Z586: DIGITAL CURATION

DR. DEVAN RAY DONALDSON

TUESDAYS, 4:00PM – 6:45PM, SPRING 2016
LI001

Preserving and providing long-term access to digital materials over time is a Grand Challenge. They require constant and ongoing maintenance. This course: 1) provides an overview of research, policy and current practices in curating and preserving digital data, and 2) gives students practical experience working with digital materials, digital curation environments, and digital curation plans.

GOALS AND OBJECTIVES

Upon successful completion of this course, students should be able to:

• Demonstrate knowledge of research, policy and current practice in curating and preserving digital data;
• Reflect on researchers’ and practitioners’ real-world experiences in performing digital curation;
• Demonstrate knowledge of the Digital Curation Centre (DCC) Curation Lifecycle Model and be able to apply it as a framework for understanding digital curation in various contexts; and
• Become familiar with digital curation computing environments and curating digital materials based on hands-on experience.

FACULTY CONTACT INFORMATION

Email: drdonald@indiana.edu
Telephone: (812) 855-9723
Office: LI019
Office hours: By appointment

INDIANA PROMISE

Students are strongly encouraged to uphold the Indiana Promise regarding all aspects of this class. I promise that: I will be ethical in my academic work; I will take personal responsibility for what I say and what I do; I will respect the dignity of others, treating them with civility and understanding.

HONOR CODE

This class, as all classes at Indiana University, requires that students abide by the “Code of Student Rights, Responsibilities and Conduct.” Please familiarize yourself with this document:

http://www.iu.edu/~code/

Students found to be engaging in plagiarism, cheating, or other types of dishonesty may receive a grade of an F on the assignment in question and may be reported to the Dean's Office for appropriate action. In particular, please understand the various nuances of plagiarism and avoid engaging in this type of behavior. If you have any questions, it is your responsibility to meet with me in advance.
**Grades**

Assignments and percentage of final grade:

- Participation: 10%
- Leading Discussion: 10%
- Reflective Essays: 20%
- Lab Assignments: 20%
- Media Digitization and Preservation Initiative (MDPI) Paper: 10%
- Final paper: 25%
- Final presentations: 5%

Grades will be assigned on an A, B, C, D, F scale. The following definitions of letter grades have been defined by student and faculty members of the Committee on Improvement of Instruction and have been approved by the faculty (November 11, 1996) as an aid in evaluation of academic performance and to assist students by giving them an understanding of the grading standards at ILS.

<table>
<thead>
<tr>
<th>Grade</th>
<th>GPA</th>
<th>%</th>
<th>MEANING</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4.0</td>
<td>96-100</td>
<td>Outstanding achievement. Student performance demonstrates full command of the course materials and evinces a high level of originality and/or creativity that far surpasses course expectations.</td>
</tr>
<tr>
<td>A-</td>
<td>3.7</td>
<td>90-95</td>
<td>Excellent achievement. Student performance demonstrates thorough knowledge of the course materials and exceeds course expectations by completing all requirements in a superior manner.</td>
</tr>
<tr>
<td>B+</td>
<td>3.3</td>
<td>87-89</td>
<td>Very good work. Student performance demonstrates above-average comprehension of the course materials and exceeds course expectations on all tasks as defined in the course syllabus.</td>
</tr>
<tr>
<td>B</td>
<td>3.0</td>
<td>84-86</td>
<td>Good work. Student performance meets designated course expectations, demonstrates understanding of the course materials and is at an acceptable level.</td>
</tr>
<tr>
<td>B-</td>
<td>2.7</td>
<td>80-83</td>
<td>Marginal work. Student performance demonstrates incomplete understanding of course materials.</td>
</tr>
<tr>
<td>C+</td>
<td>2.3</td>
<td>77-79</td>
<td>Unsatisfactory work. Student performance demonstrates incomplete and inadequate understanding of course materials.</td>
</tr>
<tr>
<td>C</td>
<td>2.0</td>
<td>74-76</td>
<td></td>
</tr>
<tr>
<td>C-</td>
<td>1.7</td>
<td>71-73</td>
<td>Unacceptable work. Coursework performed at this level or below will not count toward the MLS or MIS degree. For the course to count towards the degree, the student must repeat the course with a passing grade.</td>
</tr>
<tr>
<td>D+</td>
<td>1.3</td>
<td>69-70</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>1.0</td>
<td>67-68</td>
<td></td>
</tr>
<tr>
<td>D-</td>
<td>.7</td>
<td>65-66</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0.0</td>
<td>&lt;65</td>
<td>Failing. Student may continue in program only with permission of the Dean.</td>
</tr>
</tbody>
</table>

