Sandi Caldrone
Purdue University Research Repository (PURR)
scaldron@purdue.edu
April 22. 2020

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

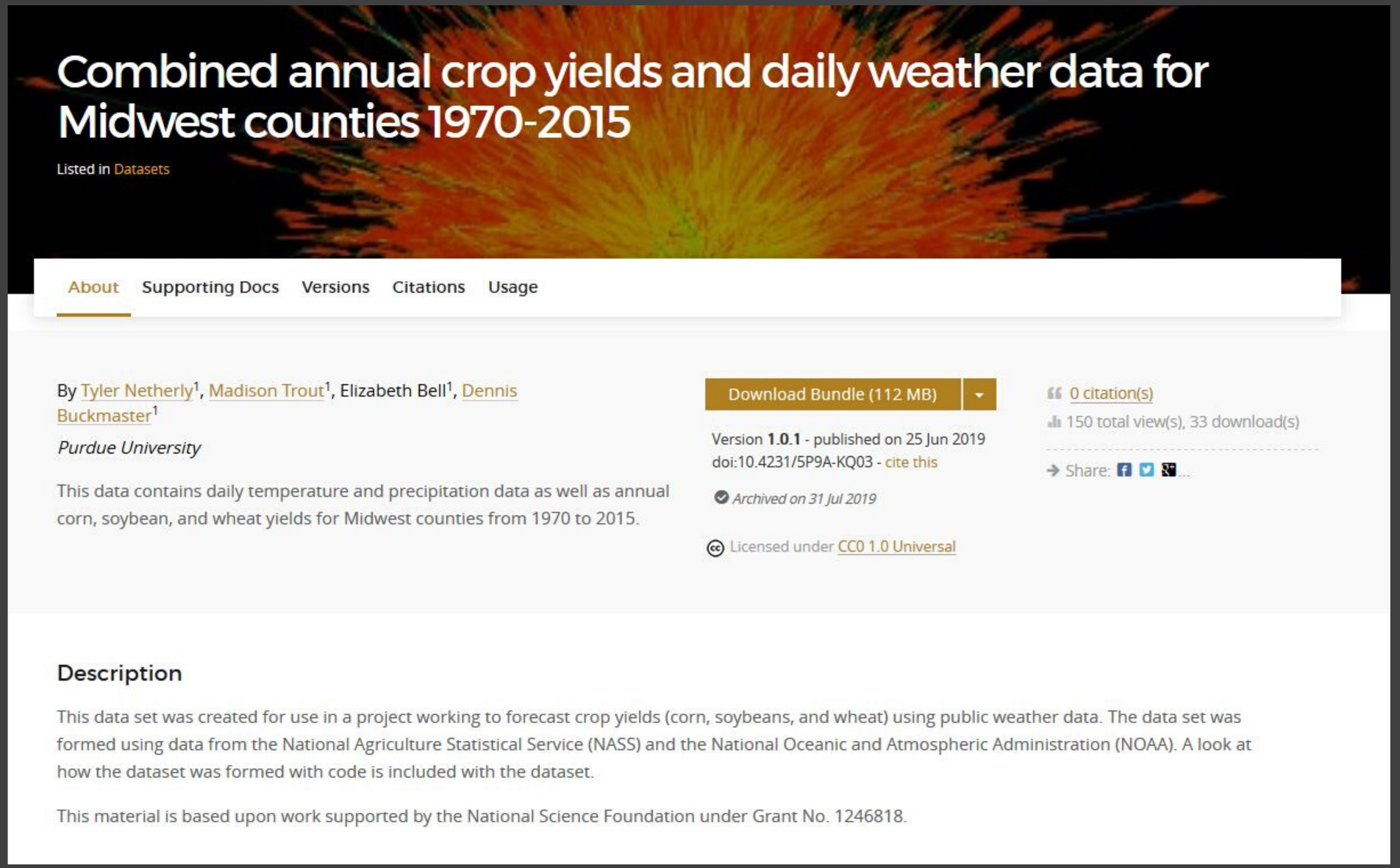
(noun)

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

For example: Netherly, T. G., Trout, M. E., Bell, E., Buckmaster, D. (2019). Combined annual crop yields and daily weather data for Midwest counties 1970-2015. Purdue University Research Repository. doi:10.4231/5P9A-KQ03



For example: Netherly, T. G., Trout, M. E., Bell, E., Buckmaster, D. (2019). Combined annual crop yields and daily weather data for Midwest counties 1970-2015. Purdue University Research Repository. doi:10.4231/5P9A-KQ03

