Data and Culture

Mondays and Thursdays, 10:20–11:40 a.m., in Academic Building 4450
Professor Andrew Goldstone (andrew.goldstone@rutgers.edu)
Office hours: Wednesdays 12–1 p.m. in Murray 019, and Zoom by appointment
Professor Meredith McGill (mlmcgill@english.rutgers.edu)
Office hours: Mondays 12–1 p.m. in Murray 105, and Zoom by appointment

The digitization of wide swaths of the print record has opened up new challenges and opportunities for researchers in the humanities. This course introduces students to some of the key techniques used by humanities scholars to organize, manipulate, and analyze digital sources—attending both to longstanding scholarly institutions and practices that shape our understanding of digital texts (critical editions, brick-and-mortar archives, and quantitative methods within social, political, and cultural history) and to new methods for studying texts, cultural geography, and relations between and among producers and consumers of culture.

Students who complete this course will develop facility in the use of digital tools for the representation, curation, and analysis of digital text. In each case, however, we will place these relatively new tools within a longer history of humanistic inquiry and will ask: what insights can these tools provide, and what questions (and texts) do they marginalize or occlude? Our aim throughout is to examine how digitization and data science have changed the questions that humanists can ask of their sources. What does it mean to think of culture as data? What new histories do these tools and methods help us uncover? In what ways has digitization helped and hindered the ability of humanities disciplines such as history, literary studies, and art history to provide an understanding of the past that can speak to urgent questions in the present moment?

Learning goals

This course is designed to meet the following Core Curriculum goals:

CCO-2: Analyze the relationship that science and technology have to a contemporary social issue.

AHp: Analyze arts and/or literatures in themselves and in relation to specific histories, values, languages, cultures, and technologies.

Through a combination of lectures, hands-on lab sessions, digital projects, and reflective essays, students will:

1. Explore the differences between manuscript, printed, and digital sources for humanities scholarship;
2. Analyze some of the structures used to organize and preserve culture, including library infrastructure, classification schemes, and item-level metadata;
3. Reflect on gaps and unevenness in the digital record;
4. Develop familiarity with a variety of approaches to analyzing digitized and born-digital texts both singularly and in aggregate;
5. Practice formulating humanistic arguments using digitized evidence, paying close attention to the analytical capacities and limitations of that evidence and of the methods that transform culture into data.

Requirements

Participation: Informed attendance and in-class writing (10%)

Informed attendance. We expect you to come to class prepared to discuss the assigned reading and participate in lab activities. You are permitted no more than **two** unexcused absences without penalty. Further absences will affect the participation grade: a student with three unexcused absences will receive no higher than a B for participation; with four, a C. Five or more unexcused absences will ordinarily lead to an F for the course as a whole.

If you need to miss class for some reason, please use the [Self-Reporting Tool](#) to let us know that you'll be absent. Missing class due to religious observance, illness, or emergency will be excused; in such cases you should **both** use the Self-Reporting Tool **and** email us before class to let us know why you will be absent. With excused absences, we will work with you to make up what you've missed; with unexcused absences, you're on your own. (We highly recommend consulting with a classmate!)

In-class writing. We will often begin class by asking you to write for 5 or 10 minutes in response to a provocation of some kind. These short writing exercises won't be graded, but they will stand as evidence that you have come to class prepared to discuss the material assigned for the day.

Office Hour Visits. You are required to visit office hours at least once before mid-term. We will ask you to sign up in groups no larger than three students; you can also just swing by unannounced.

Short exercises (20% total)

We will occasionally assign short exercises building on readings and in-class work. These will be collected and graded.

Two Essays (15% each)

You will complete two short papers (4–5 pages each), one reflecting on a digital archive and one reflecting on an example of data analysis.

Final Paper or Project (40%)

You will develop a final paper or digital project in close consultation with us. Critical essays may take up any of the major themes and topics of the course; digital projects may present new or existing cultural data in the context of a thoughtful interpretation. Projects will be primarily evaluated based on the quality of reflection and analysis, not technical accomplishment. A project proposal, not separately graded, is due after the Thanksgiving recess.

