

2 Syllabus for Open Source Software Development



I 320D - Topics in Human-Centered Data Science : Open Source Software Development

AUTHOR

James Howison

2.1 Syllabus

Official, updated version of syllabus always online at:

https://howisonlab.github.io/open_source_software_course/current/oss_syllabus.html

Jump to [Course Outline](#)

Course	I 320D Topics in Human-Centered Data Science : Open Source Software Development
Professor	James Howison
Meeting Time	Mondays and Wednesdays 3:30 PM - 04:45 PM
Location	SZB 4.508
Semester	Spring 2024
Unique No.	27450
Office Hours	By appointment booked at http://james.howison.name/calendar/
Contact Email	jhowison@ischool.utexas.edu

3 Objectives and Assignments

This course explores “open source software development” which is a name for the open collaborations that produce open source software. Open source software is a thing that is built by people, an artifact. But the way that it is developed, the way that people work together to build it is of great interest. “The open source way” is of practical interest for everyone building software, because open source development has lead the creation of ways of working used throughout software work. For social and organizational scholars, “the open source way” is of interest in advancing our theories of how people can work together and how technology matters.

The “open source way” is also known as “peer production” and that way of working extends beyond open source, to places like Wikipedia. Perhaps not coincidentally the Wikipedia page on [Peer Production](#) is quite useful:

A way of producing goods and services that relies on self-organizing communities of individuals. In such

communities, the labor of a large number of people is coordinated towards a shared outcome.

As we learn about open source and peer production we'll learn to distinguish it from different kinds of online collaborations, such as crowdsourcing, citizen science, question and answer sites, and mere sharing of code. Near synonyms for the way of working taught in this course are: "The open source way", "Open collaboration", "Open mass production".

This is a course is about a sociotechnical phenomenon and it takes a sociotechnical approach. In practice this means that we'll be learning both conceptual insight and practical skills. The course weaves together learning how to use key technologies of collaboration (e.g., git, github, travisCI, markdown) at the same time that we learn social and organizational theory about peer production (e.g., the role of copyright licenses, motivations of participants, governance models, coordination theory, models of collaboration risk, cultures of collaboration).

There are no prereqs for this course. While we will be discussing software development students will not be required to program. We will use the command line as we learn to use [git](#) and [github](#) and everything will be covered in class. I try to build [a set of analogies for git and github](#), going beyond teaching the commands to give ways to think about git.

Students will need access to a computer for classes, any version of Windows, Mac, or Linux will do. Students facing difficulties with their IT should contact the iSchool help desk (via help@ischool.utexas.edu) who can refer on to other resources as available.

3.1 Learning Objectives

Students will be able to:

- Understand what open source software is as an artifact
- Understand what is distinctive about the way that it is produced
- Know how to interact with open source software projects
- Know how the open source way has informed modern software development work (including in data science)
- Conceptually distinguish open source from other organizational forms or phenomena, such as corporations, crowdsourcing, open access publishing, and communities of practice.
- Reason about how, why, and when open source peer production works (and when it does not!)
- Engage critically with published research and popular discourse about open source

Practically students will be able to:

- Install and use [git](#) to manage versions in their own work
- Participate in [github](#) hosted peer production (making and receiving pull requests)
- Create and publish documents in [markdown](#) and [wikitext](#) format
- Ask technical questions that people want to answer
- Analyze trace data from open source software projects

3.2 Assessments

Assignment	Percentage of Course Grade	Due Date
Class Participation (discussion and activities)	10%	Throughout course
Technical challenges	40%	Weekly homework challenges throughout the semester
Git Analogies Paper	20%	Start of Spring Break
UT Austin open source projects	5%	Two weeks after Spring Break
Open Source Trace Data analysis Project report	25%	Start of last class week.

There is no final exam for this course. 100-90:A, 90 > grade > 85: A-, 85 > grade > 80 B- and so on.

