LIS 693: Digital Libraries  
Fall 2009

Class Hours: Thursdays 6:30 – 8:50p  
Office Hours: Thursdays 4:00 – 6:00p and by appointment  
Credits: 3  
Prerequisites: LIS 652 and LIS 654, or by permission  
Location: PMC 608

Bulletin Description:
This course will examine the current state of digital libraries in a new context. We will look at: the history and background of digital libraries; particular areas of digital librarianship including digitization, preservation and subscription and other third party resources. Finally we will look at communities of practice that can be served by digital libraries, with emphasis on the shifting world of learning, scholarship and play.

Detailed Description:
This course will cover the theoretical, practical and technical aspects involved in creating, using, and deploying digital libraries. Students will study the evolution of digital libraries, consider the relationship between digital libraries and their socio-technical environment, and collaboratively design a digital library or a new program or service related to digital libraries. Students will be asked to think creatively and critically about the future of digital libraries and where to best direct future development effort.

Course Objectives:
- To become familiar with the history and evolution of digital libraries, particularly with respect to the changing socio-technical environment that digital libraries are situated.  
- To understand the current and emerging tools and methods used to curate, facilitate access to, and preserve digital objects.  
- Apply the latest research and one’s creativity to collaboratively design an innovative digital library or related service.

Course Schedule and Readings

9/3 – Introduction
- Overview of Syllabus and Design Project  
- Ice-breaker activity  
- Assign weekly presenters of class readings  
- Post a photo of yourself and fill-out your profile on Moodle

9/10 - Foundations and History of Digital Libraries


Further Reading:


Essential Questions:

To what extent do you think Bush (1945) predicted the development of digital libraries and the Internet? What needs did Bush think his imagined invention addressed? What needs do we have today that could be addressed by new innovations/inventions? Would you be so bold as Bush to predict the future of information access, and if so, what would it look like?

How does Witten (2002) define a digital library? Witten outlines a set of features (pages 26-28) that are available in a particular Digital Library software package. Given that his book was written nearly eight years ago, are there other features or functions that you feel should be included in a more contemporary digital library? How might have new developments over the near decade have changed how we view digital libraries (e.g., social media such as Facebook for example).

9/17 – Digitization


Further Reading:


Design Project: Form groups

Essential Questions:

What are some of the choices that a digitization project has to make? What affects the answers if you are scanning:

- a famous manuscript (e.g. the Declaration of Independence)
- large collections of manuscripts (e.g. the papers of some Senator)
- printed 18th or 19th century books
- recent printed material
- flat works of art (paintings, posters, ....).
Technical Question: A collection of 96,000 4 X 5-inch transparencies is scanned at 400 dpi, 24-bit color, and then losslessly compressed at a 1.3:1 ratio. Calculate the cost of hard disk storage (at .75 cents/GB) needed for this collection.

9/24 – Metadata and Identifiers


Design Project: Brainstorm ideas for design project

Essential Questions:

Gilliland (2008) argues that “Effectiveness of searching can be significantly enhanced through the existence of rich, consistent, carefully crafted descriptive metadata.” However, it could be argued that searching the World Wide Web became more effective once search engines started ignoring meta-data. Gill (2008) notes that the only meta-data that is used to power Google is the “title” tag; the full-text and links are given full consideration where all other meta-tags are ignored. Gill discusses some of the arguments as to why meta-data has not worked on the web, such as Doctrow’s contention that “people lie”, “people are lazy,” and “people are stupid.” Despite Doctrow’s assertion, meta-data is seemingly resurging through the use of folksonomies. How can you reconcile these divergent opinions and trends? Does meta-data actually add value, and in what cases? Should digital libraries allow user-created meta-data where such efforts have had mixed and arguable successes (e.g., meta-tags versus folksonomies). For a digital library of your own imagining, what role would user-created meta-data play (if at all)? If you desire both rich meta data and no user-contributed meta-data, how do you respond to Gilland, who notes that: “Creation and ongoing maintenance of such [very rich] metadata, however, is complex, time consuming, and resource intensive and may only be justifiable when there is a legal mandate or other risk management incentive or when it is envisaged that the content and metadata may be reused or exploited in previously unanticipated ways, such as in digital asset management systems.”