Grading

Grades will be given on the four-point scale as specified in the Undergraduate Catalog. In converting the final numerical score to a letter, the equivalents in the Catalog are taken as the maxima of intervals open on the left and closed on the right. Thus A corresponds to scores strictly greater than 3.5 and less than or equal to 4.0, B+ to scores greater than 3.0 and less than or equal to 3.5, B to scores greater than 2.5 and less than or equal to 3.0, and so on. There are no "minus" grades. The general standards for grades are as follows:

A range (3.5, 4.0]: Outstanding work, demonstrating thorough mastery of course materials and skills.

B range (2.5, 3.5]: Good work, demonstrating serious engagement with all aspects of the course but incomplete mastery of course materials and skills.

C range (1.5, 2.5]: Satisfactory work, meeting requirements but indicating significant problems mastering the course materials and skills.

D (0.5, 1.5]: Poor or minimally passing work, meeting the basic course requirements, but frequently unsatisfactory in several major areas.

F [0, 0.5]: Failure due to unmet course requirements or consistently unsatisfactory work.

The final grade will be based on a numerical score but is subject to our discretion. Unsatisfactory work in all areas of the course will result in an F even if the numerical score corresponds to a passing grade. It is not possible to pass the course without turning in both short papers and the final paper or project.

Academic integrity

Students and instructors have a duty to each other and to our community to abide by norms of academic honesty and responsibility. To present something as your own original writing when it is not is plagiarism. Plagiarism and other forms of cheating are serious violations of trust. Academic dishonesty, including plagiarism, will have severe consequences, in accordance with the University Policy on Academic Integrity and the Code of Student Conduct. For details about the University's academic integrity policies, please see academicintegrity.rutgers.edu.

While we expect **all exercises and papers** to be written by the student himself or herself, fear of inadvertent plagiarism should not keep students from talking (and arguing) with one another about the texts we study, helping each other to interpret difficult passages, or reporting on such disputes or collaborative understandings in their written work. When in doubt, note the assistance of your friends and teachers. It's ethical *and* polite.

Students with disabilities

All reasonable accommodation will be given to students with disabilities. Students who may require accommodation should contact the Office of Disability Services (ods.rutgers.edu; 848-445-6800) and speak with us at the start of the semester.

Schedule

Readings are to be completed by the day they are listed under. All readings are available on Canvas.. The schedule may change as term goes on; the most up-to-date version will always be on Canvas.

Part I. Documents, Texts, and the Raw Material of Humanities Scholarship

What counts as evidence in humanities disciplines, and where does it come from? What are the principles that govern the preparation of scholarly editions of historic manuscripts? How do editors handle textual variants in printed editions of literary works? What strategies have editors of digital editions adopted to remediate aspects of handwritten and printed texts?

Thursday, September 8. Introduction: What is Data? What is Culture?

Monday, September 12. Textual Criticism and the Problem of Variants

> Kelemen, *Textual Editing*, 3–25.

(Tuesday, September 13. Last day to drop the course with a "W.")

Thursday, September 15. Lab: Comparing Editions, Establishing Authority

> [Walt Whitman Archive](#).

> Selections from the Whitman Variorum.

Monday, September 19. Guest Lecture: Introduction to Text Encoding (Francesca Giannetti)

> Flanders et al., "Text Encoding."

> In-class activity: marking-up on paper.

> **Exercise 1 due:** Whitman revisions response.

Thursday, September 22. Guest Lecture: Introducing the Rutgers College War Services Bureau project (Francesca Giannetti)

> Biographical notes and letters for ["Ainsworth, William P. E.,"](#) ["Costa, Joseph Louis,"](#) and ["Jackson, Morris Bacon"](#) in Giannetti, ed., *[Personal correspondence from the Rutgers College War Service Bureau](#)*.

> In-class activity: transcribing a primary source.

Monday, September 26. Lab: Close Reading with the TEI

Rau, "Introduction," Letters, *Willa Cather Archive*.

Please also use the search interface to find and read one letter on a topic of your choosing, e.g. "flu," "travel," "armistice." Scroll down or click on "More about this letter…" to examine the underlying XML encoding ("Source File").

Thursday, September 29. "Plain" and/or Edited Text?

Cordell, "Q i-jtb the Raven."

Part II. Archives

How have the institutions that protect and transmit the historical record been shaped by the needs of the state, the university, and other communities? What are the differences between and among archives, repositories, and collections, and how has digitization shifted these definitions? What is metadata and how does it reflect the organization of knowledge?

