**ANTH 592 / ILS 595 - Data Management & Curation *for Qualitative Research***
*Spring 2019*
*Tuesday 4:30-7:20pm*
*WALC 3045*

Instructor: Kendall L. Roark
Email: roark6@purdue.edu
Office: STEW174
E-mail: roark6@purdue.edu
Phone: 765-494-2637
Office Hours: day/time tba & by appointment

COURSE DESCRIPTION
This course offers an interdisciplinary introduction to data management and curation with a focus on the use, value and organization of data, materials, infrastructure, tools and scholarly communication. This course is aimed at both qualitative researchers and those interested in the curation and stewardship of qualitative research data and materials within library, museum, archival and gallery settings. The course will both introduce literature concerning ethical and legal considerations of data management and curation, and  provide the opportunity for hands-on data and digital literacy skills development.

Data and digital literacy include the cleaning, documentation and organization of primary research data and materials collected and/or accessed from other producers/sources (including a text corpus, social media and/or other data sources). Other data and digital literacy topics covered will include: finding and evaluating the quality and provenance of data; preparing data for publication and/or creating digital exhibit; writing a data management and/or sharing plan; creating and understanding metadata; and the principles of data organization (including systems for the management of both physical and digital data, common file naming conventions, implications of database design and distributed data storage and management systems).

COURSE LEARNING OBJECTIVES
● Demonstrate skill sets necessary to develop and curate data in the arts, humanities and/or social science scholarship.
● Demonstrate an understanding of the historical and emerging theoretical and practice-based implications of data curation for the arts, humanities and/or social sciences scholarship.
● Demonstrate skill sets necessary to critically evaluate the use of data, infrastructure, tools and scholarly communication within arts, humanities and/or social science research and practice.

COURSE REQUIREMENTS:
Please note that the syllabus may be modified throughout the course of the semester. Each student will sign up for a PURR (Purdue University Research Repository) account by the end of the first day of class. Updated information will be posted on the class PURR project page and emailed to the class in advance of any change. You are responsible for ensuring that you are familiar with this syllabus and keep up with class notices  and assignment materials posted to PURR and via email. https://purr.purdue.edu

READING/COURSE MATERIALS:
Corti, Liz. (2014) Managing and Sharing Research Data Handbook: A guide to good practice.
Sage.

Corti, Liz. (2014) Companion to Managing and Sharing Research Data Handbook: A guide to
good practice. Available at: https://www.ukdataservice.ac.uk/manage-data/handbook/

@eReserves

GRADES (.5 and above will be rounded to the next whole number, ex. 89.5 – 90% A-)
*Late assignments will drop one letter grade, ex. A- à B+ for each day past due)
A+ (97-100)
A (92-96)
A- (90-91)
B+ (88-89)
B (82-87)
B- (80-81)
C+ (78-79)
C (72-77)
C- (70-71)
D+ (68-69)
D (62-67)
D- (60-61)
F (0-59)

PARTICIPATION (20%)
Participation in lectures, discussion and in-class activities are critical to your success in this
course. If you are unable to attend a class session, please make alternate arrangements to
complete in-class assignments and activities. Please stay up-to-date with the readings and other
assigned materials. Please try to arrive on time to class out of respect for your peers, instructor
and/or guest lecturer. Your overall grade may drop by up to a letter grade (ex. A to A-) if you
miss more than two classes (barring major emergencies or absences prearranged with the
instructor).

WEEKLY PRACTICE ASSIGNMENTS (40%)
Each week we will begin and complete an in-class hands-on assignment. While we will work on
assignments in-class, it may often be necessary to dedicate time outside of class to successfully
complete the assignment by the weekly due date. Each assignment will be graded on critical
engagement with the assignment and attention to detail including completing all components of
the assignment.

TERM DATA CURATION PROJECT (40%)
Each student will develop and present a novel term data curation project related to an ongoing or
new research project. Students will have the freedom to choose the type of project, but must have
their term data curation project proposal approved by the instructor. Students should use
software, platforms and other tools including programming languages (if applicable) that are
appropriate for their project and that they are already familiar with or have the time to learn
during the course of the term. Projects can include, but are not limited to designing and
deploying a database prototype for use in their research; designing and deploying a prototype

digital archive, exhibit or data sharing portal for your research;  preparing a data publication or research study for deposit in a research repository or archive.

ACADEMIC INTEGRITY & DISHONESTY

Purdue Honor Pledge
"As a boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together - we are Purdue."

