Louisiana State University
School of Library and Information Science

LIS 4900: Fundamentals of Data Curation

Fall 2020
Instructor: Seungwon Yang

Catalog Description:
Introduction to definitions, principles, and practices of data curation; data curation lifecycle model; relevant technologies for data repositories.

Office: 272 Coates Hall

Contact information

- Seungwon Yang, Ph.D., Assistant Professor in SLIS & CCT
- Email: seungwonyang@lsu.edu
- Office hours: by appointment

Course Description
Data curation can be thought of as an entire process of management and care for data content, which may have diverse formats adopted in different fields of study. Ensuring long-term preservation, access, shareability, and reuse of datasets is a major concern both in academia and in the industry due to wide-spread, data-driven approaches for research and various service/product developments. In this course you will explore aspects of data curation, including its definition, curation of lifecycle, relevant technologies, data repositories, intellectual property rights, and current practices.

Course Objectives:
Upon satisfactory completion of this course, students should be able to:

- Describe fundamental principles, concepts, and practices of data curation
- Apply data curation lifecycle processes for developing a Data Management Plan (DMP) to manage diverse types of datasets
- Demonstrate the knowledge of legal aspects and application scenarios of data curation

Textbook

- Louise Corti, Veerle Van den Eynden, Libby Bishop, and Matthew Woollard. *Managing and Sharing Research Data: A Guide to Good Practice*. SAGE. 2nd Edition. Published in 2020. ISBN-13: 978-1-5264-6025-7.

Course Topics and Schedule:

- The table below presents weekly topics in data curation which will be covered in this course. The posting of discussion topics, homework assignments, project descriptions,

and their due dates are marked as 'D (discussion)', 'H (homework)', or 'P (project)' (See the 'Keys' section following the table).

- Students are required to complete and submit their homework assignments, discussions, project deliverables by the due dates, unless otherwise specified.
- Homework assignments may involve reading assigned papers, and then creating concept maps to visually present the key concepts and their relationships. Other types of homework will also be assigned as well.

No.	Week	Topic	Posted	Due
1	Aug 27	Introduction to Data Curation	.	.
2	Sep 4	Data Curation and Roles of Libraries	H1	
3	Sep 11	Data Curation Part 1: overview	.	.
4	Sep 18	Data Curation Part 2: Ingest	H2	H1
5	Sep 25	Data Curation Part 3: processing	PD	.
6	Oct 2	Data Curation Part 4: repository management	H3	H2
7	Oct 9	Data Curation Part 5: dissemination	.	P1
8	Oct 16	Datasets, types of data collections, data management plans	H4	H3
9	Oct 23	Metadata, Vocabularies, Ontologies, and Identifiers	D1	H4
10	Oct 30	Policy for Data Management and Use	.	D1
11	Nov 6	**Midterm project deliverables DUE**	.	P2
12	Nov 13	Curation for Big Data	D2	.
13	Nov 20	Intellectual Properties in Data, Legal Aspects	H5	D2
14	Nov 27	Thanksgiving Break (no class)	.	.
15	Dec 4	Data Curation Practices in Science, Engineering & Humanities	.	H5
16	Dec 11	**Final project deliverables DUE**	.	P3

* Keys:

- D = Posting to the Discussion Forum in the Moodle.
- H = Homework assignment.
- P = Project:
    - PD: Project description and timeline posted to Moodle
    - P1: Project proposal document due
    - P2: Mid-term project deliverables due
    - P3: Final project deliverables due

Homework Assignments:

- The homework assignments are designed to help students learn important concepts in the lectures and reading materials, as well as skills for applying software tools. As part of homework assignments, a concept-mapping tool might be used to visually present key concepts and their relationships. Students may discuss together about their assignments, but all of the material that is turned in for grading must be an individual work.

Concept Map

- As part of homework assignments or discussions, a concept-mapping tool will be used to visually present key concepts and their relationships found from the reading material. Students may discuss together about their assignments, but all of the assignments that are turned in for grading must be an *individual work*.
- IHMC CmapTools is a software tool for creating concept maps. You can find more details about the concept maps from this Wikipedia article: (https://en.wikipedia.org/wiki/Concept_map).
- Students are expected to learn how to use CmapTools by themselves. To download CmapTools, visit this link: (https://cmap.ihmc.us/products/), and then select 'CmapTools' icon. In the next page, fill in the blanks and click 'Submit' button to activate the download links on the bottom of the page. Then, you should select the download link that matches your computer's operating system. If you have trouble downloading and installing CmapTools, please contact the instructor.

