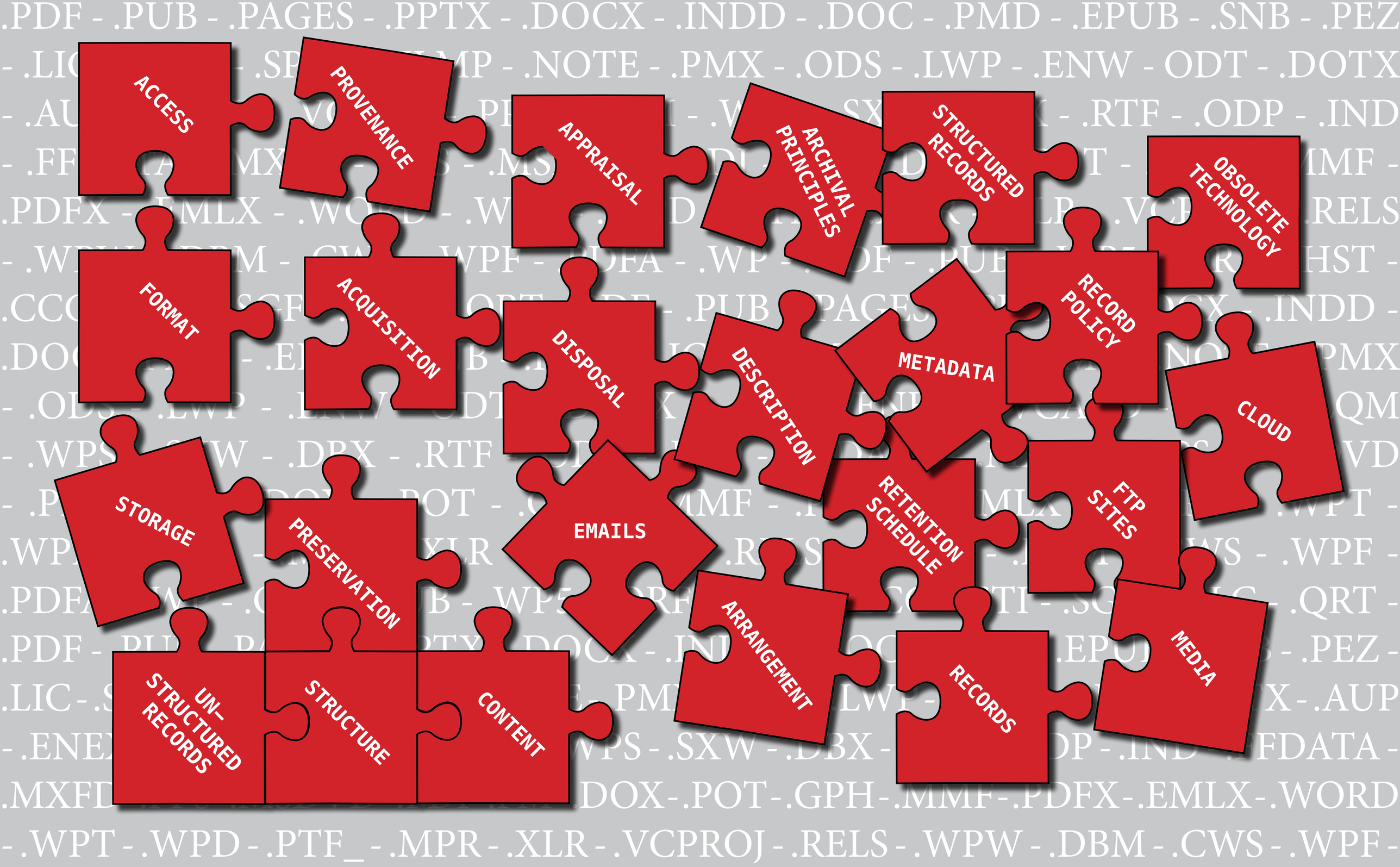

PIECING TOGETHER THE ELECTRONIC RECORDS PUZZLE

Missouri Historical Records Advisory Board

National Archives
National Historical Publications & Records Commission

Missouri State Archives

John R. Ashcroft
Secretary of State

Principles of Electronic Records Management

The Information Age has ushered in a massive shift from the creation and storage of analog records (such as paper and microfilm) to electronic records. Digital storage overtook analog capacity in 2002 and continues to grow exponentially. Some experts claim that by 2020 the world will have produced 44 zettabytes (44,000,000,000,000,000,000,000 bytes) of data. To put that into perspective, a single zettabyte equals one trillion gigabytes, or the approximate storage capacity of over 62.5 billion iPhones or 9.4 trillion DVDs! What would once fill a warehouse can now be stored on a tiny microchip.

