Digital Literacy for Moving Image Archiving and Preservation – Fall 2016

INSTRUCTOR

Sally Hubbard
Sah11@nyu.edu
917-771-6693 (mobile)
Office Hours: By Appointment

ABSTRACT

Moving images have are, and have always been, technical as much as artistic; their creation, management, and enjoyment requires the mediation of various technologies. Today, the technological ecosystem in which moving image collections are organized, preserved, and made available to users is increasingly a digital one. This situation is in many ways advantageous, as it means that the contemporary moving image archive professional shares a toolset with other digital communities. It also means that it is essential that professionals entering the field today are not only comfortable working with the tools of this ecosystem, but have sufficient understanding of how the tools work and may impact collections that they can better communicate archival risks and requirements to senior management, technologists and vendors; are better able to make informed decisions about collection management; and have the skills to act on those decisions.

This class will prepare incoming first year MIAP students for working with digital technologies throughout their academic and professional careers. The course will focus on media files, web applications, databases, and data management tools — technologies that play a fundamental role in moving image collections management today. The course will emphasize digital literacy so that students will be equipped to make informed technology decisions in the future. By introducing these topics in their first semester, this course provides students with core competencies that will be utilized in subsequent classes in the MIAP program.

COURSE OBJECTIVES

Upon completion of this course, students will:

- Have improved their ability to communicate with technologists and vendors;
- Understand digital moving image content (essence and metadata)
- Understand data/media management eco-systems and be familiar with basic data management tools and practices, including the importance of data quality, integrity, and portability
- Understand database functionality, and how databases compare to spreadsheet applications
- Understand network infrastructures, standards, protocols, and technologies including operating systems, networks, file systems, and storage systems
- Understand the role of APIs
- Understand application administration fundamentals, including rights and permissions
- Understand the role of programming and computing languages including HTML, CSS, JavaScript, and SQL
- Be familiar with software version control repositories such as Github
- Have been introduced to navigation and basic text and file manipulation through command line interfaces
- Understand that technology alone cannot provide solutions; best practices, policies, and procedures are key ingredients to successful technology usage for archiving and preservation of moving image collections

COURSE EXPECTATIONS

Laptop

Students are required to bring their own laptops to class each week. Both Windows and Mac are acceptable; pending they meet the following minimum requirements:

Mac
OS 10.6.8 or later
Intel Processor
At least 2 GB RAM
At least 30 GB available disk space

Windows
XP or later
Preferably 64 bit
At least 2 GB RAM
At least 30 GB available disk space

If you do not have access to a laptop, or do not have one that meets these minimum requirements, you may be loaned one for use during class. Please see the instructors.

Assignments

Each class will have one or more learning objectives and accompanying activities. Activities may be group or individual, and may be completed during class or as homework. Students will be required to turn in or otherwise demonstrate the results of weekly assignments before the start of the following class meeting, regardless of whether the activity was completed in class or at home as homework.

There will not be a final graded paper, but the class will conclude with each student creating an online exhibition that will compliment projects in the Introduction to Moving Image Archiving and Preservation course (MIAP students), or another course of the student's choice (non-MIAP students). By the end of the semester, students will have gone through all stages in the setup of a web application, including server setup, application installation, database creation and configuration, and front-end design. Students will also prepare media content and datasets for search and display.

Grading

Grades will be determined according to the following breakdown:
- Regular Assignments/Homework: 40%
- Participation and Attendance: 40%
- Final project: 20%

Attendance

Attendance at all classes is expected; more than one unexcused absence will affect grading.

Texts and Other Resources

There is no required textbook 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.

In lieu of purchasing textbooks, students will be required to license a hosted Virtual Private Server throughout the duration of the course. The service we will be using is called Digital Ocean (https://www.digitalocean.com/). The cost for the semester will be approximately $20-$30, but may be shared if students decide on a joint project. This can be determined during the course.

The course will also 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.

We will be making use of NYU's subscription to Lynda.com for online tutorials through nyu.edu/lynda as well as other freely available online training modules.

