ARST 5100 · ARCHIVES AND TECHNOLOGY
SYLLABUS : FALL 2015

Version 2: 19 August 2015
Subject to revision

INSTRUCTOR

Joshua Kitchens
Office location: Georgia State Archives
Office phone: 678-466-4427, 478-238-3976
Email: joshuakitchens@clayton.edu
Office hours: Face-to-face, phone, Collaborate: Monday through Thursday, 1:00 pm – 5:00pm EST, as available. Evening appointments are possible by appointment.

PREREQUISITES

ARST 5000, co-requisite.

CREDIT

3 weekly contact hours

SCHEDULE AND LOCATION

Lecture: Wednesday, 6:30 – 8:30 P.M. Via Colaborate and Recordings
Discussions, Assignments: Throughout the week via Georgia View (D2L).
Direct link: https://clayton.view.usg.edu/d2l/home/744076

FORMAT OF COURSE

Online course including weekly, synchronous lectures, asynchronous discussions, and tech labs. (Asynchronous options discussed below)

DESCRIPTION

An introduction to fundamental concepts and application of technology in an archival context, including hardware, operating systems, programming languages, data storage and backup, networking, and security. Overview of a wide range of digital record formats and their distinguishing characteristics, as well as common archival standards for metadata, including MARC format, Encoded Archival Description, and Dublin Core. Investigates the challenges of acquiring, preserving, and providing access to digital records. Students will be able to design and implement a simple database.

As increasing numbers of records are created in digital formats, archivists must have a fundamental understanding of technology and the practical skills to work with digital records. The core functions of the profession remain the same, but the manner archivists perform those activities is significantly different. Surveying records for appraisal will still result in a summary of record series, including titles, abstract of content, quantity, and location. For example, while archivists will continue to survey physical records, in the digital era they must also be able to survey the contents of hard drives and tapes. Digital records are not acquired in boxes, but on disks, drives, and file transfer. Arrangement and description can be based on software that harvest or generate metadata automatically. The instability of digital media and rapid technological obsolescence requires new approaches to preservation.

In addition, archivists must understand the characteristics of the digital materials in their collections in the same way they understand traditional media. Traditional documents have a support (paper, film), a carrier (ink, pencil, or pigment), encoding (alphabets and words), and intellectual organization (staples, folders). Digital records have similar characteristics; support may be disks or tape, carriers may be magnetic, optical, or physical signals; encoding includes ASCII and EBSDIC, and structure such as XML and other file formats.

This course will give students a rich understanding of the nature of digital materials. They will be exposed to the basics of systems administration and server applications. They will also get experience using tools to work with those materials, including an introduction to database design, hashing, and file integrity. These tools will enable archivists to be more self-reliant, to be able to accomplish many basic technical tasks, and – most important – to have the insights and vocabulary to talk to and collaborate with technologists on advanced projects.

COURSE LEARNING OUTCOMES

Students will be able to
- Describe the hardware components of a computer and their interactions, including virtualization.
- Describe a variety of operating systems, with an emphasis on the Linux operating system.
- Describe a variety of applications and to demonstrate the installation and configuration of software in a Linux environment.
- Describe the basics of a computer network, including hardware and protocols.
- Design and create a simple, relational database.
- Discuss a variety of techniques for troubleshooting implementation and configuration of computer systems and applications.

READINGS

No textbooks are required for this course. We will be making extensive use of online available resources. In the few cases where readings do include a chapter from a published monograph they will be scanned and placed in GAView (D2L). An *optional* text you should consider buying is *UNIX and Linux System Administration Handbook*, 4th ed. (Prentice-Hall, 2011), by Nemeth, Evi, Garth Snyder, Trent R. Hein, and Ben Whaley. This is a large and informative book that can be used as a secondary reference supporting the course content.

