INLS 756: Data Curation and Management

The Instructors.

Dr. Helen R. Tibbo
tibbo at email dot unc dot edu
Alex Poole
poole dot alex at gmail dot com
Thu-Mai Christian
thumai at email dot unc dot edu
Office: 201 Manning Hall
FAX #: (919) 962-8071
(919) 962-8063(w); (919) 929-6248(h)
Class Listserv: INLS756-Spring2014@sakai.unc.edu

Course Description

This class explores the full range of data curation lifecycle activities, from the design of good data through content creator management, metadata creation, ingest into a repository, repository management, access policies and implementation, to data reuse. Speakers from the arts, humanities, social sciences, and sciences will discuss data requirements in their fields and data curators will discuss challenges and current best practices as well as resource allocators. Data management plans for various funders such as NSF and NIH will be explored and federal funders will join in this discussion.1

Objectives and Learning Outcomes

By the end of the course, the student should be able to:

- Grasp the urgency of and need for long-term data curation in public and private venues
- Define data and identify their roles in organizational, technological, legal, cultural, and business contexts
- Understand both theoretical and practical issues in data curation from a broad range of perspectives
- Describe central data curation concepts, terms, tools, and technologies
- Be familiar with the current state of knowledge in the field and its key research issues as well as past trends and emerging priorities
- Understand the data curation lifecycle and identify the activities associated with each stage and their social, legal, ethical, and policy implications
- Understand the archival continuum model and how it relates to data curation.
- Understand the variety of data types generated and used by various disciplines, sub-disciplines, and research communities
- Understand how data curation relates to and has changed scholarship and research
- Examine the challenges associated with data sharing for collaboration and re-use
- Discuss roles for libraries and archives in data curation
- Understand the role of archival principles in data curation

1 Thanks to Dr. Christine L. Borgman for many ideas for this syllabus along with Alex Poole and Thu-Mai Christian.
• Delineate roles and responsibilities for data curators
• Understand the importance of planning to successful data curation
• Discuss public policy for data (scientific, governmental, and scholarly)
• Identify optimal data curation standards and best practices
• Understand crucial issues associated with storage, backup, and security of data
• Engage with challenges surrounding sustainability, auditing, risk management, trust, and costing
• Become familiar with current data curation education efforts

Office Hours.
I will be in my physical office after INLS 755, Thursdays 4:45 – 5:30. For Chapel Hill students, please drop by at this time, or send me an e-mail message for other scheduling opportunities. If I am not available when you try to see me, please send an email, which is my primary means of communication. Also, feel free to call me at home in the evening before 9:00 PM. I will so poll everyone to set up a weekly virtual office hour.

Email.
Please check the course listserv on a regular basis if not daily. This will be our primary means of communication. The address is: INLS756-Spring2014@sakai.unc.edu. If you have a question of general interest to the class such as “What do you mean by X in assignment Y?” please send this to the class list and I will answer it so that everyone can benefit.

How We Will Conduct “Class.”
While online courses afford great convenience, they also demand extra effort from both instructors and students. Because there is no face time for lectures, discussions, group work, or other activities, all this must take place through the course site on Sakai. This involves extensive writing and creation of slides, videos, and other media we will use to communicate our ideas and questions. The syllabus, assignments, and many of the readings (unless available online) will be posted on Sakai. Each week I will provide slides and brief “lectures” on important points but much of the work of the course will take place in the forums via discussion of the readings, videos, and slides.

Course Timeline.
First Week of Classes: January 8, 2014.
Spring Break: March 8 – 16, 2014.
Last Week of Classes: April 21, 2014.
Assignments & Evaluation.

<table>
<thead>
<tr>
<th>Assignment</th>
<th>% of Grade</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forum Posts (Analysis of Readings)</td>
<td>20%</td>
<td>On-going</td>
</tr>
<tr>
<td>Curation Profile Progress Report (Peer Review)</td>
<td>5%</td>
<td>Feb. 11</td>
</tr>
<tr>
<td>Data Curation Approach for Your Personal Digital Materials</td>
<td>10%</td>
<td>Feb. 18</td>
</tr>
<tr>
<td>Data Confidence Exercise</td>
<td>5%</td>
<td>Feb. 25</td>
</tr>
<tr>
<td>Workflow Tool Analysis</td>
<td>5%</td>
<td>Mar. 4</td>
</tr>
<tr>
<td>Critique Data Management Plan</td>
<td>10%</td>
<td>Mar. 18</td>
</tr>
<tr>
<td>Metadata Identification and Characterization Tools Exercise</td>
<td>5%</td>
<td>March 25</td>
</tr>
<tr>
<td>Curation Profile Presentation (Video to Class)</td>
<td>10%</td>
<td>April 8</td>
</tr>
<tr>
<td>DRAMBORA Exercise</td>
<td>5%</td>
<td>April 22</td>
</tr>
<tr>
<td>Curation Profile Written Report (Post to Class)</td>
<td>25%</td>
<td>Friday, May 2, by 4:00 PM</td>
</tr>
</tbody>
</table>

