

Research Data Management



Digital Tools for the Lab



Instructors



Julie Goldman

Research Data Services Librarian
Countway Library of Medicine
Julie_Goldman@hms.harvard.edu



Meghan Kerr

Archivist and Records Manager
Center for the History of Medicine
Meghan_Kerr@hms.harvard.edu



Slides: <https://datamanagement.hms.harvard.edu/class-materials>



HARVARD
MEDICAL SCHOOL

Data Management
Working Group



Countway Library of Medicine

An Alliance of the Harvard Medical School and Boston Medical Library



Center *for the History of Medicine*

**Harvard Chan Bioinformatics
Core**



hms | hsdm

office for postdoctoral fellows



HARVARD
MEDICAL SCHOOL

OFFICE FOR
Academic and
Research Integrity



Department of
Systems Biology



HARVARD
MEDICAL SCHOOL

Research Information Technology Solutions - RITS

HMS Information Technology

ICCB-Longwood Screening Facility

DRSC/TRiP Functional Genomics

The Neurobiology Imaging Facility

in the Neurobiology Department of Harvard Medical School

Hi+S

Harvard Program in Therapeutic Science

Harvard Biomedical Data Management

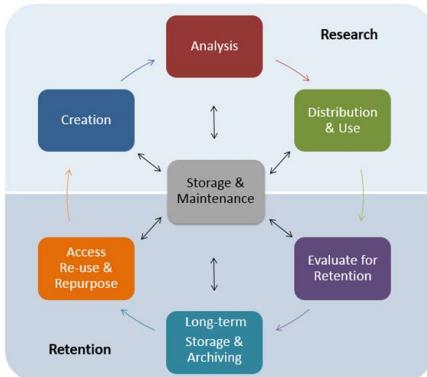
Best practices & support services for research data lifecycles

About ▾ Best Practices ▾ Plan ▾ Store ▾ Share ▾ Resources & Support

Data Management

Data Management is the process of providing the appropriate labeling, storage, and access for data at all stages of a research project. We recognize that best practices for each of these aspects of data management can and often do change over time, and are different for different stages in the data lifecycle.

Early and attentive management at each step of the data lifecycle will ensure the discoverability and longevity of your research.



← December 2018 →

S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					



FEATURED NEWS



DMWG Featured in
Nature Article: How
to pick an electronic
laboratory notebook

Thursday, August 9, 2018

[Submit Questions and Feedback](#)

[Upcoming Trainings & News](#)

[Receive Data Management Updates](#)



HARVARD
MEDICAL SCHOOL



COUNTWAY
LIBRARY

FEATURED ONLINE TRAINING:



[Best Practices for Biomedical
Research Data Management](#)

An open online course aimed at a broad audience on recommended practices for managing research data. Take at your own pace, earn badges and interact with students from around the world!

FEATURED ONLINE TRAINING:



[Research Signer & Reproducibility:
Understanding the Data
Lifecycle for Research Success](#)

An online supplement to an in-person workshop, specifically tailored for Post-Docs. If you are affiliated with Harvard, you may receive a course certificate to promote your time taken on this topic.

Licensed under the Creative Commons Attribution Non-Commercial License unless otherwise noted.

Last Updated: 2018-08-09

Harvard Biomedical Data Management Website

<https://datamanagement.hms.harvard.edu>

Introduce Yourself!