Definition of Academic Dishonesty
Purdue prohibits "dishonesty in connection with any University activity. Cheating, plagiarism, or knowingly furnishing false information to the University are examples of dishonesty" (*Section B.2.a of the* *Student Regulations*). Furthermore, the University Senate has stipulated that "the commitment of acts of cheating, lying, and deceit in any of their diverse forms (such as the use of ghostwritten papers, the use of substitutes for taking examinations, the use of illegal cribs, plagiarism, and copying during examinations) is dishonest and must not be tolerated. Moreover, knowingly to aid and abet, directly or indirectly, other parties in committing dishonest acts is in itself dishonest" (University Senate Document 72-18, December 15, 1972).

Academic Integrity
"Purdue University values intellectual integrity and the highest standards of academic conduct. To be prepared to meet societal needs as leaders and role models, students must be educated in an ethical learning environment that promotes a high standard of honor in scholastic work. Academic dishonesty undermines institutional integrity and threatens the academic fabric of Purdue University. Dishonesty is not an acceptable avenue to success. It diminishes the quality of a Purdue education which is valued because of Purdue's high academic standards" (S. Akers, *Academic Integrity, A Guide for Students*, 1995, revised 1999).

Purdue University Code of Honor
(*Student Regulations*, 2002-03)

"The purpose of the Purdue University academic community is to search for truth and to endeavor to communicate with each other. Self-discipline and a sense of social obligation within each individual are necessary for the fulfillment of these goals. It is the responsibility of all Purdue students to live by this code, not out of fear of the consequences of its violation, but out of personal self-respect. As human beings we are obliged to conduct ourselves with high integrity. As members of the civil community we have to conduct ourselves as responsible citizens in accordance with the rules and regulations governing all residents of the state of Indiana and of the local community. As members of the Purdue University community, we have the responsibility to observe all University regulations.

To foster a climate of trust and high standards of academic achievement, Purdue University is committed to cultivating academic integrity and expects students to exhibit the highest standards of honor in their scholastic endeavors. Academic integrity is essential to the success of Purdue University's mission. As members of the academic community, our foremost interest is toward achieving noble educational goals and our foremost responsibility is to ensure that academic honesty prevails."

COURSE POLICY

Plagiarism may result in an F for the course at the discretion of the instructor. Do not reproduce or share the work of the instructor (including syllabus) or your fellow students without prior permission from the author/creator. See Purdue's student guide for academic integrity: https://www.purdue.edu/odos/academic-integrity/ and University policy on use of copyrighted material: http://www.purdue.edu/policies/academic-research-affairs/ia3.html

MENTAL HEALTH

• On the recommendation of the University Senate, we are also asked to share the following on your syllabus about resources for mental health.

o CAPS Information: Purdue University is committed to advancing the mental health and well-being of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, such individuals should contact Counseling and Psychological Services (CAPS) at (765)494-6995 and http://www.purdue.edu/caps/ during and after hours, on weekends and holidays, or by going to the CAPS office of the second floor of the Purdue University Student Health Center (PUSH) during business hours.

STUDENTS WITH DISABILITIES

• The Disability Resource Center (DRC) is a resource for students and instructors. Students may present a "Letter of Accommodation" to you at any point in the semester. Should you have questions about accommodations, please contact the DRC at: 494-1247 or drc@purdue.edu. In many cases the DRC can partner with you to develop inclusive teaching strategies that benefit all students in your class.

• Accessibility and Accommodations Syllabus Statement: The DRC recommends the following statement be included in your syllabus. "Purdue University strives to make learning experiences as accessible as possible. If you anticipate or experience physical or academic barriers based on disability, you are welcome to let me know so that we can discuss options. You are also encouraged to contact the Disability Resource Center at: drc@purdue.edu or by phone: 765-494-1247." http://www.purdue.edu/drc/faculty/syllabus.html

NONDISCRIMINATION POLICY

*Purdue University prohibits discrimination against any member of the University community on the basis of race, religion, color, sex, age, national origin or ancestry, genetic information, marital status, parental status, sexual orientation, gender identity and expression, disability, or status as a veteran.*

Please refer to Purdue University's nondiscrimination statement: http://www.purdue.edu/purdue/ea_eou_statement.html

