ARST 5300: Digital Preservation

Syllabus: Fall 2016

Subject to revision

Administrivia

Instructor

- Seth Shaw
- Office location: Georgia Archives, 5800 Jonesboro Road, Morrow, GA 30260
- Office phone: 678-466-4410
- Email: sethshaw@clayton.edu
- Office hours: Face-to-face & virtual (phone, email): 10:00 a.m. - 2:00 p.m. daily, as available; appointment recommended; evenings available by appointment.

Prerequisites

- ARST 5000: Principles and Practices
- ARST 5100: Archives and Technology

Credit

3 weekly contact hours

Schedule and Location

- Course site: https://clayton.view.usg.edu/d2l/home/1178368. Log in using your Clayton State userid and password.
- Lectures: Monday, 6:30 - 8:30 P.M. via Lync/Skype for business.

Format of Course

Online course including weekly, synchronous lectures by the professor and asynchronous discussion of the readings by the students.

Description

Builds on the introduction to electronic records presented in preceding courses and will address the particular challenges of digital materials. Introduction to archival aspects of media longevity, technological obsolescence, and compatibility; information representation and experience; trust, authenticity, and integrity; preservation and
security; system documentation; and metadata. Application of the Open Archival Information System Reference Model and the Trusted Repository Audit & Certification: Criteria & Checklist, and the knowledge to develop policies and procedures needed to address administrative responsibilities, organizational viability, technological suitability, system security, and procedural accountability. Comparison and evaluation of different implementations of digital archives systems, and application of current best practices to demonstrate the authenticity of digital records and to manage collections of digital records.

**Background**

In contemporary archives, archivists are regularly challenged by the diversity of record types and formats for which they are responsible. Most, if not all archives, contain important electronic records and archivists have a moral and ethical responsibility to ensure that these records, to the extent possible, remain trustworthy, authentic, and reliable over time. Unlike most analog records, this means that archives and archivists must play an active role in storing and preserving digital material.

In the context of archives, digital preservation is the care and management of original electronic records and files. These records are sometimes described as 'born-digital.' These are different from scanned images of original photographs or paper records and include emails, websites, databases and digital images and video. The preservation of these digital records presents a number of challenges because:

- The equipment and software needed to view digital records can quickly become obsolete;
- Media such as tapes and discs can deteriorate quickly even if they do not appear to be damaged; and
- The context of a digital record and its relation to other records can easily be lost.

This course will provide an introduction to digital preservation and provide guidance about the theory and practice of digital preservation. It will build on the fundamental concepts of appraisal, selection, arrangement, and description. It will provide hands-on experience with current digital preservation tools and systems, providing archivists with not only the theory, but also the practice behind actively preserving digital files. Future archivists will conclude the class with the knowledge needed and options possible for maintaining trustworthy, authentic, and reliable records over time.

**Course Learning Outcomes**

1. Discuss fundamental concepts of digital systems, including layered representation, dependencies, and interfaces.
2. Describe the threats to the long-term preservation of digital objects.
3. Discuss trustworthiness, authenticity, reliability, usability, and comprehensiveness, as well as form, uniqueness, and quantity in the context of digital records.
4. Understand standards and community best practices for digital repositories and archives, including the Open Archival Information System (OAIS) and Trustworthy Digital Repositories Audit & Certification (TRAC).
5. Understand the use of digital repository software such as DSpace or Fedora; knowledge of digital storage methodologies (LOCKSS, iRODS); and be able to utilize commonly used digital repository software.
6. Understand the range of options for preserving records in their original format or structure, and when to utilize other preservation strategies and action plans for digital files.
7. To describe a workflow for digital archives, including objects (submission, archival, and dissemination packages), roles, and functions.
8. Discuss approaches to preserving common, challenging record formats, including social media, email, and websites.
9. Describe and explain the use of a variety of digital preservation tools.

Readings

There is one required book that must be purchased: Steve Marks. *Module 8: Becoming a Trusted Digital Repository*. Society of American Archivists (2015). You may purchase it from SAA directly and SAA members can receive a discount. The print and ebook versions are both acceptable.

