, or trivial**. Helping agencies apply good records management policies and acquire tools to help locate duplicate, trivial, and disposition-eligible records would help to reduce the volume, improve the ability to find the right records at the right time, and save effort and money in the long run.

DIGITAL PRESERVATION

Digital preservation is the application of a deliberate set of policies and practices to electronic records to keep them viable in the long-term. The practices also include the critical work of classifying and the creation of other metadata to fully describe and access those records. Digital preservation is an important component of any long-term electronic records management program and archival practice. Typical IT practice for systems storing electronic records usually involves regular data backups and updating hardware/software platforms to maintain system

functionality. File format updates may also be done for actively used records and keystone items such as high value data sets in actively used databases.

Unfortunately, for electronic records requiring long-term retention where the file may not be accessed frequently enough to justify regular file format updates, the level of attention is insufficient and digital preservation efforts are necessary. As can be seen with legacy computer systems, it is very easy for file formats to become inaccessible to modern software systems. For this reason, robust digital preservation relies on effective technological tools and practices as well as the fundamental knowledge management component of classifying and creating related finding aids and retrieval tools.

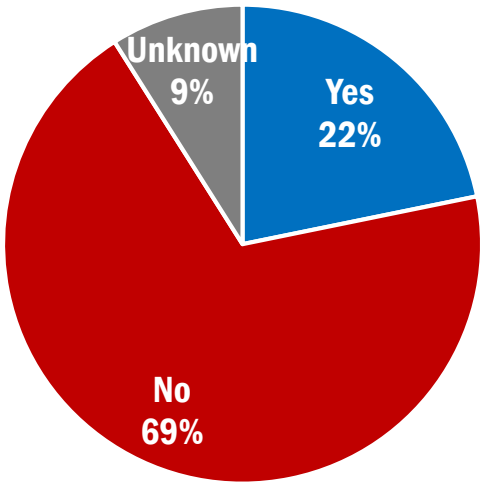
The rapid technological innovation cycle makes it critical to continually protect and preserve electronic state records. New advances can suddenly leave processing, storage, or backup resources out of date. If permanent records are stored on outdated storage media in outdated software, they may become difficult or impossible to access when needed.

DISPOSITION

Disposition is the final processing of state records by either destruction or archival preservation in a secure manner by TSLAC, a state agency, or an alternate archival institution. Authorizations in the disposition process ensures that the retention period is properly applied and there is no legal, audit, public information request, or valid reason to put the records on hold temporarily. Often, without training, cross-divisional coordination, and oversight, agency staff is unsure of what to do. They may either dispose of records too soon or hold on to them longer than required. Both early or late disposition increases an agency's risks.

Only 22% of agencies state that they track the volume of digital records disposed (destroyed, transferred to archives, etc.). The administrative rules for state agency records retention scheduling include a requirement to document the final disposition of records. Only nine agencies provided any volume information about digital data dispositions. A total of 104,138 GB was reported as dispositioned in FY17, which is .00069% of total storage volume in Texas agencies.

Figure 7: Percent of Agencies that Track Volume of Digital Records Disposed



Source: 2018 Information Resources Deployment Review

AUTOMATED TOOLS TO ENFORCE RETENTION

Some automated tools and systems exist that could help agencies fulfill their duties to manage electronic state records. However, few agencies have been able to budget, purchase, implement, and maintain these tools. Employee turnover and convincing leadership of the value proposition is a challenge for many agencies of any size.

Furthermore, the tools and market continue to evolve: electronic recordkeeping systems and electronic records management systems have moved to Enterprise Content Management systems. Agencies sometimes need experts for how the features are implemented and, possibly, they will need third party add-ons to make it work for them. Like all software, too much customization can make it hard to keep up with new versions and features due to compatibility issues.

Approximately one-third of agencies use automated tools to enforce their records retention policies, with another third planning to do so. Of those who do, the top uses for these automated tools are implemented for e-mail, databases, and enterprise file shares.

