NYU Moving Image Archiving and Preservation

DIGITAL PRESERVATION Fall 2016

CINE-GT 1807
Tuesdays, 6 – 10PM
721 Broadway, room 652
http://www.digiprez.com

INSTRUCTOR

Nicole Martin

e: **nmartin@nyu.edu**

m: (914) 266-2165

Office hours: by appointment

ABSTRACT

This course address the use of digital files and infrastructure as preservation media, and will investigate current theories and practices for the conservation and preservation of both digitized and born digital materials. Students will learn about digital preservation environments and repositories, and what infrastructure, policies and procedures are integrated into a repository to make it preservation compliant. Students will gain practical skills with identification, analysis, handling, and risk assessment for works as a whole, their component parts, and associated software and metadata. Initiatives and research and development efforts by the national, international, regional, and cooperative organizations will be explored. Digital literacy will be emphasized, and through a combination of lectures, discussion, demonstrations, hands-on activities and lab exercises, students will develop an increased understanding of digital technologies, ecosystems, and requirements.

OBJECTIVES

• Understand and articulate the requirements, responsibilities, and functions of digital preservation environments;

• Understand the computing environment in which digital objects or collections were created, and the optimal environment in which they will be preserved;

• Identify and characterize digital objects and become familiar with format specifications and associated metadata;

• Demonstrate an understanding of digital infrastructures, standards, protocols, and technologies including operating systems (UNIX/Linux/macOS/Windows), network sharing protocols (AFP, NFS, SMB), networks (HTTPS, SFTP) file systems (Mac OSX, FAT32, EXFAT, ext), and storage architectures (SAN, NAS, DAS);

• Understand strategies for digital preservation management

EXPECTATIONS

Students are expected to do required readings each week in order to be prepared for class exercises and discussions. Readings assigned for the week should be read in *advance* of the class session. Additionally, tutorials or homework exercises will be assigned to help practice and develop skills.

Attendance at all classes is expected; more than one unexcused absence will affect grading. Grades will be based on a combination of class preparedness and participation (35%); and assignments (40% for the final project, and 25% for assignment 1). Class sessions will include a mix of lecture, discussion, demonstrations, and group or individual exercises. Your participation in these in class activities is a required part of your grade. Many in-class exercises and homework assignments will require a computer. Student laptops will be required most weeks for in-class exercises, so please plan on bringing your laptop to class. If you can't bring a laptop, you can work with others, or possibly borrow one from the department. Please inform the instructor if you regularly cannot bring a laptop so that arrangements can be made.

Part of your participation grade may include volunteering to research and present briefly on specific topics addressed throughout the semester.

ASSIGNMENTS & STUDENT WORK

There will be two assignments: an interview with an archivist or practitioner who works in the field of digital preservation, and an individual final project (of your choice) to be negotiated with the professor. Students are strongly encouraged to submit a draft of their final project in advance of the deadline. Assignments will be submitted electronically in the PDF file format via email to the instructor at nmartin@nyu.edu. Feedback will be provided within three weeks after the (on-time) assignment is submitted. Student works containing sensitive or proprietary information will be archived by the Moving Image Archiving and Preservation program department, but never published. Works that do not contain sensitive information will be considered for publication on the MIAP website.

SKILLS

- Introductory bash scripting

- File identification (mediainfo, exiftool, DROID)

- Data storage, formatting and analysis

- Safe file transfer

- Fixity checks for files and collections

- Data packaging and disk imaging

- Web archiving

- HDV tape and DVD digitization

- Read/Write data to LTO tape using LTFS

- Basic web archiving using Archive-IT and Web Recorder

TEXTS

There are no required texts for this class. Most readings can be found online. Texts that are not available online will be on reserve in the Bobst Library and the Cinema Studies Film Study Center, or provided by the instructor.

The course will require the use of free or pre-licensed software, which students may be required to download in order to complete in-class and homework assignments. Please come prepared with software downloaded and installed in advance of the class when instructed.

CLASS SCHEDULE

01 :: Sept 6 :: Introduction to Digital Technology and Preservation

Topics

- Introduction: welcome, goals, syllabus/assignments/course review
- Introduction to digital technology and preservation
- How computers work: Part 1, Hardware
- Brief history of digital technology & computing 1930-1995

Lab

- Class exercise: Technology timeline, 1995 and beyond

Readings

1. [ALA Freedom to Read Statement](#)
2. [ALA Library Bill of Rights](#)
3. [Clark & Steadman: Alan Turing's Legacy (WIRED)](#)
4. [DP Workshop Digital Technology and Preservation Timeline](#)

02 :: Sept 13 :: History of Digital Preservation, Introduction to OAIS Model

Topics

- Brief history of digital preservation efforts and initiatives
- OAIS reference model, history, context, and model itself

Lab *Bring Laptops to Class*

- Booting from the digital preservation class disk image USB thumb drive
- Follow along exercise: Navigate file systems with the command line, create files, create folders, list files and directories (pwd, cd, mkdir, touch, ls, etc.)
- Individual exercise: Create your own directory structures and files based on a theme of your choice and create an "ls" report.