RESOURCES

- Computer Science E-1: Understanding Computers and the Internet. An extension course offered by Harvard professors Dan Armendariz and David J. Malan. "This course is all about understanding:

understanding what's going on inside your computer when you flip on the switch, why tech support has you constantly rebooting your computer, how everything you do on the Internet can be watched by others, and how your computer can become infected with a worm just by being turned on. Designed for students who use computers and the Internet every day but don't fully understand how it all works, this course fills in the gaps." http://computerscience1.tv/

- Computer Science 50: Introduction to Computer Science I. "A course offered by Harvard professor David J. Malan. "Introduction to the intellectual enterprises of computer science and the art of programming. This course teaches students how to think algorithmically and solve problems efficiently. Topics include abstraction, algorithms, encapsulation, data structures, databases, memory management, security, software development, virtualization, and websites. Languages include C, PHP, and JavaScript plus SQL, CSS, and HTML." http://cs50.tv/

GRADING

Online discussion and participation in lectures: 20%
Midterm: 20%
Assignments: (5% each, 40% total)
1. System Diary
2. Encoding
3. Setup a Virtual Computer
4. Shell Scripting
5. File system Survey
6. Network Discovery
7. Data Modeling
8. Create & Query a Database
Final Exam: 20%

EXPECTATIONS

Students are responsible for their own education. Throughout the course, students should assess their progress towards the course objectives and outcomes. At the same time, the course instructor is responsible to facilitate students' learning by structuring content, by providing a foundation of information through readings and lectures, by guiding discussion, and by answering students' questions.

Students should bring curiosity and creativity to the course. They are expected to think critically about the course content – the readings, the lectures, and discussion. Students are encouraged to (respectfully) challenge the ideas presented in the course. Those challenges must be justified based on the literature, empirical evidence, or other authoritative source. When evidence is contradictory, students should develop a synthesis that finds commonalities, identifies differences, and notes how a particular context may influence that synthesis. As such, there is seldom a "right" answer, but well-reasoned and well-informed points of view. Students are given credit for correcting course materials submitted by the instructor.

Students are expected to find, read, and share additional, relevant works and incorporate the ideas into class discussions.

COURSE POLICIES

GENERAL POLICY

Students must abide by policies in the Clayton State University Student Handbook.

ATTENDANCE POLICY (ASYNCHRONOUS OPTION)

Students are expected to attend online lectures unless prior arrangements have been made. If you miss a lecture, you should write a short paper (approximately 500 words) on that day's readings and topic to indicate that you have mastered the material.

Students seeking to complete the course asynchronously should watch the recorded lectures and post to that week's discussion board by the Friday after the scheduled class. These post should be a critical reflection of that week's lecture and discussion. These post should deal with the following topics.
- Major issues discussed during the week.
- Discuss your own experiences/thoughts about that weeks content
- Critical Reviews/Reflections of that weeks readings
- Problems had during assignment completions

LATE AND MISSED WORK

Assignments are due by 11:59 P.M. on the day indicated on the assignment. Work submitted a few minutes late will be accepted without penalty to allow for reasonable variation in clocks.

Without excuse, a grade of zero points will be assigned for missed work. If you have a conflict, discuss it with me prior to the due date for the assignment so other arrangements might be made.

DISABILITIES

Students with disabilities who require reasonable accommodations need to register with Disability Services in order to obtain their accommodations. You can contact them at 678-466-5445 or disabilityservices@clayton.edu. If you are already registered with Disability Services and are seeking accommodations for this course, please make an appointment with me to discuss your specific accommodation needs and give me your accommodations letter.

ACADEMIC DISHONESTY

Any type of activity considered dishonest by reasonable standards may constitute academic misconduct. Common forms of academic misconduct are cheating and plagiarism. All instances of academic dishonesty will result in a grade of zero for the work involved and may result in a failing grade in the course. All instances of academic dishonesty will be reported to the Office of Student Life/Judicial Affairs. Judicial procedures are described at http://www.clayton.edu/student-conduct/home.

STUDENT CODE OF CONDUCT

Behavior that disrupts the teaching–learning process during class activities will not be tolerated. While a variety of behaviors can be disruptive in a classroom setting, more serious examples include belligerent, abusive, profane, and/or threatening behavior. A student who fails to respond to reasonable faculty direction regarding online behavior and/or behavior while participating in course activities may be dismissed from class. A more detailed description can be found in the Student Code of Conduct. See also the disciplinary policy.