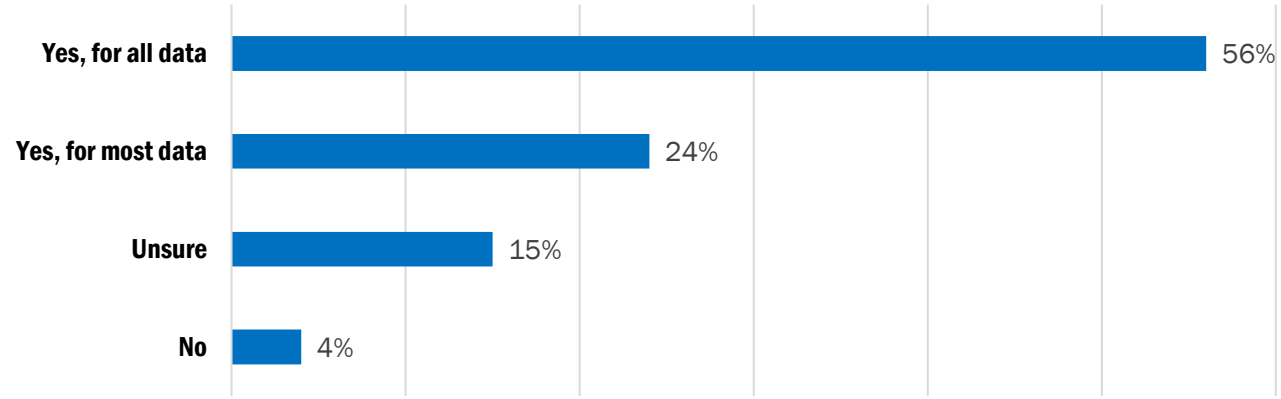
SECURITY AND CONFIDENTIALITY

By law, agencies are required to identify and take adequate steps to protect confidential and vital state records. Additionally, any technological component for electronic state records developed, used, or acquired by a state agency must provide security to ensure the authenticity of the records in accordance with statewide security rules. Agencies have implemented a variety of security controls for locally-hosted data and data housed in the cloud, including:

- anti-virus software
- operating system patching
- network Intrusion Prevention Services
- background checked employees
- continental US only operations
- web Application Firewall services (public-facing)
- virtual Local Area Network
- security Information and Event Management
- host Intrusion Detection Services
- appropriate vendor certifications (e.g. HIPAA, FERPA, CJIS compliance)
- data Loss Prevention software
- encryption at rest and in transit
- host Intrusion Protection Services
- virtual Data Center connectivity

Given the challenges of records management and digital storage, agencies still feel confident in their ability to meet security requirements of their storage solutions, with 80% stating their storage solutions allow the agency to meet the service delivery and security requirements of their data.

Figure 8: Percentages for Agency Storage Solutions that Meet Service Delivery and Security Requirements



Source: 2018 Information Resources Deployment Review. Percentages may not total 100% due to rounding.