Grading Scheme

- The course grade and scores for each course requirement will be assigned on a 100-point scale. A letter grade will be assigned according to the following policy:
    - A-: 90-93, A: 94-97, A+: 98-100
    - B-: 80-82, B: 83-86, B+: 87-89
    - C-: 70-72, C: 73-76, C+: 77-79
    - D-: 60-62, D: 63-66, D+: 67-69
    - F: below 60

Grading Policies and Evaluation Methods:

Course grades will be assigned based on homework assignments, participation in the discussion board, and a term project.

Scores on each component will be combined to produce a single overall score for each student as follows:

Component	Percentage
Homework assignments (9% x 5)	45
Discussions (7.5% x 2)	15
P1: Project proposal (report 10%)	10
P2: Midterm project report (14%)	14
P3: Final project report (16%)	16
Total	100

Late Submissions:

- After a due date, the score of the submitted homework/activities will be weighted 10 percent less per each day. For example, if you received 100 points out of 100 from a homework assignment, which was submitted one day past the due date, 90 points will be

recorded as your final score (10% off). If you received 100 points and it was late two days, you will receive 80 points (20% off), etc.

- For assignments and homework submitted 7 or more days after the due date, only 30% of the score will be considered. For example, if you scored 100 points, but it was submitted 3 months after the due date, you receive 30 points.

Description of activities that will be graded

- Discussions:
    - Both undergrad and grad students are expected to participate in discussions (usually posting your thoughts or summaries of given topics) via the Moodle discussion forums:
        - Discussion topic 1: Metadata preparation, ontologies, and semantic Web
        - Discussion topic 2: Examining issues in curating Big Data
- Homework assignments:
    - HW#1: Creating a concept map about data curation
        1. Undergraduate students: identify at least 20 concepts
        2. Graduate students: identify at least 30 concepts
    - HW#2: Comparing the appraisal and ingest processes among various data archives
        1. Undergraduate students: 2 data archives will be assigned
        2. Graduate students: 3 data archives will be assigned
    - HW#3: Comparing repository management practices for multiple data repositories
        1. Undergraduate students: 2 data repositories will be assigned
        2. Graduate students: 3 data repositories will be assigned
    - HW#4: Preparing Data Management Plans (DMPs) given a realistic research scenario
        1. Undergraduate students: 1 research scenario will be assigned
        2. Graduate students: 2 research scenarios will be assigned
    - HW#5: Examining the legal/intellectual property aspects for data in multiple data repositories
        1. Undergraduate students: 2 data repositories will be assigned
        2. Graduate students: 3 data repositories will be assigned

- Term Project
    - There are three phases in the class project. The project should be conducted individually.
        - P1 (project proposal)
        - P2 (midterm report)
        - P3 (final report)
    - Sections to be included in P1: proposal document
        1. Title of the project (1-2 lines, make it succinct yet informative)
        2. Description of the project (minimum 500 words)
        3. References (undergrad: minimum 5; grad: minimum 8)

- Sections to be included in P2: midterm report
  - Undergrad student:
    1. Title of the project
    2. Description of the project
    3. Methodologies (minimum 250 words)
    4. Initial results (minimum 250 words)
    5. References
  - Graduate student:
    1. Title of the project
    2. Description of the project
    3. Methodologies (minimum 250 words)
    4. Initial results (minimum 500 words)
    5. Initial conclusion (minimum 250 words)
    6. References

- Sections to be included in P3: final report
  - Undergrad student:
    1. Title of the project
    2. Description of the project
    3. Methodologies
    4. Results (minimum 750 words)
    5. Conclusion (minimum 250 words)
    6. References
  - Graduate student:
    1. Title of the project
    2. Description of the project
    3. Methodologies
    4. Results (minimum 750 words)
    5. Discussion and implications (minimum 500 words)
    6. Conclusion (minimum 250 words)
    7. References

Expectations

LSU's general policy states that for each credit hour, you (the student) should plan to spend at least two hours working on course related activities outside of class. Since this course is for three credit hours, you should expect to spend a minimum of six hours outside of class each week working on assignments for this course. For more information see: http://catalog.lsu.edu/content.php?catoid=12&navoid=822.

Academic Integrity

Louisiana State University adopted the Commitment to Community in 1995 to set forth guidelines for student behavior both inside and outside of the classroom. The Commitment to Community

charges students to maintain high standards of academic and personal integrity. All students are expected to read and be familiar with the LSU Code of Student Conduct and Commitment to Community, found online at www.lsu.edu/saa. It is your responsibility as a student at LSU to know and understand the academic standards for our community.