Name

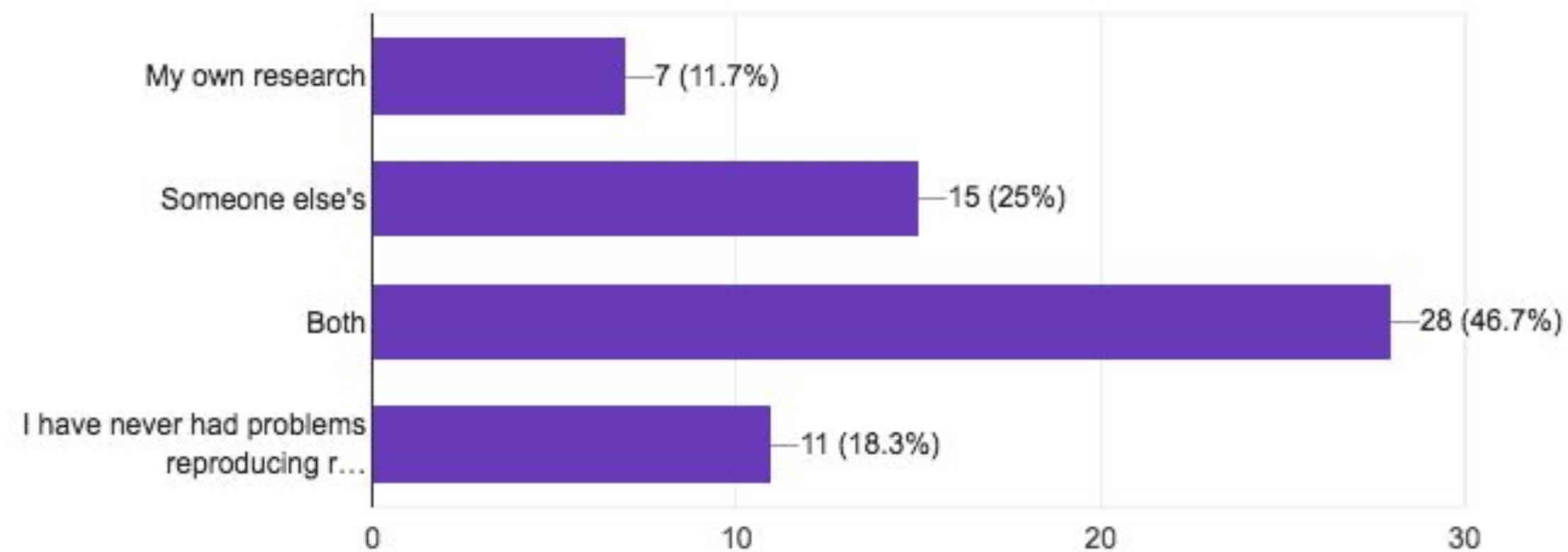
School / Department

Do you or your lab already use digital and collaborative tools? Has this helped you?

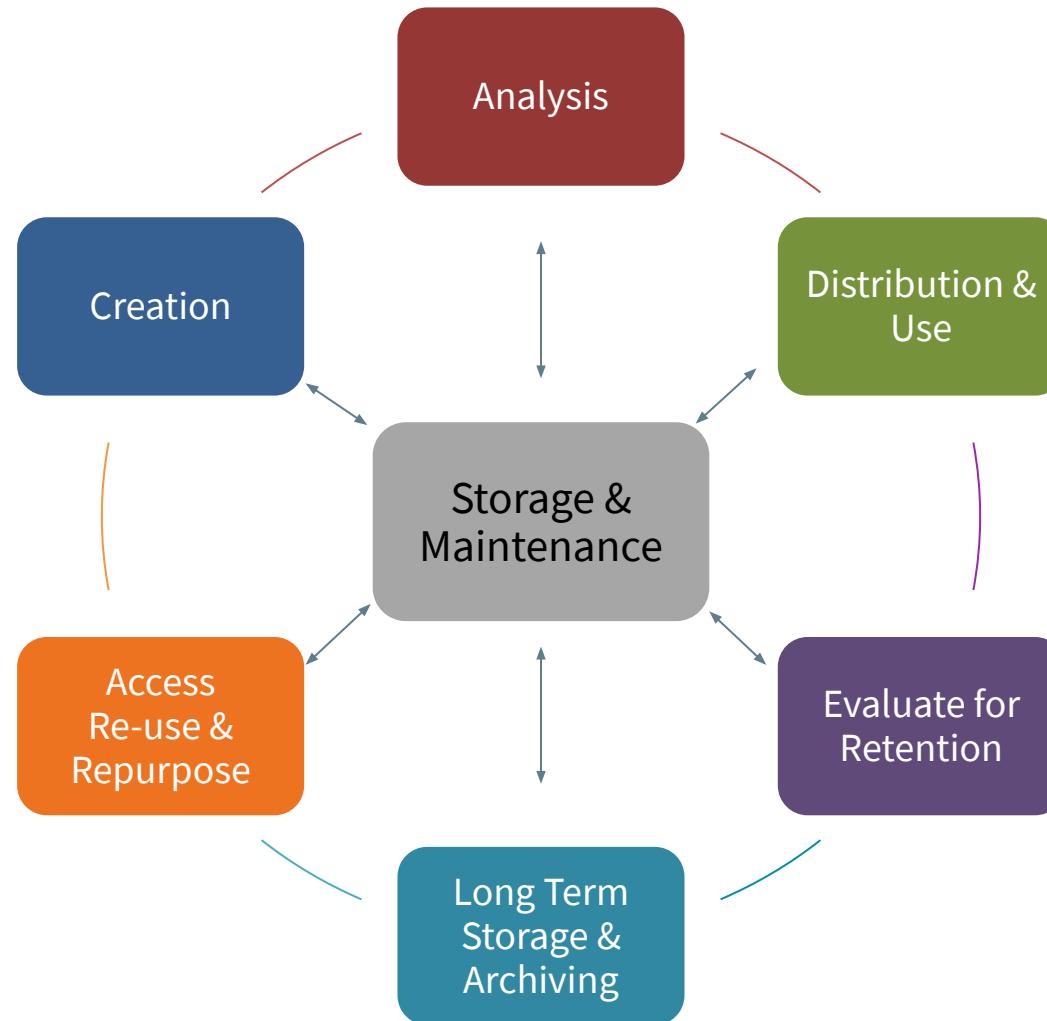
(ELNs, GitHub, OSF, Jupyter, etc.)

