

Data Management Fundamentals

Wendy Kozlowski and Leah McEwen

2023-01-12

✉ rdmsg-help@cornell.edu

🌐 data.research.cornell.edu



Session Objectives

1. Review what data is ... what data management is and why we care
2. Identify good practices related to:
 - a. Planning
 - b. Organization
 - c. Storage
 - d. Metadata
 - e. Sharing and Preservation
 - f. Documentation/Record Keeping



"You can't keep coming in here and demanding data every two years!"

What is data?

“Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate researching findings...”

OFFICE OF MANAGEMENT AND BUDGET [OMB Circular A-110](#), “Uniform Administrative Requirements for Grants and Agreements With Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations”

Categories of “Data”	
CONTEXT	<ul style="list-style-type: none">• Experimental• Observations
RAW	<ul style="list-style-type: none">• Alphanumeric• X,Y (often binary)• Images
PROCESSED	<ul style="list-style-type: none">• Analysis
DERIVED	<ul style="list-style-type: none">• Graphical representations
CODE	<ul style="list-style-type: none">• Scripts used in analysis and processing

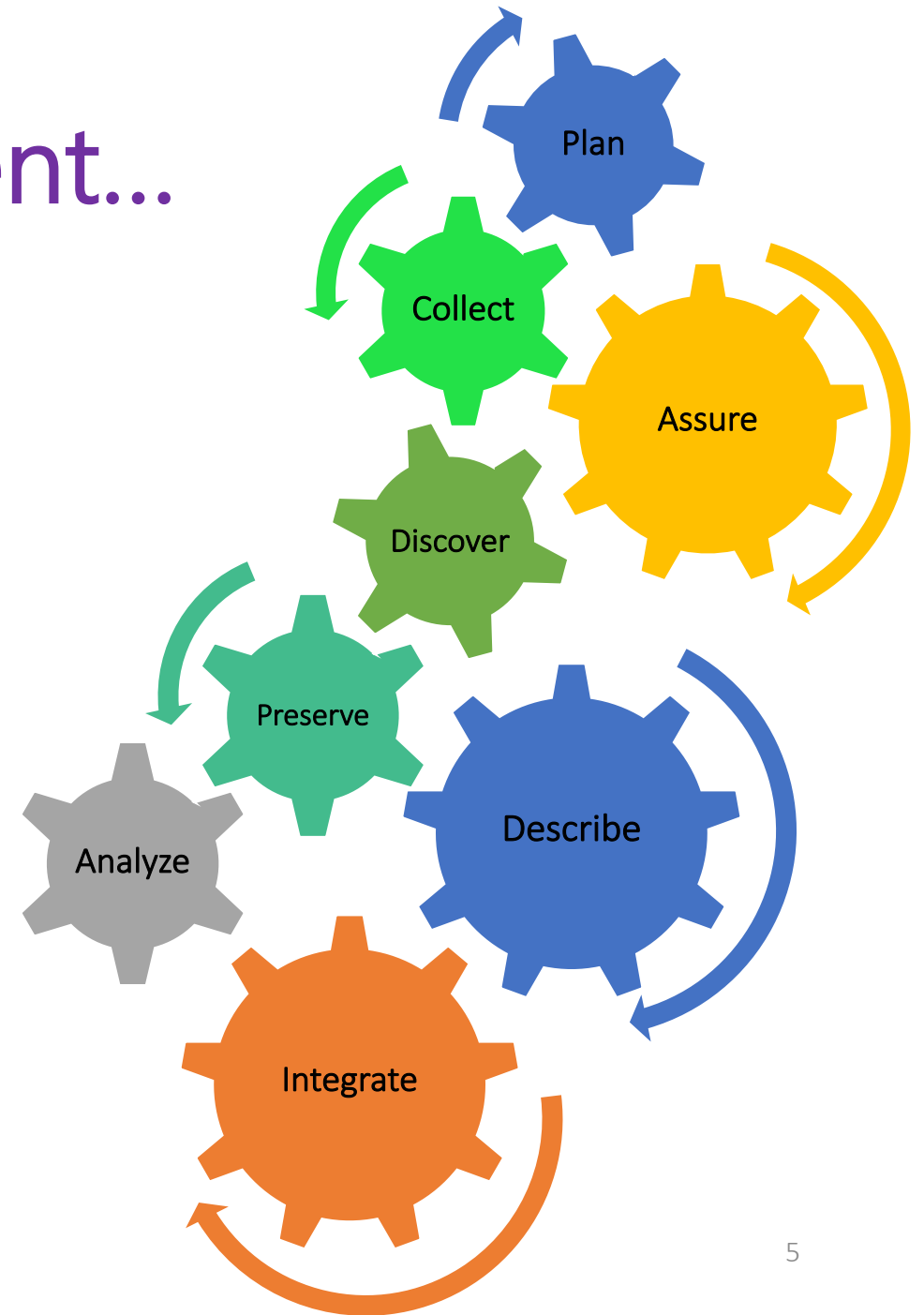
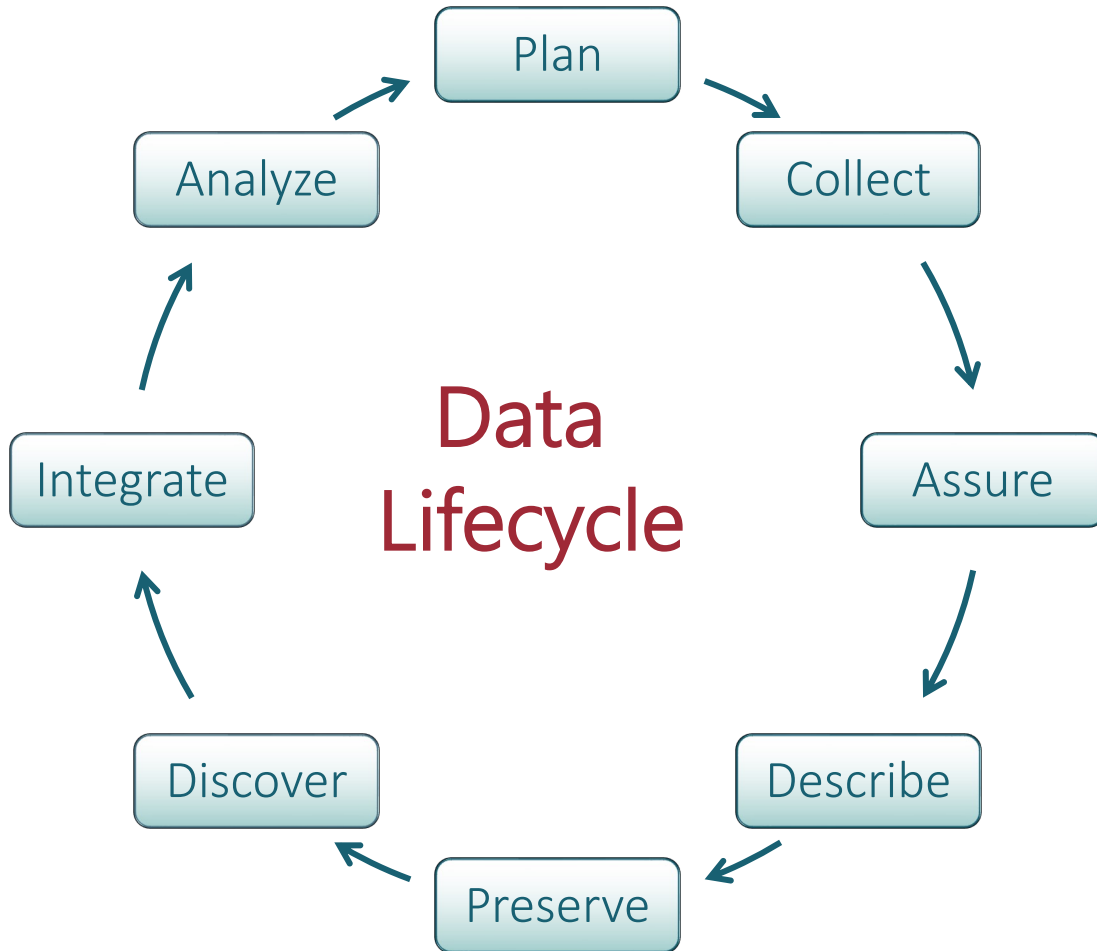
Research Data Management...

is a term that encompasses the care and maintenance of data created or used throughout the research process.