10/1 – Emerging developments in Information Organization: Folksonomies, Tags and Links


Further Reading:


Essential Questions:

Shirky (2005) argues:

It comes down ultimately to a question of philosophy. Does the world make sense or do we make sense of the world? If you believe the world makes sense, then anyone who tries to make sense of the world differently than you is presenting you with a situation that needs to be reconciled formally, because if you get it wrong, you're getting it wrong about the real world.
If, on the other hand, you believe that we make sense of the world, if we are, from a bunch of different points of view, applying some kind of sense to the world, then you don't privilege one top level of sense-making over the other. What you do instead is you try to find ways that the individual sense-making can roll up to something which is of value in aggregate, but you do it without an ontological goal. You do it without a goal of explicitly getting to or even closely matching some theoretically perfect view of the world. (p. 21)

Do you believe his argument that professional cataloging and user-generated tagging require different philosophical views? Can these views be reconciled, and if so, what would this philosophy be specifically?

10/8 - Preservation of Digital Content


**Essential Questions:**

The Digital Preservation Management tutorial highlights that technology is not the only obstacle to digital preservation but that it is an assemblage that has three components:

Organizational Infrastructure includes the policies, procedures, practices, people—the elements that any programmatic area needs to thrive, but specialized to address digital preservation requirements. It addresses this key development question: What are the requirements and parameters for the organization's digital preservation program?

Technological Infrastructure consists of the requisite equipment, software, hardware, a secure environment, and skills to establish and maintain the digital preservation program. It anticipates and responds wisely to changing technology. It addresses this key development question: How will the organization meet defined digital preservation requirements?

Resources Framework addresses the requisite startup, ongoing, and contingency funding to enable and sustain the digital preservation program. It addresses this key development question: What resources will it take to develop and maintain the organization’s digital preservation program?

Have you had any digital preservation challenges in your past experiences (e.g., unable to access digital content) and what were they? How do you imagine that this kind of challenge might by amplified within a library or institutional context? Would you be interested in being a “Chief Digital Preservation Officer,” and if so, what would your agenda be (for some given institution of your choosing)?

10/15 - Digital and Technical Infrastructure I: Introduction to Storage, Databases, Networks, and Cloud Computing

**Design Project Proposal Due**

**Storage:**

Simply Storage: Platforms - [http://www.youtube.com/watch?v=M-6IBHK4mjM](http://www.youtube.com/watch?v=M-6IBHK4mjM)
Simply Storage: RAID - [http://www.youtube.com/watch?v=a7UXt3MceyI](http://www.youtube.com/watch?v=a7UXt3MceyI)
Simply Storage: Security - [http://www.youtube.com/watch?v=iHMUHGYq05g](http://www.youtube.com/watch?v=iHMUHGYq05g)

**Databases:**
Essential Questions:

This week’s readings and videos discuss the core infrastructure of digital libraries: storage, databases and networks. Most (if not all) digital libraries make extensive use of these components. In addition, this course section discussed cloud computing, which is an emerging method for acquiring these components (e.g., Library of Congress’ cloud computing pilot project). Of these four discussion topics, which do you feel the most comfortable with? For example, have you ever created or used a relational database (MySQL, Microsoft Access, Oracle)? If so, describe the project and how you used a relational database. How comfortable are you with your understanding of how the Internet works (or other networks such as Ethernet)? How comfortable are you with your understanding of data storage and some of the related concepts, such as RAID? Which component(s) would you like to discuss more fully in the next class section?

10/22 – Digital and Technical Infrastructure II: Introduction to Digital Library Applications


Essential Questions:

The DSpace system is described by Smith et al. (2002):

So what is DSpace? It is an attempt to address a problem that MIT faculty have been expressing to the Libraries for the past few years. As faculty and other researchers develop research materials and scholarly publications in increasingly complex digital formats, there is a need to collect, preserve, index and distribute them: a time-consuming and expensive chore for individual faculty and their departments, labs, and centers to manage themselves. The DSpace system provides a way to manage these research materials and publications in a professionally maintained repository to give them greater visibility and accessibility over time.

