

Data Collection & Management

LAS 6292 Spring 20024



Instructor

Dr. Emilio M. Bruna
embruna@ufl.edu
(352) 846-0634

Class Sessions

Fridays 12:50-3:50
Grinter Hall 376

COURSE OVERVIEW

This course is a practical introduction to methods, tools, and best practices for collecting, organizing, managing, and visualizing qualitative and quantitative data. It is designed for graduate students from all disciplines at any stage of their program. At the conclusion of the 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 data set organization;
6. Describe public policies and agency requirements for data management and sharing;
7. Articulate the major legal and ethical considerations regarding data collection, use, and storage (e.g., privacy/human subjects, intellectual property, international law);
8. Create and Implement Data Management Plan in funder-specific formats;
9. Identify and properly use tools for more efficient and secure data collection in the field.

Course Format: This course is taught (mostly) as an active-learning workshop. Students are expected to complete reading or watch some short video lectures prior class. The in-class session will typically include an opportunity to ask questions about the pre-class materials and for me to demonstrate challenging concepts; occasionally there will be a class discussion about the reading. Most of the session, however, will be spent working - sometimes in groups, sometimes individually - on exercises that reinforce the session's concepts and techniques. During class I will be circulating between groups to assist with the assignment, work through mistakes, and discuss how the techniques can be applied to your research.



Required Course Materials

Students are not required to purchase any textbooks or course materials. All class materials & assignments are on the class website (<https://las6292.netlify.app/>). Some assigned readings may come from the *New York Times* and have dynamic multimedia data visualizations and video that can't be appreciated in the posted .pdf format. *Students in this class should sign up for free online access to the New York Times* by following the instructions at [this UF Libraries Website](#).

Materials and Supplies Fees: None.



Instructor Office Hours

Time and Signup: Wednesday & Thursday 10:30-11:45 am (in-person & online). Drop by anytime or sign up for a specific time here: <https://embruna.youcanbook.me>. If you can't make it these days/times please let us know - we will find a time to meet that works for you.

- **Location - in-person:** The Tropical Ecology & Conservation Lab is located next to the Rawlings Hall bus stop (711 Newell Drive; to find a map click the "Contact" link at [BrunaLab.org](https://embruna.youcanbook.me)).
- **Location - online:** use the zoom link on the course Canvas page. I am online the entire session.

COURSE CALENDAR, ASSIGNMENTS, & GRADES

Week	Dates	Topic
1	1/12	'Data' across disciplines and the Research Data Life Cycle
2	1/19	File Formats, Naming Conventions, Data Storage & Security
3	1/26	Structure & Format of Data & Datasets
4	2/2	Reproducible Data (Re)organization
5	2/9	QA/AC 1: Data Entry & Validation
6	2/16	QA/QC 2: Correction & Synthesis with Open Refine
7	2/23	QA/QC 3: Visualizing Data (to Find Mistakes)
8	3/1	Documentation: Metadata & Codebooks; Data Sharing, Reuse, & Archives
9	3/8	Data Management Plans
10	3/15	No Class - Spring Break
11	3/22	Efficient Data Collection
12	3/29	Transcription & Translation
13	4/5	Paperless Data Collection
14	4/12	Automated Data Extraction
15	4/19	Legal & Ethical Issues
Finals Week	4/30	Submission of Final Projects by 5 pm



Final Course Grades

The final course grade is based on the assignments below. 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. Note also that 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 before 9 am the day of the following class session*. Late assignments will lose 10 pts. Finally, **there is no Final Exam in this course, and the Individual Project is due on the scheduled date of the final exam.**

Assignment	Points (%)	Due
Weekly in-class exercises	350 (35%)	Following Friday at 9 am
Data Management Plan	150 (15%)	23 March
Individual Data Cleanup Project	500 (50%)	30 April
TOTAL	1000 (100%)	