COMPUTING REQUIREMENTS AND 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. Also, each student is responsible for monitoring the Clayton State issued email address on a daily basis for official communications from faculty and administrators. Online courses require a high-speed Internet connection and a headphone with a microphone.

This course uses software that requires a minimum of 2GB RAM; more is preferred. Students are responsible for their own backups; an external hard drive is highly recommended.

Plan to stay after class on Wednesday nights for group discussions to help solve problems. All assignments are due at 11:59 on Fridays.

Assignment 1. System Diary
    Due: 9 Dec
    Keep a record of the process to install, configure, and maintain your virtual computer.

Assignment 2. Encoding
    Due: 04 Sep
    A short exercise to help understand binary, hex, and decimal encoding of digital information.

Assignment 3. Setup a Virtual Computer
    Due: 25 Sep
    Download and install Oracle VirtualBox software on your computer. Download the Xubuntu operating system and create a virtual computer running the Linux operating system.

Assignment 4. Shell Scripting
    Due: 16 Oct
    A short exercise to walk through programming concepts (variables, conditionals, and looping).

Assignment 5. Filesystem Survey
    Due: 30 Oct
    Explore and document the contents of a directory structure.

Assignment 6. Network Discovery
    Due: 06 Nov
    Locate and document information about your home network.

Assignment 7. Data Modeling
    Due: 20 Nov
    Design a catalog database, including identifying data elements and typical values; establish the relationships between elements; define a table.

Assignment 8. Create & Query a Database
    Due: 04 Dec
    Create a database based on a model, insert data, and create SQL queries to find records in your database.

COURSE READING AND TOPIC SCHEDULE

19 Aug ~ 1. Introductions and Orientation

 Students and instructor will introduce themselves. Review syllabus and discuss the expectations and objectives for the course. Overview of topics to be covered in the course and how they relate to each other. How to use Georgia View. How to prepare for discussion.

Learning outcomes
- Distinguish technical fluency (knowledge), competency (basic skills), and proficiency (advanced skills).
- List different kinds of technical skills, including programming, application development, system administration, and database design.
- Discuss the skills a digital archivist needs to work with electronic records and information systems.
- Discuss the impact of legacy systems on archives.
- How to read technical manuals.

Discussion
- Introduce yourself to the class. In addition to the introduction in ARST 5000, please describe your experience and knowledge of computers. Are you a novice, or do you have significant experience using a wide range of technologies? What kind of computer are you using for this class?

Readings
- Wood, Roy. It's World Backup Day — What Are You Doing About It?, *Wired* (31 March 2011). http://archive.wired.com/geekdad/2011/03/its-world-backup-day-what-are-you-doing-about-it/

26 Aug ~ 2. Characteristics of Digital Information

 Digital materials have different characteristics and affordances than other media. The prevalence of these materials demands that archivists understand the nature of the medium and see parallels between characteristics of traditional and digital record formats.

- Discuss similarities and differences between analog and digital characteristics in terms of paleography; carrier, content, and signal.
- List and define techniques for digital encoding, including binary and hex, ASCII and Unicode and how they are layered.
- List common file formats use to store digital information, including flat text; delimited (e.g. CSV); fixed-field length; raster v. vector graphics; XML; containers; & compression.
- Name common proprietary formats and associated software.

Readings
- Evans, David, "Computing," in *Introduction to Computing*, p. 1-18.

- Evans, David, "Language," in *Introduction to Computing*, p. 19-33 (optional).
- Erway, Ricky. "Defining Born Digital" (OCLC, 2012).
  http://www.oclc.org/research/activities/hiddencollections/borndigital.pdf
- Sellen, Abigail J., and Richard H. R. Harper, "What's Wrong with Paper?" in
  The Myth of the Paperless Office (MIT Press, 2002), p. 23-49.
- Matt Ottweill, "How do audio analogue to digital converters work?" (Planet
  of Tunes, n.d.), http://www.planetoftunes.com/digital-audio/how-do-
  analogue-to-digital-converters-work.html.
- Ciaran Trace, "Beyond the Magic to the Mechanism: Computers, Materiality,
  and What It Means for Records to Be 'Born Digital'," *Archivaria* 72 (Fall 2011),
  p. 5-27.
  http://journals.sfu.ca/archivar/index.php/archivaria/article/viewArticle/13358
- Vaughan-Nichols, Steven. "The History of the Floppy Disk." Input Output.
  2012-08-27. (Optional.)
  http://web.archive.org/web/20131101192108/http://h30565.www3.hp.com/t5/F
  eature-Articles/The-History-of-the-Floppy-Disk/ba-p/6434