**Due Dates**

All assignments must be handed in on their due dates. If you cannot deliver an assignment or a project on the date it is due, it is your responsibility to discuss your situation with the instructor in advance of the due date and request an extension. If the extension is granted, late assignments will automatically lose one letter grade from the grade they would have received had they been turned in on time. For all due dates, see the schedule.
Assignments

**Participation (10%)**

To get the most out of the class, you need to be in attendance and be engaged. A general rubric for assessing your participation in this class is provided below:

<table>
<thead>
<tr>
<th>%</th>
<th>MEANING</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Perfect attendance; displays knowledge of having read and synthesized all of the readings; engages in in-class activities and discussions; provides thoughtful discussion of the readings and provides respectful responses to classmates' comments.</td>
</tr>
<tr>
<td>8</td>
<td>Missed or was late for one class; displays knowledge of the majority of the readings and provides thoughtful commentary in class; engages in in-class activities and discussions; shows respect for peers.</td>
</tr>
<tr>
<td>6</td>
<td>Missed or was late for two or three classes; reads some of the readings and makes at least one comment in class; some engagement in in-class activities and discussions; shows respect for peers.</td>
</tr>
<tr>
<td>4</td>
<td>Missed or was late for four or five classes; does not display a thorough knowledge of the readings and does not add to the discussion; little engagement in in-class activities and discussions; shows respect for peers.</td>
</tr>
<tr>
<td>2</td>
<td>Missed or was late for more than five classes; does not display a knowledge of the readings and does not add to the discussion; no engagement in in-class activities and discussions; shows respect for peers.</td>
</tr>
<tr>
<td>0</td>
<td>Did not attend; does not discuss; shows no respect for peers.</td>
</tr>
</tbody>
</table>

**Leading Discussion (10%)**

Each week, a student will lead discussion on the assigned topic and readings. During the first class, we will divide the number of presentations by the number of students so that each student presents an equal amount of times. In the case that there are more presentations than students, extra credit will be given to students who volunteer to lead additional sessions so that all sessions are covered. As discussion leader, you will do the following:

1) Read the readings for the weeks you’re responsible.
2) Consider the readings in relation to the concepts from the DCC Curation Lifecycle Model that are the focus for the weeks that you’re responsible (e.g., Create and Receive, Appraise and Select, Ingest, Preservation Action, Store, Access, Use and Reuse, or Transform) (See Higgins (2008) for descriptions of what each of the concepts mean in the context of the DCC Curation Lifecycle Model).
3) Develop a 30-minute PowerPoint presentation on the readings and how they relate to the concepts that are the focus of the week. You can discuss other relevant readings, case studies, or current events in addition to the required readings, but not as a substitute for the required readings. Include a slide or slides with 5 to 10 discussion questions in your presentation.
4) Submit your PowerPoint presentation on Canvas before class.
Reflective Essays (20%)

Throughout the course of the semester several researchers and practitioners will share with you their real-world experiences in performing digital curation. The week after each guest speaker presents to the class, you will submit a 2-page reflective essay in which you are to address the following components:

- Provide a brief summary of what the guest speaker(s) discussed.
- What, if anything, did you learn that you did not know before as a result of their presentation?
- Describe the way(s) in which their work relates to digital curation.
- What, if anything, surprised you about their work?
- What, if any, challenges to performing digital curation work did you notice or the guest speakers identify and how, in your opinion, might those challenges be addressed?
- Your general reactions to what the guest speakers discussed.

All reflective essays are due 24 hours before the class period following each guest speaker’s presentation (i.e., Mondays at 4pm).

Lab Assignments (20%)

Increasingly, various solutions for digital curation environments are being created. During these in-class lab assignments, you will learn about specific types of digital curation environments and gain hands-on experience with them. This will require you to read tutorials, watch demonstration videos, read the accompanying wikis, read discussion boards and user forums, and interact with software in order to figure out how to use it. We will learn together about these various different software packages to become familiar with them. During the course of the semester, we will have multiple lab assignments/sessions on:

- DSpace (www.dspace.org)
- BitCurator (www.bitcurator.net)
- iRODS (www.irods.org)

Media Digitization and Preservation Initiative (MDPI) Paper (10%)