Have you ever had problems reproducing your own or someone else's research?

60 responses



Data Lifecycle for Biomedical Data



Pre-publication v. Post-publication Data

- Lot of management work done in second area
- This talk focuses on first area
- Data management very important in pre-publication
 - Help scientists find their own data, saving time & effort
 - Help scientist share data
 - Even within a lab this is important!
 - Help PI's keep track of data long-term as students leave
 - Makes post-publication management smoother

Current Collection Practices

- Practices vary lab-to-lab AND researcher-to-researcher
- Very difficult to locate another's data currently
- Common practice in lab is to test an experiment against another's data!

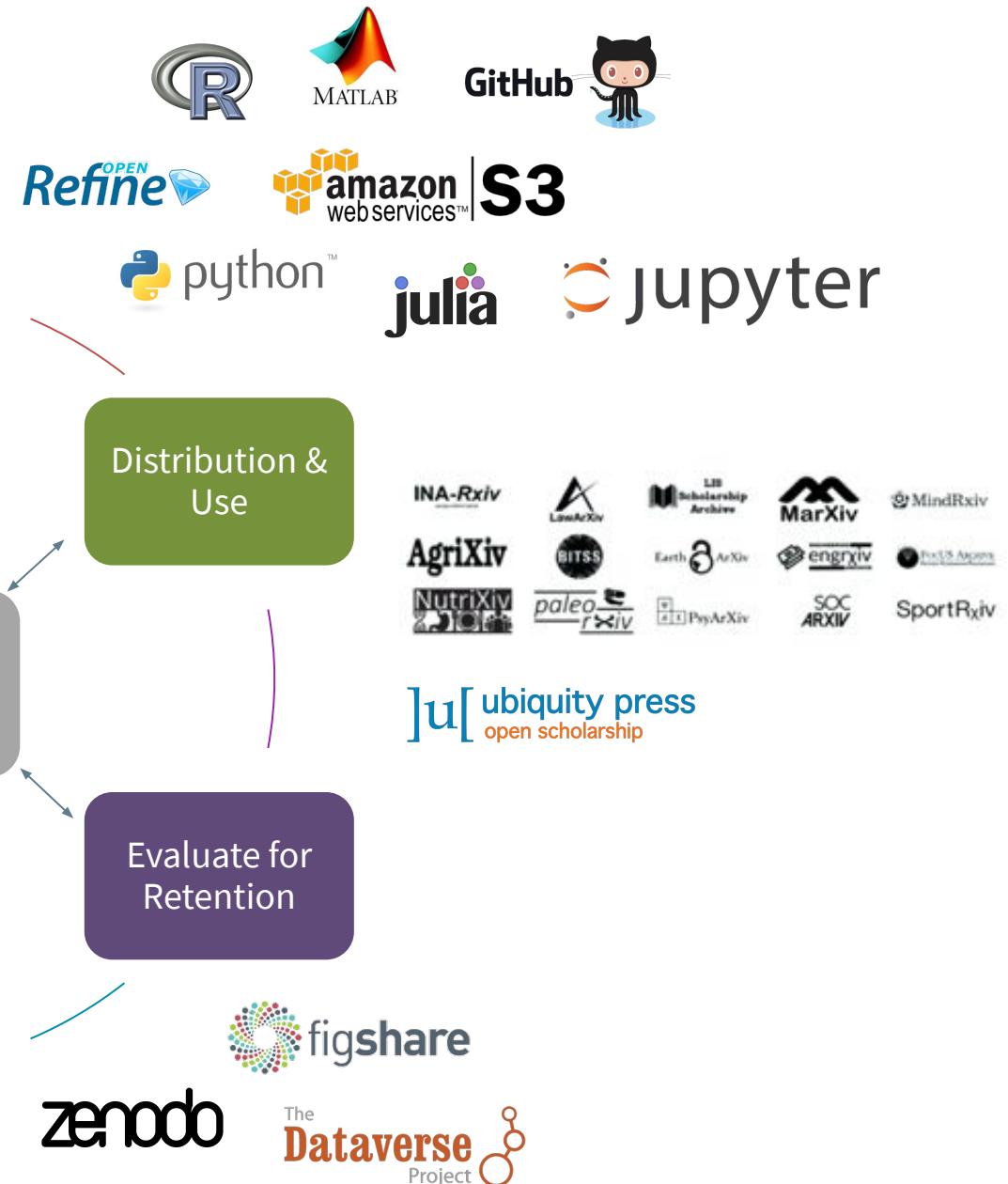
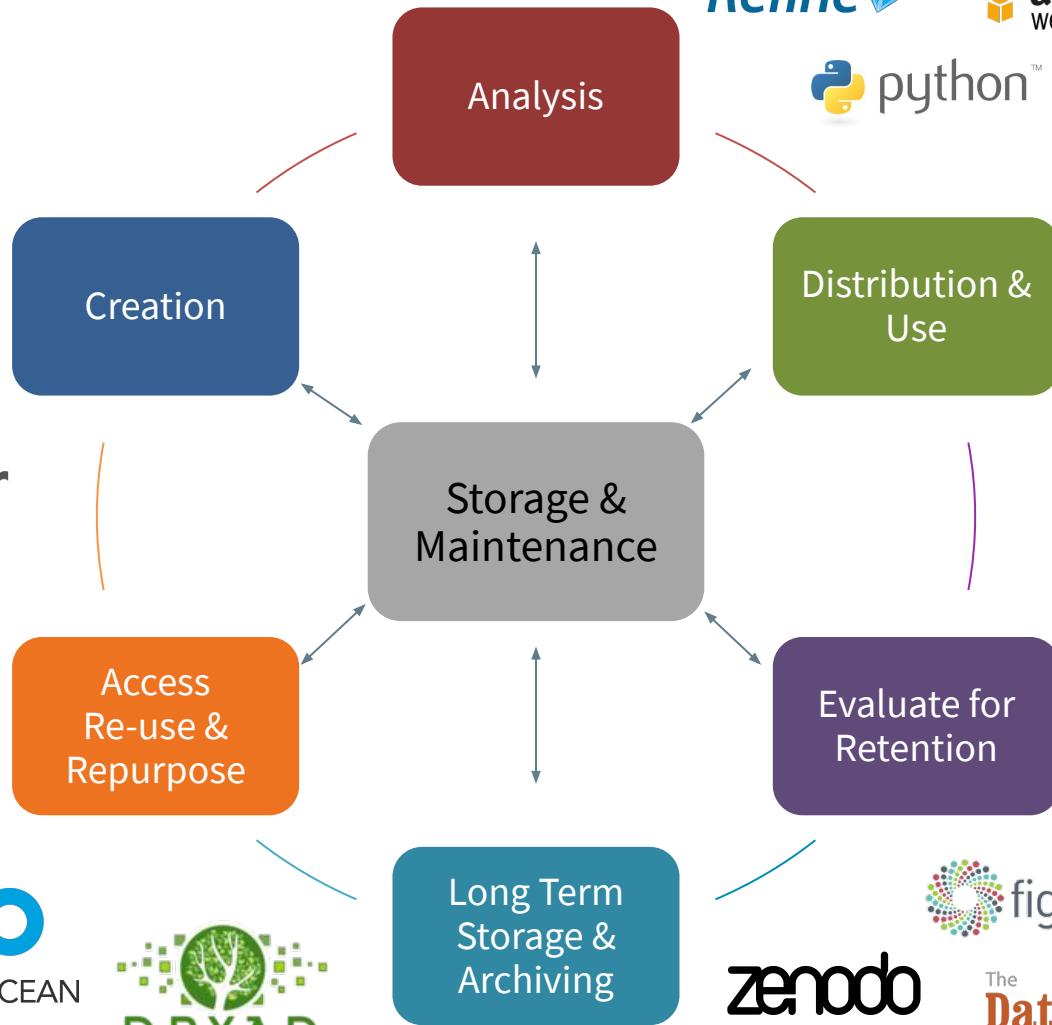
Current Collection Problems