Readings are listed on the syllabus for each week and, except for one, are freely available either online, in the course site, or through the Clayton State Library.

Grading

Grade Breakdown

- Participation (50%)
  - In-class (25%)
  - Discussion board (20%)
  - Tutorials (5%)
- Quizzes & Exam (50%)
  - Take-home Theory Quiz (10%)
  - Take-home Tools Quiz (10%)
  - Final Exam (30%)

In-class Participation

- **Exceeds expectations** (4): Participates with some comments, questions, or answers showing greater depth of understanding, preparation, leadership, engagement, or other quality.
- **Meets expectations** (3): Participates with multiple on-topic comments, questions, or answers.
- **Acceptable** (2): Participates with at least one on-topic comment, question, or answer.
- **Unacceptable** (1): Present but with little or no engagement, no demonstrated understanding of the material.
- **Missing** (0): Unexcused absence

Discussion Board Participation

Each week consider how the readings apply to the listed description and list of topics. Students are strongly encouraged to introduce other, relevant topics. An exploration of unfamiliar terms and concepts can be invaluable to clarify these ideas for yourself and others. Some weeks may include additional questions to prompt further reflection.

Posts must be made before class to receive credit.
• **Exceeds expectations** (4): Posts are particularly helpful/insightful or the presence of additional substantive posts.

• **Meets expectations** (3): *At least two posts* that either offers your own analysis and evaluation of the ideas in article or substantially builds on/answers someone else's comment or question. You are encouraged to offer your opinions in support or contrary to the ideas in the article if supported by your experience or ideas from other authorities (readings, practical experience).

• **Acceptable** (2): *At least one* on-topic comment, question, or answer.

• **Unacceptable** (1): Posts show little or no demonstrated understanding of the material.

• **Missing** (0): No posts

**Tutorials**

Tutorials are guided exercises providing hands-on experience with tools used for digital preservation tasks. Instructions will be posted to the course site. Tutorials are graded as pass/fail.

**Take-home Quizzes**

There will be two take-home quizzes, one for each of the first two course modules, due before class the week after they are given. Each will consist of 6 short-essay questions and should take 30-45 minutes to complete. The exam is open book, lectures, and notes. Please do not consult other materials, resources, or individuals while taking the quiz.

**Final Exam**

The Final Exam will be made available on it's scheduled day via the course site at 6:30 P.M. and is due in the corresponding drop-box at 9:00 P.M.

The exam is open book, lectures, and notes. Please do not consult other materials, resources, or individuals during the course of the exam. The instructor will be available via phone, text, and email for technical assistance but will not answer questions regarding the content of the exam.

**General Policies**

See the [General Course Policies](#) on my website.

**Course Schedule**

**W01 (08/15) - Introduction**

Students will be introduced to the course structure and requirements. Students will also receive an introduction to the challenges facing digital preservation as well as the main elements of a digital preservation program and the Open Archival Information System Reference Model.

**Learning Outcomes**

• Describe the structure of the course and it's requirements.
• Describe the elements of a digital preservation program: technology, policy, and administration

Readings


**W02 (08/22) - Trust & Digital Repositories**

The authority archives and archivists have to be trusted custodians of records has been built over the course of centuries. We can visit an archives and see the physical infrastructure and activity that supports the preservation of physical records but we cannot necessarily the server rack with a few blinking lights for digital repositories. How does the notion of trust—along with our claim to trustworthiness—and the notion of a repository change with the introduction of digital records?

**Learning Outcomes**

• Explain concerns that can make digital records seem less trustworthy and how they can be countered.
• Describe the high-level agents and functions of the Open Archival Information System Reference Model
• Describe the relationships between TRAC & the related International Standards Organization (ISO) standards

Readings

• Christopher Hilton, Dave Thompson, and Natalie Walters, “Trust Me, I’m an Archivist: Experiences with Digital Donors,” *Ariadne* no. 65 (October 30, 2010), http://www.ariadne.ac.uk/issue65/hilton-et-al/.