Students who are suspected of violating the Code of Conduct will be referred to the office of Student Advocacy & Accountability. For undergraduate students, a first academic violation could result in a zero grade on the assignment or failing the class and disciplinary probation until graduation. For a second academic violation, the result could be suspension from LSU. For graduate students, suspension is the appropriate outcome for the first offense. The most recent version of the Code of Student Conduct is available at http://www.lsu.edu/saa

Information for Students with Disabilities

LSU policy requires a student who claims disability status to make a formal request for accommodation through the Office of Disability Services, 115 Johnston Hall, phone 225-578-5919. This office provides the necessary evaluation and recommendations to ensure full participation in the course. For more information, go to http://www.lsu.edu/disability.

LSU Student Code of Conduct

The LSU student code of conduct explains student rights, excused absences, and what is expected of student behavior. Students are expected to understand this code as described here: http://students.lsu.edu/saa/students/code. Any violations of the LSU student code will be duly reported to the Dean of Students.

Attendance

Policy Statement 22 governs what will be accepted as an excused absence in this course. A student is required to notify the instructor in advance of the deadline if the student is unable to complete the assignment by the deadline due to an excuse consistent with Policy Statement 22. In the event of an emergency, a student must notify the instructor within five days and request an extension for any missed assignments. The instructor reserves the right to request documentation before granting approval for a make-up assignment.

Reading List:

('U' denotes readings for undergraduate students; 'G' denotes readings for graduate students)

Week 1

- [U, G] Chapter 1 of the textbook.
- [U, G] DCC_curation_lifecycle_model(Higgins,2008).pdf Abstract link: http://ijdc.net/index.php/ijdc/article/view/69
- [G] Digital_curation_for_science,digital_libraries,individuals(Beagrie,2006).pdf Abstract link: http://ijdc.net/index.php/ijdc/article/view/6
- [G] DigCurV_curriculum_framework(MolloyandGow,2014).pdf Abstract link: http://ijdc.net/index.php/ijdc/article/view/9.1.231

Week 2

- [U, G] Heidorn, P. B. (2011). The emerging role of libraries in data curation and e-science. *Journal of Library Administration, 51*, 662-672. doi:10.1080/01930826.2011.601269
- [U, G] Witt, M., Carlson, J., Brandt, D. S., & Cragin, M. H. (2009). Constructing data curation profiles. The International Journal of Digital Curation, 3(4), 93-103.doi:10.2218/ijdc.v4i3.117
- [G: up to page 40] Lord, P., & Macdonald, A. (2003). *E-Science curation report: Data curation for e-Science in the UK: An audit to establish requirements for future curation and provision. Bristol, UK: JISC.*
- (G: Optional) Anderson, W. L. (2004). Some challenges and issues in managing, and preserving access to, long-lived collections of digital scientific and technical data. Data Science Journal, 3, 191-202.

Week 3

- [U, G] Chapter 2 of the textbook.
- [G] Guide to social science data preparation and archiving: best practice throughout the data lifecycle (5th edition), ICPSR (2012)
- [G] Akmon, D., Zimmerman, A., Daniels, M. and Hedstrom, M. (2011). The application of archival concepts to a data-intensive environment: working with scientists to understand data management and preservation needs. File in Moodle: Reading_datamgmt(Akmon, 2011).pdf

Week 4

- [U, G] UK Data Service Appraisal Criteria and Grid. http://ukdataservice.ac.uk/media/455175/cd234-collections-appraisal.pdf
- [G] The Selection, Appraisal, and Retention of Social Science Data. Gutmann et al. Data Science Journal (2004) 3:209-221. http://www.digitalpreservation.gov/partners/documents/data-pass_selection_ data.pdf

Week 5

- [U, G] Chapters 4-6 of the textbook.
- [G] ICPSR Guide to Social Science Data Preparation and Archiving (PDF) http://www.icpsr.umich.edu/files/deposit/dataprep.pdf
- [G] How We Curate Data: Our Quality Control, UK Data Archive (short webpage) http://data-archive.ac.uk/curate/archive-quality See the level-based approach at UKDA.
- [G] Review Procedures for UK Data Service ReShare self-deposit data repository. http://reshare.ukdataservice.ac.uk/reshare-review-procedures/

Week 6

- [U] Core Trust Seal. https://www.coretrustseal.org/why-certification/
- [G] Digital Repository Audit Method Based on Risk Assessment (DRAMBORA), Digital Curation Centre & Digital Preservation Europe. http://www.repositoryaudit.eu/img/drambora_flyer.pdf
- [G] Trusted Repositories Audit and Certification (TRAC): Criteria and Checklist, Center for Research Libraries (2007). Read pages 1-49. http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Week 7

