

University of Maryland iSchool
INST 448 -- Section 0101

Digital Curation Research in Cultural Big Data Collections

Course Syllabus Spring 2019

Ken Heger, Ph.D.

Richard Marciano, Ph.D.

4110 Hornbake South

Class time: TuTh 11a.m.-12:15p.m.

Classroom: EDU 1121 & DCIC Lab

E-mail: kheger@umd AND marciano@umd.edu

Office hours: By Appointment

Pre-requisite: INST 201 Introduction to Information Science and INST 311 Information Organization

This course is repeatable. It can be taken up to two times for up to six credits in total.

Catalog Description: This course focuses on introducing students to the principles, methods, and technologies involved in the digital curation of large cultural data collections. Students will learn these concepts in class lectures, discussions, and participating on project teams in the Digital Curation Innovation Center (DCIC).

Extended Course Description: This course provides an overview for students interested in learning the theory and practices involved in digital curation, and how this is applied in managing and accessing information in large cultural data collections. The digital curation lifecycle will be used as the foundation for understanding how records/information are created, managed throughout active use, and preserved for future access. Cyber-infrastructure development and cultural Big Data collections accessed in the DCIC will form the basis for instruction, research, and learning. Students will participate in class lectures, discussions, and complete reading assignments.

Student learning will be reinforced by active engagement on DCIC project teams focused on cyber-infrastructure projects and large data collections involving justice, human rights, and cultural heritage documentation. Examples of projects include:

- Students will participate in an IMLS-funded 2-year project for research and development of Fedora-compatible software that achieves web scale and is open to computational approaches to improving collections and access.

- A partnership with the National Archives, students will focus on “Internal Security Case Reports” index cards from WWII Japanese American Internment Camps to create an automated processing workflow for the incident cards.
- In partnership with Johns Hopkins University, Virginia Tech, and the University of Richmond, students will work with digitized Federal records to curate relevant collections in order to develop a comprehensive national dataset of “redlining” data that capture racial, ethnic, and economic conditions.

(These are example projects; students who are taking this course may be assigned to other projects that are not listed here.)

Project teams will have expected outcomes, with student contributions clearly identified.

Learning Outcomes: Upon completion of this course, students will be able to:

- Demonstrate competency in the managing of core digital curation lifecycle processes.
- Apply the technical, intellectual, and social awareness required to successfully contribute to digital research projects.
- Develop effective research strategies to resolve or advance digital curation solutions to digital information challenges.
- Show insight and capability in applying digital curation tools in addressing research questions.
- Demonstrate mastery of project management, project team dynamics and performance.

Course Requirements:

- **Class Lectures and Discussions:** Students are expected to actively participate in class discussions and demonstrate mastery of assigned topics and readings.
- **Team Projects:** Each student will be on a small team. The teams may be posting a **blog** at designated team reports during the course documenting the project topic, goals and deliverables, progress to date, challenges encountered and strategies to overcome them. TBD. Blogs will be one to two pages in length. **Each** team member, on a rotating basis, will prepare and post the weekly blog on the ELMS/Canvas course site.
- **Final Presentation:** Each team in the last two class sessions will present a poster that documents the topic of the project, team members, work accomplished, technologies employed, and the impact of the project. **Each** team member will present a segment of the final presentation.

Grades: Students will be evaluated on the basis of:

- Class participation 10%
- Course review blogs 10%
- Assessments 20% (10% for each of the 2)
 - Deadlines: Tue. Sep 11 & Thu. Oct. 4
- Team Project 40% (20% for each of the 2 phases)
 - Deadlines: Tue. Oct. 30 & Tue. Nov. 20

- Final presentation 20%
 - Deadline: Tue. Dec. 4 & Thu. Dec. 6

Final grades will be assigned using the following categories:

A+	97-100 pts.	C	73-76.9
A	93-96.9	C-	70-72.9
A-	90-92.9	D+	67-69.9
B+	87-89.9	D	63-66.9
B	83-86.9	D-	60-62.9
B-	80-82.9	F	less than 60
C+	77-79.9		

Course Materials: There is no required text for this course. Additional readings will be assigned as needed.

Attendance: Attendance in class is required. If you are unable to attend class for reasons of illness, religious observance, participation in University activities at the request of University authorities, or compelling absences beyond the student's control, please let the instructor know [prior to class for non-emergencies].

