TCD RESEARCH METHODS: DATA COLLECTION & MANAGEMENT (LAS 6292)

Spring 2021 | Class No. 26033

INSTRUCTOR: Dr. Emilio M. Bruna (he/him) [website]

CONTACT INFORMATION: Phone: (352) 846-0634 | email: embruna@ufl.edu | Twitter: @BrunaLab

CLASS SESSIONS: Friday, Periods 6-8 (12:50-3:50) in Grinter 376 & online

OFFICE HOURS: Monday and Wednesday 10:30 am -12:00 pm via zoom. You can guarantee a specific time slot by scheduling online at http://brunalab.org/teaching/office-hours. You can email me to arrange a more convenient time to meet.

OBJECTIVES: This course is designed for graduate students from any discipline – social sciences, humanities, biophysical sciences – and at all stages of their graduate program. It is an introduction to methods for collecting, organizing, managing, and visualizing both qualitative and quantitative data. Students will gain hands-on experience with best practices and tools. ***At the conclusion of this course students will be able to***:

1. Describe the different types of research data;
2. Explain the need for and benefits of data management and sharing;
3. Describe and implement best practices for the collection, storage, management, archiving, and sharing of research data;
4. Find, download, and analyze publicly available data from repositories;
5. Carry out simple and reproducible data corrections and dataset organization;
6. Describe public policies and agency requirements for data management and sharing;
7. Articulate the major legal and ethical considerations regarding the collection, use, and storage of research data (e.g., privacy/human subjects, intellectual property;
8. Create and Implement a Data Management Plan;
9. Identify and properly use tools for more efficient and secure data collection in the field.

COURSE FORMAT: I believe there is no better way to learn than by doing, which is why this course is taught (mostly) in a 'flipped-course' format. Students are expected to complete each week's assigned reading and watch the short video lectures prior class. The class session will typically include an opportunity for students to ask questions about the pre-class materials and for the instructor to briefly summarize material or demonstrate challenging concepts; occasionally there will be a class discussion about the assigned reading. However, most of the class session will be spent working in small groups on exercises that reinforce that week's concepts and techniques. Throughout the session I will be circulating between groups to assist with the assignment, work though mistakes, and discuss how the techniques can be applied to each student's research.

COURSE MATERIALS:

1. **Required Textbook:** *None*. All readings will be provided on the course Canvas page [https://elearning.ufl.edu].
2. **Course Web Page:** All course materials, communication, and assignment submission will be via Canvas [https://elearning.ufl.edu].

COURSE OUTLINE & CRITICAL DATES

Week	Dates	Topic
1	1/15	'Data' across disciplines and The Research Data Lifecycle
2	1/22	File Formats, naming conventions, data storage, & security
3	1/29	Structure & format of Data & Datasets
4	2/5	Reproducible data (re)organization
5	2/12	Data validation & correction 1
6	2/19	Data validation & correction 2
7	2/26	Documentation: Metadata, Codebooks
8	3/5	Data Management Plans
9	3/12	Efficient data collection
10	3/19	Transcription & Translation
11	3/26	'Paperless' data collection
12	4/2	Automated data extraction
13	4/9	Legal and Ethical Issues
14	4/16	Data Sharing, Reuse, & Archives
15	4/23	Reading Days – no class
Final	4/28	Submission of Final Projects by 5 pm

ASSIGNMENTS & EVALUATION

Assignments: The course grade is based on the completion of the following assignments:

1. Weekly in-class exercises: 14 x 50 each = 700 total
2. Individual data management project: 800 points
3. **Course total: 1500 points**

Due Dates:

1. **In-class Assignments:** Most of the in-class assignments involve hands-on practice with data collection or manipulation. In some weeks, however, assignment will be the submission of questions for group discussion or brief reflection on the issues from the readings. Most in-class assignments are designed to be completed during the class session, but to ensure students master the concepts rather than rush through them *they can be submitted anytime until 9 am the following Monday*.
2. **Individual Project:** due no later than 5 pm on the scheduled date of the final exam.