DSpace has been used by many research universities. However, Davis and Connolly (2007) note that “Cornell’s DSpace is largely underpopulated and underused by its faculty. Many of its collections are empty, and most collections contain few items.”

Other installations of DSpace also suffer from low use, as indicated in Table 1 from the same article. What strategies do you think academic libraries could develop to create a more participatory institutional repository? How would you fix Cornell’s problem, given what you know about it from the piece by Davis and Connolly and what you know about higher education?

10/29 –Subscription and Third-party resources


Further Reading:


Essential Questions:

Last week, we discussed institutional repositories, which are primarily a digital library that is created, supported and maintained by an academic library. However, most digital libraries that academic libraries provide to their patrons are purchased or made available from third-party subscription, which may be for-profit or non-profit organizations (for example, eBrary for electronic books—for-profit or JSTOR for back-runs of scholarly journals—non-profit). And some materials may be made from third-parties without a subscription, such as through Open Access Journals. The authors this week discuss some of the issues within this area, especially as they relate to the way that the Internet has changed the economics of publishing. For example, Varian (1996) suggests that scholarly societies provide added-value options to individual subscribers to prevent loss of subscribers (this is in further reading). Willinsky (2005) suggests
that scholarly work ought to be available openly (without a paid-subscription). Arms (2000) discusses the various digital libraries that have emerged from the publishing industry. Given what you have read, what do you think the future of third-party and subscription services will be, and what might it mean for libraries (for example, if prices of subscription sites continue to climb, or more journals “go the open access path”)?

11/5 - Integration and Interoperability; Computation and Automation in Digital Libraries

Integration and Interoperability


Computing and Automation in Digital Libraries


Further Reading:

OAI for Beginners - the Open Archives Forum online tutorial http://www.oaforum.org/tutorial/

Open Archives Initiative homepage - http://www.openarchives.org/

Essential Questions:

This week’s readings discuss an example of a) Integration and Interoperability in Digital Libraries, as well as an example of b) Computing and Automation in Digital Libraries. The piece on (a) is highly technical and includes many acronyms that may be difficult to follow. It is not important to be able to follow everything that is going on in this article, but rather to know: 1) why would you want to allow harvesting of metadata, 2) why would you want to use OAI-MPH to allow harvesting of metadata? The further reading on this topic may also make this reading more clear.

The article on (b) discusses a strategy for automating quality control in a digital library. How does this system work? Do you think quality control in a digital repository will move to an algorithmic approach (rather than the peer review approach that was described in the article)?

11/12 - Social Media and Web 2.0 in Digital Libraries; User Interface, Usability and Human Factors in Digital Libraries

Social Media and Web 2.0 in Digital Libraries


User Interface, Usability and Human Factors in Digital Libraries


Further Reading:


Essential Questions:

This week we will be discussing user interfaces to digital libraries. A variety of approaches are discussed in the readings with respect to designing a user interface, from a Web 2.0 approach (Cocciolo et. al., 2007) to an information visualization approach (Börner and Chen, 2002). Blandford and Buchanan (2003) discuss usability of interfaces, and Norman (1988) discusses the psychology of usability. Discuss the readings and a user interface that you love OR hate. Why does it provoke such feelings of love or hate? How did your feelings about it change (or not) after using it for a period of time?

11/19 – Socio-technical Challenges to Digital Libraries


Further Reading:


Essential Questions:

This week’s readings discuss social and technical transformations that go “to the very foundations of how liberal markets and liberal democracies have coevolved for almost two centuries” (Benkler, 2007, para. 1). Benkler describes how a “series of changes in the technologies, economic organization, and social practices of production in this environment has created new opportunities for how we make and exchange information, knowledge, and culture” (para. 2). What does Benkler think these transformations mean for individuals and society?