It includes a broad range of activities, including organization, description, storage, preservation and sharing...

and needs to happen from when you begin planning, through collection, analysis, sharing and archiving.

Research Data Management...



Make a Plan

A Data Management Plan (DMP) outlines how you will handle your data during and after your research.

A DMP is a formal, but living document to help ensure the current and future security of, and access to, the data and the documentation needed to understand that data.

Funder Requirements – DMPs

NSF National Science Foundation
WHERE DISCOVERIES BEGIN

Contact | Help

Search

NSB Research Areas Funding Awards Document Library News About NSF

Home › Budget Finance & Award Manage... › Institution and Award Support

Email Print Share

Dissemination and Sharing of Research Results

NSF DATA SHARING POLICY Since at least 03 Aug 1995

Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See [Proposal & Award Policies & Procedures Guide \(PAPPG\) Chapter XI.D.4.](#)

NSF DATA MANAGEMENT PLAN REQUIREMENTS Announced 01 Oct 2010

Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See [PAPPG Chapter II.C.2.j](#) for full policy implementation.

Office of Budget Finance & Award Management (BFA)

- Office of Budget, Finance, & Award Management
- Budget Division
- Division of Acquisition and Cooperative Support
- Division of Financial Management
- Division of Grants & Agreements
- Division of Institution & Award Support
- CAAR Branch
- Policy Office
- Systems Office

<http://www.nsf.gov/bfa/dias/policy/dmp.jsp>





What goes into a DMP (generally)??

1. Expected data & research products

What types of data, samples, physical collections, code, software, curriculum materials and other materials will be produced in the course of the project?

2. Data & metadata formats & standards

What formats and standards will be used for your files and metadata?

3. Policies for access & sharing (dissemination)

For the data you are sharing, when will you make it available, and under what conditions?

What methods will be used for securing and sharing data and metadata during and after the award period?

4. Policies & provision for re-use, re-distribution & production of derivatives

How will you meet funder requirements to provide public access to your data while protecting privacy, confidentiality, security and intellectual property rights?

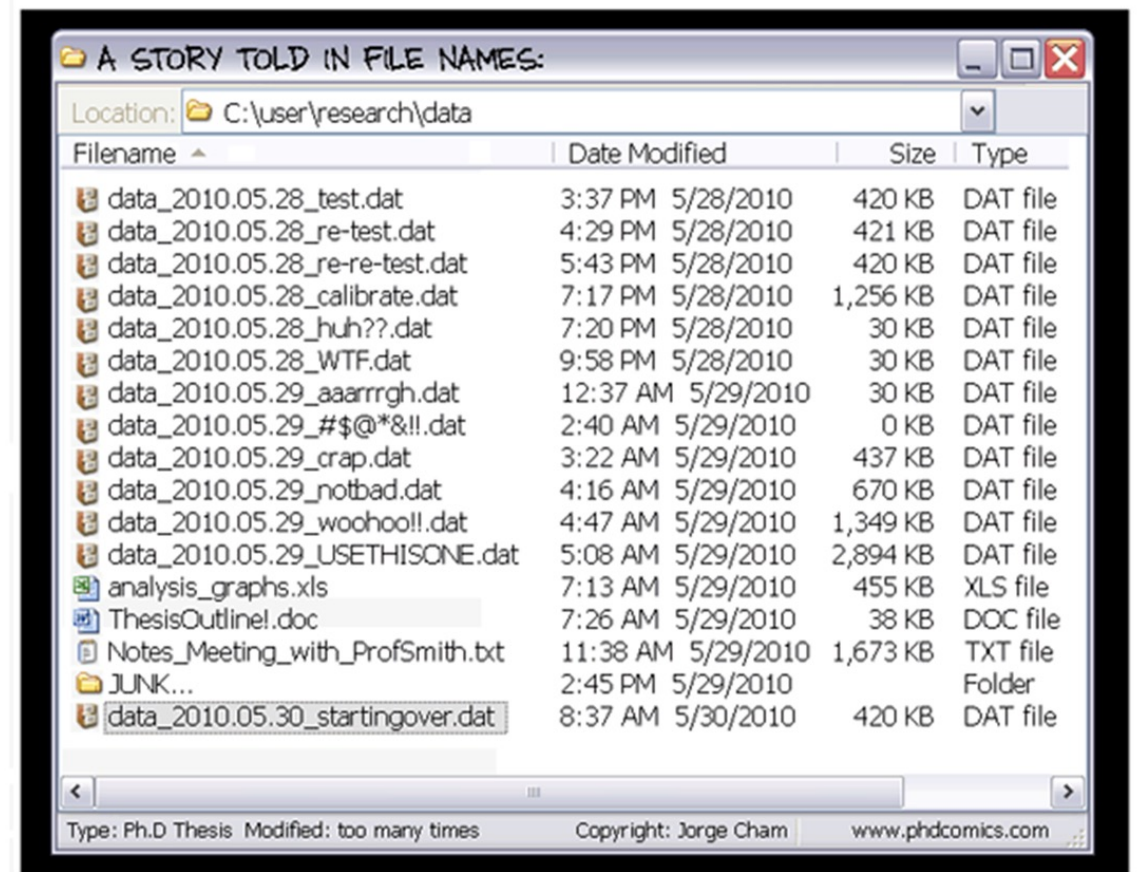
5. Plan for archiving & preservation of access

How will you preserve the integrity of your data over time?

6. Roles & responsibilities

What are the roles and who has responsibilities for managing data?

Organizing Your Work



I've not even begun to think about this yet

My practices could use some improvement

I've got this

Organizing Your Work – Basic Principles



Work is messy. Over time, you'll create a lot of material, in different formats, with multiple version, using different protocols. Often, you can't be immediately sure what is going to be valuable what what's not.



Spending time organizing up front is worth the effort.

