

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



# Combined annual crop yields and daily weather data for Midwest counties 1970-2015

Listed in [Datasets](#)

[About](#) [Supporting Docs](#) [Versions](#) [Citations](#) [Usage](#)

By [Tyler Netherly](#)<sup>1</sup>, [Madison Trout](#)<sup>1</sup>, Elizabeth Bell<sup>1</sup>, [Dennis Buckmaster](#)<sup>1</sup>

*Purdue University*

This data contains daily temperature and precipitation data as well as annual corn, soybean, and wheat yields for Midwest counties from 1970 to 2015.

[Download Bundle \(112 MB\)](#)

Version **1.0.1** - published on 25 Jun 2019  
doi:10.4231/5P9A-KQ03 - [cite this](#)

✓ Archived on 31 Jul 2019

© Licensed under [CC0 1.0 Universal](#)

🗨️ [0 citation\(s\)](#)

📊 150 total view(s), 33 download(s)

→ Share: [f](#) [t](#) [s](#) ...

## Description

This data set was created for use in a project working to forecast crop yields (corn, soybeans, and wheat) using public weather data. The data set was formed using data from the National Agriculture Statistical Service (NASS) and the National Oceanic and Atmospheric Administration (NOAA). A look at how the dataset was formed with code is included with the dataset.

This material is based upon work supported by the National Science Foundation under Grant No. 1246818.

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://purrr.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	<i>Escherichia coli</i> 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](mailto:scaldron@purdue.edu)



[Home](#)

[Datasets](#)

[Projects](#)

[Help](#)



[Login](#)

## The PURR Team

### PURR Staff



**Sandi Caldron**  
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  
Libraries



# Check out

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

Home Attention Getters **Datasets by Subject** Datasets by Format Large Data Collections

## Agriculture

### Agriculture

- Climate Time Series Analysis using R (Mehan 2019). doi:10.4231/R77H1GTX
- Hyperspectral images of Purdue University farm plots (Baumgardner 2015). doi:10.4231/R7RX991C

### Agronomy

- Tables recording the impact of phosphorus and potassium on alfalfa nutrition (Volenec 2012). doi:10.4231/D3251F
- Transcripts of interviews with agricultural advisors in Indiana, Iowa, and Nebraska about climate issues (Dunn 2017)

### Forestry

- Floodplain maps of the United States (Merwade 2017). doi:10.4231/R7F769KQ
- Water quality data from the Wabash River (Peel 2016). doi:10.4231/R7RR1W7B

## Engineering

### Biomedical Engineering

- Neural encoding and decoding with deep learning for dynamic natural vision tests (Wen 2017). This data series from [Brain Imaging](#) includes video fMRI data, the stimulus video used in the experiment, and source code related to the

### Civil Engineering

- Bridge in a minute (Bunnell 2016). doi:10.4231/R7N58JBD
- Mosaic image constructed from 220 images taken from the air of a staged crash site (Bullock 2017). doi:10.4231/R
- Time-lapse video of the demolition of the old Purdue power plant (Bullock 2015). doi:10.4231/R72F7KCC
- Time-lapse video of the demolition and construction of the Wilmeth Active Learning Center (Bullock). doi:10.4231/R
- Video data of trucks hitting a bridge overpass (Connor 2014). doi:10.4231/R7PC308C

### Electrical and Computer Engineering

- An engineering classroom exercise using Python to analyze snake feeding data from a local zoo (Witt 2019). doi:10
- Low-complexity images of street signs (Bouman 2018). doi:10.4231/R7ZP44BW