Duranceau (2008) applies Benkler’s ideas to the case of the MIT Libraries. She finds that if one were to use Benker’s analysis, “the success of any services or product will in large part be driven by how that service or product fits into this new economy’s commons-like structure” (p. 245). She discusses how the MIT Libraries has responded to these large-scale changes through technology, staffing, and alternations in how librarians view their roles. She indicates that the move from the “industrial information economy” to the “networked information economy” need not necessarily be a threat to libraries but can be an opportunity to reevaluate and reconceive the role and function of libraries. How has the MIT libraries changed to accommodate this transformation?

11/26 – No Class, Thanksgiving Holiday

12/3 - Evaluating Digital Libraries

Further Reading:


Essential Questions:

This week’s readings discuss the evaluation of digital libraries. Saracevic (2004) suggests a number of ways that individuals can evaluate a digital library, from usability studies, log analysis, surveys, interviews and observations. He notes that there is no best method to evaluate a digital library and includes an extensive bibliography of digital library evaluations from 2004 going back. He suggests that “Users are from Venus and digital libraries are from Mars,” and perhaps the best way to evaluate a digital library is to measure the adversity between users and the digital library. Under this scenario, an evaluation that found a “lessening of the adversity” between the digital library and its users may have reason to rejoice in success (p. 9). What do you think of his “versus hypothesis?”


12/10 – Future of Digital Libraries


Further Reading:


Essential Questions:
In our first week of class, digital libraries were defined as a “focused collection of digital objects, including text, video, and audio, along with methods for access and retrieval, and for selection, organization, and maintenance of the collection” (Witten & Bainbridge, 2002, p. 6). Over the past 15 years, our collective capacity to provide these functions has grown tremendously. With this increased expertise and ease, many who have been involved in digital library projects have begun asking, “what more can be done?” This has broadened the scope of digital libraries, leading several scholars to suggest that the concept of a “digital library” may be subsumed by a more expansive concept, such as “cyberinfrastructure.” Borgman et al. (2008) suggest a movement away from a “providing access to” perspective toward one that emphasizes learning and the active construction of knowledge. What are the recommendations her committee makes to the National Science Foundation, and can you see how this perspective may be an outgrowth of a digital library perspective?

12/17 – Design Project Presentations.

Design Project Documents are due.

Textbooks, Readings, and Materials

No textbook is required for this course. All readings are available online via Moodle.

Course Requirements

Students’ course grades will be determined by performance on the following activities:

1. Class Participation (20%)
2. Weekly Responses (30%) – 11 responses required over the course of the semester
3. Design Project (50%)
   3a. Proposal (2-5 pages) (15%) – due Oct. 15
   3b. Design Document (15-20 pages) (25%) - due Dec. 17, last day of class
   3c. Presentation (10%) – present on Dec. 17, last day of class

Class Participation

Students are expected to be prepared and to contribute to class discussions each week with scholarly analyses and insights. In addition, each week one student or a team of two students will present their understanding of the readings to the class. This is an opportunity to consolidate your (or your team’s) understanding on a topic, to present your perspective, to make novel connections to other domains, and to relate the readings to real-world experience. Presenters may use the essential questions posed (available on Moodle) to guide their presentations, or may choose their own direction in discussing the readings. Presenters should be prepared to make around a 10-15-minute presentation, and conclude with some questions or issues they would like to discuss more thoroughly.

The schedule of presenters will be decided on the first day of class.

Weekly Responses

Each week, students are expected to write at least two paragraphs in response to the essential questions posted on Moodle. Students should respond to the question on Moodle by noon (at the latest) on the day of class (late responses will receive a reduced grade). Please do not bring in a hard-copy or email unless Moodle is unavailable. The purpose of these responses is to allow students the opportunity to reflect on the readings and share their reflections with the other members of the class. Students are encouraged to read the responses by their fellow classmates (this is, however, not a requirement). Based on interests,
students may choose two weeks NOT to do a weekly response. This means by the end of the semester, each student should have posted 11 responses.

Please note that the instructor will refer to these responses during class discussion and may ask students to further clarify or expand on their response.