Classroom Environment: The classroom environment should be professional and respectful. Punctual arrival at classes contributes to the smooth operation of the class and the quality of the learning experience. Late arrivals, early departures, and wandering in and out of the classroom are disruptive and distracting. If you know you will be unavoidably detained or must leave early, please let the instructor know in advance. Please turn off or mute all phones and other communication devices during each class session. If you use your laptop in the classroom, limit the usage of the computer to course-related reasons [i.e., taking notes, working on projects].

Academic Integrity: The University of Maryland, College park has a nationally recognized Code of Academic Integrity, administered by the Student Honor Society. This Code sets standards for academic integrity at Maryland for all undergraduate and graduate students. As a student, you are responsible for upholding these standards for this course. The Code of Academic Integrity prohibits students from cheating on exams, plagiarizing, submitting fraudulent documents, forging signatures, submitting the same paper for credit in two courses without authorization, and buying papers. It is very important for you to be aware of the consequences of academic dishonesty. Instances of any suspected academic dishonesty will be reported and handled according to University policy and procedures. For more information on the Code, visit <http://www.shc.umd.edu>

Plagiarism is of particular concern in the networked digital environment. Students must write their essays and assignments in their own words. Whenever students take an idea or a passage of text from another author, they must acknowledge their source both by using quotation marks where appropriate and by proper referencing using footnotes or in-text

citations. For further information about proper citation of sources, consult the UMD Libraries website at <http://www.lib.umd.edu/guides/honesty.html> and <http://www.lib.umd.edu/PUBSERV/citations/index.html>.

Extensions: Late submissions of written assignments [including the final exam] will carry a penalty unless prior arrangements are made with the instructor. If an extension is granted, the work must be submitted within the extension period to avoid grade penalties. Unexcused delays in submission of the assignment will result in a reduction of the grade by one category for each day the paper is late; for example, a paper that would have received a B+ if submitted on time will receive a B if it is submitted a day late, a B- if it is two days late, and so on.

Students with Disabilities: Students with disabilities who require academic accommodations must inform the instructor of their needs and provide written documentation about the appropriate academic accommodations from Disability Support Services [<http://www.counseling.umd.edu/DSS>] at the beginning of the term.

=====

PRELIMINARY SETUP

Teams: Each team will have 3-5 students working together. Teams have to be formed by the end of Session 4.

Projects: By the end of Session 13, students will be introduced to 5 projects, from which each team needs to choose 1 to work on for the rest of the semester. Below are the 5 projects:

1. Mapping Inequality
2. Japanese-American WWII Camps
3. Legacy of Slavery
4. Global Journeys, Local Communities
5. Managing White House Correspondents Association (WHCA) Pool Reports

Course Schedule & Content

SESSIONS 1 - 12 (LECTURES AND CASE STUDIES)

These sessions will include lectures delivered by Drs. Heger and Marciano supported by one or more case studies (Case studies will be introduced after Assessment 1). The instructors will teach the concepts and then work with students through case studies where these concepts have been previously applied on real-world collections. The case studies will build the foundation for students to work on their team projects for the rest of the semester, as they now have an understanding on what Big Cultural Data collections are and what treatments can be applied on these collections to make them accessible and valuable to the end users.

SESSIONS 13 - 27 (TEAM PROJECTS)

Students will be introduced to the 5 projects and by the end of Session 14, each team will need to choose one project that they will work on for the rest of the semester. The project work will be divided into 2 phases, with clear deliverables defined for each phase, after which there will be a final presentation.

Project Clinics: Some sessions during this time will be at the DCIC Labs instead of the classroom, since students will mostly be working on projects amongst themselves. These sessions are called “Project Clinics” and an instructor will be available to guide students. However, Phase 1, 2 and the final presentations will be in the classroom.

SESSIONS 28 - 29 (FINAL PROJECT PRESENTATIONS)

During the last two sessions of the course, each team will present the work done on their projects.

.	Week	Date	Topic	Deliverables
Aug.	Week #1	#1: Tue. Aug. 28	Course Intro.	
		#2: Thu. Aug. 30	Digital Curation and Cultural Collections as Data	
Sep.	Week #2	#3: Tue. Sep. 4		Open Assessment 1
		#4: Thu. Sep. 6		Form student teams
	Week #3	#5: Tue. Sep. 11	Introduce 1st Case Study	Deadline to submit Assessment 1
		#6: Thu. Sep. 13		Lectures and Case Studies
	Week #4	#7: Tue. Sep. 18		Lectures and Case Studies
		#8: Thu. Sep. 20		Lectures and Case Studies
	Week #5	#9: Tue. Sep. 25		Lectures and Case Studies
		#10: Thu. Sep. 27		Open Assessment 2
Oct.	Week #6	#11: Tue. Oct. 2		
		#12: Thu. Oct. 4		Deadline to submit Assessment 2
	Week #7	#13: Tue. Oct. 9	Introduce 5 Projects	
		#14: Thu. Oct. 11	Instructors to provide details about Project Phase 1	Deadline for teams to provide project Preferences