Many people today are apprehensive about electronic records because they do not understand them or, worse yet, they do not consider them records at all. The below information from the Missouri State Archives and the Missouri Historical Records Advisory Board (MHRAB) provide guidance for managing electronic records.

Electronic Records

So what are electronic records? Electronic records are really just a new record format. The only difference between analog and electronic records is their composition. A record is defined as information fixed on a medium (paper, magnetic, stone, etc.) used to record human memory or to ensure accountability. While analog records are comprised of things like ink, paper and film, electronic records are strings of ones and zeros (known as "bits"). These bits are converted by computers into information readable to humans. Like their analog counterparts, electronic records have three essential characteristics: content, structure and context. To be classified as trustworthy they must demonstrate authenticity, reliability, integrity and usability.

- **Authenticity**: A record must be what it says it is.
- **Reliability**: A record must be a full and accurate representation of the transactions, activities or facts.
- **Integrity**: A record must be complete and unaltered.
- **Usability**: A record must be able to be located, retrieved, presented and interpreted.

In order to establish and maintain these characteristics, basic archival principles should be applied to electronic records. The two most important of these are provenance and original order. Provenance refers to the individual, family or organization that created or received the items in a collection, and dictates that records of different origins are kept separate. Original order, meanwhile, states that records of a single creator should retain the arrangement established by the creator to preserve the existing structure and context.

Metadata

Metadata documents the identification, management, nature, use or location of a resource. It is often referred to as "data about data," and generally provides information about the "what" and the "where" of a given resource. The computerized catalog system in your local library is a form of metadata.

Metadata usually falls into five categories:

1. **Descriptive metadata** – the "who, what, where or when" of the item.
2. **Structural metadata** – the number of pages, collection information or other data about the physical nature of the item.
3. **Technical metadata** – record characteristics (e.g., file format, size, software, checksum).
4. **Administrative metadata** – when the item was digitized, rights and reproduction information, file path location, etc.
5. **Preservation metadata** – support information, such as media migration dates. (Often grouped with administrative metadata.)

Metadata Example

Descriptive:
Title: Missouri Waltz
Recording Date: c. 1946
Performer(s): Al Goodman and his orchestra
The Mullen Sisters (vocal refrain)

Structural:
Track: 1
Length: 4:17

Technical:
File Type: Audio
File Format: wav
File Size: 500 KB

Administrative:
Source: 78 rpm record
Digitized Date: 01/02/2015
Copyright Date: 1914
Agent(s): John Valentine Eppel; Frederic Knight Logan; James Royce Shannon
Rights: F.J.A. Forster

Preservation:
Scheduled Migration Date: 01/02/2020

Note: *Creating detailed descriptions is costly, so care must be taken to balance the time and expense of creating additional metadata with access and preservation needs.*

Structured and Unstructured Electronic Records

Electronic records are either structured or unstructured. Structured records are files containing information defined by tabs or characters, written using Extensible Markup Language (XML) or fixed in databases. Unstructured records, on the other hand, do not have a pre-defined data model and are not organized in a controlled manner. The distinction between structured and unstructured records is important for preservation because structured data is much easier to manipulate than unstructured data. Below are examples of common structured and unstructured records.

Structured	Unstructured
Delimited: MS Excel, comma separated values and tab separated values	**Images:** JPEGs and TIFFs
Extensible Markup Language: XML	**Audio and Video:** WAVs, MPEGs, MP3s and MP4s
Defined Databases: MS Access, MySQL and dBase	**Text:** ASCII, Unicode, PDFs, MS Word, HTML pages and Word Perfect

Maintaining Electronic Records

From creation, through use, to destruction or archival preservation, electronic records must be actively maintained throughout their lifecycle. Digital records do not survive by accident, so it is vital for record keepers to know their institutional holdings; be aware of common preservation challenges, such as technological dependency, hardware and software obsolescence and media deterioration; and formulate policies to preserve these records for long-term access. Policy considerations should include:

- **Acquisition and Accessioning** is the process of formally accepting and documenting the receipt of records into archival custody. Upon accessioning, the records become the physical and intellectual property of the institution. It is highly recommended that institutions require a records schedule/policy for accepting electronic records and that the creators identify the appropriate record series/collections and dates. To ensure long-term preservation and accessibility, institutions should consider standardizing file formats, media formats and transfer methods.
- **Arrangement and Description** is the practice of keeping collections in the order in which they were originally created or used, and describing their relationship to the creator and other records. This is vital to understanding how the creator used the records and the context in which the records were created. It is also essential to capture metadata about the records documenting the identification, management, nature, use and location of the data.
- **Appraisal and Disposition** is the evaluation of source material and selection of the portions that will be preserved or destroyed. Aspects of the records that can be appraised include, but are not limited to, the value of the data, its technical characteristics, the importance of the creator, its relation to other records and its completeness. When records reach their designated retention time (or lose their original value), they can either be destroyed or preserved. Destruction can occur through the re-use of magnetic tape or the overwriting, or physical destruction of, optical disks and hard drives.
- **Preservation and Storage** is required to ensure long-term access. Records should be kept in conditions which are cool, dry and seasonably stable, with minimum exposure to natural or artificial light and protection from pests, pollution and unauthorized access. Furthermore, electronic records should be stored using a proven archival format with several separate copies maintained to ensure survivability. This is illustrated by the LOCKSS (Lots of Copies Keeps Stuff Safe) System at Stanford University and can be carried out using a variety of storage media, such as the "cloud," network storage, optical media, external drives and flash drives. It is also important to have backup files physically separated in different geographic regions to increase survivability in the case of a natural disaster or catastrophe.
- **Reference and Access** is perhaps the most valuable of these considerations as it allows users to read the information. What good is a record if you cannot read it? Records should be locatable through the use of physical or electronic finding aids, indexes, catalogs or other tools. It is also important to define access restrictions to ensure that only authorized persons are able to view sensitive information.

Preservation Challenges

The three most common challenges involved in the preservation of electronic records are technological dependency, obsolescence and media deterioration. The majority of analog records do not require technology to be accessed; one can usually pick up a piece of paper and read what is written on it. This is not the case with electronic records. They are completely dependent on technology for both their creation and storage. They must therefore be managed over time in a computerized environment. Computer hardware and software evolves at a lightning pace, so to ensure access, steps must be taken to maintain records compatibility.

Electronic records are also vulnerable to storage media deterioration. This can occur through corruption, bit rot (deterioration of data on storage media), viruses or hard drive failure.

These problems can be mitigated by staying abreast of the latest technology trends; keeping record systems up-to-date; and preserving records through methods such as refreshing, migration, replication, emulation, the encapsulation of information with its requisite metadata and redundancy.

Storage Media

It is important to weigh all available options when deciding on a storage medium. Cloud storage is inexpensive, as long as frequent access is not required. In-house network storage is advantageous for access, but requires robust disaster recovery precautions. External drives can store vast amounts of information, but need to be refreshed on a regular basis and are susceptible to electronic magnetic resonance (EMR) damage. Optical media (CDs and DVDs) are widely available, but only have a 5 to10 year maximum life span. Other significant factors to weigh are longevity (should be usable for at least 10 years), capacity (higher is better), viability (to ensure integrity), stability (standards-based), cost (per megabyte) and tolerance (low susceptibility to damage).

Preservation Policy

A preservation policy for electronic records can greatly increase effectiveness by mandating records management from creation, thus lessening the risk of damage or loss.

Effective policies:

- Ensure only essential records are created, records of continuing value are preserved and records of temporary value are promptly and systematically disposed;
- Encompass all record formats;
- Capture appropriate metadata;
- Create an effective method for identifying and implementing temporary destruction holds for legal and financial purposes;
- Encourage policy compliance via training;
- Provide for regular policy review;
- Establish accountability and impose sanctions for compliance failures; and
- Provide guidelines for the disclosure and disposal of records.

Electronic Records Systems

The difficulty in managing large volumes of electronic records can be greatly lessened by choosing an effective automated electronic records application/system. One must be thorough when deciding on an electronic records system. Early steps in the selection process should include creating a working group, conducting a needs assessment and comparing needs to specifications (such as open-source vs. propriety software and retention/disposition components). Strongly consider both short- and long-term needs, costs and technical capacity and, when possible, test software before purchasing.

Open Archival Information System Reference Model

Perhaps the biggest challenge in preserving electronic records and creating a digital repository is simply getting started. The Open Archival Information System (OAIS) is a useful and informative tool to help jump start the process. OAIS is a standard model for digital repositories that provides a framework for accepting, preserving and accessing digital assets. The model avoids using technical jargon and instead condenses the issue down to a set of concepts, relationships and processes common to all digital preservation projects. The model is also purposely context-neutral to allow for flexibility and use in a wide variety of environments.

The OAIS model is represented below.

Conclusion

Electronic records are dynamic documents that must be managed with even more diligence than paper records. Planning is essential. A good plan executed today is better than a great plan executed tomorrow. Remember to keep implementation simple and periodically review the system and policies to ensure compliance and safeguard vital information.

Resources

Model Requirements for the Management of Electronic Records
http://ec.europa.eu/archival-policy/moreq/index_en.htm

National Archives and Records Administration (NARA) Records Management Guidance and Regulations
www.archives.gov/records-mgmt/policy/guidance-regulations.html

OAIS model
www.oclc.org/research/publications/library/2000/lavoie-oais.html
www.paradigm.ac.uk/workbook/introduction/oais.html

For more information contact the Missouri State Archives:

- Phone: (573) 751-3280
- Email: archives@sos.mo.gov
- Website: www.sos.mo.gov/archives/