Indiana University has a strong record of commitment to audio preservation and has earned recognition as a national leader in research and development of best practices in the field of digital curation. The Media Digitization and Preservation Initiative (MDPI) is a massive project representing Indiana University’s comprehensive work to preserve historical and cultural time-based media for the research, education, and enrichment of future generations. The project involves digitizing time-based media that has been deemed to be of scholarly value by experts. Since MDPI’s ultimate goals are long-term preservation and access, the initiative will also store backup files remotely, and generate and store preservation and production masters and mezzanine files. IU is collaborating with public broadcaster WGBH to develop a software platform known as HydraDAM2 to manage these files for preservation in the Scholarly Data Archive. Other critical components for long-term preservation of digital files, such as unambiguously identifying, managing and migrating those files into the future, are also in development. During the course of the semester, we will hear from representatives of the MDPI, including its executive director (Laurie Antolovic) and director of technical operations (Mike Casey). We will also take
a tour of the Innovation Center, which houses the MDPI facility where the digitization and preservation work takes place. For this assignment, you will write a 10-page paper (double-spaced) in which you:

- Provide background on the MDPI project;
- Summarize the work and structure of the MDPI project (i.e., what work is done in the project and how?) (take notes during Laurie Antolovic’s talk and the tour of the Innovation Center to prepare to write this);
- Provide examples of the type of work done at the Innovation Center for the MDPI project (be sure to take notes during the Innovation Center tour so that you will be prepared to discuss some of these examples in your paper);
- Choose 2 to 5 different topics/concepts related to digital curation or digital preservation that either Laurie Antolovic discusses during her presentation or Mike Casey or any of his colleagues discuss during the Innovation Center tour. After choosing the topics/concepts, address the following in your paper: 1) Define the topics/concepts, 2) Incorporate relevant literature on the topics/concepts (at least 5 different projects, research articles, papers, and/or case studies), 3) Discuss how Mike Casey or his colleagues address the topics/concepts in the context of the MDPI project, and 4) Discuss how the MDPI project compares/contrasts in their handling of the concepts/topics as compared to how the concepts/topics are discussed in the literature.

It is your responsibility to conduct research to find the relevant literature for your paper. Possible sources for finding literature include, but are not limited to, the following sources:

- Archiving Conference (request articles through InterLibrary Loan)
- International Digital Curation Conference (IDCC)
- International Journal of Digital Curation (IJCC)
- Open Repositories (OR)
- International Conference on Digital Preservation (iPRES)
- ACM Digital Library
- IU’s indexed version of Google Scholar (https://libraries.indiana.edu/resources/googlescholar)

Observe the following technical specifications when writing your paper:

- Use the APA style guide to format your paper and references (https://owl.english.purdue.edu/owl/resource/560/01/).
- Provide a separate reference page.
- Give your paper a title. Whatever title you choose, it should reflect clearly and concretely the content of your paper.
- Double-space your paper, use one-inch margins and 12-point font.
- Proof-read and copy-edit your paper. Points will be deducted for gratuitous spelling and grammatical errors.
- Your review should include at least 5 sources; these sources can include books, white papers, research articles, conference papers, review articles, reports, and case studies.
- When complete, your paper should be at least 10 pages in length (excluding reference and title pages).

The paper is due on Tuesday, March 22, 2016 at 4pm.
Final Paper (25%)

Digital repositories play a key role in digital curation. We will be discussing and interacting with various types of digital repositories throughout the semester. For the final assignment, you will write a 20-page paper (double-spaced) on a topic of your choosing related to digital repositories and digital curation. Examples of topics include, but are not limited to:

- Trustworthy Digital Repositories (TDRs)
- Users of Digital Repositories
- Interoperability of Digital Repositories
- Scientific Data Reuse and Data Sharing

After choosing your topic, you need to develop an outline for your paper. This outline will be due during Week 6. Make sure you include a description of your topic at the top of the page of your outline before you submit it on Canvas during Week 6. I will provide feedback to you on your outline to help you strengthen it. During Week 12, you are to submit a rough draft of your final paper. You need to have completed at least 50% (i.e., 10 pages) of your paper so that I can review what you have and provide feedback.

Your paper must include the following:

- An introduction in which you introduce and describe the general area and the specific topic. This is a good time to provide definitions for the main topics you will be discussing (be sure to cite relevant literature). Include references. Your introduction should include a description of why the topic is important and also identify short-comings and/or limitations of the current state of knowledge and research about your topic.
- A literature review section in which you review the literature on your topic. Summarize and synthesize the arguments and ideas that you find as you read the literature on your topic. You should include at least 20 sources in your literature review. These sources can include books, white papers, research articles, conference papers, review articles, reports, and case studies. You should summarize, synthesize, and critique your sources by discussing common themes or issues that you discover as you read the literature. Analyze the information you glean from the sources: areas of overlap or repetition indicate established approaches to the research; conflicting or differing perspectives indicate some manner of disagreement; equal representation among several perspectives may indicate competition or a lack of established research priorities. Use your analysis to group the literature into topical or thematic categories. Then use your categories to develop an outline for the review. DO NOT simply provide summaries of the articles or present an annotated bibliography (if you find that each sentence is beginning with an author name, you are likely summarizing—avoid this). Find a broad way to functionally describe and relate the literature. Use subheadings. This not only aids the reader, but provides a good outline for you when writing the review. Provide Introduction and Conclusion sections for your literature review. Use the introduction to describe the goal of the literature review and an overview of what you’ll present. In the conclusion, summarize what you found in the literature review. What are the major take-home messages, especially with respect to your topic? What gap(s) in the research do you think someone in the digital curation field needs to address? For additional information on how to conduct a literature review, see: http://writingcenter.unc.edu/handouts/literature-reviews/
- A conclusion in which you focus on the most salient findings from your literature review and suggest future directions for research on your topic.
Possible sources for finding literature to include in your paper include, but are not limited to, the following:

- Archiving Conference (request articles through InterLibrary Loan)
- International Digital Curation Conference (IDCC)
- International Journal of Digital Curation (IJCC)
- Open Repositories (OR)
- International Conference on Digital Preservation (iPRES)
- ACM Digital Library
- IU’s indexed version of Google Scholar (https://libraries.indiana.edu/resources/googlescholar)

Observe the following technical specifications when writing your paper:

- Use the APA style guide to format your paper and references (https://owl.english.purdue.edu/owl/resource/560/01/).
- Provide a separate reference page.
- Give your paper a title. Whatever title you choose, it should reflect clearly and concretely the content of your paper.
- Double-space your paper, use one-inch margins and 12-point font.
- Proof-read and copy-edit your paper. Points will be deducted for gratuitous spelling and grammatical errors.
- Your review should include at least 20 sources; these sources can include books, white papers, research articles, conference papers, review articles, reports, and case studies.
- When complete, your paper should be at least 20 pages in length (excluding reference and title pages).

Final papers are due by 11:59PM on Saturday, April 30, 2016.

**Final Presentations (5%)**

During the last class session, you are to present your final paper to the class. Use of PowerPoint is required. You are expected to dress appropriately and approach the presentation in a professional manner. You will be graded on both your presentation style and your ability to present your topic in an approachable and clear manner.
<table>
<thead>
<tr>
<th>WEEK</th>
<th>DATE</th>
<th>DISCUSSION</th>
<th>WORKSHOP</th>
<th>LAB</th>
<th>DUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/12</td>
<td>Overview of Class and Digital Curation</td>
<td>Introductions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1/19</td>
<td>Create or Receive (Part I)</td>
<td>The National Digital Newspaper Program (Deborah Thomas, LOC)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1/26</td>
<td>Create or Receive (Part II)</td>
<td>Getting materials “ready” for digital curation (Juliet Hardesty, IU)</td>
<td>DSpace (Part I)</td>
<td>Reflective Essay #1</td>
</tr>
<tr>
<td>4</td>
<td>2/2</td>
<td>Appraise and Select (Part I)</td>
<td>Web Archiving (Abigail Grotke, LOC)</td>
<td>DSpace (Part II)</td>
<td>Reflective Essay #2</td>
</tr>
<tr>
<td>5</td>
<td>2/9</td>
<td>Appraise and Select (Part II)</td>
<td>Architectural, Engineering and Landscape Records (Kit Arrington, LOC)</td>
<td>DSpace (Part III)</td>
<td>Reflective Essay #3</td>
</tr>
<tr>
<td>6</td>
<td>2/16</td>
<td>Ingest</td>
<td>American Folklife Center Audio and Visual Collections (Julia Kim, LOC)</td>
<td>BitCurator (Part I)</td>
<td>Reflective Essay #4; Outline</td>
</tr>
<tr>
<td>7</td>
<td>2/23</td>
<td>Preservation Action (Part I)</td>
<td>Media Digitization and Preservation Initiative (MDPI) (Laurie Antolovic, IU)</td>
<td>BitCurator (Part II)</td>
<td>Reflective Essay #5</td>
</tr>
<tr>
<td>8</td>
<td>3/1</td>
<td>WE WILL MEET AT THE IU INNOVATION CENTER THIS WEEK (2719 E. 10th. St.); Media Digitization and Preservation Initiative (MDPI) (Mike Casey)</td>
<td></td>
<td>Reflective Essay #6</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3/8</td>
<td>Preservation Action (Part II)</td>
<td>Digital Preservation (Heidi Dowding, IU)</td>
<td>BitCurator (Part III)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3/15</td>
<td>SPRING BREAK</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3/29</td>
<td>Store</td>
<td>Brainstorming/outlining session for final paper</td>
<td>iRODS (Part II)</td>
<td>Reflective Essay #8; Rough Draft</td>
</tr>
<tr>
<td>13</td>
<td>4/5</td>
<td>Access, Use and Reuse (Part I)</td>
<td>SEAD Project (Beth Plale, IU)</td>
<td>iRODS (Part III)</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>4/12</td>
<td>Access, Use and Reuse (Part II)</td>
<td></td>
<td></td>
<td>Reflective Essay #9</td>
</tr>
<tr>
<td>15</td>
<td>4/19</td>
<td>Transform</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>4/26</td>
<td></td>
<td>FINAL PRESENTATIONS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Readings**