Readings

[Besser: Moving from Digital Collections to Interoperable Libraries](#)
[1996 Task Force on Archiving of Digital Information](#)
[Rosenthal: Requirements for Digital Preservation Systems](#)
[Digital Preservation Management Workshop: Sections 1–3](#)
[Lavoie: The Open Archival System Introductory Guide](#)

03 :: Sept 20 :: Operating Systems and Computing Environments

Topics

- How computers work: Part 2, Software
- Computing foundations and digital preservation: UNIX, Linux, the Free Software Movement and Open Source Software

Lab *Bring Laptops to Class*

- Demo: nano and cron
- Follow along exercise: Write a bash script using nano
- Individual exercise: Write your own bash shell script using ATOM (GUI) with a prepared script to perform a batch process (ffmpeg, disk usage, df, etc.)

Readings

[Finley: Linux Took Over the Web. Now, It's Taking Over the World (WIRED](#))
[Kelty: Two Bits: The Cultural Significance of Free Software](#) (Chapter 3 only)
[Lyons: Introduction to Using the Command Line](#)
[UNIX tutorial 1 & 2](#)

04 :: Sept 27 :: Digital Format Identification and Characterization

Topics

- NDSA Levels of Preservation
- Characteristics of digital files
- File Format Registries (PRONOM) and Format Analysis Tools (DROID, JHOVE)

Lab *Bring Laptops to Class*

- Partner Exercise: Build your own "files"
- Individual Exercise: File identification with exiftool, ffprobe and mediainfo
- Individual Exercise: File identification and validation with DROID and JHOVE
- **Unix tutorial output 1 & 2 review**

Readings

[NDSA Levels of Preservation](#)
[Peltzman: Expanding NDSA Levels of Preservation](#)
[Lacinak: A Primer on Codecs](#)
[Jackson: Formats Over Time, UK Libraries](#)
[Rosenthal: Formats Through Time, DHSR Blog](#)

Rosenthal: Cleaning Formats Through Time, DHSR Blog

05 :: Oct 4 :: Digital Asset Management Tools & Strategies, Data Storage Architecture

Topics

- Digital Asset Management Strategies
- The Digital Curation Center Lifecycle Model
- Digital File Systems and Storage Media
- Servers and Storage Architecture

Lab *Bring Laptops to Class*
- Follow-along exercise: Hard drive formatting and partitioning
- Follow-along exercise: Setting up macOS file share
- Old Media Lab with Ethan Gates

Readings

Higgins: DCC Lifecycle Model (International Journal of Digital Curation)
Google – Pinheiro, Weber, Barroso: Failure Trends in a Large Disk Drive Population
Backblaze – Klein: What Can 49,056 Hard Drives Tell Us?
Altman: NDSA Storage Report
Glicksman: Storage Architectures and Networks
Unix tutorial 3 & 4

06 :: Oct 11 :: Data Integrity, Fixity and Transfer

Topics

- Fixity and maintaining data integrity
- History of checksums and use for preservation
- Checksums for audiovisual media
- Safe file transfer

Lab - Bring Laptops to Class
- Follow along exercise: Fixity and checksums (md5, sha1, hashdeep, framemd5)
- Follow along exercise: File transfer (cp, mv, rsync)
- **Unix tutorial output 3 & 4 review**

Readings

Baily: Protect Your Data (The Signal)
Rice: Reconsidering Checksums
Havemeyer-King: Trojan Dots and DIY Solutions (NDSR Blog)

07 :: Oct 18 :: Intro to Web Archiving and Archive-IT

Topics

- Introduction to the Internet and World Wide Web
- Internet Architecture and Digital Preservation
- Internet Preservation Initiatives

Lab - Bring Laptops to Class
- Rhizome Webrecorder workshop
- Internet Archive-led Archive-IT training

Readings

Summers: Web as a Preservation Medium (Medium)
McKeehan: Symmetrical Web Archiving with Webrecorder (NDSR Blog)
Fino-Radin: Rhizome Preservation
Lasar: 25 Years of Hypercard (Ars Technica)
Berners-Lee: WorldWideWeb Executive Summary
Archive.org readings about Archive-IT
Five Challenges of Web Archiving
Archiving Social Media Sites
Glossary of Web Archiving Terms

08 :: Oct 25 :: Digital Physical Format Digitization, LTO Tape Demo

Topics

- **Guest Speaker: Kara Van Malssen**
- Digital physical format digitization (DV, HDV, DVD)
- LTO backup media and LTFS

