

LIS654

Building Digital Libraries

2012-08-22

See the course web site at <http://openlib.org/home/krichel/courses/lis654b12a> for the latest online version of this file.

Course description

This course introduces students to the “still evolving” concepts of digital libraries. The core subject matter is repository building. Each student builds a public-access repository on the web. This will store born-digital or digitized materials. The course covers related technical material and organizational issues. Potential topics include digitization, the representation of text, file formats for media such as images and sound, identification strategies and technologies, metadata, markup languages, databases, contents delivery technologies, and repository implementation software.

The course is mostly non-technical. Students will be introduced to technological concepts. They will learn what they are. Most of the interaction with the systems will be through web interfaces. For a small amount of work, usage of the ssh protocol is required. This will be covered separately.

Course objectives

After taking this course students:

- will understand the conceptualizations of digital libraries;
- will have built their own image-based digital library;
- will have been introduced to management aspects of digital libraries;
- will have seen content identification strategies as a crucial aspect of this;
- will have been introduced to the XML format and simple metadata schema in plain (as opposed to RDF) XML;
- will have been introduced to the benefits and problems of metadata interoperability;
- will have a basic conceptual introduction to the other repository software;
- will have seen some cross-repository user services

The Palmer School objective covered by this course is 2.E. “Students will ... build information systems and/or records used in such systems”. Students in this course will achieve a developing level with respect to this objective.

Prerequisites

Students should also be able to use a web browser. Everything that goes beyond this is explained in class or by personal tuition from the instructor.

A limited amount of work is conducted on computers. If students want to use lab computers, they are required to have basic skills on computers using the Microsoft operating system, because this is the type of machines used in the lab. There include e.g. click on an icon to run a program, cut and paste between applications, and copy files from one location to another.

Students are provided with free http/ssh accessible space where they can design their own digital library. This space continues to be available after the course ends. In order to operate that web space students only need a computer with an Internet connection and http/ssh client software.

Instructor

Thomas Krichel

Palmer School of Library and Information Science
C.W. Post Campus of Long Island University
720 Northern Boulevard
Brookville, NY 11548-1300
krichel@openlib.org
work phone: +1-(516)299-2843
skype: thomaskrichel

Class structure

Classes are held on the Brentwood Library on Fridays between 10:00 and 15:00. Classes are hopefully not running in the computer lab. Students should bring their laptops. Between classes, support via the class mailing list, or on a person-to-person basis, via Skype.

Each class has a presentation by the instructor. Students give two presentations. Students are expected to do most of the the work on their digital libraries at home.

Class details

Here are some class details:

| | | |
|------------|----------------|---|
| 2012-09-07 | 10:00 to 15:00 | introduction to digital libraries, the ssh protocol |
| 2012-09-15 | 10:00 to 15:00 | early digital library history, Omeka installation |
| 2012-09-21 | 10:00 to 15:00 | Omeka internals |
| 2012-09-28 | 10:00 to 15:00 | representation of images, and project presentations |
| 2012-10-05 | 10:00 to 15:00 | representation of text, plugin presentations II |
| 2012-10-12 | 10:00 to 15:00 | repository interoperability |

Slides for all classes can be downloaded from the course web site. The slides on the course website are drafts until the commencement of the class.

Reading

If students want to consult a textbook, there are two, but none cover the topic at sufficient technical depth. A simple, dated, but well-written book is Arms (2005). The contents is simple but not completely trivial.

Reese Jr and Barnerjee (2008) has an overview over the class topic. Unfortunately the contents is mostly trivial and the book is expensive. Students are advised to contact the instructor about access options before buying the book.

Web sites used

Omeka is a project of the Roy Rosenzweig Center for History and New Media at George Mason University. There is an official list of sites using omeka. Students should also consult the previous edition of this course, as available on the edition listing page to get an idea of minimal standards. The Omeka Showcase may also provide inspiration for wider usage.

The National Information Standards Organization has a Framework of Guidance for Building Good Digital Collections. Cornell University bring us the *Digital Imaging Tutorial*¹.

For images, there are useful web sites Aware Systems provide Information about TIFF. Greg Roelofs provides information about PNG.

Background readings

Digital libraries are built by building them, not by reading about them. Nevertheless, we have to note some background readings in the syllabus.

