

# IS430 Tutorial For Organizing Files

# Agenda

- What Is the Main Topic of This Unit?
- How I Chose a File Organization Project
- [Illinois Data Bank](#)
- File Naming Recommendations to Consider
- File Grouping Recommendations to Consider
- Tour of the Current Project State
- Potential Next Steps for This Project (TODO)
- Project Lessons Learned So Far

# What Is the Main Topic of This Unit?

- This unit is about file organization like the Zelle Chapter 11 unit is about simulation.
- The main goal of Zelle Chapter 11 is to learn design, coding, and testing of complex programs.
- Simulation is interesting but incidental.
- The main goal of this unit is to learn about choosing and managing a project.
- File organization is interesting but incidental.

# How I Chose a File Organization Project

- File organization is a common practitioner headache.
- Sweigart provides a good overview of Python file organization features.
- Organizing files is common activity in data curation and data librarianship.
- I was inspired by a student's story about their GA job.
- I am expecting to be organizing research data files in my volunteer work at the Plant Science Center at Chicago Botanic Garden.

# Illinois Data Bank

- The Illinois Data Bank is a public access repository for publishing research data from the University of Illinois Urbana-Champaign.
- [Submission Guides](#)
- I decided to concentrate on two parts of the guide:
  - File Naming
  - File Grouping
- Please note that while the coding of metadata into filenames is common practice, filenames are a weak finding aid.

# File Naming Recommendations to Consider

- Date Format
- **Use Letters, Numbers, Underscores, and Hyphens**
- **File Extensions**
- Leading Zeroes for Numbers
- Keep Names Short
- Use Alphanumericals
- **Use Underscores Not Spaces**
- **Use Lowercase Characters**

# File Grouping Recommendations to Consider

- Organize Files into Logical Groups
- Add a Note Explaining Grouping
- Limit Number of Files in Each Group

# Tour of the Current Project State

- Tour the data preparation notebook.
- Tour the file organization common module.
- Tour the synthetic file generation program.
- Tour the filename cleaning program.
- Tour the broken automated unit test module.

# Potential Next Steps for This Project (TODO)

- Generate test files with names that must be cleaned manually.
- Include more **file naming** features consistent with advice in Illinois Data Bank Guide.
- Implement **file grouping** features consistent with advice in Illinois Data Bank Guide.
- Address Illinois Data Bank guidelines beyond naming and grouping.
- Develop a full set of Pytest automated tests for `file_organization_common.py`
- Revise tutorial to reflect new developments.
- Consider developing coding assignment for this unit.

# Project Lessons Learned So Far

- Project ideas can be exciting.
- Projects often spawn side projects.
- Exciting project ideas can lead to an infeasible project scope **very quickly**.
- Big projects can be slimmed down to become manageable projects.
- Project scope can be reduced by committing to the highest priority features first.
- Engaging early with high-priority features helps you understand your project better.
- Keep your eye on the calendar/clock and be ready to wind down your work.
- Leave behind a TODO list to guide your work when you return to this work and start the **next project**.

**Last Revised 2026-04-13**