- Pointing metadata to data
- Long-term concerns with keeping data and metadata together
- Back-ups are an issue
- No standard for digital file organization

Ideally...

- Metadata are digital
- Data are digital and stored on computer
- Notebooks store or point to data

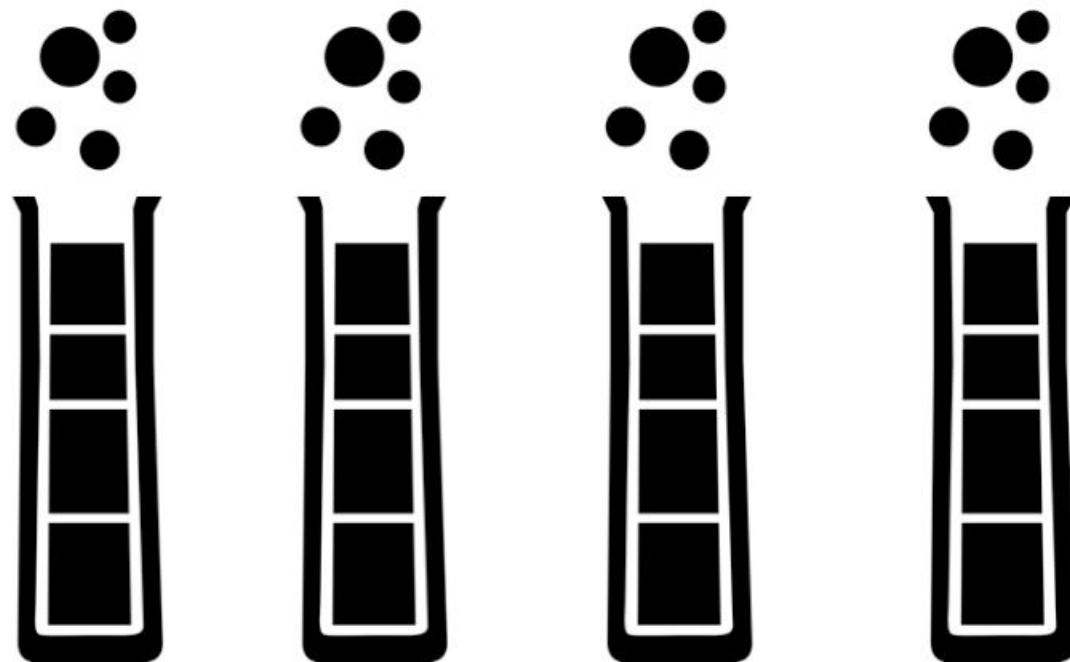
Digital Workspaces



Ideal Collection Process

- Electronic lab notebook
 - Metadata and copies of data embedded in same file
 - Integration of outside material/files
 - Plugins like a chemical structure creator
 - Digital signatures and date stamps
- Easily searchable
- Shareable, but control who has access to data
 - Other issues here like de-identifying data, etc

Goal = Reproducibility



Science!



