

EDS 213 - Databases and Data Management

Spring 2024

Instructors

Julieb Brun (jb160@ucsb.edu), Greg Janée (gjanee@ucsb.edu) and Renata Curty (rcurty@ucsb.edu)

Teaching Assistant

Jamie Miller (jkmiller@ucsb.edu)

Course Lecture Times

Monday & Wednesday 9:30-10:45 AM, NCEAS
Discussion - session 1: Thur 1-1:50PM, Bren Hall 1510
Discussion - session 2: Thur 2-2:50PM, Bren Hall 1510

Office Hours:

Monday 11-12 pm (NCEAS)

Overview

This course will teach students the fundamentals of relational databases and data management. Students will learn the principles of database modeling and design, and gain practical experience in applying SQL (Structured Query Language) to manage and manipulate relational databases. The course also introduces the role and application of data documentation and metadata standards for interoperability and effective data management. By the end of the course, students will be equipped to make informed decisions about how to manage databases and data ethically and responsibly with a particular focus on issues such as bias, data privacy, sharing, ownership, and licensing.

Learning Objectives

- Understand the fundamental principles of relational databases, including table structures, primary and foreign keys, relationships between tables, and data normalization.
- Understand how to use the Unix command line and how to manage DuckDB databases from the command line.
- Use SQL to retrieve, manipulate, and manage data stored in a relational database.

- Demonstrate proficiency in querying, filtering and sorting, as well as programmatically accessing and interacting with relational databases from R and Python.
- Become familiar with advanced database topics such as concurrency, transactions, indexing, backups, versioning, and publication.
- Understand the role of good data documentation and metadata standards for interoperability, effective data management and reproducibility.
- Operationalize the FAIR principles into data management practices.
- Produce a metadata record in EML (Ecological Metadata Language) and apply metadata crosswalks to programmatically convert metadata schemas.
- Evaluate ethical and responsible data management practices, including issues related to bias, data privacy, sharing, ownership and licensing.

Course Content Areas

- Unix command line
- Relational databases (using DuckDB as an example)
- Data modeling
- SQL
- Accessing relational databases from programming environments (R & python)
- Ethical and responsible data management
- Documentation and metadata standards according to the FAIR principles
- Sensitive data and de-identification
- Data publication and licensing

Course Assessment

Your performance in this course will depend on:

- 90% on weekly homework assignments
- 10% on participation

Homework turned in late will be docked 25% per day.

Course Schedule (subject to change)

See [course website](#) for specific lecture topics each week, reading materials/slides, assignments, and due dates.

Code of Conduct

All students are expected to read and comply with the UCSB code of conduct. We expect cooperation from all members to help ensure a welcoming and inclusive environment for everybody. We are determined to make our courses welcoming, inclusive and harassment-free for everyone regardless of gender, gender identity and expression, race, age, sexual orientation, disability, physical appearance, or religion (or lack thereof). We do not tolerate harassment of class participants, teaching assistants, or instructors in any form. Derogatory,

abusive, or demeaning language or imagery will not be tolerated.

Student Support

We understand that ongoing crises impact students differently based on experiences, identities, living situations and resources, family responsibilities, and unforeseen challenges. We encourage you to prioritize your well-being. We are here to help you reach your learning and career goals. You are always welcome to reach out to our teaching team so that we can best support you. Please see the UCSB Campus Resource Guide for campus student support and services.

Disabled Students Program

Students with disabilities and/or alternative learning needs are encouraged to work with the Disabled Students Program at UCSB to ensure we can best support your learning and success.

Bibliography

Burnette. (2022). Managing environmental data: principles, techniques, and best practices. CRC Press.

https://search.library.ucsb.edu/permalink/01UCSB_INST/1aqck9j/alma9917095897506531

Chapman, AD, & Grafton, O. (2008). Guide to Best Practices for Generalising Sensitive Species-Occurrence Data (v1.0). <https://doi.org/10.15468/doc-b02j-gt10>

Crystal-Ornelas, R., Varadharajan, C., O’Ryan, D., Ramírez-Muñoz, J., Jones, M. B., Lehnert, K. A., ... & Servilla, M. (2022). Enabling FAIR data in Earth and environmental science with community-centric (meta)data reporting formats. *Scientific Data*, 9(1), 700.

<https://doi.org/10.1038/s41597-022-01606-w>

Jones, M. B., O’Brien, M., Mecum, B., Boettiger, C., Schildhauer, M., Maier, M., Whiteaker, T., Earl, S., & Chong, S. (2019). Ecological Metadata Language version 2.2.0. KNB Data Repository. <https://doi.org/10.5063/F11834T2>

Labastida, I., & Margoni, T. (2020). Licensing FAIR Data for Reuse. *Data Intelligence*, 2(1-2), 199-207. https://doi.org/10.1162/dint_a_00042

McGovern, A., Ebert-Uphoff, I., Gagne, D., & Bostrom, A. (2022). Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science. *Environmental Data Science*, 1, E6. doi:10.1017/eds.2022.5

Recknagel, F., & Michener, W. K. (Eds.). (2018). *Ecological informatics: Data management and knowledge discovery* (3rd ed.). Springer.

Templ, M., Meindl, B., Kowarik, A., & Gussenbauer, J. (Jan. 2, 2023). Package 'sdcMicro' (Version 5.75) [Manual]. Retrieved March 1, 2023, from <https://cran.r-project.org/web/packages/sdcMicro/sdcMicro.pdf>