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

Read Me

example data dictionary

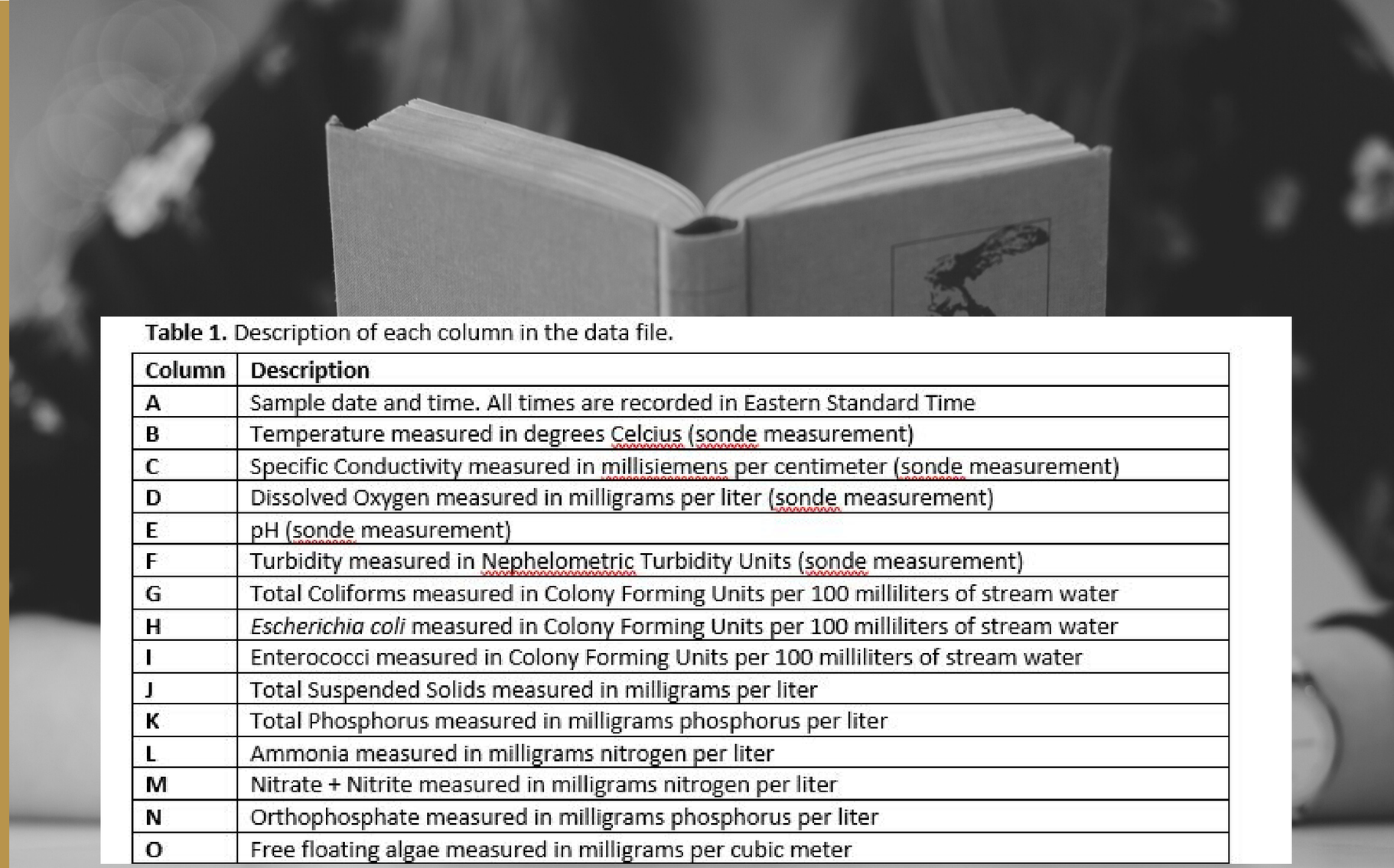


Table 1. Description of each column in the data file.

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

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

02 Consider Publication

To publish or protect?