3.2.1 Participation in class discussion and activities

Students are expected to attend class and to participate in discussion and activities. Students should email the professor prior to class if they cannot make it. Material throughout the course builds on earlier material (both technical and conceptual). If you cannot make class you should refer to the online materials first and then consult with your classmates. Office hours are not for personal replays of teaching, nor can they compensate for not participating in discussion.

Hints on participation:

- Useful participation can come from asking questions if you don't understand the point someone is making. If you have questions, very likely others do to.
- You can also summarize discussion which helps everyone by giving them something to test their understanding of the discussion.
- You can challenge or disagree with people, sometimes that can be introducing a counter-example or questioning whether evidence really means what the speaker originally thought.
- I really welcome examples from your own experience. For example if there is an organization, movement, or line of thought (modern/traditional) in your culture that relates to open source development, I would love that to be part of the discussion.

3.2.2 Technical Challenges

The course will have assignments based on the technical topics we are learning, including use of the DataCamp course (see below).

3.2.3 Papers and Projects

3.2.3.1 Analogies paper

How would you explain what we have been learning about git to your future work teammates?

In this assignment you will describe three different analogies for git-based collaboration (including branching, GitHub, and PRs). No more than two analogies can be found online (use references!), at least one you must create yourself. Each analogy should cover as much of git and GitHub collaboration as possible.

For each, include both text and pictures. You should have approximately 2 pages per analogy (~500-750 words). Pictures can be photos you find online, or diagrams you create yourself (digital collage can be useful). You should include connections between your analogy and the specific git and GitHub commands we have learned (e.g., init, add, commit, push, pull, create PR, merge, etc.). In particular, you must include a discussion of "splitting a pull request" including git cherry-pick. Your analogies should cover multiple commands (not a one analogy for one command); each should cover as much of git and GitHub collaboration as possible.

In a final 1-2 pages of writing (~350-750 words), Compare and contrast your three analogies in reference to this question: How would each analogy help your teammates? What might it make clear, what might it hide or obfuscate? How might using a combination of your analogies help? Could that combination make things more difficult?

Your paper must be submitted as a PR at https://github.com/howisonlab/oss_spring2024_analogies

Include a note at the top of the file telling me which formal citation approach you are using (e.g., APA, MLA, etc). Students are cautioned that as a scholarly paper proper citation and use of sources is required otherwise students will face academic misconduct proceedings. See UTexas materials on Academic Integrity at <https://deanofstudents.utexas.edu/conduct/academicintegrity.php>

GenAI: For this assignment, I give you specific permission to use GenAI systems like ChatGPT. The requirement, however, is that you must include the sequence of prompts you used, your analysis of the analogy provided by the system, discuss how you adapted the analogy for presentation in the paper, and how you added images/diagrams (as well as including this analogy in your compare and contrast).

The structure of the paper should be as follows (a page is ~300-500 words):

```
# Introduction (~ 1-2 paragraphs/100-200 words)

# Analogy 1 (~1 page including images/~500 words)

# Analogy 2 (~1 page including images/~500 words)

# Analogy 3 (~1 page including images/~500 words)

# Compare and Contrast Analogies (1-2 pages of writing, ~350-750 words).

# References (excluded from any word counts).
```

3.2.3.2 Final Paper or project

For this assignment you will research open source projects that involve faculty at UT Austin. You will:

1. Identify at least five projects using a search strategy (and describe your strategy)
2. Describe each project's online profile. How do the projects describe themselves?
3. Identify the main contributors to each repository. Who are they, what role do they have at the university (faculty, research staff, graduate students). How do their websites compare to the websites of other open source projects you have looked at?
4. Use the concepts introduced in this course to compare and contrast how these projects are organized. You should discuss both how they use git and GitHub, and social organizational aspects of the projects.

3.2.3.3 Final Paper or project

Pick one of the following two options:

1. Conduct a trace data analysis that compares at least 3 projects contained in this provided dataset:

https://github.com/howisonlab/transition_augur/raw/main/oss_data.zip

The code used to pull these data is here: https://github.com/howisonlab/transition_augur

The ER diagram for the underlying data is linked in the notebook in that repo.

