ARST 5300: DIGITAL PRESERVATION
Sample Syllabus (Based on Fall 2014)

Instructor
Seth Shaw
Office location: Baker University Center 316C
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
Lectures: Monday, 6:30 - 8:30 P.M. via WebEx

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 associated with each of the archival domains. Introduction to archival aspects of media longevity, technological obsolescence and compatibility; information representation and experience; authenticity and integrity; backup, 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, financial sustainability, 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**

Students will be able to:

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. Create institutional preservation plans and understand preservation action plans for digital archival content.
7. 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.
8. To describe a workflow for digital archives, including objects (submission, archival, and dissemination packages), roles, and functions.
9. Discuss approaches to preserving common, challenging record formats, including social media, email, and websites.
10. Describe and explain the use of a variety of digital preservation tools.

**Readings**

Required readings are listed in the Course Schedule section of this syllabus, as are optional readings for each class session. You are expected to read all required readings before the start of the class session for which these readings are assigned; you may choose to read the optional readings for the particular class session as well, though this is not a requirement. Optional readings are provided to encourage and develop understanding of topics of particular interest to you.

**Grading**

Participation:
- In-class: 20%
- Online: 15%

Assignments & Labs: (5% each, 35% total)
- Equipment & Media Inventory
- Software & File-type Survey
- Organizational Framework
- Repository Workflow Lab
- Forensics & Network Transfer Lab
- Migration Lab
- Emulation Lab

Exams: (15% each, 30% total)
- Theory
- Final
Online Discussions

Over the semester, you will be asked to discuss assigned readings. The discussion board is similar to a seminar. The class will have a chance to explore the ideas in the readings. Working as part of the group, you will better understand what the authors have to say and may discover a variety of interpretations.

The discussion boards often contain suggested questions intended as a primer for conversation. Generally speaking, however, consider how each week’s readings apply to the listed learning outcomes. 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.

To get full credit for discussion, you submit at least one post each week that provides some context (information about the article or assignment) and offers your own analysis and evaluation of the ideas in article or substantially builds on/answers another’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). Posts must be made before class starts to receive credit.

Posts to the discussion boards are assessed primarily on content. Your point should be clear and whenever possible they should be supported with evidence from the literature or your experience. Formal footnotes are not necessary, but your sources should be clear. Your style may be more conversational and colloquial. However, posts should be appropriate for a professional discussion list.

Computing Requirements & Responsibilities

Each student enrolled at Clayton State University is required to have ready access throughout the semester to a notebook computer that meets faculty-approved hardware and software requirements for the student’s academic program. Students may use either a Windows or Macintosh computer. Computers should have at least 1GB of RAM, and 2GB or more is highly recommended. Computers must have at least 50GB free disk space.

Students in the MAS program must also have:

- High-speed Internet.
- A headset with a microphone (preferred) or earphones during synchronous classes to minimize feedback. Typical price is $25 and up. Headsets with a USB connection are recommended. For purposes of comparison, not endorsement, see Logitech’s H390.
- A webcam. Typical price is $25 and up. For purposes of comparison, not endorsement, see Microsoft LifeCam Cinema 720p HD Webcam or LifeCam VX-3000.

Course Schedule

Wk 01 (8/18) - Introduction and Orientation

Students and instructor introductions. Review syllabus and discuss the expectations and objectives for the course. How to prepare for discussion and labs. Overview of units to be covered in the course, including:

- Digital Preservation Theory (challenges, strategies, and tactics).
- Preservation Infrastructure & Trusted Repositories
- Preservation Strategies & Tactics

Learning Outcomes

- Elements of digital preservation: technology, policy, and administration

Readings

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

Readings

Discussion
- Consider your own collection of computer equipment & media. What would they tell someone else about you?
- What might you need to know about older (obsolete?) technology to be able to preserve records created not that long ago?

Assignment
   Equipment & Media Inventory. *Due: 08 Sep.*

**Labor Day (09/01) - No Class**

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.