PUBLISH

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

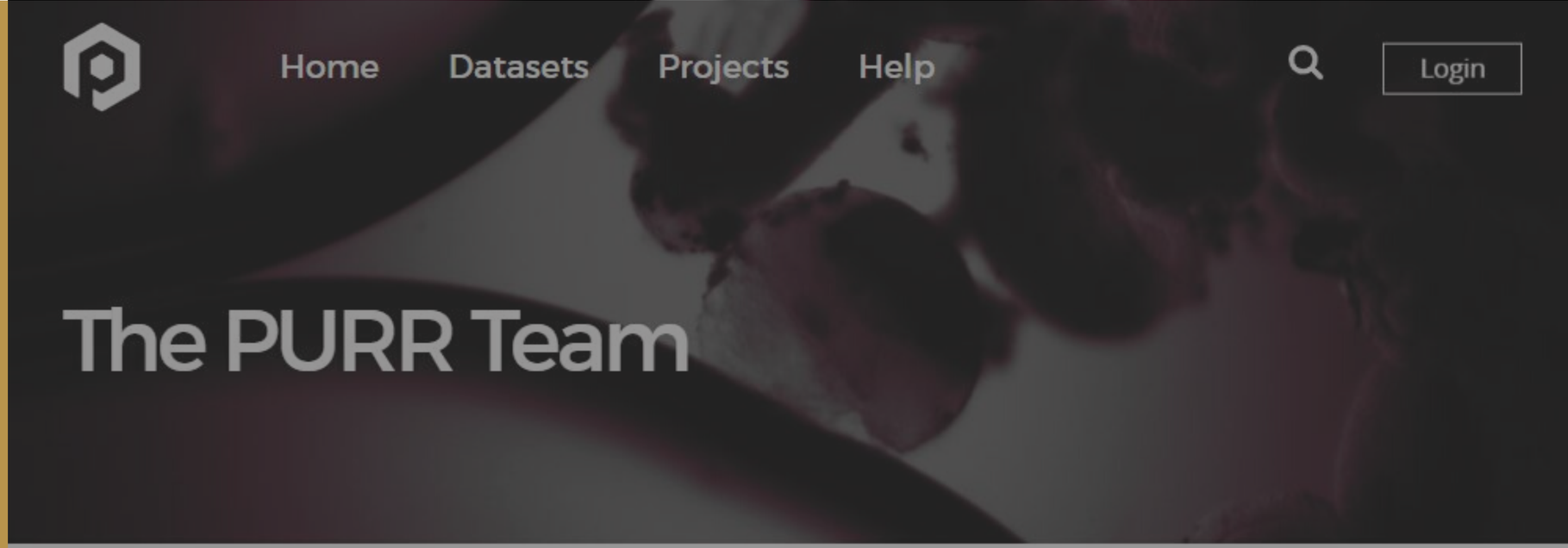
PROTECT

- Confidentiality
- Intellectual property
- Legal restrictions

More information on sensitive data: http://guides.lib.purdue.edu/sensitivedata