**Design Project**

**Overview**

The primary assignment for the course is to participate in a team that will draw on the digital libraries research and literature to design a digital library or a new program or service related to digital libraries (e.g., facilitating access to and/or preserving digital objects). Groups should use this opportunity to be innovative and think creatively and critically about digital libraries (What is a digital library? What counts as a digital library? Why do we need this?). Groups should aim to design a project that could be “picked-up” by an outside agency (a library, university, venture-capitalist, foundation, technology company, non-profit, think-thank, etc.) and fully implemented using your group’s design materials as a guide. In order to ensure innovativeness, teams should consider the work being done by outside agencies and ask: does our project have something to offer that these projects don’t? Why is our project innovative?

The class will be divided into design groups with around four members. Time will be provided in class for groups to meet; however, meeting out of class time may be required. Each group will be expected to deliver a project proposal, a project design document, and a presentation on the last day of class. Details on these aspects are below:

**Proposal**

The Proposal should be 2-5 pages and outline the idea for your project. The proposal should be considered a less fully-fleshed-out version of the project design document (see below). The instructor will provide feedback on the proposal which you can use in further refining your project.

**Design Document**

The project design document should be 15-20 pages (this page count can include figures, but not references or other appendices). The design document should address:

a) What is the purpose of your project?

1) Why do we need it?
2) What extent (if at all) does your digital library make use of the following? What extent (if at all) does your project use or re-imagine the following?
   
   - meta-data, identifiers, folksonomies, tags, preservation, storage, databases, networks, applications, subscriptions, third-party resources, integration, interoperability, computation, automation, social media, Web 2.0, user interface, usability, human factors, law, economics, technology

3) What educational or learning goals will motivate this effort, if any?
4) What populations of users (if any) will be served?
5) What type of community (if any) will be fostered by this effort?
6) What role (if any) will librarians play in this project?
7) What will be the size of this effort?
8) What resources will be required?
9) How will the project be assessed?
b) What are the features and functions of the project? Please be specific.

c) Include one or more prototypes of the project. These prototypes can come generated electronically (Adobe Illustrator, Photoshop, Powerpoint, etc.) or by-hand (drawings on paper, etc.). These prototypes should strive to be more than sketches; they should be visibly assembled with care.

i) Screen shots: What would a user see when interacting with this system? What elements make up the user interface (if any)?

ii) Diagrams: Illustrations that convey flows or networks of interaction.

iii) Visualizations: How would you convey the design’s social and/or human interactions? Be creative.

d) Implementation: What do you think would be involved to make this design a reality? Provide estimations.

e) What does the literature and research on digital libraries offer in thinking about this project?

**Presentation**

Each group will get 20 minutes to present, and a 10 minute question and answer period. Each group should:

a) Make it fun and educational! Be creative! We have all been subject to ill-prepared or low-energy presentations- avoid it!

b) Discuss the goals, why your project is needed, and what makes your project innovative.

c) Provide a way of demonstrating your prototypes to the class. These may include electronic illustrations (Powerpoint), an interactive simulation, or large paper/drawing presentations. You may also want to consider handouts for the class.

**Assessment and Evaluation**

1. All assignments must completed in order to receive a passing grade in the course
2. Assignments must be turned in during class in hard copy (except for the weekly responses which should be completed on Moodle). Late assignments will receive a reduced grade
4. Late papers will receive a grade but no comments
5. Pratt policy: Students with extensive absences (three or more for any reason) will be required to drop the course.

**Pratt’s grading scale:**

| Superior work: | A 4.0 (96-100) | A- 3.7 (90-95) |
| Very good work: | B+ 3.3 (87-89) | B 3.0 (83-86) | B- 2.7 (80-82) |
| Marginally satisfactory: | C+ 2.3 (77-79) | C 2.0 |
| Failed: | F 0.0 (0-69) |

**Policies**

All Institute-wide policies are listed in the Bulletin under “Community Standards,” which include policies on attendance, academic integrity, plagiarism, computer, and network use. Students who require special accommodations for disabilities must obtain clearance from the Office of Disability Services at the beginning of the semester. They should contact Mai McDonald, Disability Services Coordinator, in the Office of the Vice President for Student Affairs, Main Building, Lower Level: 718-636-3711.