Evaluation: Final Grades will be assigned on the following scale (based on percentage of points out of the total): **A:** >93% **A-:** 92-90% **B+:** 89-87% **B:** 86-83% **B-:** 82-80% **C+:** 79-77% **C:** 76-73% **C-:** 72-70% **D+:** 69-67% **D-:** 62-60% **E:** 59% and below.

Absences and Make-Up Work: Requirements for class attendance and coursework are consistent with university policies found at: https://catalog.ufl.edu/graduate/regulations/#text. ***Let me know as soon as possible if you need to miss an assignment due date for any reason (e.g., conference, illness, family emergency).***

Regrades: Requests for re-evaluation of assignment grades will require a letter explaining why you think you deserve additional credit and how many additional points. The deadline for submitting these requests is **one week** after the work has been returned.

UF ACADEMIC POLICIES & SUPPORT SERVICES

- **Academic Honesty:** As a student at the University of Florida, you have committed yourself to uphold the Honor Code, which includes the following pledge: *"We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honesty and integrity."* You are expected to exhibit behavior consistent with this commitment to the UF academic community, and on all work submitted for credit at the University of Florida, the following pledge is either required or implied: *"On my honor, I have neither given nor received unauthorized aid in doing this assignment".* It is assumed that you will complete all work independently in each course unless the instructor provides explicit permission for you to collaborate on course tasks (e.g. assignments, papers, quizzes, exams). Furthermore, as part of your obligation to uphold the Honor Code, you should report any condition that facilitates academic misconduct to appropriate personnel. It is your individual responsibility to know and comply with all university policies and procedures regarding academic integrity and the Student Honor Code. Violations of the Honor Code at the University of Florida will not be tolerated. Violations will be reported to the Dean of Students Office for consideration of disciplinary action. For more information regarding the Student Honor Code, please see: http://www.dso.ufl.edu/SCCR/honorcodes/honorcode.php.

- **Services for Students Requiring Accommodations:** The Disability Resource Center (0001 Reid Hall, 352-392-8565, www.dso.ufl.edu/drc/) coordinates the needed accommodations of students with disabilities. This includes registering disabilities, recommending academic accommodations within the classroom, accessing special adaptive computer equipment, providing interpretation services and mediating faculty-student disability related issues. ***I want you to do well in this class and will gladly work with you to implement any neccessary accommodations.***

- **Course Evaluation:** Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Students will be notified when the evaluation period opens, and can complete evaluations through the email they receive from GatorEvals, in their Canvas course menu under GatorEvals, or via www.ufl.bluera.com/ufl/. Summaries of course evaluation results are available to students here.

- **Student Privacy:** There are federal laws protecting your privacy with regards to grades earned in courses and on individual assignments. For more information, please see the Notification to Students of FERPA Rights.

- **Campus Helping Resources:** Students experiencing crises or personal problems that interfere with their general well-being are encouraged to utilize the university's counseling resources. The Counseling & Wellness Center provides confidential counseling services at no cost for currently enrolled students. Resources are available on campus for students having personal problems or lacking clear career or academic goals, which interfere with

their academic performance.

- **Food insecurity**. The Pantry is a resource on the University of Florida campus committed to eradicate food insecurity (https://pantry.fieldandfork.ufl.edu/). Food insecurity is not having a reliable access to nutritious foods for yourself on a regular basis. If you, or anyone you know is experiencing food insecurity, the Pantry is a resource to visit.
- **Software Use**: All faculty, staff, and students of the university are required and expected to obey the laws and legal agreements governing software use. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against university policies and rules, disciplinary action will be taken as appropriate.
- ***Sexual harassment of any kind will not be tolerated in this course:*** UF provides an educational and working environment for its students, faculty, and staff that is free from sex discrimination and sexual harassment. For more about UF policies regarding harassment: http://www.dso.ufl.edu/studentguide/studentconductcode.php#s4041

HEALTH & WELLNESS

- **U Matter, We Care:** If you or a friend is in distress, please contact umatter@ufl.edu or 352 392-1575 so that a team member can reach out to the student.
- **University Counseling & Wellness Center**: 3190 Radio Road, (352) 392-1575, www.counseling.ufl.edu/cwc/ (Counseling Services, Self-Help Library, Groups and Workshops, Training Programs, Outreach and Consultation, Community Provider Database).
- **Sexual Assault Recovery Services (SARS):** Student Health Care Center, 392-1161.