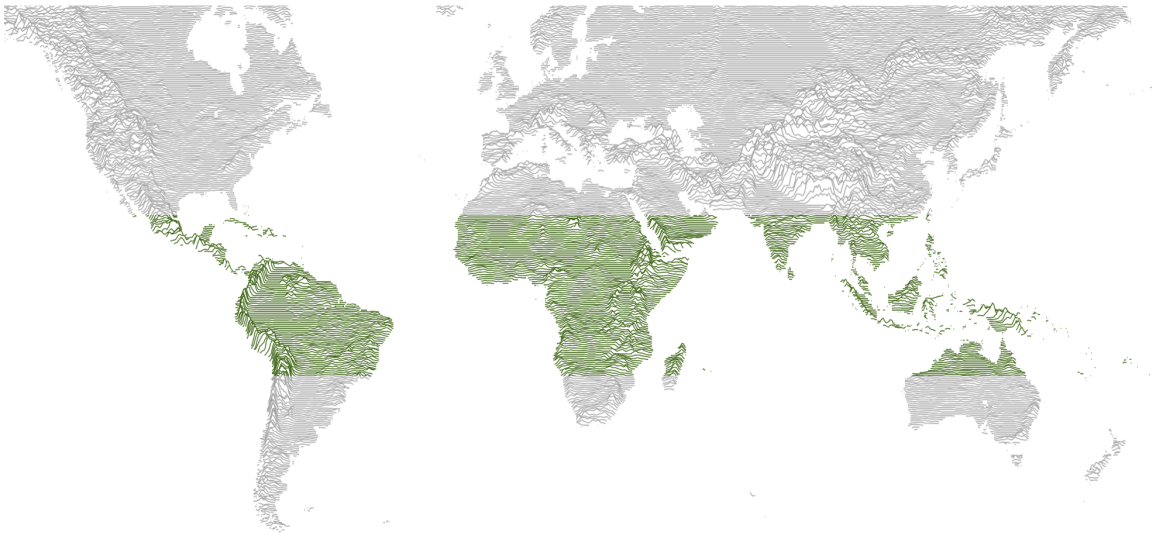
Regrades: Requests for re-evaluation of assignments must be accompanied by an explanation for why you think you deserve additional credit and the number of additional points you think you deserve. The deadline for submission is one week after the work was returned.












Final Course Grades

Course grades will be based on the percentage of total points earned and will be assigned with the following scale: **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.

Grade Points: For information on how UF assigns grade points, visit:
<<https://catalog.ufl.edu/UGRD/academic-regulations/grades-grading-policies/>>



FREQUENTLY ASKED QUESTIONS

	What is the best way to contact the instructors?	Email sent via Canvas That's also how we will respond.
	Can you give me one good reason why I should go to Office Hours?	I can give you ten. (1) To introduce yourself. (2) To get clarification on assignments. (3) To discuss topics that came up in class. (4) There is free tea, coffee, or espresso in our lab kitchen. (5) You can check to make sure you understood the key points from a class session. (6) We can give you feedback on ideas for course projects. (7) Get advice on successfully navigating college. (8) Ask questions about how to gain experience for your post-graduation goals. (9) To get help arranging a study group. (10) You don't need a good reason...just come on by.
	How will you send announcements to the class?	Canvas! Check the course Canvas page for announcements and be sure you are receiving Canvas emails and updates.
	What work should we do <i>before</i> class?	Read, watch, listen to, or review all materials assigned for the session. This material will set the stage for the in-class activities.
	What will we do <i>during</i> class? When is 'in-class' work due?	In-class exercises that reinforce key concepts, discussions of the assigned readings, and let you practice skills in other assignments. Some are completed individually, while others require working in groups or pairs. Each activity will have instructions and a rubric; most are designed to be finished in class. <i>In-class work is due one week from the date it was assigned.</i>
	What if I miss class?	Attendance is not required, but in many of the sessions we will be completing activities in class that count towards your grade. Most of these can be completed independently, but by doing them in class you will benefit from working collaboratively with the other students. Some of the in-class activities, however, can not be completed on your own. That's why only a subset of the in-class assignments count towards your grade and we offer extra credit.
	I know I will miss class on a certain date. What should I do?	Let us know as soon as possible so we can make arrangements for you to review material you will miss and complete assignments.
	Class discussions are difficult for me. Will this affect my grade?	No! If there are issues that make engaging in discussions difficult (e.g., shyness, language barriers, a medical condition), let us know and we will find alternative modes of participation.
	I have no child care today. My kid's school was cancelled. My partner can't stay with our baby. What should I do? Can I bring my kid to class?	UF does not have a policy on children in the classroom; the following is my commitment to student-parents. I never want students to feel they have to choose between feeding their baby and their education. <i>You and your nursing baby are welcome in class anytime</i> ; I also don't want parents to feel they must chose between attending class or staying home to care for kids, so <i>occasionally bringing your kid to class because of a gap in child care is also acceptable</i> . If you do bring your child to class, please sit close to the exit so that you can more easily step outside if you need to care for them (and so other students can continue learning). Non-parents in the class: please help by offering your seat near a door to parents. Finally, I understand that balancing school, childcare and perhaps also a job can be exhausting and a barrier to learning. I maintain the same high standards for all students in my class, and will work with you to find ways to balance your responsibilities as a parent and student. Let me know how I can help.

UF RESOURCES FOR STUDENTS

Health, Safety, & Wellness

Wellness and Mental Health: 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.