	Week #8	#15: Tue. Oct. 16		Project clinic #1.1
		#16: Thu. Oct. 18		Project clinic #1.2
	Week #9	#17: Tue. Oct. 23		Project clinic #1.3
		#18: Thu. Oct. 25		Project clinic #1.4
	Week #10	#19: Tue. Oct. 30		Project Phase 1 submission and presentation
Nov.		#20: Thu. Nov. 1	Instructors to provide details about Project Phase 2	
	Week #11	#21: Tue. Nov. 6		Project clinic #2.1
		#22: Thu. Nov. 8		Project clinic #2.2
	Week #12	#23: Tue. Nov. 13		Project clinic #2.3
		#24: Thu. Nov. 15		Project clinic #2.4
	Week #13	#25: Tue. Nov. 20		Project Phase 2 submission and presentation
		Thu. Nov. 22	No Class: Thanksgiving Break	
	Week #14	#26: Tue. Nov. 27	Instructors to provide details about Final Project Presentation	Project clinic #3.1
		#27: Thu. Nov. 29		Project clinic #3.2
Dec.	Week #15	#28: Tue. Dec. 4		Final Project Presentation
		#29: Thu. Dec. 6		Final Project Presentation

ASSESSMENTS AND PROJECT EVALUATIONS

Students will be given team-based assessments to enhance their knowledge in the field and also assess their understanding of the concepts. The goal is to keep the students focused in class and timed deliverables will help the continuous learning process. The assessments are take-home mini projects or reports that need to be submitted.

Assessment 1:

Students will be provided with a sample collection and asked to do an initial analysis of the documents based on a set of sample questions the instructors will provide. The goal is to make the data accessible, usable and generate value out of it. During this time, students would not have been introduced to most of the tools available at their disposal. The goal is for students to understand the difficulty in working with these collections without the kinds of state-of-the-art technologies that are leveraged at the DCIC.

Submission Deadline: Session 5

Assessment 2:

By now, students would have mastered a number of working concepts and have also worked through one or more case studies. Now that they have a better understanding of how to deal with these types of collections, they will revisit the sample collection from Assessment 1 and demonstrate how they would apply the kinds of treatments discussed in the course.

Submission Deadline: Session 12

Project Phase 1: Conceptual Analysis & Collection Preparation (4 in-class sessions)

Phase 1 deliverables (*we will use the Legacy of Slavery project as an example of how the two phases can be structured, and the same will apply for the other 4 projects*) will include the following:

- Perform basic descriptive statistics on this dataset to explore it and understand it better. For example, what are the different age groups of slaves, count of male or female, and who are the owners with the greatest number of slaves?
- Transcribe 50 Certificates of Freedom / Census records and store the information in a Google Spreadsheet, explore how the data transcribed by each student and perform any data cleaning tasks if necessary.
- Each team will submit a short report on the steps taken to identify the data, clean and explore data and the individual contributions of each team member.
- Teams will also present the work done in Phase 1 to everyone in class, talk about any challenges they had and get feedback from other students and the instructors.

Submission Deadline: Session 19

Project Phase 2: Data Analytics and Knowledge Discovery (4 in-class sessions)

The deliverables will include the following (*again, using Legacy of Slavery Project as the example*):

- Export the data that is already transcribed and archive it in a database that acts as a warehouse for easy reporting, analysis and visualization.
- Connect the database to a data visualization tool like Tableau and validate that the connection works.
- Each team will submit a report on the steps taken to create the database, connect to Tableau, and the individual contribution of each team member.
- Teams will also present the work done in Phase 2 to everyone in class, talk about any challenges they had and get feedback from other students and the instructor.

Submission Deadline: Session 23

Final Presentation: Wrapping it up with Extra Steps (2 in-class sessions)

The deliverables will include:

- The complete analysis on the collection they have been working on over the last 2 additional weeks which will include the work done after Phase 2
- Each team will submit a comprehensive report of the project with the steps they took to implement it, some interesting findings from the collection and the individual contribution of each team member.

- Teams will also deliver a 20-minute presentation to everyone in class, followed by a 5-10 minute Q&A / discussion.

Submission Deadline: Sessions 29 and 30