- [U, G] UK Data Service user advice and training http://ukdataservice.ac.uk/use-data/advice
- [U, G] UK Data Service case studies of reuse http://ukdataservice.ac.uk/use-data/data-in-use/case-studies
- [G] UK Data Service teaching with data http://ukdataservice.ac.uk/use-data/teaching
- [G] UK Data Service video tutorials http://ukdataservice.ac.uk/use-data/tutorials
- [U, G] Teaching and learning with ICPSR https://www.icpsr.umich.edu/icpsrweb/instructors/index.jsp
- [G] ICPSR Resources for students https://www.icpsr.umich.edu/icpsrweb/ICPSR/support/students
- [G] ICPSR YouTube Channel for user support, educational webinars, and official meetings http://www.youtube.com/user/icpsrweb

Week 8

- [U, G] Chapter 3 of the textbook.
- [U, G] Digital Curation Centre: Data Management Plans. HTTP://WWW.DCC.AC.UK/RESOURCES/DATA-MANAGEMENT-PLANS
- [G] Video Recording: M. Mayernik, "The Case for Data Stewardship: Preserving the Scientific Record" http://commons.esipfed.org /node/699

Week 9

- [U, G] Elings, M. W., & Waibel, G. (2007). Metadata for all: Descriptive standards and metadata sharing across libraries, archives, and museums. First Monday, 12, 3-5. Retrieved from http://firstmonday.org/article/view/1628/1543
- [G] Gilliland, A. J. (2008). Setting the stage. In M. Baca (Ed.), Introduction to Metadata. Los Angeles, CA: Getty Research Institute. Retrieved from http://www.getty.edu/research/publications/electronic_publications/intrometadata/setting.html
- [G] Hillmann, D. (2005). *Using Dublin Core*. Retrieved from http://wiki.dublincore.org/index.php/User_Guide

Week 10

- [U, G] Chapters 9-10 of the textbook.
- [G] Dietrich, Dianne, Trisha Adamus, Alison Miner, and Gail Steinhart. "De-mystifying the Data Management Requirements of Research Funders." Issues in Science and Technology Librarianship no. Summer (2012). doi:10.5062/F44M92G2.
- [G] Hswe, Patricia, and Ann Holt. "Joining in the Enterprise of Response in the Wake of the NSF Data Management Planning Requirement." Research Library Issues, no. 274 (2011): 11-17. http://publications.arl.org/rli274/12
- [G] (Video) Haywood, Jeff. Research Data Management Policies. Edinburgh, UK: University of Edinburgh. Accessed October 3, 2013. http://youtu.be/V8IIdfBAr_0

Week 11

- [U, G. Read up to page 40] Bollier, D. "The Promise and Peril of Big Data." https://www.emc.com/collateral/analyst-reports/10334-ar-promise-peril-of-big-data.pdf (11 July 2011).
- [U, G] Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute, May 2011. https://mck.co/2wLmJJZ
- [G] Boyd, Danah and Kate Crawford, "Critical Questions for Big Data," Information, Communication, and Society 15, number 5 (2012), 662-679. https://people.cs.kuleuven.be/~bettina.berendt/teaching/ViennaDH15/boyd_crawford_2012.pdf

Week 12

- No readings – Midterm project deliverables due

Week 13

- [U, G] Chapters 7, 8, and 11 of the textbook.
- [G] Task Group on Data Citation Standards and Practices, C. I. (2013). Out of cite, out of mind: The current state of practice, policy, and technology for the citation of data. Data Science Journal, 12(0), CIDCR1-CIDCR75.

- [G] AURA/NOAO Data Rights Policy. (2013). http://www.noao.edu/noaoprop/help/datarights.html

Week 14

- Thanksgiving Break – no readings

Week 15

- [U, G] Inter-University Consortium for Political and Social Research. (2012). Guide to Archiving Social Science Data for Institutional Repositories (1st ed.). http://www.icpsr.umich.edu/files/datamanagement/guide-for-irs.pdf
- [U, G] Vardigan, M., & Whiteman, C. (2007). ICPSR meets OAIS: Applying the OAIS Reference Model to the social science archive context. Retrieved from https://deepblue.lib.umich.edu/bitstream/handle/2027.42/60440/Vardigan.Whiteman.Applying%20OAIS.pdf?sequence=1&isAllowed=y
- [G] Borgman, C. L. (2009). The digital future is now: A call to action for the humanities. Digital Humanities Quarterly, 3. Retrieved from http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html

Week 16

- No readings – Final project deliverables due