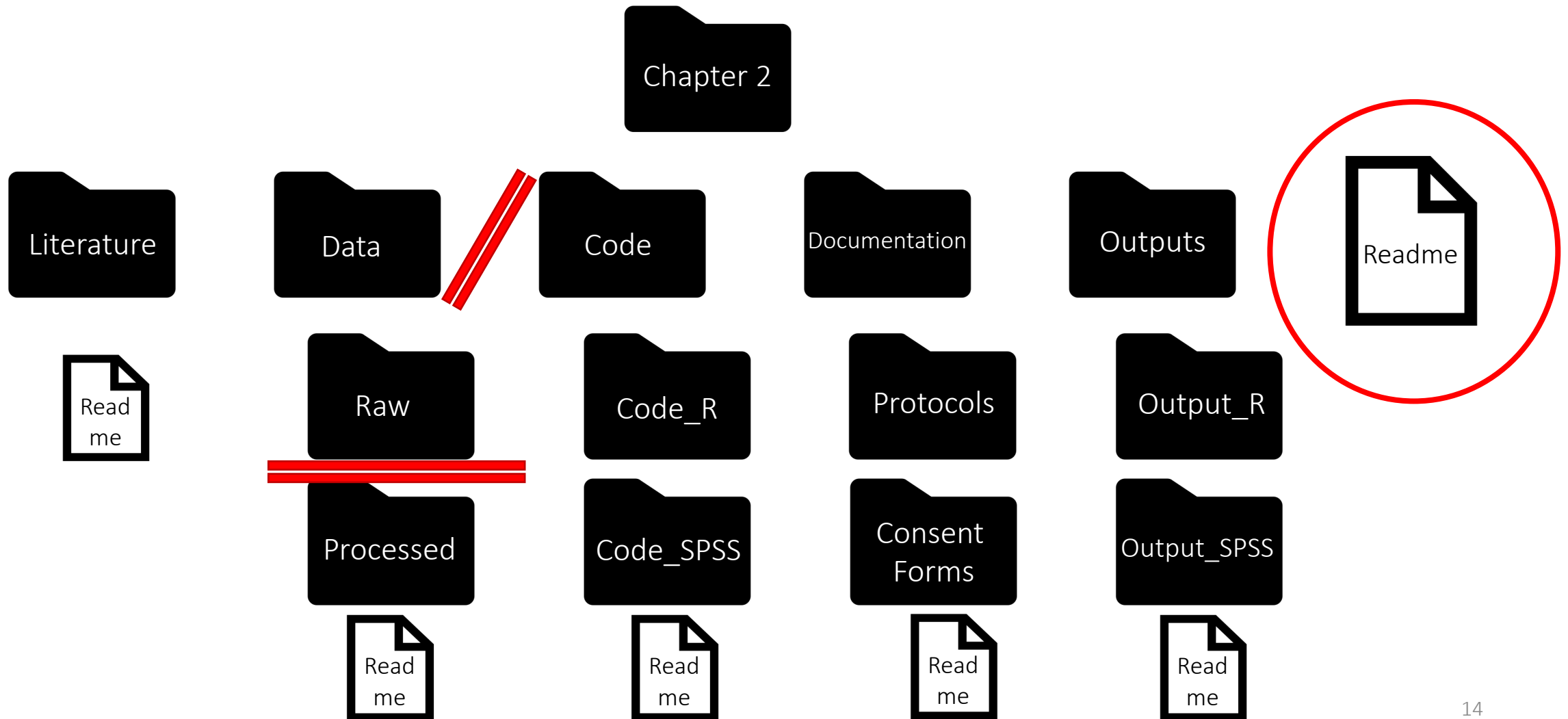


Be realistic – set a plan that's something you'll actually carry out.



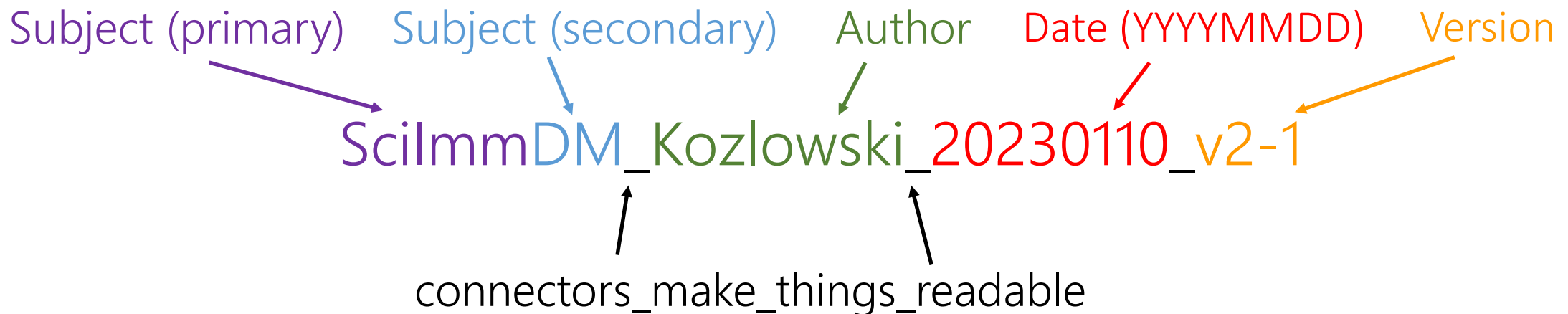
There is no *one* right way, and your own method will likely change over time. Take the extra step to write down what you do and why. Put a date on it.

File organization strategies example & best practices



File Naming

- No special characters (especially ? ; : ' *), spaces or long names (>32 characters)
 - ThisIsJustWhatThirtyTwoLooksLike.txt
- Use meaningful abbreviations and components
- Format dates and times: YYYYMMDD or YYYY-MM-DDThh-mm-ss
- Files should stand independently of folder structure
- Depending on storage system, include also version information



File naming and organization strategies

Order by date:

1955-04-12_notes_MassObs.docx
1955-04-12_questionnaire_MassObs.pdf
1963-12-15_notes_Gorer.docx
1963-12-15_questionnaire_Gorer.pdf

Order by subject:

Gorer_notes_1963-12-15.docx
Gorer_questionnaire_1963-12-15.pdf
MassObs_notes_1955-04-12.docx
MassObs_questionnaire_1955-04-12.pdf

Order by type:

Notes_Gorer_1963-12-15.docx
Notes_MassObs_1955-04-12.docx
Questionnaire_Gorer_1963-12-15.pdf
Questionnaire_MassObs_1955-04-12.pdf

Forced order with numbering:

01_MassObs_questionnaire_1955-04-12.pdf
02_MassObs_notes_1955-04-12.docx
03_Gorer_questionnaire_1963-12-15.pdf
04_Gorer_notes_1963-12-15.docx

Storage

THE FOUR STAGES OF DATA LOSS DEALING WITH ACCIDENTAL DELETION OF MONTHS OF HARD-EARNED DATA



www.phdcomics.com

title: "Stages of Data Loss" - originally published 10/30/2003

I've not even begun to think about this yet

My practices could use some improvement

I've got this

Data Storage – Rule of Three



Here

Local / Working Copy

- Laptop, workstation, shared workspace



Near

External / Remote Copy

- External hard drive, lab servers



Far

Remote Copy

- Cloud Copy (eg. Cornell Box, OneDrive, Cornell GSuite/Google Drive)

Data Storage Best Practices

Sync
≠
Backup

- Decide **what** to *backup* and **how often**
 - Consider value / reproducibility / cost / size
 - Consider full vs partial backups
- **Test your backups** – don't assume they're working
 - Use things like file size comparisons, dates and checksums
- Know how to **restore** your data
- Decide **what** to *archive* and **how often** *
- **Write down your plan** where people can find it
- **Be consistent**

Archive
(≠)
Backup

Version Control

Keeping track of your (and your collaborator's) changes, and being able to recall past versions of your code, data, and processing steps is critical!

Version Control Systems (Git, Mercurial, Subversion | *Github, Sourceforge*)

- *Back up frequently, keep changes small, share frequently, document changes, use remote repositories*

Manual Version Control

- *Maintain a change-log, use consistent naming structure that incorporates version changes, consider major/minor revision tracking*

Version Control

What about versioning in tools like Box, Dropbox and Google Drive???

Understand the fine print!

Box: 100 most recent versions can be recovered

Google Drive: All versions with no time limit (Enterprise); 30 or 90 days unless otherwise specified on Basic account

Dropbox: Depends on plan: 30 (Basic, Plus) – 180 (Prof., Business) days

OneDrive: 500 most recent versions can be recovered

Best Practice: Combine with manual version control to maintain access

Data Storage



<https://finder.research.cornell.edu>

1. What is the classification of your data?

- Public
- Sensitive / Moderate Risk
- Confidential or Restricted / High Risk
- HIPAA-Regulated

2. Do you need backups, snapshots or replication of your data?

- I need one or more backup/snapshot copies of the data, and need to be able to restore data from previous points in time (high durability).
- I need to have replicate copies of the data to minimize downtime (high availability).

3. How much data do you have and how fast will it grow?

- Unlikely to exceed 1TB in 2 years
- Greater than 1TB or likely to exceed in 2 years

4. Do you have special performance needs?

- I am likely to have more than 1,000 files in a single directory within two years.
- My data interactions demand high transaction or transfer rates.