SECURITY BENEFITS AND COSTS OF CLOUD COMPUTING SERVICES

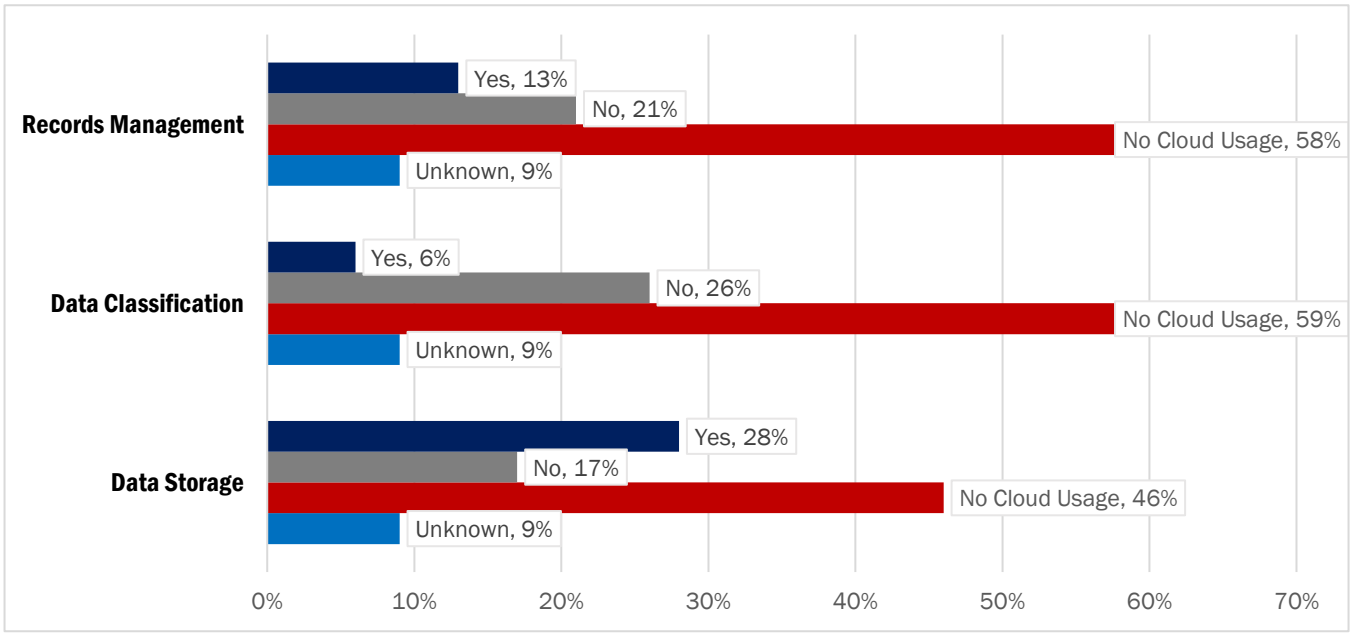
Increased reliance on cloud services is changing the way agencies view and utilize storage. Cloud architecture allows data security controls, such as data confidentiality, data access controllability, and data integrity. These controls can be difficult to achieve with traditional, locally-managed architecture. With cloud services, whether they are a public, private, or hybrid cloud, an organization is responsible for the security level of the data that resides in the cloud, while the cloud provider is responsible for managing the security of the cloud itself.

Cloud provider agreements need appropriate contract language to identify and protect the records in the cloud. Cloud contracts should address ownership, security, and authorized access for the records. Contracts must also detail what will happen to the records either at the end of the contract or at the end of their lifecycle. Terms must spell out what it would cost to access or return the records to the agency and address how records will be securely destroyed or transferred to TSLAC when they meet their retention period. Agencies should have approval processes to confirm destructions or records holds as appropriate. Unmanaged records repositories cost more and increase an agency's risk exposure.

Any time records are stored in the cloud or with a contractor, agencies must address the cloud's location (inside Texas, inside the U.S., etc.) for storage and backups, background clearances for all appropriate cloud staff at all locations (including subcontractors), secure disposition of authorized records for destruction or transfer including backups, and similar critical components that ensure protection in the cloud.

