Managing Your Data

What to do when a research project ends
To-Do

1. Capture metadata
2. Consider publication
3. Back up and archive
4. Update your resume or CV
01 Capture Metadata
Metadata
(noun)

data about other data (e.g. labels, titles, units, tags, etc.)
Dataset Metadata

- Title
- Author(s)
- Abstract
- Key terms
- If sharing or publishing
  - License
  - Citation

ReadMe File

(noun)

a text file provided by the author with the background information necessary for someone else to understand and use the dataset
What to include in a ReadMe file?

1. RESEARCH DESCRIPTION
   Purpose, data collection methods, analyses conducted, and any connection to larger projects.

2. INSTRUMENTS AND SOFTWARE USED
   All tools used to collect and analyze the data including instrument calibrations and software versions.

3. FILE MANIFEST
   List with a brief description of each file (or group of files), file type, and the software used to create it.

4. DATA DICTIONARY
   Define all column labels, abbreviations, acronyms, key terms, and units of measurement.

More information on creating a ReadMe file: [https://purr.purdue.edu/kb/metadata](https://purr.purdue.edu/kb/metadata)
Table 1. Description of each column in the data file.

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Sample date and time. All times are recorded in Eastern Standard Time</td>
</tr>
<tr>
<td>B</td>
<td>Temperature measured in degrees Celsius (sonde measurement)</td>
</tr>
<tr>
<td>C</td>
<td>Specific Conductivity measured in millisiemens per centimeter (sonde measurement)</td>
</tr>
<tr>
<td>D</td>
<td>Dissolved Oxygen measured in milligrams per liter (sonde measurement)</td>
</tr>
<tr>
<td>E</td>
<td>pH (sonde measurement)</td>
</tr>
<tr>
<td>F</td>
<td>Turbidity measured in Nephelometric Turbidity Units (sonde measurement)</td>
</tr>
<tr>
<td>G</td>
<td>Total Coliforms measured in Colony Forming Units per 100 milliliters of stream water</td>
</tr>
<tr>
<td>H</td>
<td>Escherichia coli measured in Colony Forming Units per 100 milliliters of stream water</td>
</tr>
<tr>
<td>I</td>
<td>Enterococci measured in Colony Forming Units per 100 milliliters of stream water</td>
</tr>
<tr>
<td>J</td>
<td>Total Suspended Solids measured in milligrams per liter</td>
</tr>
<tr>
<td>K</td>
<td>Total Phosphorus measured in milligrams phosphorus per liter</td>
</tr>
<tr>
<td>L</td>
<td>Ammonia measured in milligrams nitrogen per liter</td>
</tr>
<tr>
<td>M</td>
<td>Nitrate + Nitrite measured in milligrams nitrogen per liter</td>
</tr>
<tr>
<td>N</td>
<td>Orthophosphate measured in milligrams phosphorus per liter</td>
</tr>
<tr>
<td>O</td>
<td>Free floating algae measured in milligrams per cubic meter</td>
</tr>
</tbody>
</table>

02 Consider Publication
To publish or protect?

PUBLISH

- Public good
- Validation
- Funder or publisher requirements
- Author credit

PROTECT

- Confidentiality
- Intellectual property
- Legal restrictions

More information on sensitive data: [http://guides.lib.purdue.edu/sensitivedata](http://guides.lib.purdue.edu/sensitivedata)
Ask us

scaldron@purdue.edu
Check out

The Teaching with PURR Data LibGuide has a directory of sample published datasets

LibGuide: https://guides.lib.purdue.edu/c.php?g=899358
PURRR Publication Process

1. Login with Purdue credentials
2. Create a private project space
3. Upload data files, ReadMe file, and any other supporting documents to your private space
4. Use PURR's publication wizard to add publication metadata (title, description, etc.)
5. Submit for review by the PURR team

Step-by-step video tutorials: https://purr.purdue.edu/guides
03 Back Up and Archive
Don't just save. Preserve.
Digital Preservation
(noun)
a series of managed activities, policies, strategies and actions to ensure the accurate rendering of digital content for as long as necessary, regardless of the challenge of media failure and technological change.
Archival File Formats

1. TEXT
   plain text, comma separated values, tab delimited, OpenDocument Text, PDF/A

2. SPREADSHEETS
   OpenDocument spreadsheets, comma separated values, tab delimited

3. IMAGES
   TIFF, JPG 2000

4. AUDIO
   WAVE

More recommendations: https://purr.purdue.edu/legal/file-format-recommendations
If you're using proprietary software...

When it comes time to share, publish, or archive your data, save files in two formats: the proprietary format native to the software, and an archival format like plain text, csv, or tiff.

Also, be sure to keep a record of the software and version you used to create your files.
3-2-1 Back Up Strategy

3
Copies of important files

2
Different kinds of storage

1
At a remote location
Not a back up

- Keeping your USB drive next to your PC
- Backing up Google Drive files to another folder on Google Drive
- Assuming ITaP is doing it for you
- Only keeping 1 version of active files
- Setting an auto back-up and not checking it
Where should I store my data?

Is it big data? More than a few GB?

- **Purdue Options**
  - They give you way more space than other cloud options, and you get on-campus tech support. Just remember to move any files you want to keep after you graduate.

  
  **Purdue Options**
  - OneDrive
    - portal.office.com
    - 1 TB
    - 100 GB
    - purr.purdue.edu
      - 100 GB private storage. Publish and archive 1 GB. Share with anyone.
    - box.com
      - purdue.box.com
      - 5 TB
      - Shift 10 GB to a personal account after graduation. Share with anyone.
  
  **Public Options**
  - Google
    - google.com/drive
    - 15 GB
  - Dropbox
    - google.com/drive
    - 2 GB
  - iCloud
    - icloud.com
    - 5 GB

Do you want to keep it after you leave Purdue?

- **Purdue Options**
  - They have tons of space. They're all free while you're at Purdue, and they have on-campus tech support.

Do you think other researchers could use it?

- **Purdue Options**
  - Consider publishing your data in PURR, Purdue's data repository. You'll help advance research in your field, and you'll get an author credit to put on your resume.

- **Public Options**
  - They all offer some free space to start with, and there's no need to move files after you graduate.
04 Update Your Resume or CV
What have you gained? Be specific.

**KNOWLEDGE**
- Collection methods
- Security or lab protocols
- Specific tools and software

**EXPERIENCE**
- Collection
- Organization
- Cleaning
- Analysis
- Visualization
- Publication
- Preservation

The above numbers reflect non-unique views of the page, which may include multiple views in the same day by the same user.

The above numbers reflect non-unique user access to the linked content, which includes but is not limited to primary access button clicks, bundle downloads, and supporting documents.
Resources

• Step-by-step video tutorials on using PURR: purr.purdue.edu/guides
• Sensitive data LibGuide: guides.lib.purdue.edu/sensitivedata
• Directory of sample datasets: guides.lib.purdue.edu/c.php?g=899358
• Creating a ReadMe file: purr.purdue.edu/kb/metadata
• Real life example of ReadMe files and data dictionary: Peel, S., Haas, M. H., Turco, Jr, R. F. (2016). Biological, chemical and flow characteristics of five river sampling sites in the Wabash River watershed near Lafayette, Indiana – 2015. Purdue University Research Repository. doi:10.4231/R7RR1W7B
• Archival file format recommendations: purr.purdue.edu/legal/file-format-recommendations
Thank you

Send questions to scaldron@purdue.edu.