Discussion
- Explore the notion of affordances in different technologies. For example,
  what are the different affordances of traditional landline phones (consider
  rotary phones) and mobile phones? Of vinyl LP records, CDs, and MP3s?
  Consider other technologies that have analog and digital equivalents. In
  particular, consider aspects of human interaction with these technologies.
     Responses to posts may suggest other affordances of the different
  technologies. Or, challenge whether an affordance is a benefit.

2 Sep ~ 3. Digital Information Systems and Hardware
   Overview of physical components of digital information systems and how
they are integrated as a system.

Learning outcomes
- List and describe the major components, including CPU, ROM, RAM, I/O,
  storage, bus size.
- Describe the differences between different hardware platforms based on their
  size (super to micro) and function within a system (server, workstation,
  peripheral).
- Describe virtualization, including host and guest operating systems,
- List common storage media.
- Discuss the importance of backups and perform a basic backup.

Readings
- Yates, JoAnne. "Communication Technology and the Growth of Internal
  Communication," in Control through Communications (Johns Hopkins Press,
  1989), p. 21-64.
- "First Steps," in *Oracle VM VirtualBox User Manual* (Oracle, 2004-2012), p. 10-
  33. http://download.virtualbox.org/virtualbox/UserManual.pdf
- Kirschenbaum, Matthew G., Richard Ovenden, and Gabriela Redwine.
  "Introduction" in *Digital Forensics and Born-Digital Content in Cultural Heritage*

Collections (CLIR, 2010), p. 1-14.
http://www.clir.org/pubs/reports/pub149/pub149.pdf (optional).

Discussion

Use the discussion boards for the next few weeks to share your experiences configuring the virtual hardware and installing the operating system. Where did you get stuck? What did you try, and what was the solution? You may also use the boards to seek advice from other students.

Also, describe any insights about how computers work "under the hood;" is anything different from how you expected things to be? Consider the impact that technology had on 19th century communications and recordkeeping: can you see any parallels in evolving communication technologies, such as email, Twitter, Facebook, blogs?

9 Sep ~ 4. Operating Systems & System Interfaces

Introduce the components and functions of an operating system. Discuss common operating systems and interfaces (shells).

Learning outcomes
- To define an operating system and describe its function.
- To name major operating systems (Windows, Mac, Unix/Linux).
. To describe the differences between command line interfaces and graphical user interfaces and to discuss the affordances of each.
- To acquire basic familiarity of command line interfaces

Readings
- Thomas, Keir. Chapter 5, Hands on at the Command-Line. *Ubuntu Pocket Guide and Reference*, p. 49-96.
  http://www.ubuntupocketguide.com/index_main.html *Note: you may skip any portion not relevant to your operating system. You will also want to be familiar with the table of contents so you'll know when to refer back to this book.*
- Stephenson, Neal. In the Beginning was the Command Line (1999). Online at http://www.cryptonomicon.com/beginning.html, checked 1 Aug 2012.
- Birkel, Garrett. "The Command Line in 2004."
  http://garote.bdmonkeys.net/commandline/index.html. (Optional)
- "Configuring Virtual Machines," in *Oracle VM VirtualBox User Manual* (Oracle, 2004-2012), p. 42-55. (Parts of this may not make sense; read for concepts.) http://dlc.sun.com.edgesuite.net/virtualbox/4.1.18/UserManual.pdf

16 Sep ~ 5. System Administration

Common system tasks (and commands). Creating batch commands, capturing output from commands.

Learning Outcomes
- To discuss the role of system administrator
- To describe basic OS-level operations, such as login, exit, shutdown, restart.
- To identify the common important OS directories.
- To describe the nature and use of system configuration and log files.