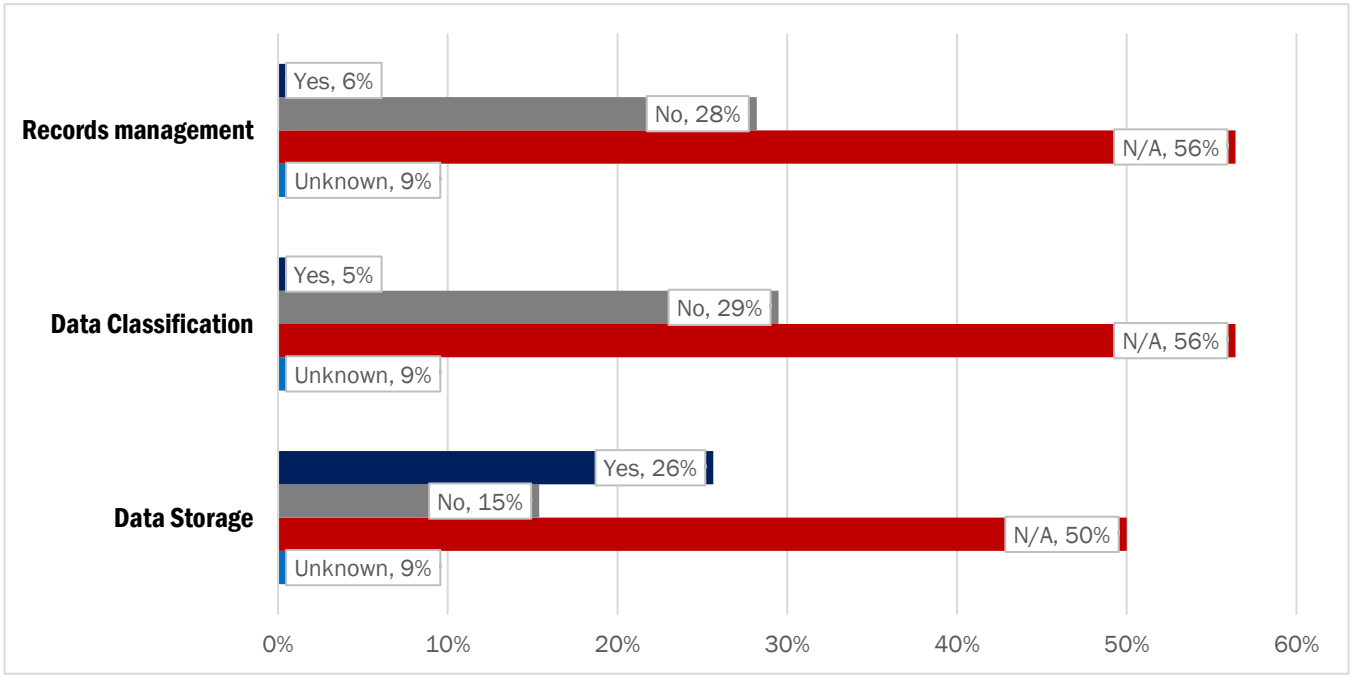
As shown in Figures 9 and 10, most agencies said cloud services were not used in their agencies for records management, data classification, and data storage.

Figure 9: Agency Experience *Security Benefits* of Cloud Computing



Source: 2018 Information Resources Deployment Review. Percentages may not total 100% due to rounding.

Figure 10: Agency Experienced *Cost Savings* of Cloud Computing Services



Source: 2018 Information Resources Deployment Review. Percentages may not total 100% due to rounding.

ON THE HORIZON

Records management is not a singular type of project. Agencies must continue to focus and work for an ongoing and active records management program. Records management is a necessary beginning and governing foundation for lifecycle planning for all types of technology efforts. There is always innovation in technology and that generates new types of content that agencies must address in policies to apply records management to records. The market continually sees acquisitions and merging of products. Some experts are looking to assess the records management possibilities of artificial intelligence, robotic processing, blockchain, cloud content services, and more.

Along with low cost long-term storage options currently being offered through TSLAC's Texas Digital Archive and being considered in the next generation of the DCS Program, which will include both public and private cloud, there are several modern archival solutions available that would allow agencies to be more efficient at managing their data storage through targeted backup and retention schedules. In addition to realizing cost savings through decreases in the volume of data stored, the cost generated by access and/or the transaction of stored data can be reduced through a more well-defined classification of data.

The two disciplines, records management and the management of information resources, can help build in records management to policies, procedures, and compliance for IT purchases, whether they are off-the-shelf commercial software, in-house developed, or hybrid projects. Focus on the inventory, management, policies, and disposition of records can provide the key to meeting the influx of data that is inevitably on the horizon.

RECOMMENDATIONS

1. IMPLEMENT A STATEWIDE INFORMATION GOVERNANCE COORDINATOR POSITION WITHIN TSLAC'S STATE AND LOCAL RECORDS DIVISION TO PRIORITIZE INFORMATION GOVERNANCE AND INCREASE RESOURCES AND OUTREACH FOR STATE AGENCY RECORDS MANAGEMENT PROGRAMS.

Agencies cite lack of dedicated personnel, competing priorities, and lack of qualified staff as barriers to implementing a successful data management and governance program, and continually request more dedicated and robust assistance from TSLAC. With the Statewide Data Coordinator enhancing the data sharing and coordination efforts for agency data, there is now a need for an information governance coordinator to ensure information and record management practices are implemented within state agencies. The coordinator will help agencies ensure compliance, increase opportunities for agency records management officers, and increase overall awareness and outreach for the information governance program. Without this resource, we see increased costs for all phases of the records' lifecycle, increased time to locate records for operations and transparency, and increased security risks. By increasing awareness, collaboration and model policies, executives and staff will have the ability to properly manage records throughout their lifecycle which reduces costs, time, volumes, and security risks.