On early history, Bush (1945) is a standard reference but it is overrated. Licklider (1965) is much better work in the same direction. You may also enjoy Licklider (1960), Licklider (1968), and Hauben (2005). On more modern approaches to such early systems, see Schatz (1997). On repository history, you may consult historic papers of Van de Sompel, Krichel, Nelson, Hochstenbach, Lyapunov, Maly, Zubair, Kholief, Liu, and O'Connell (2000) and Davis and Lagoze (2000).

Harter (1996) is an early paper on digital libraries. Borgman (1999) has more about different digital library concepts. Lynch (2005) demonstrates good thinking about directions in digital libraries that are worth discussing if they have become reality. On XML, see the lecture notes of Thomas Krichel lis650 course, as well as the wonderfully short piece of Taylor (2004a). For more theoretical perspective, musings of Janisch (2002) and Pitti (2002) are available.

¹<http://www.library.cornell.edu/preservation/tutorial/>

For metadata, Taylor (2004b) gets it right again. Foulonneau and Riley (2008) provides a general introduction. On METS we have a good background paper by McDonough (2006). Taylor (2004a), Taylor (2004b), On the topic of Omeka, we mainly use its web site, but some favorable review by Kucsma, Reiss, and Sidman (2010) may provide inspiration.

When we turn to digital repositories, we will mainly use Lynch (2003) Brown, Griffiths, and Rascoff (2007) for background and Inc. (2011) for how to get things done.

For the process of digitization and its requirements, see ALCTS Preservation and Reformatting Section (2012).

To see how complicated persistent and digital preservation can be, a peek at Consulative Committee for Space Data Systems (2002)

On identifiers, see Kunze (2008) Lynch (2001) Berners-Lee (1998) and Tonkin (2008)

Last, but by no means least, there is a bunch of home-grown resources <http://openlib.org/home/krichel/courses/lis654>.

Mailing list

There is a mailing list for the course at <https://lists-1.liu.edu/mailman/listinfo/cwp-lis654-krichel>. All students are encouraged to subscribe. As a rule, answers to email sent to the instructor are copied to the list. There are exceptions to this rule:

- if the question writer requests the answer not to be posted;
- if the question is a purely private matter.

Assessment

Students have a quiz every class bar the first one. The worst quiz performances will be discounted. The average of all quizzes accounts for 5/13 all of the final grade. Students will give two presentation of 10 to 20 minutes. Student upload their presentation materials to their web site (as opposed to their omeka sites). The first presentation is the `project presentation`² on 2012-09-28. The presentation will count for 1/13 of the final grade. The second presentation is the `plugin presentation`³ on 2012-10-05. Each student introduces an Omeka plugin of the student's choice, but different for each student. For this a roster is available, where the plugin per student is recorded on first-come-first-served basis.

For the remaining 6/13 of the final grade, students prepare a small image-based digital library, of 20 to 100 images. The material must be of some cultural value. A purely personal collections is not admissible, even though it was in earlier editions of the course. The site has to be finished by 2012-10-19.

References

- ALCTS Preservation and Reformatting Section (2012, June). Minimum Digitization Capture Recommendations. available at http://openlib.org/home/krichel/courses/lis618/readings/salton88_term.pdf.
- Arms, William Y. (2005). Digital Libraries. available at <http://www.cs.cornell.edu/wya/diglib/MS1999>.
- Berners-Lee, Tim (1998). Cool URIs don't change. available at <http://www.w3.org/Provider/Style/URI>.
- Borgman, Christine L. (1999). What are digital libraries? Competing visions. *Information Processing and Management* 38(3), 227-243. available at http://local.openlib.org/home/krichel/courses/read/borgman99_what.pdf.
- Brown, Laura, Rebecca Griffiths, and Matthew Rascoff (2007). University Publishing In A Digital Age. available <http://www.ithaka.org/ithaka-s-r/strategyold/Ithaka%20University%20Publishing%20Report.pdf>.
- Bush, Vannevar (1945). As We May Think. *The Atlantic Monthly* 176(1). available at <http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/3881/>.
- Consulative Committee for Space Data Systems (2002). Reference Model for an Open Archival Information System (OAIS). available at <http://ddp.nist.gov/refs/oais.pdf>.
- Davis, James R and Carl Lagoze (2000). NCSTRL: design and deployment of a globally distributed digital library. *Journal of the American Society for Information Science* 51(3), 273-280. available at http://openlib.org/home/krichel/courses/lis654/readings/davis00_ncstr.pdf.