- To define patches and know how to apply them.
- To install software using Windows Control Panel and Synaptic package manager (or aptitude).

Readings
- Sobell, Mark G. "The Filesystem" in *A Practical Guide to Linux Commands, Editors, and Shell Programming*, 2nd ed. (Prentice-Hall, 2010), p. 77-115.
- Thomas, Keir. Chapter 4: Users and the Filesystem. *Ubuntu Pocket Guide and Reference*, p. 49-65.
- Garfinkel, Simson, Daniel Weise, and Steven Strassman. The Unix-Haters Handbook (IDG Books, 1994). http://web.archive.org/web/20121228012221/http://m.simson.net/ugh.pdf (Optional, but see Appendix B, "Creators Admit C, Unix Were Hoax," p. 307-308.)

23 Sep ~ 6. Software

Users typically interact with specific applications to perform practical tasks. Software may be design for an enterprise or individual. Basic software by functionality: word processing, communication (email, IM), web, spreadsheets, databases, workflow and business process, enterprise.

Learning outcomes
- To name functional categories of software: word processing, spreadsheet, database, enterprise, communication, graphic processing; application, productivity, groupware.
- To distinguish the affordances of different text processors, including WYSIWYG (Word, Open Office), plain text (NotePad++, Textpad)
- To define open source software and name and describe common licensing schemes (e.g., GNU Public License).

Readings
- Software. *Wikipedia*. https://en.wikipedia.org/wiki/Software
- Owens, Trevor. "When Is Open Source Software the Right Choice for Cultural Heritage Organizations? An Interview with Peter Murray." *The Signal: Digital Preservation*, January 18, 2013. http://blogs.loc.gov/digitalpreservation/2013/01/when-is-open-source-software-the-right-choice-for-cultural-heritage-organizations-an-interview-with-peter-murray/.
- Team Digital Preservation and the Deadly Cryptic Conundrum. http://youtu.be/Yun9hkPPF9M

30 Sept ~ 7. Midterm exam

A combination of short and long answer questions. The exam (in Word format) will be available for download from GAView at 6:30 p.m. It must be submitted in Georgia View by 9:00 p.m. Contact the instructor immediately if you have technical problems.

7 Oct ~ 8. Programming

The act of creating software to be run on computers is called programming.   Assignment 4. Shell

There are a few basic concepts that govern most programming languages.

Scripting.
Due: 17 Oct.

Learning Outcomes
- Describe how source-code is interpreted into byte-code
- To describe basic programming concepts:
    o Variables
    o Conditionals
    o Loops

Readings
- Holowczak, Richard. "Programming Concepts: A Brief Tutorial for New Programmers." http://holowczak.com/programming-concepts-tutorial-programmers/
    Note: The Introduction through Loops is required; the rest is optional although "Hashing Functions" is recommended for future reference.
    Also, don't be intimidated by tables of language examples; it simply serves to illustrate that programming languages use different syntaxes to do exactly the same thing.
- Nemeth, Evi, Garth Snyder, Trent R. Hein, and Ben Whaley. "Chapter 2: Scripting and the Shell." in *UNIX and Linux System Administration Handbook*, 4th ed. (Prentice-Hall, 2011), pp. 29-48, 73-75.
    Note: The whole chapter will be scanned but only the pages listed are required; however, the rest is recommended.
- Evans, David, "Programming," in *Introduction to Computing*, p. 35-52. (Optional)

9 Oct ~ Midterm

Last day to withdraw from any course and receive a passing grade. Withdrawal after this date will receive an automatic 'F' unless the student is granted an exemption for hardship.

14 Oct ~ 9. Storage

Digital information may be stored in a number of places, including removable media, local storage, central servers, and the cloud. Each has specific affordances and limitations. Digital archivists must consider the appropriate technology to support the specific needs of the repository and the data.

Learning outcomes
- To list and describe media commonly found and use in archives, including signal encoding (physical, magnetic, optical); media (paper, plastics, metal, laminates); connectivity (removable, local, system, cloud).
- To define backup and describe backup strategies; to define online, nearline, and offline storage.
- To distinguish 'deleting' and 'erasing' files completely, define slack space.
- To discuss mechanisms to transfer files using sftp, ssh.
- To discuss LOCKSS and SRB as a means to store content.