Ask us

scaldron@purdue.edu



PURR Staff

Sandi Caldrone
Data Repository Outreach Specialist
Purdue University Libraries

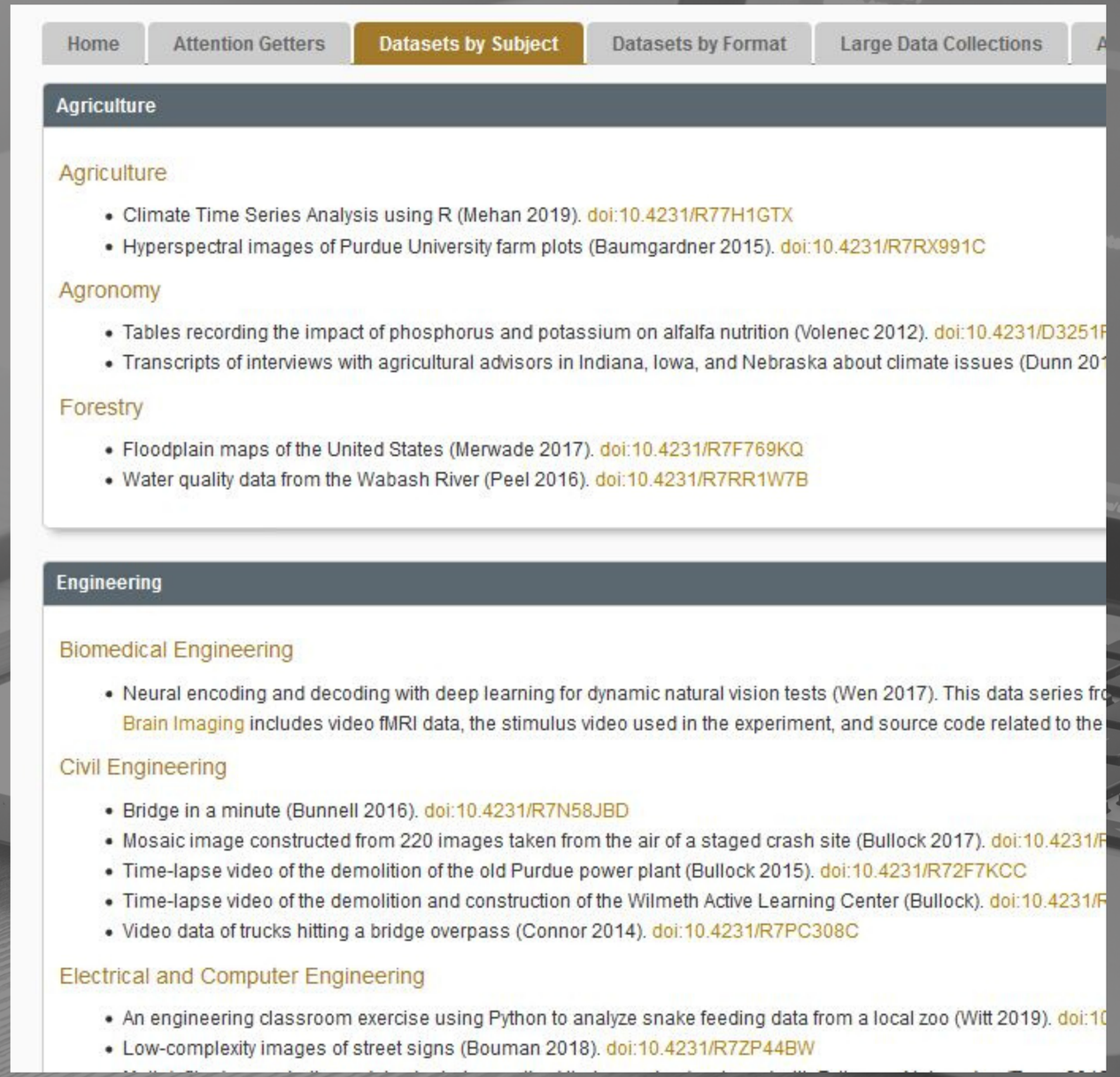
Jerry Kuang
Digital Library Software Developer
Purdue University Libraries