Lab - Bring Laptops to Class
- Demo: LTO tape recovery from BRU and LTFS
- Partner Exercise (Stations): MiniDV (HDV/DV) digitization with DVHSCap , DV analysis with DV Analyzer software, DVD Digitization with MPEGStreamclip

Readings

Rice: Digitization Software Obsolescence, Too?
Van Malssen: Tools for Smaller Budgets
Clipper Notes: LTO tape advantages over disk
Pease, Amir, et al: The Linear Tape File System

:: Oct 28 :: ASSIGNMENT #1 FINAL PAPER DUE at 6pm

09 :: Nov 1 :: Preservation Metadata

ASSIGNMENT #1 PRESENTATIONS DUE

Topics
- Introduction to preservation metadata
- PREMIS, METS, and XML

Lab
- **Assignment 1 Presentations**

> Lavoie, Gardener: Preservation Metadata
> Caplan: Understanding PREMIS
> PREMIS Data Dictionary (Review)
> Amaral: METS for Transferable Metadata

:: Nov 8 :: AMIA – Students in Pittsburgh for Association of Moving Image Archivists Conference

NO CLASS

10 :: Nov 15 :: Sustainability, Policy and Trust

Topics
- **Guest Speaker: Seth Anderson**
    - Sustainability and policy for digital preservation environments
    - TRAC and the TRAC Checklist
- LOCKSS and CLOCKSS

Lab
- **Assignment #2 Workshop (Project Approvals)**
- Old Media Lab with Ethan Gates – Netscape and Hypercard

Readings

> TRAC Checklist (read pages 1–8, skim/review remainder of document)
> What is LOCKSS? and LOCKSS Preservation Principles
> The CLOCKSS Archive and What is the Difference Between LOCKSS and CLOCKSS?
> Center for Research Libraries: Certification Report on CLOCKSS
> Rosenthal: TRAC Audit, Lessons Learned (DSHR Blog)

Smith: Sustainable Economics for a Digital Planet (Exec. Summary & Chapter 1)

11 :: Nov 22 :: Business Case for Preservation & Digital Asset Management Basics

Topics

- **Guest Speaker: Sally Hubbard - Business Case for Digital Preservation**
- Business cases, institutional readiness
- Digital preservation systems and DAMS

Lab

- Digital preservation applications demo

Readings

Digital Preservation Business Case Toolkit
Kenney, McGovern: Five Organizational Stages of Digital Preservation
Vermaaten: SPOT Model
Smith: Sustainable Economics for a Digital Planet (Chapters 2 & 3)

12 :: Nov 29 :: Post-Custodial Archiving and Economically Sustainable Digital Preservation (Thanksgiving Week)

Topics

- **Guest Speaker: Rachel Mattson on Post Custodial Archiving**
- Long-term preservation strategies and economically sustainable digital preservation

Lab

- **Assignment #2 Workshop (Final Questions)**
- Group exercise: Preservation metadata exports: Analysis of XML files from technical metadata programs (ffprobe, mediainfo, exiftool)

Readings

Rosenthal: Half-Empty Archive (DSHR Blog)
Rothenberg: How far we have come and what's next? VIDEO and SLIDESHOW
Rosenthal: Half life of digital formats
Lavoie: The fifth blackbird
Smith: Sustainable Economics for a Digital Planet (Chapter 4 & 5)

13 :: Dec 6 :: LAST CLASS – Presentation of Final Projects

FINAL PROJECT PRESENTATIONS

:: Dec 16 :: LAST FALL CLASSES ASSIGNMENT 2 FINAL PAPERS DUE at 6pm

Assignment 1: Practitioner Interview – Final Paper Due Friday, October 28th, 6pm

For your first assignment, you will write a 6-8 page paper based on an interview you will conduct with an archivist, digital asset manager, preservationist or other practitioner in the field of digital preservation. Applying topics learned in class, you will ask a series of questions of the practitioner to investigate the inner-workings of digital repositories and archives, and to gain a better understanding of preservation practices. Topics will include but are not limited to the OAIS model, computing environments, file format identification, digital asset management tools and strategies, data storage and architecture, data integrity and fixity checks.

Ask your subject if it is okay to contact them with follow up with questions after the interview.

Due Dates
Wednesday, October 5th through Wednesday October 12th – Conduct practitioner interview
Friday, October 28th, 6pm – Final 6-8 page papers due (single spaced)
Tuesday, November 1st – Class presentations due

Questions

1. General questions: What digital information do you collect? What are the contents of your collections? Tell me more about the history of your collections and your archive. If you are a vendor, what types of collections to your clients send to you? What services do you perform?