**W03 (08/29) - Hardware & Materiality**

The digital ecosystem may be virtual, but it exists on tangible hardware, including computers (processors, RAM, I/O), storage, and networks, integrated to work together through software. Digital archivists must understand our relationships with technology and how the different components influence digital preservation.
Learning Outcomes

- Describe affordances of computer equipment and how humans might interact directly or indirectly with them.
- Review:
  - To name the major components of a computer and describe what they do (CPU, RAM, disks, I/O).
  - To describe the architecture of storage media.
- Describe the long-term challenges of storage media
- Discuss the impact of legacy systems on archives.

Readings


Labor Day (09/05) - No Class

W04 (09/05) - Recordness & Representations

The pathway information takes between analog storage or generation and the analog output is governed by software. Preservation of recordness and experiential value require close attention to the software stack and the choices made along the way.

Learning Outcomes

- Explain how representation information control the performance of digital objects using the OAIS Reference Model, PREMIS Data Model, and National Archives of Australia Performance Model.
- Explain how significant properties and affordances inform the selection of digital object formats.

Readings

- Reference Model for an Open Archival Information System (OAIS), sec. 4.2.
- Lavoie, The Open Archival Information System (OAIS) Reference Model, sec. 5.4.
- Marks, Becoming a Trusted Digital Repository, secs. 4.1.1–4.1.2.
W05 (09/19) - Ingest — Media Capture

For data to persist it needs to be written to some form of media from early paper cards to modern solid-state drives. Unfortunately these media usually have a short lifespan and the data on them must be captured to ensure data longevity.

Learning Outcomes

- Explain the relationship between computer forensic science and digital preservation practice, including the techniques used.
- Explain the use of hashing algorithms to verify file integrity over time.
- Explain the practical and ethical implication of both forensic and operating system mediated data capture and when one method might be chosen over another.

Readings


W06 (09/26) - Ingest — Network & API Capture

Networked systems switch the focus from media and file systems to communications protocols and advanced programming interfaces (APIs). How can archivists acquire records using these technologies?

Learning Outcomes

- Compare and contrast media-based acquisition with protocol and API-based acquisition.

Readings

- Best Practices for the Capture of Social Media Records (National Archives and Records Administration,


W07 (10/03) - Ingest — Collection Analysis

After capturing content it must be analyzed to identify requisite representation information (usually file formats), potential threats (such as viruses and other malware), and content of interest (especially restricted content) to determine appropriate courses of action. Even if file formats can be readily identified, sufficient documentation is not always readily available. How can we determine (with relative confidence) a digital object’s representation information?

Learning Outcomes

- Explain how file formats can be identified and where to find format documentation.
- Survey a collection of files to identify formats, threats, and content of interest.
- Determine the degree of file format “endangerment.”

Readings

- Reference Model for an Open Archival Information System (OAIS), sec. 4.1.1.2.
- Lavoie, The Open Archival Information System (OAIS) Reference Model, sec. 5.3.1.

FALL BREAK (10/10) - No Class

W08 (10/17) - Migration

Migration transforms one of an intellectual entity’s representations into another to support different architectures.

Learning Outcomes
Explain how migration is performed.
Explain the advantages and disadvantages of format migration as a preservation methodology.
Evaluate potential migration paths.

Readings


W09 (10/24) - Emulation

Emulation, “the use of one system to reproduce the functions and results of another system” Richard Pearce-Moses “Emulation,” *Glossary of Archival and Records Terminology*, Archival Fundamentals Series. II (Society of American Archivists, 2005), http://www2.archivists.org/glossary/terms/e/emulation. to preserve the source and as much of the original processes as possible.

Learning Outcomes

- Explain how emulation is performed.
- Explain the advantages and disadvantages of emulation as a preservation methodology.
- Evaluate potential emulation solutions.

Readings


W10 (10/31) - Archival Storage & Technology Infrastructure

Just as physical archival materials need storage space and physical facilities digital materials need digital storage space and other hardware to properly preserve them.