- *University Counseling & Wellness Center*, 3190 Radio Road, 352-392-1575. They can provide Counseling Services, Groups and Workshops, Outreach and Consultation, a Self-Help Library, and Wellness Coaching. <http://www.counseling.ufl.edu/>.
- *U Matter, We Care*. If you or someone you know is in distress, please contact umatter@ufl.edu, 352-392-1575, or visit U Matter, We Care website (www.umatter.ufl.edu/) to refer or report a concern and a team member will reach out to the student in distress.

Student Health Care Center: Call 352-392-1161 for 24/7 information to help you find the care you need, or visit the Student Health Care Center website.

University Police Department: Visit UF Police Department website or call 352-392-1111 (or 9-1-1 for emergencies).

UF Health Shands Emergency Room / Trauma Center: For immediate medical care call 352- 733-0111 or go to the emergency room at 1515 SW Archer Road, Gainesville, FL 32608; Visit the UF Health Emergency Room and Trauma Center website.

Field and Fork Pantry: The Hitchcock Pantry can provide food and toiletries for students experiencing food insecurity. <https://pantry.fieldandfork.ufl.edu/>.

Academic Services

The Writing Studio: The writing studio is committed to helping University of Florida students meet their academic and professional goals by becoming better writers. Visit the writing studio online at <http://writing.ufl.edu/writing-studio/> or in 2215 Turlington Hall for one-on-one consultations and workshops.

Career Connections Center: First Floor JWRU, 392-1601, <https://career.ufl.edu/>.

Student Success Initiative: <http://studentsuccess.ufl.edu>.

Student Complaints:

- for Residential Courses: <https://sccr.dso.ufl.edu/policies/student-honor-code-studentconduct-code/>.
- for Online Courses: <https://pfs.tnt.aa.ufl.edu/state-authorization-status/#student-complaint>

UF POLICIES



STUDENT ACCOMMODATIONS: Students with disabilities or learning barriers that would like to request academic accommodations should connect with the Disability Resource Center by visiting <https://disability.ufl.edu/students/get-started/>. Please share your letter with me and discuss access needs as early as possible in the semester so that I can do whatever is necessary to ensure your participation and learning.



COURSE EVALUATIONS: Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online via GatorEvals. Guidance on how to give feedback in a professional and respectful manner is available at <https://gatorevals.aa.ufl.edu/students/>. 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 <https://ufl.bluer.com/ufl>. Summaries of course evaluation results are available to students at <https://gatorevals.aa.ufl.edu/public-results/>.



UNIVERSITY HONESTY POLICY: UF students are bound by The Honor Pledge which states, “We, the members of the University of Florida community, pledge to hold ourselves and our peers to the highest standards of honor and integrity by abiding by the Honor Code. On all work submitted for credit by students 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. The Honor Code (<https://www.dso.ufl.edu/sccr/process/student-conduct-honor-code/>) specifies a number of behaviors that are in violation of this code and the possible sanctions. Furthermore, you are obligated to report any condition that facilitates academic misconduct to appropriate personnel. If you have questions or concerns, consult with the instructor or TAs.



SOFTWARE USE: All faculty, staff & students are required & expected to obey 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 UF policies & rules, disciplinary action will be taken as appropriate.



ATTENDANCE: Requirements for class attendance & make-up exams, assignments, and other work are consistent with UF policies found at: <https://catalog.ufl.edu/ugrad/current/regulations/info/attendance.aspx>.



IN-CLASS RECORDING: *Students are allowed to record video or audio of class lectures. However, the purposes for which these recordings may be used are strictly controlled. The only allowable purposes are: (1) for personal educational use, (2) in connection with a complaint to the university, or (3) as evidence in, or in preparation for, a criminal or civil proceeding. All other purposes are prohibited. Specifically, students may not publish recorded lectures without the written consent of the instructor. A 'class lecture' is: an educational presentation intended to inform or teach enrolled students about a particular subject, including any instructor-led discussions that form part of the presentation, and delivered by any instructor hired or appointed by the University, or by a guest instructor, as part of a UF course. A class lecture does not include: lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving solely student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session. Publication without permission of the instructor is prohibited. To 'publish' means: to share, transmit, circulate, distribute, or provide access to a recording, regardless of format or medium, to another person (or persons), including but not limited to another student within the same class section. Additionally, a recording, or transcript of a recording, is considered published if it is posted on or uploaded to, in whole or in part, any media platform, including but not limited to social media, book, magazine, newspaper, leaflet, or third-party note/tutoring services. A student who publishes a recording without written consent may be subject to a civil cause of action instituted by a person injured by the publication and/or discipline under UF Regulation 4.040 Student Honor Code and Student Conduct Code.*

Weekly Reading/Viewing

Please review the assigned material before class.

All items are posted on the Canvas site. Topics & readings are subject to change based on current events; changes will be announced via Canvas.