2. How does this collection fit into the OAIS model? How doesn't it fit? Which computing environments were used to create the collection items? Is there a pattern (like in an archive that serves a production environment)? What is the computing environment used to archive the collection and within the repository? Describe the computing environment.

3. Which computing environments were used to create the collection items? Is there a pattern (like in an archive that serves a production environment)? What is the computing environment used to archive the collection and within the repository? Describe the computing environment.

4. What digital formats are represented in the collection? How are they identified? Are different file types treated differently within the archive, and if so, how? Can you give an example or two? Does the archive contain any obsolete or at-risk formats, and if so, what? Do you treat at-risk formats differently than other formats, and if so, how? Does the archive contain formats that are not at risk by may be inherently more fragile than other formats? If so, how? Does the archive contain digital physical formats such as DV tapes or DVDs?

5. How are collections stored? Are collections stored on a server, and if so, what kind of storage media does the server use (flash or hard drive)? How do users and administrators connect to the server (over ethernet or fiber, and how fast are speeds)? Does the repository rely on LTO tape (LTFS or proprietary?), cloud storage (which service?), external hard drives, or other forms of storage? How do these systems work (hardware, software specifications, etc.)? Does the repository use geographical separation? Have any of these systems ever failed or experienced outages? Do you have redundant backups of your data, and if so, do store data in duplicate, triplicate… even quadruplicate?

6. How do you maintain data integrity in your collection? Do you use fixity checks? If so, what kind? How often do you run them? What do you do when you find bit rot or irregularities in your fixity checks? If you have audiovisual media in your collection, do you use special tools for fixity checks or to verify data integrity? What method or program do you use to transfer data within your repository?

Assignment 2: Final Project – Due Friday, December 16th, 6pm

A substantive, in-depth, individual project or 8 – 12 page paper on digital preservation or access. You should try to discuss your proposed topic with the instructor in September or October. You must turn in a formal project/paper proposal no later than Oct 21, and the instructor must approve your topic and scope before you start work on it. You are strongly encouraged to turn in a draft or outline by November 21st.

Due Dates
Tuesday, October 18th – One-page proposal due
Tuesday, November 21st – Draft paper/outline (optional)
Tuesday, December 6th – Twenty minute in-class presentation, including Q&A
Friday, December 16th, 6pm – Final 8-12 page paper (single spaced) + bibliography/webography

----

Tisch Policy on Academic Integrity
The core of the educational experience at the Tisch School of the Arts is the creation of original work by students for the critical review of faculty members. Any attempt to evade that essential transaction through plagiarism or cheating is educationally self-defeating and a grave violation of Tisch's community standards. Plagiarism is presenting someone else's original work as if it were your own; cheating is an attempt to deceive a faculty member into believing that your mastery of a subject or discipline is greater than it really is. Penalties for violations of Tisch's Academic Integrity Policy may range from being required to redo an assignment to dismissal from the School. For more information on the policy--including academic integrity resources, investigation procedures, and penalties--please refer to the Policies and Procedures Handbook on the website of the Tisch Office of Student Affairs.

Non-Discrimination and Anti-Harassment Policy

NYU is committed to equal treatment and opportunity for its students and to maintaining an environment that is free of bias, prejudice, discrimination, and harassment. Prohibited discrimination includes adverse treatment of any student based on race, gender and/or gender identity or expression, color, religion, age, national origin, ethnicity, disability, veteran or military status, sexual orientation, marital status, or citizenship status, rather than on the basis of his/her individual merit. Prohibited harassment is unwelcome verbal or physical conduct based on race, gender and/or gender identity or expression, color, religion, age, national origin, ethnicity, disability, veteran or military status, sexual orientation, marital status, or citizenship status. Prohibited discrimination and harassment undermine the character and purpose of NYU and may violate the law. They will not be tolerated. NYU strongly encourages members of the University Community who have been victims of prohibited discrimination or prohibited harassment to report the conduct. MIAP students may make such reports to MIAP Associate Director, Scott Statland, or directly to Marc Wais, Senior Vice President for Student Affairs.

Sexual Misconduct, Relationship Violence, and Stalking Policy

NYU seeks to maintain a safe learning, living, and working environment. To that end, sexual misconduct, including sexual or gender-based harassment, sexual assault, and sexual exploitation, are prohibited. Relationship violence, stalking, and retaliation against an individual for making a good faith report of sexual misconduct are also prohibited. These prohibited forms of conduct are emotionally and physically traumatic and a violation of one's rights. They are unlawful, undermine the character and purpose of NYU, and will not be tolerated. A student or employee determined by NYU to have committed an act of prohibited conduct is subject to disciplinary action, up to and including separation from NYU. Students should refer to the online [Sexual Misconduct, Relationship Violence, and Stalking Resource Guide for Students](#) for detailed information about on-campus and community support services, resources, and reporting procedures.