5. How are you expecting to access the data?

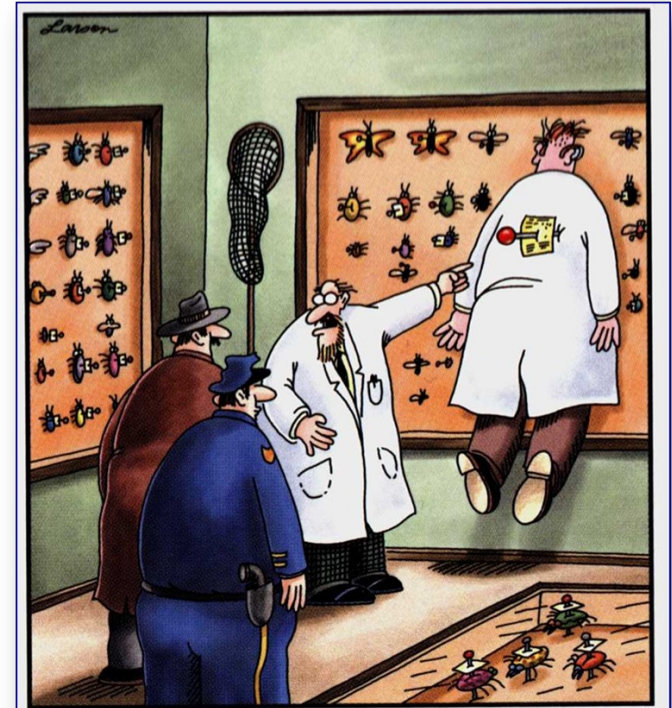
- I need easy access to this data from anywhere, even when I don't have my own computer or mobile device with me.
- I frequently need access from a mobile device such as a phone or tablet.

6. With whom do you need to share your data regularly?

- Only those with a Cornell NetID or GuestID
- Only those with a Weill CWID
- Users in and out of the Cornell community

Amazon Web Services Elastic Block Store Storage for use with Amazon EC2	Amazon Web Services Elastic File System Storage for use with multiple Amazon EC2 instances	Amazon Web Services Glacier Cloud based archival storage	Amazon Web Services S3 Flexible, scalable object storage	BioHPC Cloud Storage for BioHPC lab computing services
CAC Archival Storage Single copy, non-mountable storage	CAC Red Cloud Storage Storage for Red Cloud compute instances	CISER Research Servers Storage for CISER computing environment	CUGIR Publicly shared geospatial data storage	CUL eCommons Publicly shared data repository
Cornell Box Online file sharing and collaboration	Cornell Restricted Access Data Center Storage for CRADC (confidential) computing environment	EZ-Backup Static Storage Archival storage and backup storage	Google Drive Cornell G Suite file storage and sharing	Kaltura Video Platform Service Flexible, scalable video and multi-media storage (customizable)
Kaltura Video on Demand Video and multi-media storage	LabArchives Online electronic lab notebook	Open Science Framework Online project management repository	Shared File Services File sharing between users and computers	Shared File Services - Confidential File sharing between users and computers for confidential data (non-WCM)
WCM Block Storage High performance storage attached to centrally hosted servers (WCM only)	WCM File Sharing Secure network storage (NFS/CIFS) for research computing (WCM only)	WCM Red Cloud Secure Storage Secure storage for Red Cloud compute instances (WCM only)	WCM Secure Remote Archive Secure, single copy, non-mountable storage (WCM only)	

Documentation



"Professor LaVonne had many enemies in the entomological world, detective, but if you examine that data label, you'll find exactly when and where he was—shall we say—'collected.'"

I've not even begun to think about this yet

My practices could use some improvement

I've got this

Example: README

(<https://data.research.cornell.edu/content/readme>)

```
1 ReadMe File for Loftus_et al_MitesData.csv
2 Cornell eCommons Repository
3
4 Publication Title:
5 How Honey Bee Colonies Survive in the Wild: Testing the Importance of Small Nests
6 and Frequent Swarming
7
8 Authors:
9 J. Carter Loftus, Michael L. Smith, and Thomas D. Seeley*
10
11 *Corresponding Author Email: tds5@cornell.edu
12 Alternate Contact Email: mls453@cornell.edu
13
14 Funding:
15 This research was supported by a grant from the Eastern Apiculture Society Research
16 Fund, a Hatch Grant (Project No. NYC-191400) from the Cornell University Agriculture
17 Experiment Station (to TDS), and by a US National Science Foundation Graduate
18 Research Fellowship (DGE-1144153) (to MLS).
19
20 Geographic Location:
21 This study was performed at a site owned by Cornell University outside of Ithaca, NY
22 (42 Degrees 26'9.88" N, 76 Degrees 25'50.45" W). The site consisted of a field with two
23 mowed areas for two apiaries: one for the small-hive colonies and one for the
24 large-hive colonies.
25
26 Statement of Data Use:
27 These data are free and available for public use. The authors would appreciate if you
28 would cite both the original publication and the dataset:
29 Loftus JC, Smith ML, Seeley TD (2016) How Honey Bee Colonies Survive in the Wild:
30 Testing the Importance of Small Nests and Frequent Swarming. PLoS ONE 11(3):
31 e0150362. doi:10.1371/journal.pone.0150362
32
33 Descriptions of column headers:
34
35 ColonyNo.
36 Labels for colonies used in the experiment. Each begins with the letter c (to denote
37 colony) and then a number. The letter "a" after the number denotes colonies that had
38 been replaced in 2013.
```

Readme documentation might include:

- Title, Author(s), Contact info
- Timeframes, Geographic info, Sponsorship info
- Statement of use or License information
- Suggested citation, Related citations
- Description of file(s), types, and constructs needed to understand the dataset
- Methodology (if relevant)
- Abbreviations used, units, blank or missing data definitions
- Change logs / versioning details








Example: Code

Good Enough Practice: Thorough Commenting Line, Block or Bold Comments

Best Practice: README

```
ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics
```

A system for 'declaratively' creating graphics, based on "The Grammar of Graphics". You provide the data, tell 'ggplot2' how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.

Version: 3.3.3
Depends: R (≥ 3.2)
Imports: [digest](#), [glue](#), [grDevices](#), [grid](#), [gtable](#) (≥ 0.1.1), [isoband](#), [MASS](#), [mgcv](#), [rlang](#) (≥ 0.3.0), [scales](#) (≥ 0.5.0), [stats](#), [tibble](#), [withr](#) (≥ 2.0.0)
Suggests: [covr](#), [dplyr](#), [ggplot2movies](#), [hexbin](#), [Hmisc](#), [knitr](#), [lattice](#), [mapproj](#), [maps](#), [maptools](#), [multcomp](#), [munsell](#), [nlme](#), [profvis](#), [quantreg](#), [RColorBrewer](#), [rgeos](#), [rmarkdown](#), [rpart](#), [sf](#) (≥ 0.7-3), [svglite](#) (≥ 1.2.0.9001), [testthat](#) (≥ 2.1.0), [vdiff](#) (≥ 0.3.0)
Enhances: [sp](#)
Published: 2020-12-30
Author: Hadley Wickham  [aut], Winston Chang  [aut], Lionel Henry [aut], Thomas Lin Pedersen  [aut, cre], Kohske Takahashi [aut], Claus Wilke  [aut], Kara Woo  [aut], Hiroaki Yutani  [aut], Dewey Dunnington  [aut], RStudio [cph, fnd]
Maintainer: Thomas Lin Pedersen <thomas.pedersen at rstudio.com>
BugReports: <https://github.com/tidyverse/ggplot2/issues>
License: MIT + file LICENSE
URL: <https://ggplot2.tidyverse.org>, <https://github.com/tidyverse/ggplot2>
NeedsCompilation: no
Citation: [ggplot2 citation info](#)
Materials: [README NEWS](#)
In views: [Graphics](#), [Phylogenetics](#), [TeachingStatistics](#)
CRAN checks: [ggplot2 results](#)
Downloads:

```
1 %% Automate importing Data by Generating Code Using the Database Explorer App
2 % This code reproduces the data obtained using the Database Explorer app by
3 % connecting to a database, executing a SQL query, and importing data into
4 % the MATLAB(R) workspace. To use this code, add the password for
5 % connecting to the database in the database command.
6
7 % Auto-generated by MATLAB Version 9.5 (R2018b) and Database Toolbox
8 % Version 9.0 on 31-May-2018 13:18:18
9
10 %% Make connection to database
11 conn = database('MS SQL Server Auth','');
12
13 %% Execute query and fetch results
14 data = fetch(conn,['SELECT productName, ' ...
15     ' Quantity, ' ...
16     ' Price ' ...
17     'FROM toy_store.dbo.inventorytable']);
18 %% Close connection to database
19 close(conn)
20
21 %% Clear variables
22 clear conn
```