Class Participation

- You are expected to participate in discussion boards on a regular basis, preferably every week.
- You may create a new discussion topic, but please try to advance the discussion of an existing topic.
- Posts may address the readings and the other materials that are part of that week’s assignments as well as any other topic relevant to a given week’s content.
- Grades are assigned based on the quality of the posts rather than the quantity.
- Posts will be assessed based on the following criteria:
  - Relevance of the post to the question/conversation, i.e. the degree to which a post advances the discussion
  - Level of insight displayed by the post
  - Depth of understanding of the material, i.e. the level of preparation shown in the post.
  - Inclusion of relevant materials outside the assigned readings

The purpose of the forum discussions is to help students to think critically about issues and challenges related to digital curation and management, and to address ways that the literature may affect practice. It is helpful, when preparing each week’s work, to consider how that session’s particular theme and readings may apply to your area of professional interest. You are also encouraged to share current news and events you find informative to issues of data curation, on Sakai.
Graduate Grading Scale

- **H** (95-100): "Clear excellence", above and beyond what is required
- **P** (80-94): Entirely satisfactory; SILS recognizes subtle levels of "satisfactory" since most grades tend to cluster here:
  - **L** (70-79): Low passing
  - **F** (< 70): Failed
  - **IN**: Work incomplete (only given under extreme circumstances, such as serious illness)

Honor Code & Class Conduct.

**Honor Code:**

The University of North Carolina at Chapel Hill has had a student-administered honor system and judicial system for over 100 years. The system is the responsibility of students and is regulated and governed by them, but faculty share the responsibility. If you have questions about your responsibility under the honor code, please bring them to your instructor or consult with the office of the Dean of Students or the Instrument of Student Judicial Governance. This document, adopted by the Chancellor, the Faculty Council, and the Student Congress, contains all policies and procedures pertaining to the student honor system. Your full participation and observance of the honor code is expected.

This class follows the UNC Honor System. Information on the Honor Code can be found at: [http://honor.unc.edu/](http://honor.unc.edu/). Please read through The Honor System's Module at: [http://studentconduct.unc.edu/students/honor-system-module](http://studentconduct.unc.edu/students/honor-system-module) to become familiar with the UNC Honor Code and to understand the rights and responsibilities defined therein.

The Instrument of Student Judicial Governance, ([http://instrument.unc.edu/](http://instrument.unc.edu/)) which contains the provisions of the Honor Code, states that students have four general responsibilities under the Code:

1. Obey and support the enforcement of the Honor Code;
2. Refrain from lying, cheating, or stealing;
3. Conduct themselves so as not to impair significantly the welfare or the educational opportunities of others in the University community, and maintain the integrity of the University;
4. Refrain from conduct that impairs or may impair the capacity of University and associated personnel to perform their duties, manage resources, protect the safety and welfare of members of the University community, and maintain the integrity of the University.

The Instrument ([http://instrument.unc.edu/](http://instrument.unc.edu/)) prohibits giving or receiving unauthorized aid on examinations or in the completion of assignments. The Honor Code defines plagiarism as "deliberate or reckless representation of another's words, thoughts, or ideas as one's own without attribution in connection with submission of academic work, whether graded or otherwise." Whenever you use the words or ideas of others, this should be properly quoted and cited. You should adopt a style guide – e.g., American Psychological Association, Chicago Manual of Style, MLA, or Turabian – and use it consistently. (I do not care which one you select!) Students who are discovered
attempting to take credit for work performed by others will be referred to the Honor Court for resolution.

Please include the following pledge on all submitted work: “On my honor, I have neither given nor received unauthorized aid on this assignment.”

**Class Policies:**

- Be prepared for each week’s lessons by completing the assigned reading and exercises, enabling you to ask questions and participate in class discussion online.
- Be an active and positive participant in online discussions, characterized as:
  - Having a clear command of the readings for the week;
  - Sharing analyses and opinions based on the readings;
  - Respecting other students’ views and opinions; and
  - Freely agreeing and disagreeing with others when warranted.
- Follow the netiquette guidelines below.
- Please note: An intellectual exchange of ideas is the cornerstone of education, but any criticism should be limited to an idea and not the person specifically.
- Turn in assignments by the due date unless you make previous arrangements with the instructor. Unexcused late assignments are unfair to your classmates.

**Penalty for Late Assignments.**

I expect assignments to be passed in on time. This is important for at least two reasons: 1) the need to meet deadlines is a reality of professional life, and 2) giving some people more time for an assignment than others in the class is not equitable. However, life happens to all of us at one time or another. If you cannot meet an assignment deadline please tell me why PRIOR to the due date. I will negotiate a new deadline with anyone who has a valid reason for needing this (i.e., NOT “I just didn’t get it done.”). Otherwise, late assignments will drop 3 points for each day late.