Standa Pejša
Data Curator
Purdue University Libraries

Michael Witt
PURR Director
Associate Professor
Purdue University

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

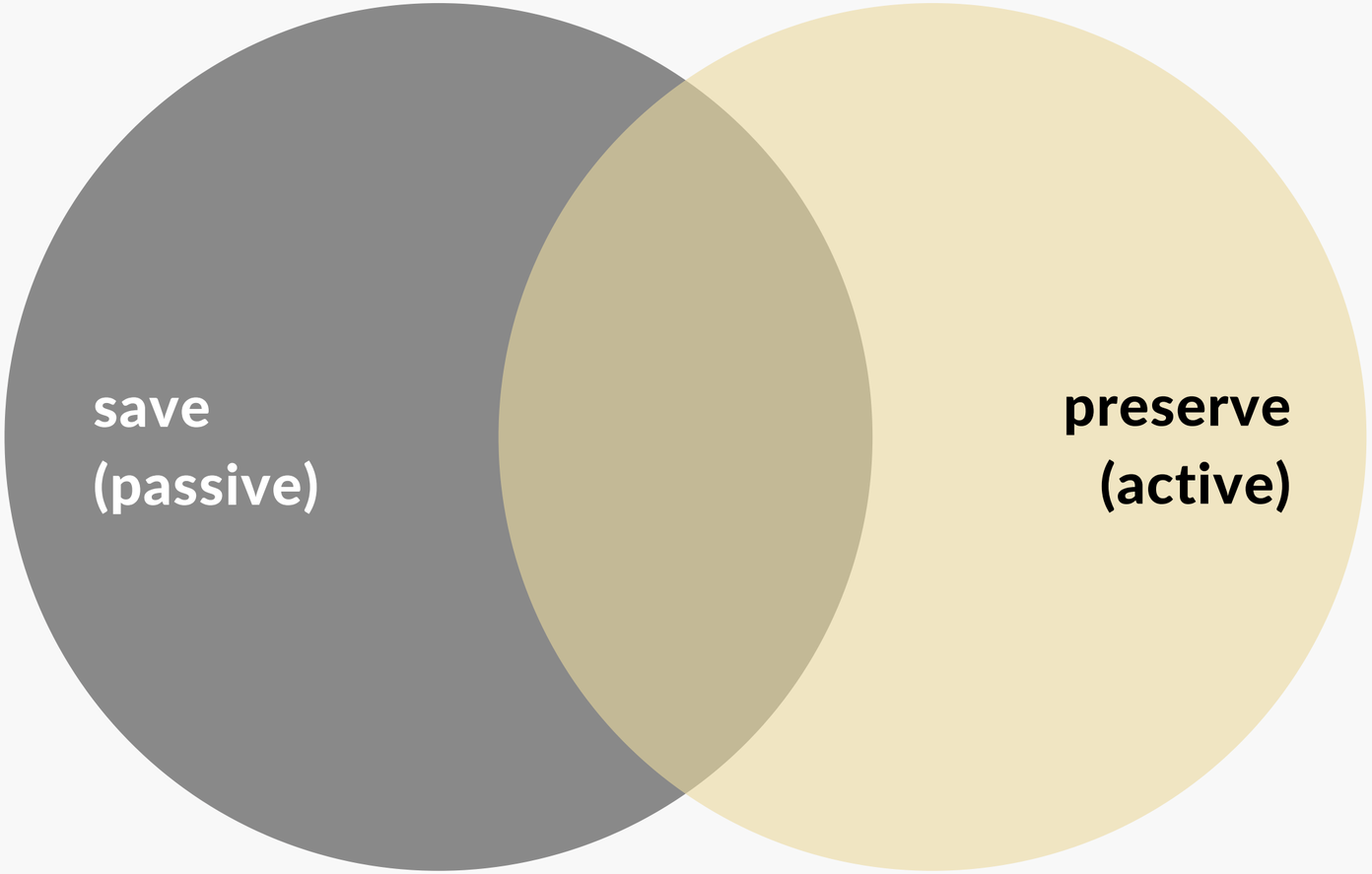
PURR Publication Process

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

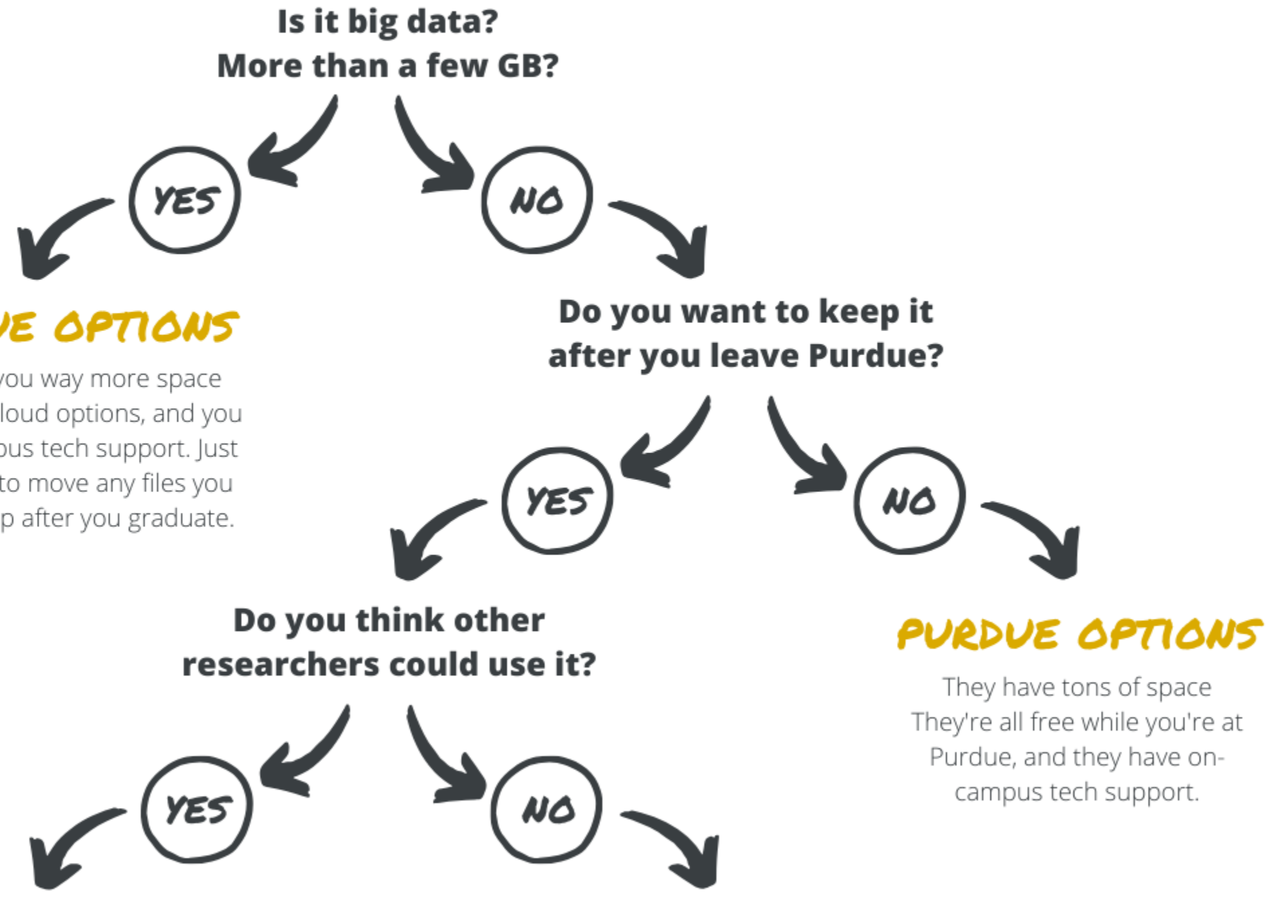
1

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?



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

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

PUBLISH AT PURDUE

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.

PURDUE OPTIONS

		On-campus support	Edit online
OneDrive	**1 TB**	✓	✓
portal.office.com			
Part of Office 365. Share with other Office users.			
PURR	**100 GB**	✓	-
purr.purdue.edu			
100 GB private storage. Publish and archive 1 GB. Share with anyone.			
box.com	**5 TB**	✓	✓
purdue.box.com			
Shift 10 GB to a personal account after graduation. Share with anyone.			