- Purpose
- Software version information
- Packages and other dependencies
- How to execute & expected outputs
- Citation & Licensing information
- Change log

Example: Standard Format (CIF)

Index

Core dictionary (coreCIF) version 2.4.5

Category ATOM_SITE

Name:
'_atom_site[]'

Definition:

Data items in the **ATOM_SITE** category record details about the atom sites in a crystal structure, such as the positional coordinates, atomic displacement parameters, and magnetic moments and directions.

Examples:

Example 1 - based on data set TOZ of Willis, Beckwith & Tozer [Acta Cryst. (1991), C47, 2276-2277].

```
loop_
_atom_site_label
_atom_site_fract_x
_atom_site_fract_y
_atom_site_fract_z
_atom_site_U_iso_or_equiv
_atom_site_adp_type
_atom_site_calc_flag
_atom_site_calc_attached_atom
O1 .4154(4) .5699(1) .3026(0) .060(1) Uani ? ?
C2 .5630(5) .5087(2) .3246(1) .060(2) Uani ? ?
C3 .5350(5) .4920(2) .3997(1) .048(1) Uani ? ?
N4 .3570(3) .5558(1) .4167(0) .039(1) Uani ? ?
C5 .3000(5) .6122(2) .3581(1) .045(1) Uani ? ?
O21 .6958(5) .4738(2) .2874(1) .090(2) Uani ? ?
C31 .4869(6) .3929(2) .4143(2) .059(2) Uani ? ?
# - - - data truncated for brevity - - -
H321C .04(1) .318(3) .320(2) .14000 Uiso ? ?
H322A .25(1) .272(4) .475(3) .19000 Uiso ? ?
H322B .34976 .22118 .40954 .19000 Uiso calc C322
H322C .08(1) .234(4) .397(3) .19000 Uiso ? ?
```



Crystallographic Information File (CIF) format includes data dictionaries and defined syntax and semantics.

File structure enhances machine readability and interoperability

https://www.iucr.org/resources/cif/dictionaries/cif_core

```
#####
#
# This CIF contains the data in a paper accepted for publication in
# Acta Crystallographica Section C. It conforms to the requirements of
# Notes for Authors for Acta Crystallographica Section C, and has been
# peer reviewed under the auspices of the IUCr Commission on Journals.
#
# Full details of the Crystallographic Information File format
# are given in the paper "The Crystallographic Information File (CIF):
# a New Standard Archive File for Crystallography" by S. R. Hall, F. H.
# Allen and I. D. Brown [Acta Cryst. (1991), A47, 655-685].
#
# The current version of the core CIF dictionary is obtainable from
# ftp://ftp.iucr.org/pub/cif_core.dic.
#
# Software is freely available for graphical display of the structure(s)
# in this CIF. For information consult the CIF software page
# http://www.iucr.org/resources/cif/software.
#
# This file may be used for bona fide research purposes within the
# scientific community so long as proper attribution is given to the journal
# article from which it was obtained.
#
#####
data_I
_audit_creation_method 'SHELXL-2016/6'
_chemical_name_systematic
;
4-[<i>trans</i>-4-(<i>trans</i>-4-Propylcyclohexyl)cyclohexyl]benzonitrile
;
_chemical_name_common
;
1-Cyano-4-[<i>trans</i>-4-(<i>trans</i>-4-propylcyclohexyl)cyclohexyl]benzene
;
_chemical_formula_moiety 'C22 H31 N'
_chemical_formula_sum 'C22 H31 N'
_chemical_formula_iupac 'C22 H31 N'
_chemical_formula_weight 309.48
_space_group_crystal_system triclinic
_space_group_name_H-M_alt 'P -1'
_space_group_name_Hall '-P 1'
loop_
_space_group_symop_operation_xyz
'x, y, z'
'-x, -y, -z'
_cell_length_a 9.6230(7)
_cell_length_b 10.1300(18)
_cell_length_c 10.1510(18)
_cell_angle_alpha 104.686(12)
_cell_angle_beta 103.935(14)
_cell_angle_gamma 98.024(15)
_cell_volume 907.6(3)
_cell_formula_units_Z 2
_cell_measurement_reflns_used 93
_cell_measurement_theta_min 3.522
_cell_measurement_theta_max 23.034
_cell_measurement_temperature 173(2)
exptl_crystal_description prism
exptl_crystal_colour 'colourless'
exptl_crystal_size_max 0.400
exptl_crystal_size_mid 0.400
exptl_crystal_size_min 0.300
exptl_crystal_density_diffrn 1.132
exptl_crystal_F_000 340
exptl_absorpt_coefficient_mu 0.064
exptl_absorpt_correction_type multi-scan
exptl_absorpt_process_details
;
```