**Online Etiquette Guidelines.**

http://www.kent.edu/dl/technology/etiquette.cfm
http://www.indiana.edu/~icy/netiquette.html
http://www.brightHub.com/education/online-learning/articles/26946.aspx

**Textbook.**


**Supplemental Resources.**

DCC Glossary: [http://www.dcc.ac.uk/digital-curation/glossary](http://www.dcc.ac.uk/digital-curation/glossary)
DCC How-To Guides: [http://www.dcc.ac.uk/resources/how-guides](http://www.dcc.ac.uk/resources/how-guides)
DCC Briefing Papers: [http://www.dcc.ac.uk/resources/briefing-papers](http://www.dcc.ac.uk/resources/briefing-papers)
Weekly Course Calendar

Week 1, January 8-14: Introduction, Syllabus, and Context

*Please watch this short video first. This sums up much of what is going on in data sharing and the challenges data curators face on a day-to-day basis. Enjoy.
http://www.youtube.com/watch?v=N2zK3sAtr-4

http://www.ijdc.net/index.php/ijdc/article/viewFile/6/2


Data Models

Digital Curation Lifecycle Model: http://www.dcc.ac.uk/docs/publications/DCCLifecycle.pdf

Data Curation Profiles


Week 2, January 15-21: All about Data


Week 3, January 22-28: Digital Data Sharing and Reuse: the Fundamentals


Week 4, January 29 – February 4: Big Data and DataNet

**Big Data**


**DataNet**


SEAD: Sustainable Environment through Actionable Data: http://sead-data.net/

The DataNet Federation Consortium: http://datafed.org/

Terra Populus: http://www.terrapop.org/

Week 5, February 5-11: Digital Data and the Evolution of Scholarship (E-Science/Cyberscholarship)

**Curation Profile Progress Report DUE.**


Week 6, February 12-18: The Role of Archives in Data Curation

**Data Curation Approach for Your Personal Digital Materials DUE.**


Week 7, February 19-25: The Roles of Libraries and Institutional Repositories in Data Curation

**Data Confidence Exercise DUE.**


Week 8, February 26-March 4: Education, Training, and Skilling Up

Workflow Tool Analysis DUE.


Renear, Allen H.; Palmer, Carole L., and John Unsworth. *Extending Data Curation to the Humanities: Curriculum Development and Recruiting.* Urbana-Champaign: Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 2013. [http://hdl.handle.net/2142/42628](http://hdl.handle.net/2142/42628)


**Week 9, March 5-18: Data Management Plans & Policy**

**Data Management Plan Critique DUE.**


**Tools to Explore:**

CDL. UC3 “Manage Your Data.” http://www.cdlib.org/services/uc3/datamanagement/


DCC. DMPT Online: Data Management Planning Tool. http://dmponline.dcc.ac.uk/


ICPSR. “Guidelines for Effective Data Management Plans.” http://www.icpsr.umich.edu/icpsrweb/ICPSR/dmp/index.jsp

NSF. Dissemination and Sharing of Research Results NSF Data Management Plan Requirements

NSF Data Management for NSF Engineering Directorate Proposals and Awards.

NSF Guide to Grant Applications (effective January 18, 2011).


**Week 10, March 19 - 25: Metadata**

**Metadata Identification and Characterization Tools Exercise DUE**


Week 11: March 26 – April 1: Data and the Social Sciences

Preservation Planning (PLATO) Exercise Due.


DDI, Data Documentation Initiative. http://www.ddialliance.org/

DVN, Dataverse Network Repository Software. www.thedata.org


King, Gary. “Ensuring the Data-Rich Future of the Social Sciences,” Science 331 no. 6018, 719-721 http://www.sciencemag.org/content/331/6018/719.short


**Week 11: April 2 – 8: Data and the Arts and Humanities**

**Curation Profile Presentation to Class (Video).**


Week 12, April 9-15: Data for Science and Health Research


Week 13, April 16-22: Auditing, Risk Management, Costing, Trust, and Sustainability

**DRAMBORA Exercise DUE**


[http://public.ccsds.org/publications/archive/652x0m1.pdf](http://public.ccsds.org/publications/archive/652x0m1.pdf)

[http://catalog.crl.edu/search~S1/?/Xtrusted+repositories&searchscope=1&SORT=R/Xtrusted+repositories&searchscope=1&SORT=R&SUBKEY=trusted%20repositories/1,15,15,B/1856-b2212602&FF=Xtrusted+repositories&searchscope=1&SORT=R&6,6,,1,0](http://catalog.crl.edu/search~S1/?/Xtrusted+repositories&searchscope=1&SORT=R/Xtrusted+repositories&searchscope=1&SORT=R&SUBKEY=trusted%20repositories/1,15,15,B/1856-b2212602&FF=Xtrusted+repositories&searchscope=1&SORT=R&6,6,,1,0)

[http://datasealofapproval.org/](http://datasealofapproval.org/)


[http://www.dlib.org/dlib/july06/ross/07ross.html](http://www.dlib.org/dlib/july06/ross/07ross.html)

The Cost of Archiving: several short video talks from the Personal Digital Archiving Conference at the Internet Archive, February, 2011. Jeff Ubois, Stephen Griffin, David Rosenthal, Brewster Kahle, and panel discussion:  

**Week 14, April 23 – 25: Forensics, the Cloud, and the Future**

http://www.clir.org/pubs/reports/pub149/reports/pub149/pub149.pdf


May 2, by 4:00 PM.

Curation Profile Written Report (Post to Class) DUE.