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

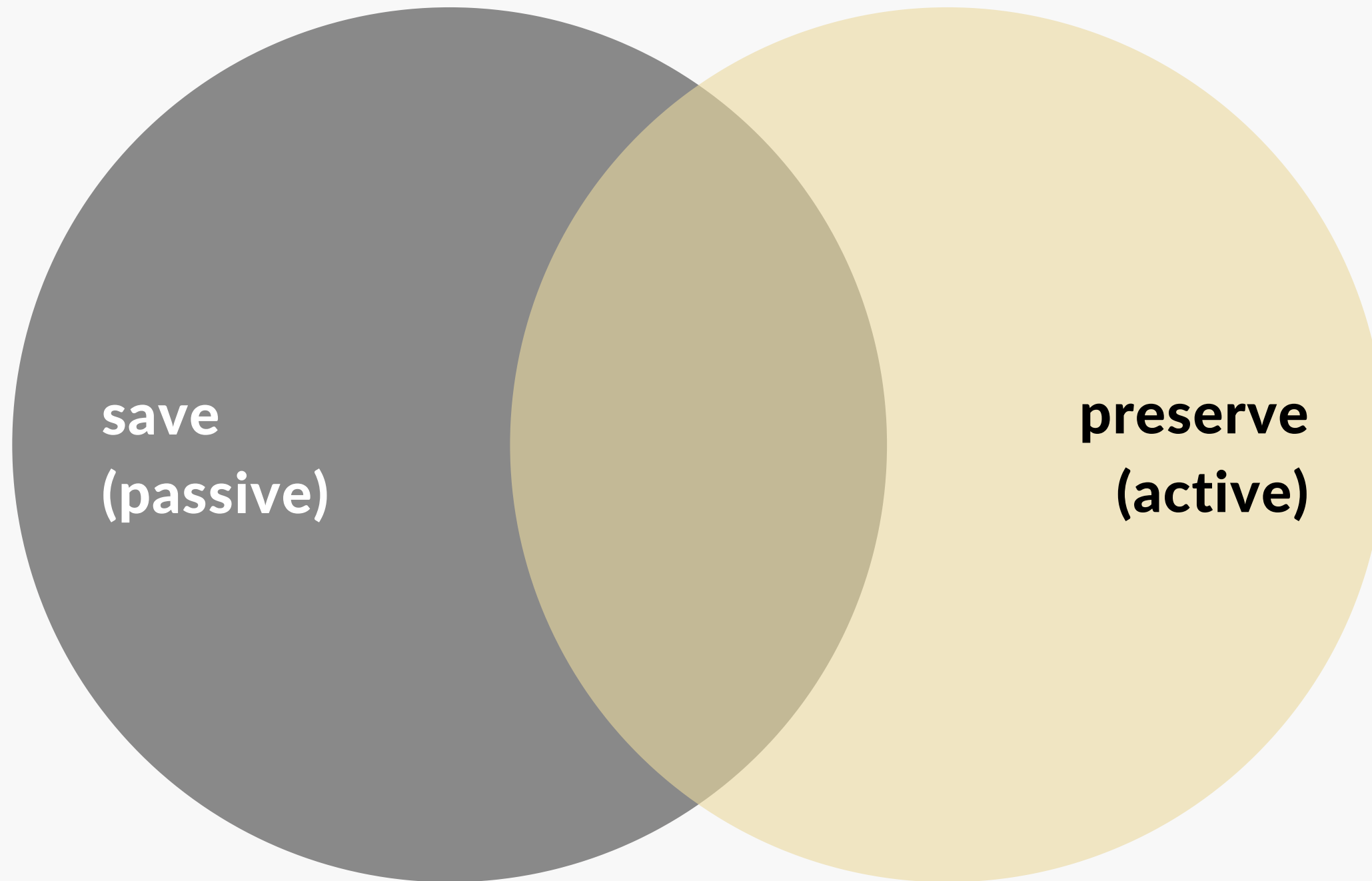
Step-by-step video tutorials: <https://purrr.purdue.edu/guides>



**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?

YES

NO

## 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.

Do you want to keep it after you leave Purdue?

YES

NO

Do you think other researchers could use it?

YES

NO

## 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

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

## PURDUE OPTIONS

<b>OneDrive</b> portal.office.com	<b>1 TB</b>	Part of Office 365. Share with other Office users.
<b>PURR</b> purr.purdue.edu	<b>100 GB</b>	100 GB private storage. Publish and archive 1 GB. Share with anyone.
<b>box.com</b> purdue.box.com	<b>5 TB</b>	Shift 10 GB to a personal account after graduation. Share with anyone.

ON-CAMPUS SUPPORT

EDIT ONLINE

✓	✓
✓	-
✓	✓

## PUBLIC OPTIONS

<b>Google</b> google.com/drive	<b>15 GB</b>	Create and edit documents. Share with anyone.
<b>Dropbox</b> google.com/drive	<b>2 GB</b>	Collaboratively edit Office files. Mount to your desktop. Share with anyone.
<b>iCloud</b> icloud.com	<b>5 GB</b>	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: [purrr.purdue.edu/guides](http://purrr.purdue.edu/guides)
- Sensitive data LibGuide: [guides.lib.purdue.edu/sensitivedata](http://guides.lib.purdue.edu/sensitivedata)
- Directory of sample datasets: [guides.lib.purdue.edu/c.php?g=899358](http://guides.lib.purdue.edu/c.php?g=899358)
- Creating a ReadMe file: [purrr.purdue.edu/kb/metadata](http://purrr.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](https://doi.org/10.4231/R7RR1W7B). Purdue University Research Repository. doi:10.4231/R7RR1W7B
- Archival file format recommendations: [purrr.purdue.edu/legal/file-format-recommendations](http://purrr.purdue.edu/legal/file-format-recommendations)



Sandi Caldrone

Purdue University Research Repository (PURR)

April 22, 2020

# Thank you

Send questions to [scaldron@purdue.edu](mailto:scaldron@purdue.edu).