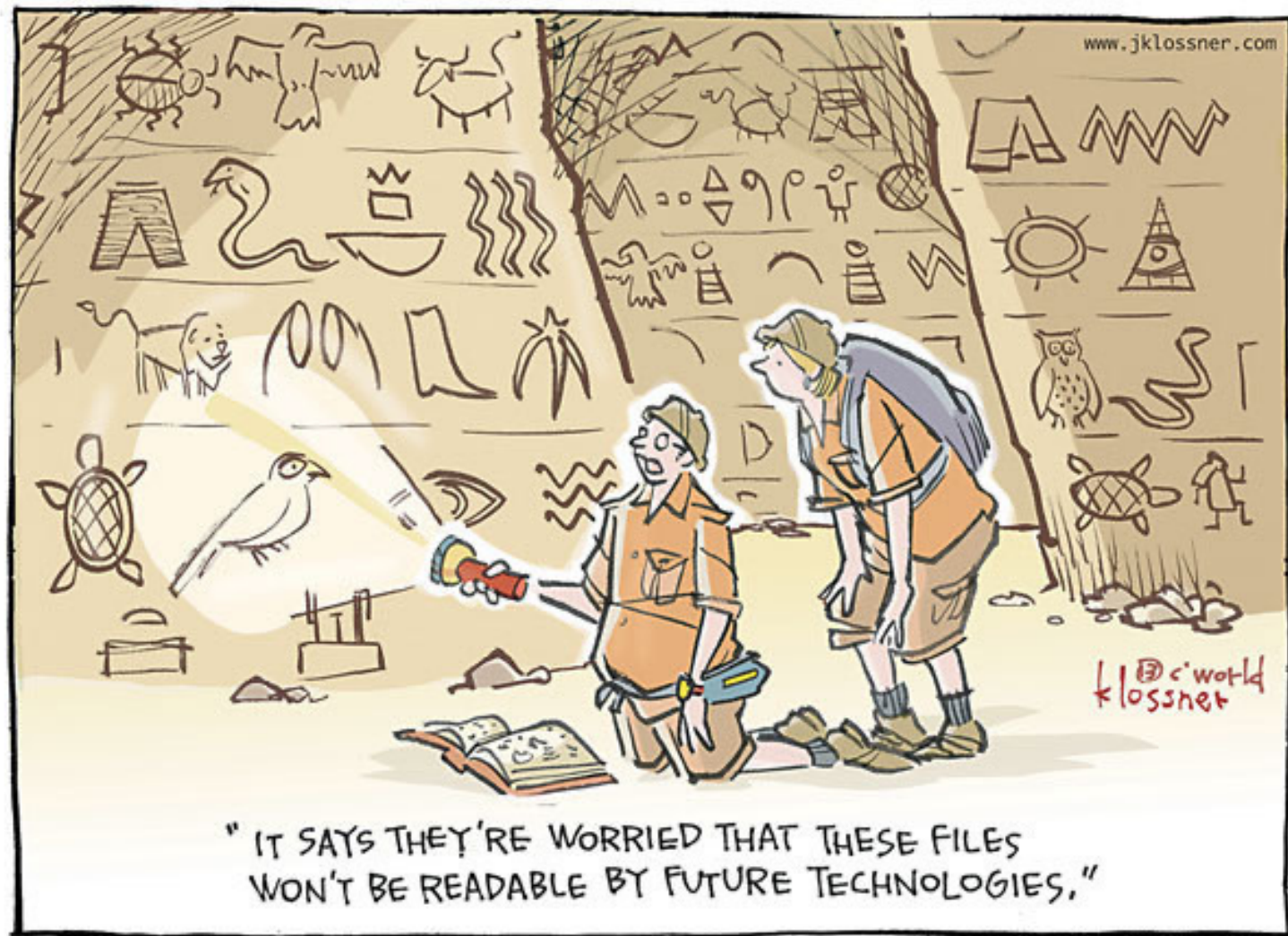
Example: Data Dictionary

(<https://pds.nasa.gov/tools/dd-search/>)

aperture	
Version ID	1.10
Class Name	Telescope
Local Identifier	urn:nasa:pds:context:attribute:0001_nasa_pds_1.pds.telescope.pds.aperture_1a00
Steward ID	pds
Type	PDS4
Namespace ID	pds
Nullable Flag	false
Submitter Name	Submitter_PDS
Definition	The aperture attribute provides the diameter of an opening, usually circular, that limits the quantity of light that can enter an optical instrument.
Registered By	RA_0001_NASA_PDS_1
Registration Authority ID	0001_NASA_PDS_1
Concept	Number

Value Domain	
Enumeration Flag	false
Value Data Type	ASCII_Real
Minimum Characters	Unbounded
Maximum Characters	Unbounded
Minimum Value	0
Maximum Value	1.7976931348623157e308
Pattern	<code>[+]?[0-9]+(\.[0-9]+)?([eE][+]?[0-9]+)?</code>
Unit Of Measure Type	Units_of_Length
Conceptual Domain	Real
Specified Unit ID	m

Share and Preserve It



John Klausner
<https://www.computerworld.com/article/2473350/it-industry/131254-2013-The-tech-year-in-cartoons.html>

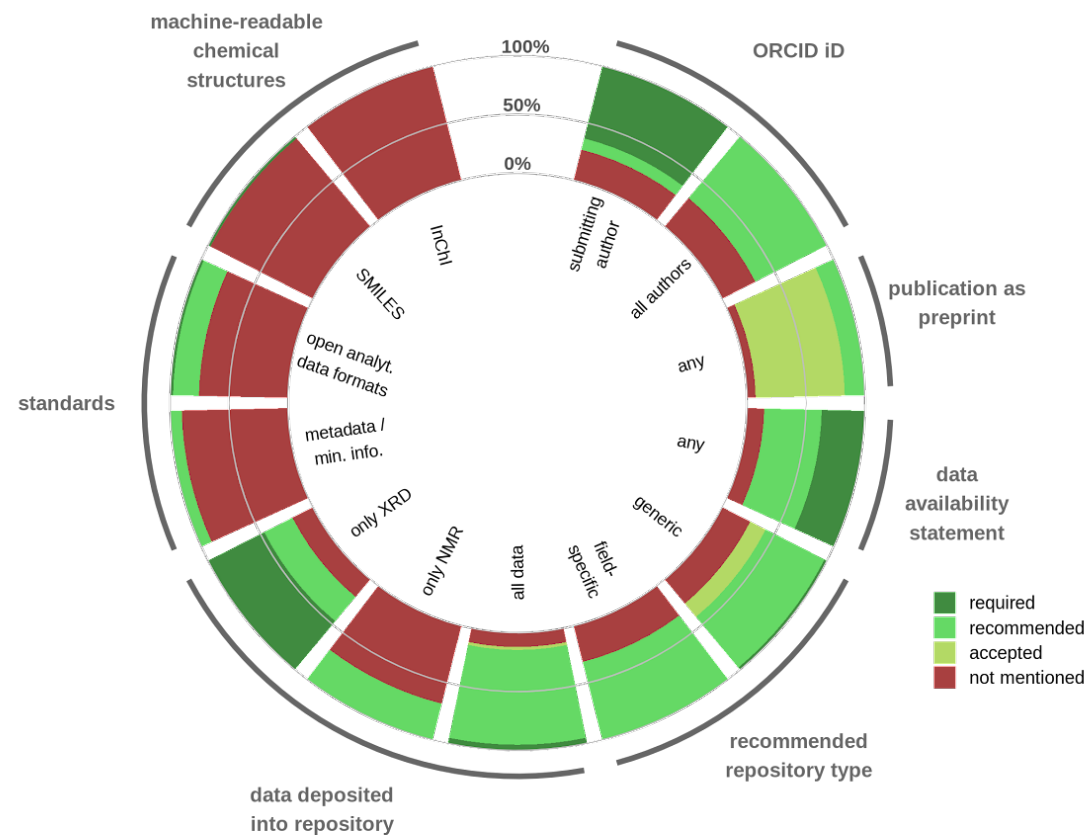
Data Sharing - Funder & Journal Requirements

NSF FAQ about data deposit:

<https://www.nsf.gov/pubs/2018/nsf18041/nsf18041.jsp#q42>

Do I have to deposit the data that support findings in my article in a public access repository?

- Mandatory deposit of data on which an article is based may be required by the journal publisher or other funders. Data collected as part of NSF-funded research, whether or not they are used to support a given publication, should be managed according to the data management plan.*



Summary of data requirements from 42 chemistry journals. Parks *et al.* (2023) *in press*

FAIR Data

“... as open as possible, as closed as necessary”

- Horizon2020 Programme, Open Research Data pilot principle

DATA SHOULD BE

Findable

Use unique and persistent identifiers.



Accessible

Post datasets in open repositories.



Interoperable

Save data in open standard file formats.