Learning Outcomes

• Explain the use of hashing algorithms to verify file integrity over time.
• List the attributes of good storage infrastructure, explain why they are, and provide examples.
• Describe the technology management concerns in building preservation infrastructure.

Readings

• Reference Model for an Open Archival Information System (OAIS), sec. 4.1.1.3.
• Lavoie, The Open Archival Information System (OAIS) Reference Model, sec. 5.3.2.
• Marks, Becoming a Trusted Digital Repository, sec. 5.

W11 (11/07) - Information Packages & Data Management

The OAIS requires information packages to be managed but does not dictate how to implement them. What do information packages look like in practice and how can we manage them?

Learning Outcomes

• Discuss packaging formats for long-term storage, such as PDF, XML, encapsulation.
• Discuss a variety of formats commonly used that serve as SIPs, including XML, tar, zip, and bags.
• Explain the challenges and possible solutions for managing digital object versioning.
Readings

- Reference Model for an Open Archival Information System (OAIS), sec. 4.1.1.4 and 4.2.
- Lavoie, The Open Archival Information System (OAIS) Reference Model, sec. 5.3.3 and 5.4.
- Marks, Becoming a Trusted Digital Repository, secs. 4.1.3–4.2 and 4.4-4.5.

W12 (11/14) - Access

The end of all preservation is access. Given the methodologies we have for preservation, how can we ensure eventual access to these records?

Learning Outcomes

- Explain the opportunities and complications of providing access to digital materials
- Describe some of the options for providing access to digital materials

Readings

- Reference Model for an Open Archival Information System (OAIS), sec. 4.1.1.7.
- Lavoie, The Open Archival Information System (OAIS) Reference Model, sec. 5.3.5.
- Marks, Becoming a Trusted Digital Repository, sec. 4.6.
- Josh Schneider, “Viewing Email Through a New Lens: Screening, Managing, and Providing Access to
W13 (11/21) - Preservation Planning

Digital preservation doesn’t “just happen.” An effective, efficient digital preservation program must be planned. A digital preservation strategy begins with an understanding of the end point, and must address issues of feasibility, sustainability, practicality, and appropriateness.

Learning Outcomes

- To list different benefits and limitations of preservation as a service and internal preservation program.
- To discuss when different strategies are appropriate.
- To describe a scenario for automated, rules-based processing.
- To describe the workflow for processing collections of digital records.

Readings

- *Reference Model for an Open Archival Information System (OAIS)*, sec. 4.1.1.6.
- Lavoie, *The Open Archival Information System (OAIS) Reference Model*, sec. 5.3.4.
- Marks, *Becoming a Trusted Digital Repository*, sec. 4.3.
- Gengenbach, “‘The Way We Do It Here’,“ 3–6, 70–88.

W14 (11/28) - Repository Systems

In order to automate and effectively manage their digital objects many repositories rely on complex repository systems.
Learning Outcomes

- Describe the use of digital repository software such as Fedora or Dspace and be able to utilize commonly used digital repository software
- Evaluate the relative strengths and weaknesses of digital repository software applications
- Choose appropriate technologies to manage an archival program.
- List third-party digital preservation options available.

Readings

- Reference Model for an Open Archival Information System (OAIS), sec. 6.
- Lavoie, The Open Archival Information System (OAIS) Reference Model, sec. 5.5.

W15 (12/05) - Administration

Digital preservation is a complex process which needs to be managed including interactions between OAIS Agents and Functions. How can we assess our current digital preservation capabilities and make plans for improvement?

Learning Outcomes

- Discuss organization readiness for digital preservation.
- Discuss the elements of a policy framework.
- Name and describe major preservation projects

Readings

- Reference Model for an Open Archival Information System (OAIS), sec. 4.1.1.5.
- Lavoie, The Open Archival Information System (OAIS) Reference Model, sec. 5.3.6.
- Marks, Becoming a Trusted Digital Repository, section 3.


**Wk 16 (12/12) - Final Exam**

The exam will be available on GAView at 6:30 p.m. and should be submitted to the drop-box before 9:00 p.m. The exam will include a combination of short and long answer questions.