## COURSE SCHEDULE*

***The instructor reserves the right to modify the course schedule, reading, assignments and due dates.  The order and content of some of the reading and lab assignments may change to accommodate guest speakers.**

| WK | Topic(s) | Activity |
|---|---|---|
| 1 | **Research & Data Lifecycle Approach**: An Introduction to Data Management & Curation<br><br>Reading:<br>Jahnke, Lori, Andrew Asher, and Spencer D. C. Keralis. (2012) The Problem of Data. Washington, DC: Council on Library and Information Resources. Available at: http://www.clir.org/pubs/reports/pub154<br><br>Palmer, Carole L., Nicholas M. Weber, Trevor Munoz, and Allen H. Renear. (2013) "Foundations of Data Curation: The Pedagogy and Practice of 'Purposeful Work' with Research Data." Archive Journal no. 3. http://www.archivejournal.net/issue/3/archives-remixed/foundations-of-data-curation-the-pedagogy-and-practice-of-purposeful-work-with-research-data/<br><br>Tibbo, Helen. (2012) Placing the Horse before the Cart: Conceptual and Technical Dimensions of Digital Curation. Historical Social Research 37 (2012), 187-200. Available at: http://www.cceh.uni-koeln.de/files/Tibbo_final.pdf. | ● in-class, break into groups to discuss research topics/questions & potential semester projects<br>LAB Introduction to the Command Line Assignment<br>LAB Setting up Python Anywhere Account |
| 2 | **Collecting/Finding, Evaluating and Utilizing data** publications, open data, digital scholarship (scholarly digital archives, portals, digital scholarship, etc.)<br><br>Reading:<br>Corti Ch. 10 | ● Discovery/search exercise<br>● 4Qs of Data<br>● Heterogeneous Data<br><br>LAB APIs<br>LAB Intro to Basic Data Manipulation with Python |
| 3 | **Project planning/research protocols**<br>Reading: | ● project planning & resource budget<br>LAB Reverse Engineering & |

| | | |
|---|---|---|
| | Geary J[1], Jardine CG, Guebert J, Bubela T. (2013) Access and benefits sharing of genetic resources and associated traditional knowledge in northern Canada: understanding the legal environment and creating effective research agreements. Int J Circumpolar Health. Aug 5;72. doi: 10.3402/ijch.v72i0.21351. eCollection 2013. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23986896# <br> Reverse engineering digital projects: Miriam Posner's How Did They Make That? [blog & video, video focuses on reverse engineering DH project] <br> ● **https://builtwith.com/** | Transparency in Qualitative research & DH projects <br> (anatomy of an NVivo project file) |
| 4 | ● databases <br> ● portal/platforms <br> Reading (tba): <br> ● visual resource review <br> ● Content Management Systems, Digital Asset Management Systems & Virtual Research Environments <br> ● Relational Database basics <br> ● SQL, noSQL | ● conceptual design & database schemas <br> ● understanding 'stacks' & research portal/ platforms <br> ● tradeoffs in using out of the box tools/solutions or developing your own <br> LAB CMS, DAMs and VREs and CAQDAS <br> LAB (potential SQL tutorial, tba) |
| 5 | **Design principles and Requirements Analysis** <br> ● ex. 1 designing digital research archives & exhibits <br> ● ex. 2 creating a data publication <br> Reading: <br> Corti Ch. 11 <br><br> Tibbo, Helen R. (2003) On the Nature and Importance of Archiving in the Digital Age." In Advances in Computing v. 57. Elsevier, 1-67. | ● conceptual design for archives & exhibits <br> ● specs for data publications <br> ● Project Planning & Resource budget Assignment <br> LAB Dig Archives or Qualitative Data Deposit (guest Lab tba) |
| 6 | **Data organization & data storage** <br> Reading: <br> Corti Ch. 1 & 6 | ● file naming conventions <br> ● file systems & inventories, <br> ● workflows and distributed storage <br> ● Data Workflow Assignment <br> LAB automating Data Workflow |
| 7 | **Data management & security plans & data sharing protocols** <br><br> Reading: <br> Corti Ch. 2 & 3 | ● following funder or institutional guidelines to write a data management and/or sharing plan <br> ● ethics case studies |