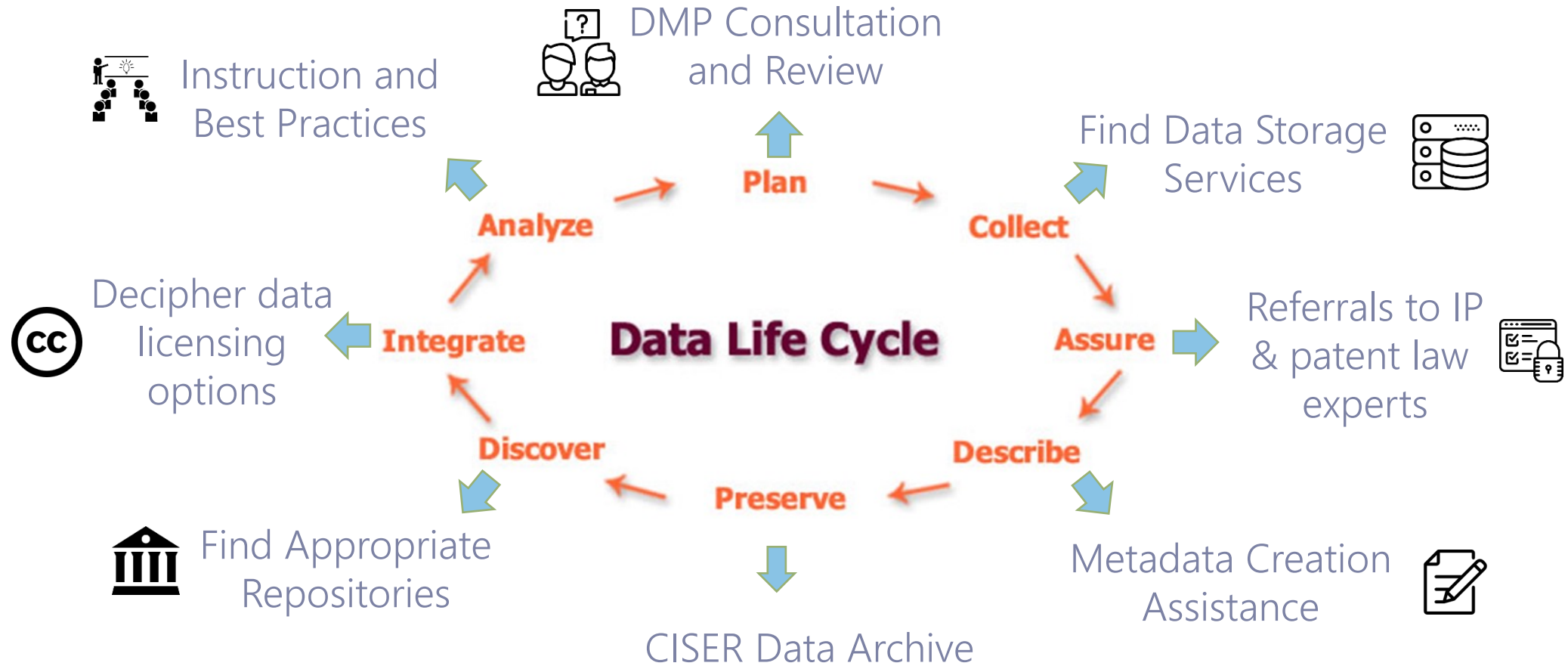
Reusable

Document data provenance and file information in README.



**BY HUMANS
AND MACHINES**

Data Management Services at Cornell



 rdmsg-help@cornell.edu

Research Recordkeeping

✉ rdmsg-help@cornell.edu

🌐 data.research.cornell.edu

2022-01-12



Research requires good recordkeeping

- Imagine you're trying to locate some work from 6 months ago. Would you be able to find your notes and associated data?

Research requires good recordkeeping

- Imagine there's a fire in your lab and all of your paper notebooks and the laptop you left there overnight are destroyed. Would you lose work? Would you be able to recover it?

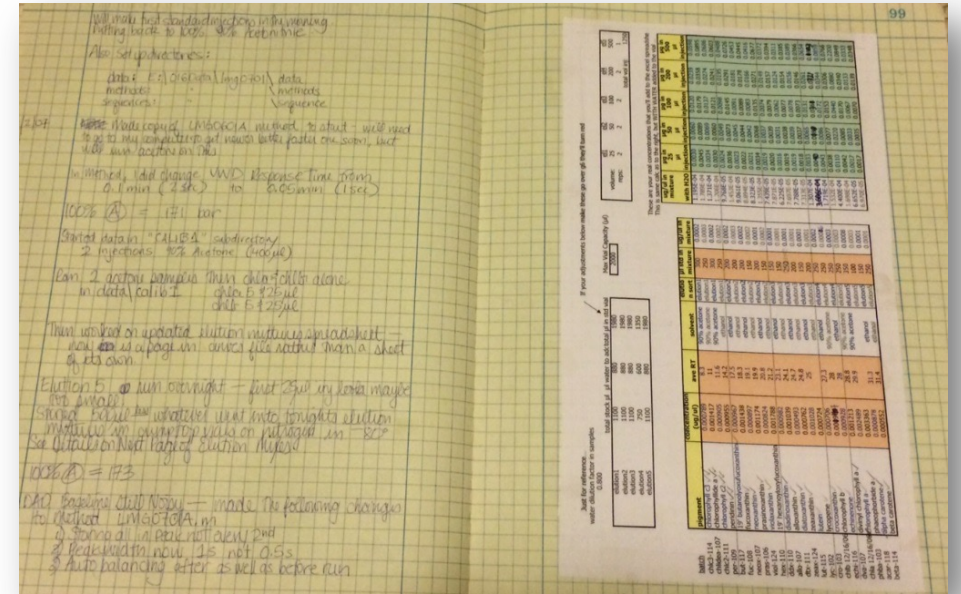
Research requires good recordkeeping

- If you fell under a bus tomorrow, and were temporarily indisposed. Would your supervisor / lab partners / colleagues / professors be able to access your work?

What is a lab notebook?

A laboratory notebook is a legal document that is the primary record of both physical and mental research activity.

Laboratory data include tangible data such as gels, slides, photographs, and computer printouts as well as intangibles such as observations and conclusions. Notes on methodology and process can also be found in lab or field notebooks.



Adapted from: Guidelines for Scientific Record Keeping in the Intramural Research Program at the NIH https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical_conduct/guidelines-scientific_recordkeeping.pdf and UMN's Electronic Lab Notebooks informational page: <https://www.lib.umn.edu/datamanagement/elb>. Photos by Wendy Kozlowski.



What is an Electronic Lab Notebook?

An ELN moves research records to a digital platform, adding functionality for collaboration, templates, searching, version control and file access.



3 minutes: Tools Roundup

Recordkeeping tools – what do you use?



What to consider when selecting an ELN?

- **Disciplinary needs**
- **Your, your lab's, and your collaborators' workflows and tools**
- **Security**
 - Access restrictions
 - Digital signatures
 - Revision management
 - Infrastructure and geographic location of servers (for cloud-based systems)
- **Cost**
- **Source (proprietary vs open source)**
 - For commercial vendors, what are their business practices, preservation policies, and sunseting practices?

Box as a Record Keeping tool (it's more than just storage)

The screenshot shows a web browser window displaying the Box.com interface for a folder named 'Notebook' within the 'Lancaster Lab' folder. The interface includes a sidebar with navigation options, a main file list, and a sharing panel on the right. A red arrow points from a text box to a specific file in the list.

File List:

Name	Updated	Size
Calculators	May 19, 2018 by Sa...	1 File
ChemDraws	Jun 1, 2018 by Sama...	1 File
Literature Preps	May 19, 2018 by Sa...	1 File
2018_04_24 - 2018_001_Synthesis of NMe... 2018_001	May 24, 2018 by Sa...	290.1 KB
2018_05_31 - 2018_002_Oxidation of NMe...	Jun 1, 2018 by Sama...	78.5 KB
2018_06_1 - 2018_003_Oxidation of NMe4_Fe...	Jun 1, 2018 by Sama...	77.4 KB
2018_06_1 - 2018_004_Oxidation of NMe4_Fe...	Jun 1, 2018 by Sama...	78.6 KB

Text Boxes:

- Notebook page - contains references to raw data, tag is used to group together raw information
- Example student notebook - folders with information, individual files are notebook pages

Sharing Panel:

- Lancaster Lab (LL) - Owner
- Samantha MacMillan (SM) - Co-owner
- ga.owner.lancasterlab (g) - Co-owner
- Kyle Lancaster (KL) - Co-owner
- +7 People

© 2018 Box Inc.

LabArchives at Cornell

Customer Log In

[LabArchives](#) is cloud-based Electronic Lab Notebook software that can be used by researchers, instructors, and students for input and organization of laboratory data, information sharing and collaboration, and saving historical versions of files. It is appropriate for use in a wide variety of laboratories, including biology, chemistry, engineering and more.

Find out more at the [LabArchives at Cornell](#) web site.

More Information

- [i Help](#)
- [i FAQ](#)
- [i Privacy](#)
- [i Terms of Use](#)
- [i Get Support](#)

Log in with your **Cornell NetID**

Click "Continue" below to login to LabArchives using your Cornell credentials.