PUBLIC OPTIONS

Google	**15 GB**	-	✓
google.com/drive			
Create and edit documents. Share with anyone.			
Dropbox	**2 GB**	-	✓
google.com/drive			
Collaboratively edit Office files. Mount to your desktop. Share with anyone.			
iCloud	**5 GB**	-	-
icloud.com			
For Apple account holders. Share files between your Apple devices.			

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

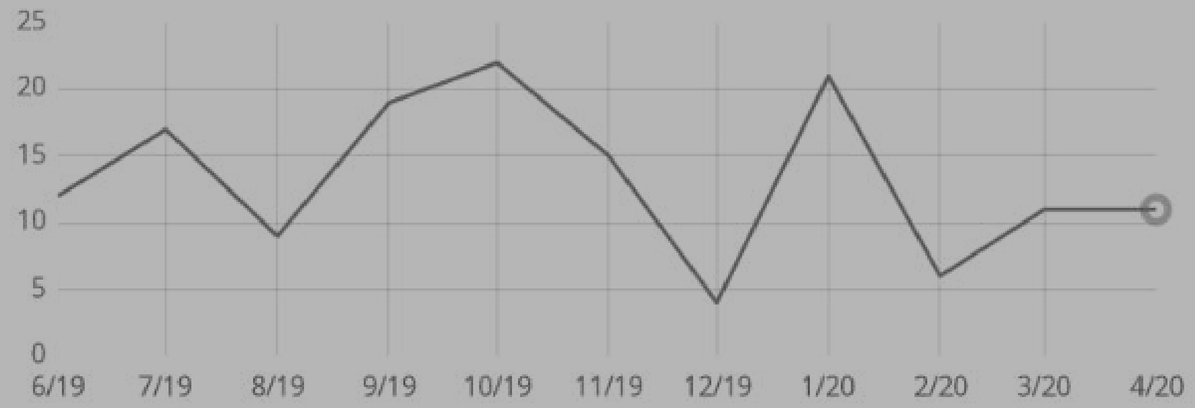
Author Stats

available for all PURR publications

Netherly, T. G., Trout, M. E., Bell, E., Buckmaster, D. (2019). Combined annual crop yields and daily weather data for Midwest counties 1970-2015. Purdue University Research Repository. doi:10.4231/5P9A-KQ03

Total Views

147



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

Total Downloads

33

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

Sandi Caldrone
Purdue University Research Repository (PURR)
April 22. 2020

Thank you

Send questions to scaldron@purdue.edu.