Readings
- Huth, Geof, and Ann Marie Przybyla. "Conducting an Inventory of Electronic Records." *New Skills for a Digital Era*, p. 47-51.

- David S. H. Rosenthal, et al. "The Economics of Long-Term Digital Storage."
  http://www.lockss.org/locksswp/wp-content/uploads/2012/09/unesco2012.pdf
- David S. H. Rosenthal and Daniel L. Vargas, "LOCKSS Boxes in the Cloud."
  http://www.lockss.org/locksswp/wp-content/uploads/2012/09/LC-final-
  2012.pdf

21 Oct ~ 10. File Systems

The organization of files in storage.

Learning outcomes
- To define a file system.
- To describe path.
- To navigate the file system using absolute and relative paths.
- To navigate and manipulate files systems using a CLI or GUI, the Unix and
  Windows commands pwd, ls (dir), mv (rename), cp (copy), cd, mkdir, rmdir.
- To describe the rules for naming files, including practical file naming
  conventions. To define filename and filetype (extension).

Readings

- Steven Sinofsky, "Disk Defragmentation – Background and Engineering the
  Windows 7 Improvements." *Engineering Windows 7* [blog], 25 January 2009.
  http://blogs.msdn.com/b/e7/archive/2009/01/25/disk-defragmentation-
  background-and-engineering-the-windows-7-improvements.aspx
- Andrew Bednarz. "Disk Fragmentation & OS X: When Does It Become a
  Problem?" *GigaOM*, 14 April 2009. http://gigaom.com/2009/04/14/disk-
  fragmentation-os-x-when-does-it-become-a-problem/

28 Oct ~ 11. Networks

Computers are frequently connected to share data and processing using the
public Internet or private intranets. Discussion of means of connecting
computers with emphasis on Ethernet and related protocols for exchanging
data securely.

Learning outcomes
- To describe common network activities: storage, communication, remote
  desktop/ssh, file transfer
- To define common network protocols: TCP/IP, http, ftp, smtp
- To describe the basic purpose components of TCP/IP and DNS, including
  packet switching for robust, post nuclear network
- To define: IP address, DHCP, static IP, MAC address
- To describe the purpose of key addresses and ports, including localhost,
  127.0.0.1, 192.168.x.x; ports 80, 8080, 21, 22
- To describe basic networking in Windows clients (ipconfig, Network and
  Sharing Center) and Ubuntu (ifconfig)
- To distinguish the functionality of bridged and NAT configuration in
  VirtualBox

- Security, including firewall ports

Readings
- Denniston, Ken. Introduction to Networking, *Building a Simple Network* (Intel Press, 2002), p. 1-14.
- "Internetworking Basics" (Cisco, 2009). http://docwiki.cisco.com/wiki/Internetworking_Basics
- Callon, Roy. RFC 1925 The Twelve Networking Truths (Network Working Group, 1996). http://www.faqs.org/rfcs/rfc1925.html
- Waitzman, David. "RFC 2549 - IP over Avian Carriers with Quality of Service" (Internet Society, Network Working Group, 1999). http://www.faqs.org/rfcs/rfc2549.html

4 Nov ~ 12. Preservation and Security

Ensuring digital information and records can be access for many years poses a number of interesting challenges. Protecting systems from accidental and malicious hazards, including backup, secure passwords, SQL injection, https, ssh. Tools can help identify common formats to help us prepare for preservation activities as those formats become obsolete.