| | | |
|---|---|---|
| | Borgman, Christine L. (2012) The Conundrum of Sharing Research Data." Journal of the American Society for Information Science and Technology 63(6): 1059–1078. doi:10.1002/asi.22634. Available at: http://onlinelibrary.wiley.com/doi/10.1002/asi.22634/pdf<br><br>Hardy, Lisa, Hughes A, Hulen E, Schwartz AL (2016) Implementing Qualitative Data Management Plans to Ensure Ethical Standards in Multi-Partner Centers. J Empir Res Hum Res Ethics, Apr;11(2):191-8. doi: 10.1177/1556264616636233. Epub 2016 Apr 13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27074911 | LAB implementing privacy & data security protocols |
| 8 | ● Project Workshop I<br>Reading: tba | ● status of project<br>● project workshop exercise<br>LAB (tba) |
| 9 | **Data & Workflow Documentation**<br>● metadata<br>Reading:<br>Corti Ch. 4<br><br>Paul Getty Trust. Introduction to Metadata, Murtha Baca, ed. Available at: http://www.getty.edu/publications/intrometadata/introduction/ | ● code books, readme files<br>● data transformation, annotations & analysis layers<br>● understanding & creating metadata<br>LAB Automating Documentation |
| 10 | **Data Cleaning & Data Quality**<br>Reading:<br>Rawson, Katie and Trevor Muños**.** Against Cleaning. Curating Menus [project] July 7, 2016. Available at: http://curatingmenus.org/articles/against-cleaning/<br><br>Van den Broeck J, Argeseanu Cunningham S, Eeckels R, Herbst K (2005) Data cleaning: Detecting, diagnosing, and editing data abnormalities. PLoS Med 2(10): e267. Available at:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1198040/pdf/pmed.0020267.pdf<br><br>van Hooland, Seth; Verborgh, Ruben; De Wilde, Max. Cleaning Data with OpenRefine, | ● data cleaning<br>● copyright and other legal issues<br>● privacy preserving techniques for qualitative data<br>● implications for data analysis & reporting<br>● provenance<br>● data veracity (quality, completeness, consistency, uncertainty)<br>LAB (tba HathiTrust & web-scraping) |

| | | |
|---|---|---|
| | Programming Historian (tutorial) https://programminghistorian.org/lessons/cleaning-data-with-openrefine<br><br>Weiss, R. Jason and Robert J. Townsend. Using Excel to Clean and Prepare Data<br><br>for Analysis (see Documentation for data cleaning, pg. 96)<br><br>https://www.siop.org/tip/backissues/Jan05/PDF/423_089to096.pdf | |
| 11 | **Data Ownership within Collaborative & Team-based work**<br>Reading:<br>Corti Ch. 9<br><br>Anderson, Jane and Kim Christen. (2013) "Chuck A Copyright On It': Dilemmas of Digital Return and the Possibilities for Traditional Knowledge Licenses and Labels" in After the Return: Digital Repatriation and the Circulation of Indigenous Knowledge, Special Issue of Museum Anthropology Review, 7: 105-126. Available at: https://scholarworks.iu.edu/journals/index.php/mar/issue/view/233<br><br>Roark, Kendall. (2015) "Data Management Plans, Collaborative Research Governance, and the Consent to Share," Re: Thinking Blog, Council on Library and Information Resources (CLIR), Available at: http://connect.clir.org/blogs/kendall-roark/2015/01/15/data-management-plans-collaborative-research-governance-and-the-consent-to-share<br><br>https://www.w3.org/TR/annotation-model/ | ● collaborative agreements<br>● copyright<br>● licensing<br>● traditional knowledge<br>● cultural heritage<br><br>LAB web annotation data model and collaboration<br>or<br>LAB revisiting graph based databases and collaboration |
| 12 | **Long-term Data Access and Preservation**<br><br>Reading:<br>Duerr Ruth, et al. (2004) Challenges in Long-Term Data Stewardship. In Proceedings of the 21st IEEE Conference on Mass Storage | ● file formats, data normalization & long-term preservation<br><br>LAB (tba guest) |

| | | |
|---|---|---|
| | Systems and Technologies. 2004. 101-121. http://storageconference.org/2004/Papers/05-Duerr-a.pdf<br><br>Rosenthal, David S. H. and Daniel L. Vargas. "Distributed Digital Preservation in the Cloud," 8th International Digital Curation Conference (January 2013), http://www.lockss.org/locksswp/wp-content/uploads/2013/01/IDCC2013.pdf | |
| 13 | **Big Data implications for the "Small Tail"** and traditional research practices in arts, humanities & social sciences<br>Reading:<br>Corti Ch. 7 & 8 | ● panel discussion, reading reflection<br>LAB time for projects/troubleshooting |
| 14 | **Critical Approaches to Research Infrastructures:** Reflections on Data Management & Curation (final data curation projects due<br>Reading:<br>Corti Ch. 9 | ● critical engagement with infrastructure / design, reflection/write up<br>LAB time for projects/troubleshooting |
| 15 | **Open Project Presentations**<br>Reading: | ● final project presentations |
| 16 | Not applicable | Exam Week |