

LIS654

Building Digital Libraries

2011–11–11

See the course web site at <http://openlib.org/home/krichel/courses/lis654n11a> for the latest online version of this file.

Course description

This course is an introduction to digital libraries. We look at commonly used concepts of digital libraries. We emphasize the concept a repository as a core component of a digital library. We finally look at digital library services through the use of examples.

The course is almost completely non-technical. Students will be introduced to technological concepts. They will learn what they are. They will not be expected to use them directly. Instead, all interaction with the information system will be intermediated through the use of web interfaces. If a non-web interaction is required, the instructor will preform it for the students.

Course objectives

The course contents will allow students to develop collection and local service strategies based on the use of off-the-shelf technologies. It will also enable them to more competently interact with staff maintaining such technologies.

After taking this course students:

- will understand the conceptualizations of digital libraries;
- will have seen the broad types of digital libraries;
- will have built their own image-based digital library;
- will have had an introduction to copyright and licensing;
- 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

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 Microsoft machines. 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. But students may bring they own laptops to class and use the wireless network.

Students are provided with free web space where they can design their own digital library. This web 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.

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 in PC1 at Bobst Library on Tuesdays between 18:20 and 20:10. Between classes, support via the class mailing list will be available.

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

Class details

Here are some class details:

2011-09-13	18:30 to 20:20	introduction: from physical to digital libraries	
2011-09-20	18:30 to 20:20	origins of digital libraries and origins of repositories	
2011-09-27	18:30 to 20:20	repository planning	
2011-10-04	18:30 to 20:30	getting to work with Omeka: installation and system overview	
2011-10-11	18:30 to 20:30	omeka themes and example plugin	
2011-10-18	18:30 to 20:30	omeka metadata and tables	
2011-10-25	18:30 to 20:30	presentation of project plan	
2011-11-01	18:30 to 20:30	digital imaging	
2011-11-08	18:30 to 20:30	presentations on plugins	
2011-11-15	18:30 to 20:30	copyright basics I	
2011-11-22	18:30 to 20:30	copyright basics II	
2011-11-29	18:30 to 20:30	software solutions for repositories and repository interoperability	(lis654n11a-11.pptx)
) 2011-12-06	18:30 to 20:30	cross-repository services)

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, Reese Jr and Barnerjee (2008) is a reasonable, but book. Consult the instructor about access options before buying it.

A simple, dated, but well-written book is Arms (2005).

Web sites used

The National Information Standards Organization has a Framework of Guidance for Building Good Digital Collections. The Library of Congress maintains the Metadata Encoding & Transmission Standard: Official Site and the Metadata Object Description Schema: Official Site.

Cornell University bring us the Digital Imaging Tutorial¹.

Omeka is a project of the Roy Rosenzweig Center for History and New Media at George Mason University. Look at the Omeka Showcase. A simple example of what is doable for a class site is Bridget Harmon's site about the Middle Village Lutheran Cemetery². and Julianna Monjeau's site about The Kosher Meat Boycott of 1902. There is an official list of sites using omeka.

DuraSpace maintain DSpace.

Long Island Memories provide sample simple digital collections we intend to reproduce.

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.

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

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

²<http://brigidharmon.com/lutherancemetery/>

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 copyright, the main reference is Hirtle, Hudson, and Kenyon (2009). We have other useful material in Copyright Office (2010), Bennett (1999) and Templeton (2008).

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.

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.

To see how complicated persistent and digital preservation can be, a peek at Consultative 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, and bar the two sessions where the presentations are due. The first presentation is on the project plan on 2011–10–11. The second introduces a plugin of the students choice, but different for each student, on 2011–11–01. Student upload their presentation materials to their web site (as opposed to their omeka sites). For this a roster is available, where the plugin per student is recorded. The four worst performances will be discounted. The average of all quizzes accounts for 10/30 all of the final grade. The quizzes related to all Palmer School objects relevant for the course. Students prepare a small image-based digital library, of 10 to 20 images. This has to be finished by week 10. It will account for 10/30. Finally, there is the assessment of an institutional repository. Each student will be given one institutional repository URL, and a set of metadata records from that repository. The student assesses the content, and the description of the contents. The paper can be quite descriptive. It should not exceed three printed page or 1000 words. It counts for 8/30. It has to be handed at the second before last class meeting.

Students

Gregory Ferguson and Gregory Ferguson's omeka site

Todd Simpson and Todd Simpson's omeka site

Judith Schwartz and Judith Schwartz's omeka site

Marta Joksimovic and Marta Joksimovic's omeka site

Lauren Dzura and Lauren Dzura's omeka site

Josephine Koncurat and Josephine Koncurat's omeka site

Tom Rose and Tom Rose's omeka site

Kelly Blanchat and Kelly Blanchat's omeka site

Daniel Reisner and Daniel Reisner's omeka site

References

Arms, William Y. (2005). Digital Libraries. available at <http://www.cs.cornell.edu/wya/diglib/MS1999>.

Bennett, Scott (1999). Authors' Rights. *Journal of Electronic Publishing* 5. available at <http://hdl.handle.net/2027/spo.3336451.0005.203>.

- 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>.
- Copyright Office, Library of Congress (2010). Copyright Basics. available at <http://www.copyright.gov/circs/circ1.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.
- 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>.
- Hirtle, Peter B., Emily Hudson, and Andrew T. Kenyon (2009). *Copyright and Cultural Institutions: Guidelines for Digitization for U.S. Libraries, Archives, and Museums*. eCommons@Cornell. available at http://ecommons.library.cornell.edu/bitstream/1813/14142/2/Hirtle-Copyright_final_RGB_lowres-cover1.pdf.
- 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>.
- Templeton, Brad (2008). 10 Big Myths about copyright explained. available at <http://www.templetons.com/brad/copymyths.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>.