These are open source projects funded by the US National Science Foundation during the last decade in a program called SI2. The data is collected from GitHub using the augur project from Linux Foundation. The files when unzipped are quite large, commits.csv has over 4.5million rows. You will not be able to work with these files in Excel (for example) but they should work ok with tools like Tableau or Pandas/Matplotlib.

Your analysis should use the trace data collected to address insight concepts that we have discussed during the course, for example: Leadership, Coordination, Motivation, Dependency (including interdependency), Governance, Bias and lack of diversity. You should provide figures (most likely multiple time-series analyses) and you must relate your work to either a discussion in the book <http://producingoss.com/> or a trace data analysis of open source data you find in the peer reviewed literature. One suggested source is the the Mining Software Repositories conference: <http://www.msrf.org/> Your project document should include multiple figures and ~300-500 words of writing. Think the equivalent of about 3 pages.

2. Write a 1,000-1,500 word scholarly comparative analysis of 3 open source projects that connect with your interests. Explain why you chose these three projects to compare. You should include in your discussion topics such as: leadership, collaboration, license choice, ownership, contribution process, how the project seeks to attract contributions, how the project uses collaboration infrastructure. You must include screenshots and you must discuss at least one PR for each project (of course you can discuss more). You should reference literature from the course and connect with the discussion in at least two additional peer reviewed articles. You may find these two review articles a useful place to identify relevant literature: <https://doi.org/10.1145/2089125.2089127> or <https://www.misqresearchcurations.org/blog/2021/12/7/information-systems-development>

Submission:

Submit by making a PR to https://github.com/howisonlab/oss_spring2023_assignments/projects/ Note that it should be a separate PR from other assignment submissions; this means you must use a different branch name for this work.

Include a note at the top of the file telling me which formal citation approach you are using (e.g., APA, MLA, etc). Students are cautioned that as a scholarly paper proper citation and use of sources is required otherwise students will face academic misconduct proceedings. See [“Academic Integrity” presentation](#) if at all unsure.

3.2.4 Late submission policy

The late policy for papers/projects/presentation is a 10% reduction for submissions up to 24 hours after the due time, but zero points after that. You can, of course, contact me if you have an emergency.

3.3 Materials

There are no required texts for this course and no materials to purchase.

Readings, tutorials, and will be provided via pages linked from the class calendar below.

I will enroll the class in [DataCamp](#), giving students free access to the DataCamp courses (including their premium courses). In particular we will be using their interactive course on [git](#) during classes in the first half of the course. I encourage students to explore their other course options during the semester.

4 Course Outline

4.1 Draft Course Schedule

Table below shows classes and topics planned. Each class has both an insight (aka theory, conceptual) and a skills (aka tech, practical) component. These will become links to materials for the class.

Week	First Class			Second Class		
	Day	Module	Topic	Day	Module	Topic
2	Mon Jan 15	No Meeting	No meeting: MLK day	Wed Jan 17	Skills	Syllabus Review
3	Mon Jan 22	Insights	What is open source? Chapter 3	Wed Jan 24	Skills	Paper Planes: Version Control <i>git add, commit via paper planes and tables</i>
4	Mon Jan	Insights	Motivations and Asking questions people want to	Wed Jan	Skills	Git basic workflow Chapter 10