Continue

<https://labarchives.cornell.edu>

LabArchives Basics

- Create Account (<http://labarchives.cornell.edu>) using NetID
- Structure: Notebook >> Folders >> Pages >> Entries
- Add and Manage Content
 - Text
 - Attachments
 - Folder Monitor / Email
 - Widgets
 - Version Control
 - Comments, Tagging, Searching
- Share / Set User Permissions

More tailored ELN options

M BOOK
Desk Search Inventory Message
👤

Status Open

User Keri Steiniger

Start 2018-04-02 15:33

End

References Files

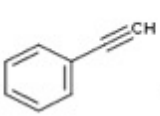
Source

No associated sources

Health/Safety

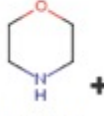
Bibliography

Other



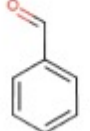
C_8H_8 (R1)

+



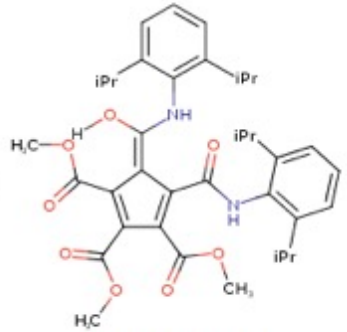
C_4H_9NO (R2)

+



C_7H_6O (R3)

+



$C_{37}H_{46}N_2O_8$ (R4)

Nickel(II) acetate (R5)

→

Compound	MW	W mg	Moles mmol	Equiv	Conc (mol/L)	Vol mL	Density (g/mL)	Pur (%)	Yield (%)
C_8H_8 (R1)	102.134	30.0	0.294	1.000		0.032	0.930	100.0	🗑️
C_4H_9NO (R2)	87.121	20.5	0.235	0.800		0.021	1.000	100.0	🗑️
C_7H_6O (R3)	106.122	20.8	0.196	0.667		0.020	1.049	100.0	🗑️
$C_{37}H_{46}N_2O_8$ (R4)	646.781	6.5	0.010	0.033				100.0	🗑️
Nickel(II) acetate (R5)	176.781	1.8	0.010	0.033				100.0	🗑️

B *I* U ABC \times_2 \times^2 [List Icons] [Undo] [Redo] [Close]

Fonts Colors

Ni(PCCP)2 was dissolved in toluene (2mL) in a 2 dram vial. R1, R2, R3 were added by syringe. Reaction was stirred overnight at 110 degrees.

NMR: kas-1-2018-4-3-1H-2a

Conclusion: No reaction

What if I don't want to switch to a NEW tool?



Browse ▾

Support



Sign Up

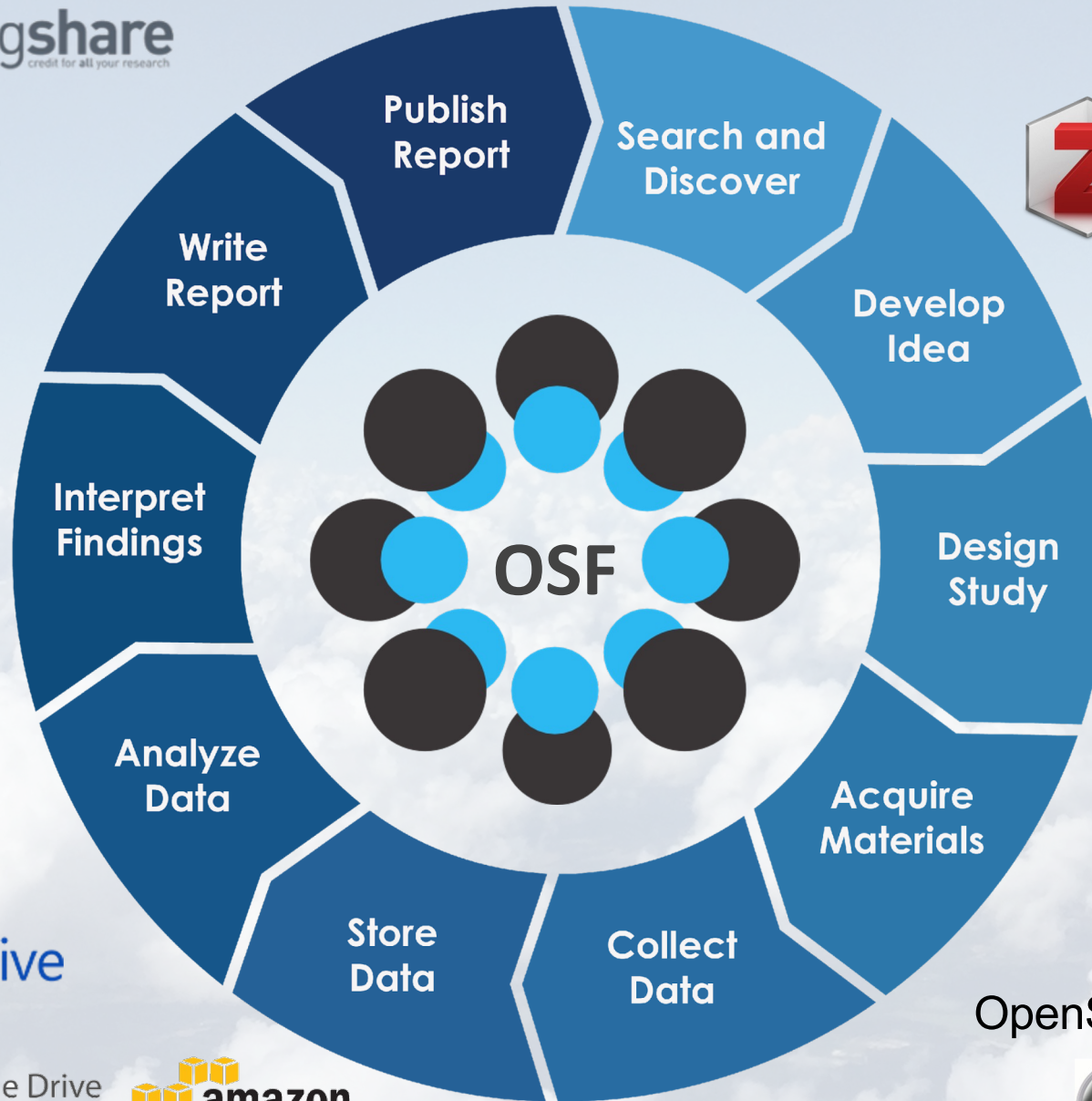
Sign In

Open Science Framework

A scholarly commons to connect the entire research cycle



<https://osf.io/institutions/cornell/>



GitHub



GitLab



Bitbucket



OneDrive



Google Drive



OpenSesame



PsychoPy
Psychology software in Python

OSF Basics

- Create Account (<https://osf.io/institutions/cornell>) using “Institution” (add ORCID later if you have one)
- Structure: Project >> Components
- Add and Manage Content
 - Wiki
 - Add-Ons
 - Tags/Comments
 - Citation support
 - Revision tracking
 - Linking projects
- Share / Set Collaborator Permissions

Resources

- Comparison table of electronic lab notebooks: <https://doi.org/10.5281/zenodo.4723753>
- NIH Guide to lab notebooks: [https://www.training.nih.gov/assets/Lab_Notebook_508_\(new\).pdf](https://www.training.nih.gov/assets/Lab_Notebook_508_(new).pdf)
- <https://www.labfolder.com/electronic-lab-notebook-elN-research-guide/>
- Research Data Management Service Group: <https://data.research.cornell.edu>



✉ rdmsg-help@cornell.edu

🌐 data.research.cornell.edu



This work is shared under a [Creative Commons 0 1.0 Public Domain Dedication](https://creativecommons.org/licenses/by/4.0/).

This work may be copied, modified and distributed without asking permission. Suggested citation: Kozlowski, Wendy and Erica Johns. "Research Recordkeeping, OSF and Lab Archives" Cornell University, Ithaca, NY. 05 November 2021.