Week 1: Course Introduction

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

Read:

1. Jan Čurn. 2014. How a bug in Dropbox permanently deleted my 8000 photos. [\[read online\]](#) [\[download pdf\]](#)
2. Panzarino, M. 2012. How Pixar's Toy Story 2 was deleted twice, once by technology and again for its own good. TNW. [\[read online\]](#) [\[download pdf\]](#)
3. Hart EM et al. (2016) Ten Simple Rules for Digital Data Storage. PLoS Comput Biol 12(10): e1005097. [\[read online\]](#) [\[download pdf\]](#)

Watch:

1. E. Bruna: [How To Name Your Files](#)
2. E. Bruna: [How to Organize Your Files](#)
3. E. Bruna: [How to Store and Backup Your Files](#)

Week 3: Data structure & format of data and datasets

Read

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

Read:

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. Alston, J. M., and Rick, J. A.. 2020. A Beginner's Guide to Conducting Reproducible Research. Bull Ecol Soc Am 00(00):e01801. [\[read online\]](#) [\[download pdf\]](#)

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. [\[read online\]](#) [\[download pdf\]](#)

Watch:

1. [A Data Sharing and Management Snafu in 3 Short Acts](#), by the NYU Health Sciences Library

Week 5: QA/QC 1: Data Entry & Validation.**Read:**

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 2: Correction & Synthesis with Open Refine.

Read: *No readings this week.*

Week 7: QA/QC 3: Finding Errors with Data Visualization.**Read:**

1. Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLoS Comput Biol 10(9): e1003833. [\[read online\]](#) or [\[download pdf\]](#)

Watch:

1. Tommy McCall: [The simple genius of a good graphic](#)
2. R. Luke DuBois: [Insightful human portraits made from data](#)

Week 8: Documentation: Metadata, Codebooks; Data Sharing, Reuse, & Archives.**Read (Metadata):**

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. [\[\[download pdf\]\]](#)

Read (Data Sharing):

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.
[\[read online\]](#) [\[download pdf\]](#)

Week 9: Data Management Plans.

Read

1. Michener, W. K. (2015). Ten simple rules for creating a good data management plan. PLoS Computational Biology, 11(10), e1004525. [\[read online\]](#) or [\[download pdf\]](#)
2. Goodman A, et al. (2014) Ten Simple Rules for the Care and Feeding of Scientific Data. PLoS Comput Biol 10(4): e1003542. [\[read online\]](#) or [\[download pdf\]](#). this paper connects what we've done with what's coming up next

Watch

1. [Content of a DMP](#)" (by RWTH Aachen University)

Week 10: Spring Break.

Week 11: Efficient data collection.

Read:

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

Watch

1. Atul Gawande's TED Talk: [The Importance & Value of the CHECK LIST](#).
2. Animated summary of [The Checklist Manifesto](#).
3. Jess Stratton's LinkedIn Learning Video Overview: "Use Google Forms to Create Surveys" (7 min.). Watch it [here](#); must be on UF computer or use UF proxy.
4. OPTIONAL: The more advanced, multi-video "Google Forms Essential Training Course". These short (1-3 min) videos explain each step in more detail. It's great, and only 39 min long from start to finish. Watch [here](#); must be on UF computer or use UF proxy.

Week 12: Transcription & Translation.

Read:

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 [\[link\]](#)

Watch

1. Dr. Jarek Kriukow: How to transcribe interviews - [Part 1: "naturalism" and "denaturalism"](#)
2. Dr. Jarek Kriukow: How to transcribe interviews - Part 2: [Which approach to use?](#)

Week 13: 'Paperless' data collection.**Read:**

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 app-rehensive. Ecology and Evolution 3(16):5268– 5278 [\[read online\]](#) [\[download pdf\]](#)

Watch

1. E. Bruna: [Paperless Data Collection](#)

Great EpiCollect tutorials by UConn's David Dickson

1. [Intro to EpiCollect 5](#)
2. [Creating an EpiCollect PROJECT](#)
3. [Creating a FORM for your EpiCollect Project](#)
4. [Collecting data in EpiCollect with your phone or tablet](#)

Week 14: Automated data extraction.**Read:**

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 15: Legal and Ethical Issues.**Read:**

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. [\[read online\]](#) [\[download pdf\]](#)

Watch

1. NIH: %5BStatistical methods for protecting data confidentiality](<https://youtu.be/3vN-6A2zbsk>)
2. Patricia Jenkinson Video Seminar: [Ethics in Human Subject Research](#)
3. US Department of Health & Human Services: [Informed Consent Overview](#)
4. Australian Council for International Development (ACFID): [Principles and Guidelines for Ethical Research and Evaluation in Development](#)