	29		answer Chapter 4	31		<i>git add, commit (locally)</i>
5	Mon Feb 05	Insights	Licenses Chapter 5	Wed Feb 07	Skills	Rewinding work Chapter 11 <i>revert, other undos, checking out old versions, rewriting history</i>
6	Mon Feb 12	Insights	Coordination Chapter 6	Wed Feb 14	Skills	Branching Chapter 12 <i>git checkout, merge.</i>
7	Mon Feb 19	Insights	Governance and decision making Chapter 7	Wed Feb 21	Skills	Sharing and collaborating via github Chapter 13 <i>git clone, remote, push</i>
8	Mon Feb 26	Insights	Knowledge sharing	Wed Feb 28	Skills	Collaboration conflicts and workflows Chapter 14 <i>github fork, pull request, pull upstream</i>
9	Mon Mar 04	No Meeting	No meeting: Project Time	Wed Mar 06	No Meeting	No meeting: Project Time
10	Mon Mar 11	No Meeting	No meeting: Spring Break	Wed Mar 13	No Meeting	No meeting: Spring Break
11	Mon Mar 18	Insights	Bias and lack of diversity Chapter 8	Wed Mar 20	Skills	Splitting PRs Chapter 15 <i>merge, conflicts, mark resolved, cherrypick, collaboration workflows</i>
12	Mon Mar 25	Insights	The stack and the stream Chapter 9	Wed Mar 27	Skills	Rebase Chapter 16 and Managing Secrets Chapter 17 <i>git rebase</i>
13	Mon Apr 01	Insights	Agile and test driven development	Wed Apr 03	Skills	Documentation and Q&A <i>github flavored markdown</i>
14	Mon Apr 08	Insights	Cybersecurity and open source	Wed Apr 10	Skills	Unit Tests <i>Test suites</i>
15	Mon Apr 15	Insights	Open software work in science and research	Wed Apr 17	Skills	Continuous Integration <i>Travis CI integration, Github Actions</i>

16	Mon Apr 22	Insights	Peer production in Art and Culture	Wed Apr 24	Skills	Creating and distributing packages <i>python packages, pypy pinning</i>
17	Mon Apr 29	Insights	NA	NA	NA	NA

4.2 Skills Readings

4.3 Insight Readings

Often I can link directly to websites or PDFs, but sometimes I will provide links to articles in journals etc. You must be able to get the article through the library, generally speaking using the [web VPN](#) is the most convenient approach. Another option that can sometimes work is [UnPaywall](#) which works to find an open access article version when looking at a publisher's page for the article.

5 Policies

5.1 Class Recordings

Class Recordings: Class recordings are reserved only for students in this class for educational purposes and are protected under FERPA. The recordings should not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings. Guidance on public access to class recordings can be found [here](#).

5.2 Academic Integrity

Each student in the course is expected to abide by the University of Texas Honor Code: "As a student of The University of Texas at Austin, I shall abide by the core values of the University and uphold academic integrity." Plagiarism is taken very seriously at UT. Therefore, if you use words or ideas that are not your own (or that you have used in previous class), you must cite your sources. Otherwise you will be guilty of plagiarism and subject to academic disciplinary action, including failure of the course. In particular, students are reminded that proper citation requires mentioning sources when you use them, not just in a general list of references at the end of a document. You are responsible for understanding UT's Academic Honesty and the University Honor Code which can be found at the following web address: http://deanofstudents.utexas.edu/sjs/acint_student.php

5.3 Student rights and responsibilities

- You have a right to a learning environment that supports mental and physical wellness.
- You have a right to respect.
- You have a right to be assessed and graded fairly.
- You have a right to freedom of opinion and expression.
- You have a right to privacy and confidentiality.
- You have a right to meaningful and equal participation, and to self-organize groups to improve your learning environment.
- You have a right to learn in an environment that is welcoming to all people. No student shall be isolated, excluded or diminished in any way.

With these rights come responsibilities:

- You are responsible for taking care of yourself, managing your time, and communicating with the teaching team and with others if things start to feel out of control or overwhelming.
- You are responsible for acting in a way that is worthy of respect and always respectful of others.
- Your experience with this course is directly related to the quality of the energy that you bring to it, and your energy shapes the quality of your peers' experiences.
- You are responsible for creating an inclusive environment and for speaking up when someone is excluded. In particular, you are responsible for ensuring that your participation does not exclude the participation of others. Office hours are available for in-depth further discussion of advanced topics or other interests that pursuing in depth during class would exclude others.
- You are responsible for holding yourself accountable to these standards, holding each other to these standards, and holding the teaching team accountable as well.

5.4 Personal Pronoun Preference

Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender, gender variance, and nationalities. Class rosters are provided to the instructor with the student's legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records.