Learning Outcomes
- To describe the differences between various user interfaces and to discuss the affordances of each.
- To describe the use of an API and it's impact on recordness or experience.
- To explain the PREMIS data model's Intellectual Entity & Objects.
- To describe how the concept of representation information applies to digital information and software.
Readings

Discussion
Consider the Court of Appeals decision in the Lake v. City of Phoenix case. Would you support the majority or the dissenting opinion? Why? (Set aside the debate about what makes a record "public.")

Assignments
Software & file-type survey. Due: 22 Sep.

Wk 04 (9/15) - Digital Preservation Threats & Tactics
A variety of long and short-term threats to digital preservation exist. What are they and how can we mitigate their effects?

Learning Outcomes
- To list and explain the general threats to digital information.
- To list and generally describe the tactics used to mitigate digital preservation threats with their advantages and limitations.
- To list and explain the methods for identifying a file's format and where to find format documentation.

Readings
- "Preserving Digital Records: Guidelines for Organizations." InterPARES 2
- "Digital Preservation Strategies," in Digital Preservation Management: Implementing Short-term Strategies for Long-Term Problems. (Optional, but good summary list of strategies.)

Discussion
Briefly describe a threat to digital preservation and an approach to mitigate it.

Wk 05 (9/22) - Documentation & Metadata
Digital objects cannot (generally) be physically inspected but instead rely on documentation and metadata to support preservation.

Learning Outcomes
- Identify appropriate documentation for a preservation program and explain why.
- Explain the differences between technical, descriptive, and interpretive metadata and list examples of each.
Readings

- Caplan, Priscilla. Understanding PREMIS (Library of Congress, 2009). Note: you read part of this earlier; now read the rest of it.

Discussion

- Why is metadata especially important for preserving digital materials?

Wk 06 (9/29) - Theory exam

The exam will be available on GeorgiaView at 6:10 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. Note: I will be in Arizona teaching a workshop this day. Although I will be teaching when the exam begins please let me know ASAP if you encounter any problems. I plan to have a workshop break ~ 6:15 pm EST and hope to trouble-shoot problems then. (You are essentially getting a 20-minute bonus in exchange for my absence.) I will be generally available after 8 p.m. EST.

Discussion

- Submit a question to the discussion board you might, hypothetically, see on the exam.
- Answer one of your classmate's questions.

Wk 07 (10/06) - Defining Repositories: Open Archival Information System (OAIS) Reference Model

OAIS is a high-level model of the roles, entities, and functions that make up a digital archives. It is the foundation of best practices for implementation.

Learning Outcomes

- To explain the fundamental agents, objects, and activities in the high-level OAIS model.
- To explain the importance of each data model's various components.

Readings

- Reference Model for an Open Archival Information System (Consultative Committee on Space Data Systems, 2012), sections 1-3. Note: You may want to review parts of OAIS not included in the assigned readings for additional information.

Discussion

- OAIS discusses three different agents. Who are they and what are their roles or stakes?

Wk 08 (10/20) - Trusted Digital Repositories

What does it mean for a digital repository to be trusted? The Trusted Repositories Audit and Certification: Criteria and Checklist (TRAC) provides an organizational context for the OAIS implementation model.

Learning Outcomes

- To describe the role of an audit in a asserting trustworthiness
- To discuss the major areas of a TRAC audit.
• To distinguish TRAC from DRAMBORA.
• To describe the relationships between TRAC & the related International Standards Organization (ISO) standards

Readings
• Trusted Repositories Audit and Certification: Criteria and Checklist (OCLC and CRL, 2007).

Discussion
• What makes a person trustworthy? How does that compare to a trustworthy record or repository?

Wk 09 (10/27) - Organizational Framework
The first step in developing a digital preservation action plan identifies the program's context within a particular organization.

Learning Outcomes
• To discuss organization readiness for digital preservation.
• To discuss the elements of a policy framework.
• To describe essential aspects of policies and procedures and their components
• Name and describe major preservation projects

Readings

Discussion
• Rephrase some of the organizational capabilities measured in Ashley and Dollar’s model as goals to be achieved.

Assignments
Organizational Framework. Due: 17 Nov.