²./lis654/admin/project_presentation.html

³./lis654/admin/plugin_presentation.html

- Foulonneau, Muriel and Jenn Riley (2008). *Metadata for digital resources*. Chandos. first part available at http://local.openlib.org/home/krichel/courses/read/foulonneau08_metad_1.pdf, second part available at http://local.openlib.org/home/krichel/courses/read/foulonneau08_metad_2.pdf.
- Harter, Stephen P. (1996). What is a Digital Library? Definitions, Content, and Issues. available at <http://php.indiana.edu/~harter/korea-paper.htm>.
- Hauben, Jay R. (2005). Vannevar Bush and JRC Licklider: Libraries of the Future 1945–1965. available at <http://www.columbia.edu/~hauben/lof-final05.doc>.
- Inc., DuraSpace (2011). DSpace 1.7.1 Documentation. available at http://www.dspace.org/1_7_1Documentation/.
- Janisch, Troy (2002). What the Hell is XML? available <http://www.alistapart.com/articles/hellxml/>.
- Kucsma, Jason, Kevin Reiss, and Angela Sidman (2010). Using Omeka to Build Digital Collections: The METRO Case Study. *D-Lib Magazine* 16(3/4).
- Kunze, John A. (2008). Towards Electronic Preservation Using ARK Identifiers. <https://confluence.ucop.edu/display/Curation/ARK>.
- Licklider, Joseph Carl Robnett (1960). Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics HFE-1*, 4–11. available at http://openlib.org/home/krichel/courses/lis654/readings/licklider60_man_comput_symbios.pdf.
- Licklider, Joseph Carl Robnett (1965). *Libraries of the Future*. M.I.T. Press. available at http://local.openlib.org/home/krichel/courses/lis654/readings/licklider65_librar_futur.pdf.
- Licklider, Joseph Carl Robnett (1968). The Computer as a Communication Device. *Science and Technology*. available at http://openlib.org/home/krichel/courses/lis654/readings/licklider68_comput_commun_devic.pdf.
- Lynch, Clifford (2001). Identifiers and Their Role in Networked Information Applications. available at <http://www.arl.org/bm~doc/identifier.pdf>.
- Lynch, Clifford (2003). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. http://scholarship.utm.edu/21/1/Lynch,_IRs.pdf.
- Lynch, Clifford (2005, July). Where Do We Go From Here?: The Next Decade for Digital Libraries. *Digital Libraries Magazine*. available at <http://www.dlib.org/dlib/july05/lynch/07lynch.html>.
- McDonough, Jerome (2006). METS: standardized encoding for digital library objects. available at <http://hdl.handle.net/2142/177>.
- Pitti, Daniel V. (2002). Introduction to XML. available at <http://www.iath.virginia.edu/~dvp4c/xmlintro.html>.
- Reese Jr, Terry and Kyle Barnerjee (2008). *Building Digital Libraries*. How-to-do-it manual no 153. Neal-Schuman. first part available at http://local.openlib.org/home/krichel/courses/read/reese08_build_digit_librar0-4.pdf, second part available at http://local.openlib.org/home/krichel/courses/read/reese08_build_digit_librar5-10.pdf.
- Schatz, Bruce R. (1997). Information Retrieval in Digital Libraries: Bringing Search to the Net. *Science* 275, 327–333. available at http://openlib.org/home/krichel/courses/lis654/readings/schatz97_infor_retriev_digit_librar.pdf.
- Taylor, Mike (2004a). All You Need to Know about XML in One Page. available at <http://www.miketaylor.org.uk/tech/xml.html>.
- Taylor, Mike (2004b). One Man’s Ceiling is Another Man’s Floor—or—Why your data may not be as meta as you think it is. available at <http://www.miketaylor.org.uk/tech/metadata.html>.
- Tonkin, Emma (2008). Persistent Identifiers: Considering the Options. *Ariadne*.
- Van de Sompel, Herbert, Thomas Krichel, Michael L. Nelson, Patrick Hochstenbach, Victor M. Lyapunov, Kurt Maly, Mohammad Zubair, Mohamed Kholief, Xiaoming Liu, and Heath O’Connell (2000). The UPS Prototype An Experimental End-User Service across E-Print Archives. *D-Lib Magazine* 6(2). available at <http://www.dlib.org/dlib/february00/vandesompel-ups/02vandesompel-ups.html>.