ACADEMIC RESOURCES

- **E-learning technical support:** 352-392-4357 (select option 2) or e-mail to Learning-support@ufl.edu.
- **Career Resource Center:** Reitz Union, 352-392-1601. www.crc.ufl.edu/. Career assistance and counseling.
- **Library Support:** https://uflib.ufl.edu/find/ask/ Various ways to receive assistance with respect to using the libraries or finding resources.
- **Teaching Center:** Broward Hall, 392-2010 or 392-6420. https://teachingcenter.ufl.edu/ General study skills and tutoring.
- **Writing Studio:** 302 Tigert Hall, 846-1138. https://writing.ufl.edu/writing-studio/ Help brainstorming, formatting, and writing papers.

WEEKLY READING ASSIGNMENTS

subject to change based on the research interests of students enrolled in the course

Week 2: File Formats, Naming Conventions, Data Storage, and Data Security

1. Jan Čurn. 2014. How a bug in Dropbox permanently deleted my 8000 photos.
   [read online] [download pdf]
2. PSA: Scrivener, Data Integrity and You. Or, How To Avoid Data Loss Heartbreak.
   [read online] [download pdf]
3. Hart EM et al. (2016) Ten Simple Rules for Digital Data Storage. PLoS Comput Biol 12(10): e1005097.
   [read online] [download pdf]

Week 3: Data structure & format of data and datasets

1. Tesi, W. 2020. An Outdated Version of Excel Led the U.K. to Undercount COVID-19 Cases. Slate. [read online] [download pdf]
2. Stolberg et al. 2020. CDC Test Counting Error Leaves Epidemiologists 'Really Baffled'. NY Times.
   [read online] [download pdf]
3. Broman, K. W., & Woo, K. H. (2018). Data organization in spreadsheets. The American Statistician, 72(1), 2-10.
   [read online] [download pdf]
4. Johnson, B. D., Dunlap, E., & Benoit, E. (2010). Organizing "mountains of words" for data analysis, both qualitative and quantitative. Substance Use & Misuse, 45(5), 648-70.
   [read online] [download pdf]

Week 4: Reproducible data (re)organization

1. Laskowski, 2020. What to do when you don't trust your data anymore.
   [read online] [download pdf]
2. Pennisi, E. 2020. Spider biologist denies suspicions of widespread data fraud in his animal personality research. Science.
   [read online] [download pdf]
3. Lewis, Keith P., Eric Vander Wal, and David A. Fifield. 2018. Wildlife biology, big data, and reproducible research. Wildlife Society Bulletin 42(1): 172-179.
   [read online] [download pdf]
4. Wilson G, Bryan J, Cranston K, Kitzes J, Nederbragt L, Teal TK (2017) Good enough practices in scientific computing. PLoS Comput Biol 13(6): e1005510.
   [read online] [download pdf]

Week 5: QA/QC: Data validation & correction 1

1. Kamentez, A. 2018. The School Shootings that weren't.NPR
[read online] [download pdf]

2. Lincoln, Matthew D. 2018. "Best Practices for Using Google Sheets in Your Data Project."
[read online] [download pdf]

Week 6: QA/QC: Data validation & correction 2

- Guest Instructor: Dr. Hao Ye, UF Reproducibility Librarian
- *No readings this week.*

Week 7: Documentation: Metadata, Codebooks

1. Michener, W.K., et al . 1997. Non-geospatial metadata for the ecological sciences. Ecological Applications 7: 330–342.
[read online] [download pdf]