5.5 DEI in classroom discussions

[Texas Senate Bill 17](#), the recent law that outlaws diversity, equity, and inclusion programs at public colleges and universities in Texas, does not in any way affect content, instruction or discussion in a course at public colleges and universities in Texas. Expectations and academic freedom for teaching and class discussion have not been altered post-SB 17, and students should not feel the need to censor their speech pertaining to topics including race and racism, structural inequality, LGBTQ+ issues, or diversity, equity, and inclusion.

5.6 Drop Policy

If you want to drop a class after the 12th class day, you'll need to execute a Q drop before the Q-drop deadline, which typically occurs near the middle of the semester. Under Texas law, you are only allowed six Q drops while you are in college at any public Texas institution. For more information, see: <http://www.utexas.edu/ugs/csacc/academic/adddrop/qdrop>

International students *must* [meet with the international office](#) before dropping a class that would put them below full-time status. Although it is worth noting that there are legitimate reasons that allow International students to be below full-time status, so if you think you are failing a course (or just performing below your expectations) don't make assumptions either way, speak with the international office to discover your options.

5.7 University Resources for Students

Your success in this class is important to me. We will all need accommodations at different times because we all learn differently. If there are aspects of this course that prevent you from learning or exclude you, please let me know as soon as possible. Together we'll develop strategies to meet both your needs and the requirements of the course. There are also a range of resources on campus, detailed below.

5.7.1 Accessible/Compliant Statement:

If you are a student with a disability, or think you may have a disability, and need accommodations please contact Disability and Access (D&A). You may refer to D&A's website for contact and more information: <http://diversity.utexas.edu/disability/>. If you are already registered with D&A, please deliver your Accommodation Letter to me as early as possible in the semester so we can discuss your approved accommodations.

5.7.2 Accessible, Inclusive, and Compliant Statement:

The university is committed to creating an accessible and inclusive learning environment consistent with university policy and federal and state law. Please let me know if you experience any barriers to learning so I can work with you to ensure you have equal opportunity to participate fully in this course. If you are a student with a disability, or think you may have a disability, and need accommodations please contact Disability and Access (D&A). Please refer to D&A's website for contact and more information: <http://diversity.utexas.edu/disability/>. If you are already registered with D&A, please deliver your Accommodation Letter to me as early as possible in the semester so we can discuss your approved accommodations and needs in this course.

5.7.3 Counseling and Mental Health Center

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. <http://www.cmhc.utexas.edu/individualcounseling.html>

5.7.4 The Sanger Learning Center

All students, including graduate students, are welcome to take advantage of Sanger Center's classes and workshops, private learning specialist appointments, peer academic coaching, and tutoring for more than 70 courses in 15 different subject areas. For more information, please visit <https://ugs.utexas.edu/slc/grad> or call 512-471-3614 (JES A332).

5.7.5 [University Writing Center free programs for grad students](#)

5.7.6 [Libraries](#)

5.7.7 [IT services](#)

5.7.8 [Student Emergency Services](#)

5.7.9 Important Safety Information

If you have concerns about the safety or behavior of fellow students, TAs or Professors, call BCAL (the Behavior Concerns Advice Line): 512-232-5050. Your call can be anonymous. If something doesn't feel right—it probably isn't. Trust your instincts and share your concerns.

The following recommendations regarding emergency evacuation from the Office of Campus Safety and Security, 512-471-5767, <http://www.utexas.edu/safety/>

Occupants of buildings on The University of Texas at Austin campus are required to evacuate buildings when a fire alarm is activated. Alarm activation or announcement requires exiting and assembling outside.

- Familiarize yourself with all exit doors of each classroom and building you may occupy. Remember that the nearest exit door may not be the one you used when entering the building.
- Students requiring assistance in evacuation shall inform their instructor in writing during the first week of class.
- In the event of an evacuation, follow the instruction of faculty or class instructors. Do not re-enter a building unless given instructions by the following: Austin Fire Department, The University of Texas at Austin Police Department, or Fire Prevention Services office.
- Link to information regarding emergency evacuation routes and emergency procedures can be found at: <http://www.utexas.edu/emergency>