2. REQUIRE IMPLEMENTATION OF DATA CLASSIFICATION, DATA SECURITY, AND DATA RETENTION REQUIREMENTS AT PROJECT INITIATION.

Today, classifying data and determining data security and retention requirements at project initiation is not a standard practice. Should this become a requirement, it can result in more

effective data management, increased quality and reliability of data, and a reduction in storage costs. It can be more cost effective to employ a proper data classification scheme at project initiation because it would allow agencies to focus on protecting its higher risk data assets. This could also help to reduce security and storage costs by protecting data and only retaining it as long as necessary as per the data classification and retention requirements that are defined at the beginning of a project. A better understanding of the types of data agencies are storing, how often and for what purpose, will better inform how we can use data to make government more efficient.

3. ESTABLISH A PERIODIC TRAINING REQUIREMENT FOR ALL AGENCY STAFF (NEW HIRE AND ONGOING) REGARDING THE AGENCY-SPECIFIC RECORDS RETENTION SCHEDULE AND RECORDS MANAGEMENT PROGRAM. AGENCIES MAY USE TRAINING MATERIALS PROVIDED OR MADE AVAILABLE BY TSLAC.

Currently, records management training is done differently by each agency, with different elements and at varying times. Agency culture was cited as one of the largest barriers to effectively implementing and adhering to a records management schedule. An effective records management program requires ongoing training within agencies for the purposes of records management, including training for new employees and yearly training for employees at all levels. Training, notifications, and coordination is crucial to figuring out what to keep, what to properly authorize for disposition, and how to dispose. In turn, training or resources could be provided to agencies to help them address how retention could be improved.

4. REQUIRE DIR, IN COLLABORATION WITH TSLAC, TO DEVELOP A STRATEGY FOR AGENCIES TO TRANSFER APPROPRIATE ELIGIBLE ARCHIVAL ELECTRONIC STATE RECORDS TO THE TEXAS DIGITAL ARCHIVE AND TO LEVERAGE THE STATEWIDE TECHNOLOGY CENTER'S LONG-TERM NON-ARCHIVAL STORAGE OPTIONS, IN COORDINATION WITH RECORDS MANAGEMENT AND ARCHIVAL PRACTICES.

Current transfers of archival electronic state records of all types represent a small volume of possible transfers. TSLAC is prepared to work with more agencies to identify and start receiving special and routine transfers of these records. Additionally, the next generation of the DCS Program, which will include both public and private cloud, plans for several modern archival solutions available that, when integrated with the long-term storage options, would allow agencies to be more efficient at managing their data storage through targeted archive and retention schedules. The collaborative strategy designed by DIR and TSLAC will better guide agencies through the records services and technology solutions available to them.

AGENCY BEST PRACTICES

TEXAS DEPARTMENT OF PUBLIC SAFETY – APPLYING RECORDS RETENTION DURING PROJECT IMPLEMENTATION

The Department of Public Safety (DPS) was charged with implementing the Compassionate Use Registry application as required by the Texas Compassionate Use Act (Senate Bill 339), which was enacted by the Texas Legislature in 2015. The bill required DPS to create a secure registry of physicians who treat epilepsy by prescribing low-THC cannabis to patients who have been diagnosed with intractable epilepsy. To ensure records retention requirements were met, DPS applied records retention within the software of the application during the initiation phase of the implementation of the new registry, which at the time, was not a standard practice. The communication and collaborative partnership between the business product owner and the project manager led to a successful implementation of the registry. DPS was pleased with the results of implementing records retention at the project initiation level and the agency plans to continue with this practice on future initiatives going forward.

TEXAS COMPTROLLER OF PUBLIC ACCOUNTS – CLOUD STORAGE INITIATIVE

To reduce electronic records storage and address other existing records management issues, the Texas Comptroller of Public Accounts (CPA) migrated data from its existing environment to a NetApp platform. Moving to the NetApp platform has provided more flexibility with protocols and eliminated the need for a host to be physically connected to a storage area network (SAN). Other benefits provided by NetApp include deduplication and compression functionality that help to reduce storage space. Since moving to NetApp, CPA has realized a 50% reduction in storage space.