All readings are required. While you do not have to read every word of these assignments per se, you do need to read/skim them with an eye toward the topic of discussion for the week in which the readings are listed, and you need to be prepared to discuss these readings in class. Readings are available from the Canvas course site under “Files > Readings” or by copying and pasting the urls for the resources.

**Week 1 – January 12, 2016**

**Topic:** Overview of Class and Digital Curation

**Required reading:**
- Syllabus

**Discussion reading:**

**Week 2 – January 19, 2016**

**Topic:** Create or Receive (Part I) and the National Digital Newspaper Program

**Discussion reading:**

**Workshop reading:**
- National Digital Newspaper Program: A partnership between the Library & the National Endowment for the Humanities (National Digital Newspaper Program) http://www.loc.gov/ndnp/
- About the Program (National Digital Newspaper Program) (Library of Congress) http://www.loc.gov/ndnp/about.html
- Chronicling America (Library of Congress) http://chroniclingamerica.loc.gov
Week 3 – January 26, 2016
Topic: Create or Receive (Part II) and Getting Materials “Ready” for Digital Curation
Discussion reading:


Lab reading:


Week 4 – February 2, 2016
Topic: Appraise and Select (Part I) and Web Archiving at the Library of Congress
Discussion reading:


Workshop reading:

- Web Archiving (The Library of Congress) http://www.loc.gov/webarchiving/
- Collections with Archived Web Sites (The Library of Congress) http://loc.gov/websites/collections/
Week 5 – February 9, 2016
Topic: Appraise and Select (Part II) and Architectural, Engineering and Landscape Records at the Library of Congress
Discussion reading:


Workshop reading:


Week 6 – February 16, 2016
Topic: Ingest and American Folklife Center Audio and Visual Collections at the Library of Congress
Discussion reading:

- Depositing and ingesting Digital Objects. Digital Curation Centre. http://www.dcc.ac.uk/resources/external/tools-services/depositing-and-ingesting-digital-objects [Note: explore all of the resources listed on the website that are designed to help digital curators incorporate new content into a digital repository, from collection-building through data transfer and the accessioning process]

Workshop reading:


Lab reading:


Week 7 – February 23, 2016
Topic: Preservation Action (Part I) and the Media Digitization and Preservation Initiative (MDPI)
Discussion reading:

on Library and Information Resources. 
http://www.clir.org/pubs/reports/reports/rothenberg/pub77.pdf


**Workshop reading:**


• MDPI Timeline. https://mdpi.iu.edu/timeline/index.php

---

**Week 8 – March 1, 2016**

**Topic:** The Media Digitization and Preservation Initiative (MDPI)

**Required reading:**


---

**Week 9 – March 8, 2016**

**Topic:** Preservation Action (Part II) and Digital Preservation

**Discussion reading:**


---

**Week 10 – March 15, 2016**

**SPRING BREAK**

---

**Week 11 – March 22, 2016**

**Topic:** Preservation Action (Part III) and Digital Preservation

**Required reading:**


**Lab reading:**  

---

**Week 12 – March 29, 2016**

**Topic:** Store  
**Discussion reading:**  
  http://public.ccsds.org/publications/archive/650x0b1s.pdf  
  http://datalib.edina.ac.uk/mantra/storageandsecurity/

---

**Week 13 – April 5, 2016**

**Topic:** Access, Use and Reuse (Part I) and SEAD Project  
**Discussion reading:**  

**Workshop reading:**  
• Sustainable Environment Actionable Data (SEAD).  
  http://sead-data.net

---

**Week 14 – April 12, 2016**

**Topic:** Access, Use and Reuse (Part II)  
**Discussion reading:**  
**Week 15 – April 19, 2016**

**Topic:** Transform

**Discussion reading:**
- File formats and transformation. MANTRA. Research Data Management Training. University of Edinburgh
  http://datalib.edina.ac.uk/mantra/fileformatandtransformation/

**Week 16 – April 26, 2016**

**FINAL PRESENTATIONS**