Wk 10 (11/03) - Preservation Infrastructure
Digital preservation plans are implemented using a variety of technologies that work together. Overview of software designed to preserve digital information. Digital preservation establishes requirements for the records as information packages, including metadata.
Learning Outcomes
- To list technologies to support OAIS functions, including submission, ingest, storage, data management, and access.
- List the attributes of good storage infrastructure, explain why they are, and describe examples.
- To explain the use of hashing algorithms to verify file integrity over time.
- To discuss a variety of formats commonly used that serve as SIPs, including XML, tar, zip, and bags.
- To discuss formats for long-term storage, such as PDF, XML, encapsulation.
- To discuss packaging for preservation v. for access.

Readings

Discussion
- What affordances would appropriate for preservation or access formats? Why?

Wk 11 (11/10) - Digital Repositories
The goal is to differentiate the approaches being used, services being created, and the applicability for various collections and/or organizations. Evaluation of digital repository software, including LOCKSS, iRODS, Dspace, Fedora, Archivematica, and micro-services based solutions.

Learning Outcomes
- To understand the use of digital repository software such as Fedora or Dspace and be able to utilize commonly used digital repository software
- To understand the relative strengths and weakness of digital repository software applications
- To use appropriate technologies to manage an archival program.
- To understand third-party digital preservation options available.
- To describe LOCKSS, Dspace, and other repositories

Readings
- Abrams, Stephen ; Cruse, Patricia ; Kunze, John ; Minor, David (2010) "Curation Micro-Services: A Pipeline Metaphor for Repositories." The 5th International Conference on Open Repositories (OR2010), Madrid, Spain, 6-9 July 2010 Note: see both attached files.

Discussions
- What are the goals of having a digital repository? How is it different than preservation infrastructure?

Wk 12 (11/17) - Developing a Digital Preservation Strategy
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

- Gengenbach, Martin J. "'The Way We Do It Here': Mapping Digital Forensics Workflows in Collecting Institutions." University of North Carolina at Chapel Hill, 2012. Note: read the Introduction, Discussion, and Conclusion (pp.3-6 & 70-88). Also, read the workflow descriptions for at least two of the nine repositories; look over the diagrams for the others.

Discussion

Explain and define in your own words a digital preservation strategy or stage. When might you select this strategy over another?

Assignment

Repository Workflow Lab. Due: 24 Nov.

Wk 13 (11/24) - Forensics & Network Transfers

The active preservation of digital materials often begins at the point of capture. Media and network-based materials require different methods for successful capture.

Learning Outcomes

- Explain the relationship between computer forensic science and digital preservation practice, including the techniques used.
- To distinguish deleting and erasing files completely, define slack space.
- Compare and contrast the acquisition of media and network-based materials.
- To list and describe aspects of techniques to transfer SIPs, including disk and network transfer.
- To explain the structure and use of BagIt bags.

Readings

- Gengenbach, Martin J. "'The Way We Do It Here': Mapping Digital Forensics Workflows in Collecting Institutions." University of North Carolina at Chapel Hill, 2012. Read the literature review (pp.6-21). (Optional.)
- BagIt: Transferring Content for Digital Preservation [video] Library of Congress. (Optional.)

Assignment

Forensics Lab. Due: 1 Dec.

Wk 14 (12/01) – 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 pros and cons of format migration.
- Evaluate potential migration paths.

Readings
- "Selecting the right preservation strategy: Migration" in Workbook on Digital Private Papers. Personal Archives Accessible in Digital Media (paradigm) Project. (Last updated Jan 2008.)

Assignment
Migration Lab. Due: 8 Dec.

Wk 15 (12/08) – Emulation
Emulation, "the use of one system to reproduce the functions and results of another system" (SAA Glossary) and aims to reproduce the original experiential nature of records as closely as possible.

Learning Outcomes
- Explain how emulation is performed.
- Explain the pros and cons of system emulation.
- Evaluate potential emulation tools.

Readings
- "Selecting the right preservation strategy: Emulation" in Workbook on Digital Private Papers. Personal Archives Accessible in Digital Media (paradigm) Project. (Last updated Jan 2008.)

Assignment
Emulation Lab. Due: 15 Dec.

Wk 16 (12/15) - 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.

Discussion
- Submit a question to the discussion board you might, hypothetically, see on the exam.
- Answer one of your classmate's questions.