HEALTH AND HUMAN SERVICES COMMISSION – DATA MAPPING INITIATIVE

As part of an organizational transformation started in September 2016, the Health and Human Services Commission (HHSC) began mapping data across disparate systems to identify security requirements, develop classification standards, and implement controls across the agency. HHSC faces challenges to the successful completion of this initiative, including the scale and scope of data housed at the agency, and the effects of the agency's reorganization, which has changed the relationship between systems, management authority, end users, and technology. In addition, security requirements for data used by HHSC vary based on the program in which the data were created, as well as the requirements established by any of the thirteen federal agencies contracting with HHSC. The Center for Analytics and Decision Support (CADS) coordinates the data mapping initiative, with special cooperation between CADS and agency divisions responsible for security, privacy, records, and legal compliance. The strong working relationship between these divisions and CADS was essential in putting the data mapping initiative on its current course, which continues toward successful completion.

OFFICE OF THE ATTORNEY GENERAL – RECORDS MANAGEMENT TRAINING

Records management practices at the OAG benefit from training provided to all staff. Records management training at OAG includes an introduction to the agency's Records and Information Management Program for all new employees at New Employee Orientation, as well as annual in-person records management training opportunities for all OAG employees. In addition, the OAG Information Governance Division provides consultation to a network of division-level Records Management Liaisons, who in turn provide records management support to individuals within their respective divisions. Lastly, OAG attributes its success to the strong relationship

between its Information Governance and Information Technology Services Divisions. Strong communication and training ensure employees handling records receive the guidance they need.

TEXAS STATE LIBRARY AND ARCHIVES COMMISSION – TEXAS DIGITAL ARCHIVE

Digital preservation takes the necessary steps to keep electronic records accessible for records long-term or permanent archival retention. Once agencies transfer archival electronic records to TSLAC, the Texas Digital Archive (TDA) takes on this preservation role and responsibilities. Because the TDA provides an economy of effort for digital preservation, in addition to shifting storage costs away from an agency, it relieves the agency of the burden of servicing Public Information Act (PIA) requests and saves agency staff time and associated funds for tasks other than digital preservation. At the same time, where applicable to the records, the TDA provides public access, helping state government fulfill its mandate under the PIA.

TEXAS DEPARTMENT OF INFORMATION RESOURCES – AUTOMATIC EMAIL RECORDS RETENTION

In January 2016, DIR initiated a project to implement a records and information management program with the goal of managing records that should be retained and disposing of all records that no longer need to be retained. In May 2016, efforts were focused on implementing an automated email management policy that would delete items in the inbox, sent folder, conversation history, clutter and junk email after 60 days, and in the deleted items folder after 7 days. Emails that needed to be kept beyond the policy dates would need to be moved to a new folder, outside of the inbox, called Archive. DIR has since decreased the number of GBs of email stored by approximately 27% since automatic enforcement of the email records retention policy went into effect.

ACKNOWLEDGEMENTS

The 2018 Digital Storage Study is a collaborative effort between the Texas Department of Information Resources (DIR) and the Texas State Library and Archives Commission (TSLAC). Thank you to the individuals from both DIR and TSLAC who developed this study on state agency digital data storage and records management practices. DIR and TSLAC would also like to acknowledge with appreciation, the representatives from the following agencies who provided valuable insight on their successes, challenges, and recommendations regarding digital data storage and records management practices:

Comptroller of Public Accounts

Health and Human Services Commission

Office of the Attorney General

Texas Animal Health Commission

Texas Department of Information Resources

Texas Department of Insurance

Texas Department of Public Safety

Texas Higher Education Coordinating Board

Texas Parks and Wildlife

Texas State Board of Pharmacy

Texas State Board of Public Accountancy

Texas Veterans Commission

Texas Workforce Commission

Texas Department of Information Resources

TEXAS STATE
LIBRARY
AND
ARCHIVES
COMMISSION