Communication

We will be using Slack for communication about class topics. Everyone will receive an invitation a Slack account, and should install the application on their computers. There will be different channels set up within the application for each topic. Students are strongly encouraged to help each other with troubleshooting; this will factor into your participation grade.

COURSE SCHEDULE

Class 1 (September 12):

Technology and the Moving Image Archivist: Digital Eco-Systems, Components And Communication

Topics: Course introduction and syllabus review. These topics will be introduced this week and reviewed in greater detail during the course.
- Why does a moving image archivist need to understand digital technologies?
- Digital vs. Analog
- What is a file?
- What is a media file?
- What is the internet?
- What is the Web?
- What is the Cloud?
- Digital media infrastructure components
- Making it work: standards, requirements and documentation
- Technical evolution: continuous change and education

Class 2 (September 19):

Digital Media Files as Complex Objects

Time-based digital media files (i.e. audio and video) are extremely complex and heterogeneous, with multiple components and parameters, all of which need to be accounted for in preservation. This class will deconstruct sample media files.
- Formats
- Codecs
- Channels

- Closed Captions and Subtitles
- File integrity

Activity:
- Guest Speaker: to be confirmed
- Download MediaInfo and review properties of video file.

Homework:
- Review and export MediaInfo results as a spreadsheet (csv file)

Class 3 (September 26):

Managing Content: Essence Data and Metadata 1

Topics: spreadsheets, databases and asset management systems. All these tools can be used to create so-called "smart content" – content married to the information needed to preserve and access it.
- Introduction to data storage formats: CSV, RDBMS/SQL
- Relational database management systems vs. flat data management systems
- Best practices for data creation in spreadsheets (Excel and Google Spreadsheets)
- Entity-relation modeling

Activity: Spreadsheet overview: Lynda.com Google Sheets Essential Training

Homework: Complete Google Sheets Essential Training
- Optional Review:
  - Open Refine: http://enipedia.tudelft.nl/wiki/OpenRefine_Tutorial
  - Lynda.com: Relational Database Fundamentals

Class 4 (October 3):

Managing Content: Essence Data and Metadata 2

Topics: NoSQL (Not Only SQL) databases include a variety of data models and systems that provide alternatives to traditional RDBMS/SQL tools, and may be more appropriate for given use cases.
- XML/document stores; triple stores, graph databases.

Activity: Guest Speaker Roy Walter

Homework: review resources at https://neo4j.com/developer/get-started/

Class 5 (October 17)

Managing Content: Essence Data and Metadata 3

Topics: Digital Asset Management and Media Asset Management systems link metadata and content together and provide a search interface to assets. They also provide additional functionality, such as linking associated files (e.g. versions and proxies); workflow or business process modeling; processing; tracking; automation, etc. This session will eview various MAM system types and capabilities.

Activity: Guest Speaker David Lipsey

Homework: To Be Assigned

Class 6 (October 24)

Media Management Environments

Topics: Media management environments, of which a MAM or DAM system is likely to be only a component, or in which there may be multiple MAMs serving different user groups or technical functional, can be very complex, and now are likely to include cloud components.
- Moving Images in an IT Environment
- Archival components (or the lack thereof)
- The Cloud
- APIs

Homework: To Be Assigned

Class 7 (October 31)

Network components and communication

Topics: Review the Open Systems Interconnectedness (OSI) conceptual model to understand the standards and protocols joining networks together
- Clients (browsers) and servers
- Network communication - TCP/IP
- File transfer - FTP, SFTP, SSH
- Web pages are just text: Introduction to HTML, CSS
- What is a CMS? WordPress and more

Activity:
- In-browser website manipulation

- Exploring the components and structure of a website
- Making a website

Tasks:
- Install
  - FTP client - Cyberduck (https://cyberduck.io/)
  - Text editor - Sublime text (http://www.sublimetext.com/)