2. Pp 446-450 in Bernard, H.R. and Bernard, H.R., 2013. Social research methods: Qualitative and quantitative approaches. Sage.
[[download pdf]]

Week 8: Data Management Plans

1. Michener, W. K. (2015). Ten simple rules for creating a good data management plan. PLoS computational biology, 11(10), e1004525.
[read online] [download pdf]

2. Goodman A, et al. (2014) Ten Simple Rules for the Care and Feeding of Scientific Data. PLoS Comput Biol 10(4): e1003542. this paper connects what we've done with what's next
[read online] [download pdf]

Week 9: Efficient data collection

1. Redman, T. 2016. Bad Data Costs the U.S. $3 Trillion Per Year. Harvard Business Review.
[read online] [download pdf]

Week 10: Transcription & Translation

1. Bakker, Rebecca. "Transcription Tools and Software" (2017). Works of the FIU Libraries. 62.
[read online] [download pdf]

2. Watch "The Text Wash team discusses text anonymization" [link]; related blog post [link]

Week 11: 'Paperless' data collection

1. Aanensen DM, Huntley DM, Feil EJ, al-Own F, Spratt BG (2009) EpiCollect: Linking Smartphones to Web Applications for Epidemiology, Ecology and Community Data Collection. PLoS ONE 4(9): e6968.
[read online] [download pdf]

2. Moylan, CA et al. 2013. Increasingly mobile: How new technologies can enhance qualitative research. Qualitative social work: research and practice, 14(1):36-47.
[read online] [download pdf]

3. Teacher, Amber G. F. et al. Smartphones in ecology and evolution: a guide for the apprehensive. Ecology and Evolution 3(16):5268– 5278
[read online] [download pdf]

Week 12: Automated data extraction

1. Drinkwater, R. E., Cubey, R. W., & Haston, E. M. (2014). The use of Optical Character Recognition (OCR) in the digitization of herbarium specimen labels. PhytoKeys, (38), 15-30.
[read online] [download pdf]

2. Joo, Jungseock and Zachary C. Steinert-Threlkeld. 2019. Image as data: Automated visual content analysis for social science.
[read online] [download pdf]

Week 13: Legal and Ethical Issues

1. de Koning M, Meyer B, Moors A, Pels P. 2019. Guidelines for anthropological research: Data management, ethics, and integrity. Ethnography. 20(2):170-174.
[read online] [download pdf]

2. Katie Fortney, UC Office of Scholarly Communication. 2016. “Who owns your data?”
[read online] [download pdf]

3. Goodluck, K. 2020. Indigenous data sovereignty shakes up research. High Country News.
[read online] [[download pdf]]

4. Woodbury, R. B., Beans, J. A., Hiratsuka, V. Y., & Burke, W. (2019). Data Management in Health-Related Research Involving Indigenous Communities in the United States and Canada: A Scoping Review. Frontiers in genetics, 10, 942.
[read online] [download pdf]

Week 14: Data Sharing, Reuse, & Archives

1. Alexander, S.M., Jones, K., Bennett, N.J. et al. Qualitative data sharing and synthesis for sustainability science. Nat Sustain 3, 81–88 (2020).
[read online] [download pdf]

2. Renaut, S. et al. 2018. Management, Archiving, and Sharing for Biologists and the Role of Research Institutions in the Technology-Oriented Age. *BioScience* 68(6)400–411
[read online] [download pdf]

3. Duke, Clifford S., and John H. Porter. 2013. “The ethics of data sharing and reuse in biology.” BioScience 63(6): 483-489.
[read online] [download pdf]

4. Tenopir, C., et al. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. PloS one, 10(8), e0134826.
[read online] [download pdf]

5. Mauthner, NS, & O. Parry O. 2009. Qualitative data preservation and sharing in the social sciences: On whose philosophical terms?. Australian J of Social Issues 44(3):291-307.
[read online] [download pdf]