Monday, October 3. Cruising the Library

Adler, *Cruising the Library*, "Preface" (ix-xvii) and Chapter 3: "Mapping Perversion: HQ71, etc." (92-119)

Exercise 2 due: response to Cordell**.**

Thursday, October 6. Guest Lecture: Documents and Documentary Editing (Paul Israel)

Readings TBA.

Monday, October 10. Guest Lecture: Versions of Cultural Memory (Andrew Parker)

Theimer, "Archives in Context and as Context."

Exercise 3 due: TEI markup of wartime letter.

Thursday, October 13. Visit to Alexander Library: Libraries and the Organization of Knowledge

Meet in Alexander Library Teleconference Lecture Hall, 4th floor.

In-class: begin library analytic exercise.

Monday, October 17. Lab: Introduction to Omeka

Follow instructions on Canvas for creating a free omeka.net account.

Thursday, October 20. Guest lecture: Digital History, Digital Memory (Kristin O'Brassill-Kulfan)

Rumsey, *When We Are No More: How Digital Memory Is Shaping Our Future* (excerpt).

Exercise 4 due: Group report on Library Analytic Exercise.

Part III. More than one (cultural) thing

What are some of the advantages and disadvantages of treating elements of culture not as individual cases, but in the aggregate? What is revealed through counting that is hidden in the singular case, and vice-versa?

Monday, October 24. Antinomies of Taste

Lieberson, *A Matter of Taste*, xi–xvi, 31–42, 77–91, and optionally 92–102.

Bourdieu, *Distinction*, 11–18.

Thursday, October 27. Lab: Exploring Taste Data with R

Follow instructions on Canvas for R set-up.

Monday, October 31. Counting, Representing, Seeing

Griswold, "American Character and the American Novel."

Tuchman and Fortin, "Fame and Misfortune."

Short essay 1 due: on digital archives.

Thursday, November 3. Lab: Visualization in R

> Manovich, "What Is Visualization?"

> Wickham and Grolemund, *R for Data Science*, chap. 3 (all exercises optional)

Part IV. Texts as data

Monday, November 7. Genre and Literary Classification

> Moretti, "The Slaughterhouse of Literature."

> Underwood, "The Life Span of Genres," chap. 2 in *Distant Horizons.*

Thursday, November 10. Lab: Exploring Genre Data

> Online supplement to Underwood, "Life Span of Genres."

Monday, November 14. Scaling Up, or Not?

> Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books."

> English, "Now, Not Now."

(Wednesday, November 16.)

> **Short essay 2 due:** on data and analysis in Underwood.

Thursday, November 17. Lab: Most of the Words in Libraries

> Optional: Rosenberg, "Data before the Fact."

Part V. Networks

What can we learn by connecting individual historical actors to one another in networks of association? What are the principles that govern the visualization of networks, and what

changes when we weigh factors differently? What can networks tell us about cultural history, and what do they occlude?

Monday, November 21. Guest Lecture: Networks and Network Analysis (Sean Silver)

Ahnert et al., *The Network Turn*, chap. 2 ("Historical Threads").

Easley and Kleinberg, *Networks, Crowds, and Markets,* chap. 2 ("Graphs") .

Tuesday, November 22 (Thursday classes today). Lab: Let's Make a Hairball!

Franco Moretti, *Distant Reading*, 212–30.

Shin-Kap Han, "The Other Ride of Paul Revere."

(Thursday, November 24. Thanksgiving recess.)

Monday, November 28. Networks and Cultural Fields

Casanova, *World Republic of Letters*, 133–37, 164–72.

Sapiro, "Translation and Symbolic Capital in the Era of Globalization."

Final project proposal due.

Thursday, December 1. Lab: Project Workshop Day

Monday, December 5. Network Reality and Network Uncertainty

Browse: Six Degrees of Francis Bacon.

Optional: Warren et al., "Six Degrees of Francis Bacon: A Statistical Method for Reconstructing Large Historical Social Networks."

Very optional: Healy, "The Performativity of Networks."

Thursday, December 8. Lab: Project Workshop Day

Monday, December 12. Retrospect and Prospect

(Monday, December 19.)

Final projects due at 12 p.m.