Homework:
- Build your own website in 3 hours – CodeAcademy Make a Website:
  http://www.codecademy.com/skills/make-a-website
- Set up your Digital Ocean server: https://www.digitalocean.com/
- If you are on Windows, download Putty
- Recommended: Review HTML tutorial -
  http://www.w3schools.com/html/default.asp

Class 8 (November 7):

Under the Hood: Command Line Interface

Topics: Command Line Interface (CLI) vs Graphic User Interface (GUI). While GUI is immediately more user-friendly, CLI can offer many functional and practical advantages.
- Introduction to command line navigation
- Connecting to the server via SSH
- Introduction to package management
- Introduction to Git and GitHub

Activity:
- Working with command line on your own VPS (Virtual Private Server)
- Using Git and Github

Homework:
- Code Academy Command Line Tutorial:
  https://www.codecademy.com/courses/learn-the-command-line
- Read Anthony Cocciolo (2014), "Unix Commands and Batch Processing for the Reluctant Librarian or Archivist," Code4Lib journal, Issue 23, 2014-01-17. Available from http://journal.code4lib.org/articles/9158

Class 9 (November 14):

Creating an Online Exhibit - Application Installation

Topics: creating an online exhibition in class will allows us to build experience with both the technical requirements and other aspects of creating a successful application. This session will lay the application groundwork.
- Domain creation
- Database setup
- Working with configuration files
- PHP basics
- Introduction to Omeka, a free and open source web-publishing platform for the display of library, museum, archives, and scholarly collections and exhibitions.

Activity:
- Installation of Omeka

Homework:
- Read articles in the "Getting Started with the Software Application" section of the Omeka Documentation. Available from: http://omeka.org/codex/Documentation
- Start planning your Omeka site - send instructors your idea (1 paragraph) before the next class session

Class 10 (November 21):

Creating an Online Exhibit – Functional Requirements, Use Cases, and Design

Topics: no project is purely technical, and the more work put in understanding requirements and developing use cases before you build anything the more you will be prepared for problems that inevitably arise and the more elegant your final solution is likely to be:
- Meeting users needs and expectations
- Documenting requirements
- Understanding rights and permissions
- Planning features of an exhibition

Activity:
- Developing user personas and use cases
- Develop and plan for an online exhibition

Homework:
- Read articles in the "Getting Started with Project Planning" section of the Omeka Documentation. Available from: http://omeka.org/codex/Documentation

Class 11 (November 28):

Creating an Online Exhibit – Adding and Enhancing Functionality + Adding Data

Topics:
- Plugins and Templates
- Dublin Core and metadata practice
- Creation of files suitable for access
- File import
- Troubleshooting

Activities:
- Installing templates and plugins
- Configuring User roles, rights and permissions
- Creating/gathering files for exhibition and upload
- Metadata creation and editing
- Troubleshooting using StackExchange and other tech Q&A sites

Homework:
- Ludovico Fischer (2013), "A Beginner's Guide to HTTP and REST," Tuts+. Available from http://code.tutsplus.com/tutorials/a-beginners-introduction-to-http-and-rest--net-16340

Class 12 (December 5):

Creating an Online Exhibit – Adding Data (Part 2) + Look and Feel:

Topics:
- Setting up nightly MySQL backups
- The relationship between metadata and search engines
- Introduction to web programming: PHP and JavaScript
- Reading source code
- DIY usability testing: http://peek.usertesting.com

Activities:
- Basic metadata creation using Dublin Core

Homework:
Finalizing online exhibit for final class (December 13)

Class 13 (December 12):

Technical Best Practice from an Archival Perspective

Topics: Archivists are necessarily thinking long-term and aiming to build robust infrastructure and content that can survive into the future. Balancing this against real world short-term technical requirements and emergency work-arounds can be tricky, and one way to do it is to build best practices into everyday activity, and to disseminate them across institutions.
- Sponsorship
- Stewardship programs
- Training
- Incentives

Activities:
- Review of any outstanding issues with final project

Homework:
> Finalizing online exhibit for final class (December 13)

Class 14 (December 13):

Final Project Presentation + Pizza