Learning Outcomes
- To describe the risks and hazards of digital preservation, including software and hardware obsolescence, media and signal degradation.
- To discuss the component of a digital preservation system.
- To list key digital preservation standards, including OAIS, TRAC, and PREMIS.
- To describe basic techniques for system security, including passwords, physical and network security.
- To describe best practices for system backup.
- To describe approaches for system recovery.
- To describe hacker attacks, defenses, and methods to research potential attacks (CERT).
- To define hash values as digital thumbprints for security

Readings
- Erway, Ricky. "You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media." (OCLC Research, 2012). http://www.oclc.org/research/publications/library/2012/2012-06r.html
- Honan, Mat. "How Apple and Amazon Security Flaws Led to My Epic Hacking." *Wired* (6 August 2012). http://www.wired.com/ 2012/08/apple-amazon-mat-honan-hacking/all/.
- Mat Honan, "Kill the Password: Why a String of Characters Can't Protect Us Anymore," *Wired* (15 Nov 2012). http://www.wired.com/ 2012/11/ff-mat-honan-password-hacker/all/
- Stoll, Clifford. "Stalking the Wily Hacker," *Communication of the ACM* 31:5 (May 1988), p. 484-500. http://pdf.textfiles.com/academics/wilyhacker.pdf

- Team Digital Preservation and the Aeroplane Disaster (Digital Preservation
  Europe, 2009). 3' 37". http://youtu.be/EKnsZZzuUr4 (Optional)
- Kirschenbaum, Matthew G., Richard Ovenden, and Gabriela Redwine.
"Challenges" in *Digital Forensics and Born-Digital Content in Cultural Heritage
Collections* (CLIR, 2010), p. 14-48.
http://www.clir.org/pubs/reports/reports/pub149/pub149.pdf (Optional)

11 Nov ~ 13. Relational Data Modeling

Introduction to concepts. Common database software (Access, SQL   .
Server, Oracle, MySQL, PostgreSQL). Spreadsheets are not databases. Data
modeling

Learning outcomes
- To distinguish structured and unstructured data.
- To describe data types (integer, string, date, currency).
- To define relational databases, tables, columns, and rows
- To identify data elements and determine the appropriate data type.
- To read a simple UML Class diagram
- To describe normalization.
- To relate tables using primary and foreign keys.

Readings
- Kroenke and Auer. "Introduction." *Database Concepts*, 11th ed. (Prentice-
  Hall, 2010), p. 2-25. *In GAView.*
- Mariano Casanova. "BASE TUTORIAL: From Newbie to Advocate in a one,
  two... three!" (2010). Chapters 1-4 (p. 3-55).
  https://wiki.documentfoundation.org/images/0/02/Base_tutorial.pdf

18 Nov ~ 14. Using Relational Databases

Implementing the data model in SQL. Demonstration of MS Access and
LibreOffice as front ends to MySQL databases.

Learning outcomes
- Install MySQL, MySQL Workbench.
- To use management tools to create a simple database schema.
- Insert information into the database using the update query and from a file
(insert data infile).
- To write simple SQL queries that retrieve select data element (*select* [*columns*]),
with limits (where), in order (ordered by) using a simple join.

Readings
- "SQL Tutorial" (w3schools, checked 2 Aug 2012).
    http://www.w3schools.com/sql/. *Read all of SQL Basic, including Intro,
    Syntax, Select, Distinct, Where, And & Or, Order By, Insert, Update, and Delete.*
- Mariano Casanova. "BASE TUTORIAL: From Newbie to Advocate in a one,
  two... three!" (2010). Chapters 7 & 9 (p. 71-84 and 111-131). *Note: Chapters 8 &
  10 are optional but recommended.*
  https://wiki.documentfoundation.org/images/0/02/Base_tutorial.pdf

25 Nov ~ Thanksgiving

Holiday, no class

2 Dec ~ 15. Additional Data Modeling Techniques

An introduction to other means of modeling & querying data.

Learning outcomes
- To describe different types of databases (e.g. Graph/Triple-stores, object dbs,
& document dbs)

Readings
- NoSQL. *Wikipedia.* http://en.wikipedia.org/wiki/NoSQL
- Antoniou, Grigoris and Frank van Harmelen. Chapters 2 (Structured Web
   Documents in XML) & 3 (Describing Web Resources in RDF). *A Semantic Web
   Primer*, 2nd edition. MIT Press (Cambridge, MA). 2008. pp 23-105.
- Myer, Tom. "A Really, Really, Really Good Introduction to XML." *Sitepoint.* 24
Aug 2005. http://www.sitepoint.com/really-good-introduction-xml/ (Optional.)

9 Dec ~ Final Exam

No late work accepted after this date.