Not Science.

Practical Tips for Reproducibility

1. Plan for reproducibility before you start
2. Keep track of things
3. Report your research transparently
4. Archive & share your materials

Plan for reproducibility before you start

Write a study plan or protocol and track new versions.

Set-up a reproducible project using an electronic lab notebook to organize and track your work. Avoid saving proprietary file formats.

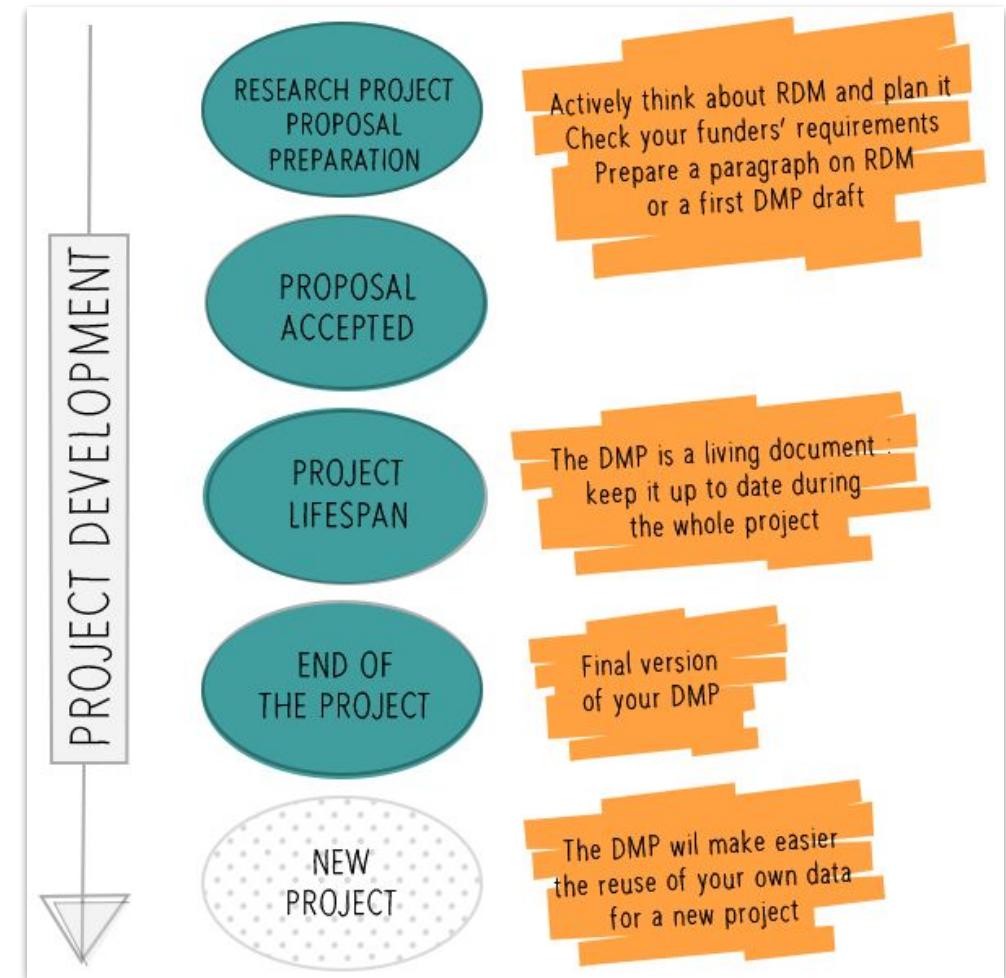
Why Manage Data?

- Easier to analyze organized, documented data
- Avoid accusations of fraud & misconduct
- Don't lose data
- Find data more easily
- Get credit for your data
- Don't drown in irrelevant data



Data Management Plan

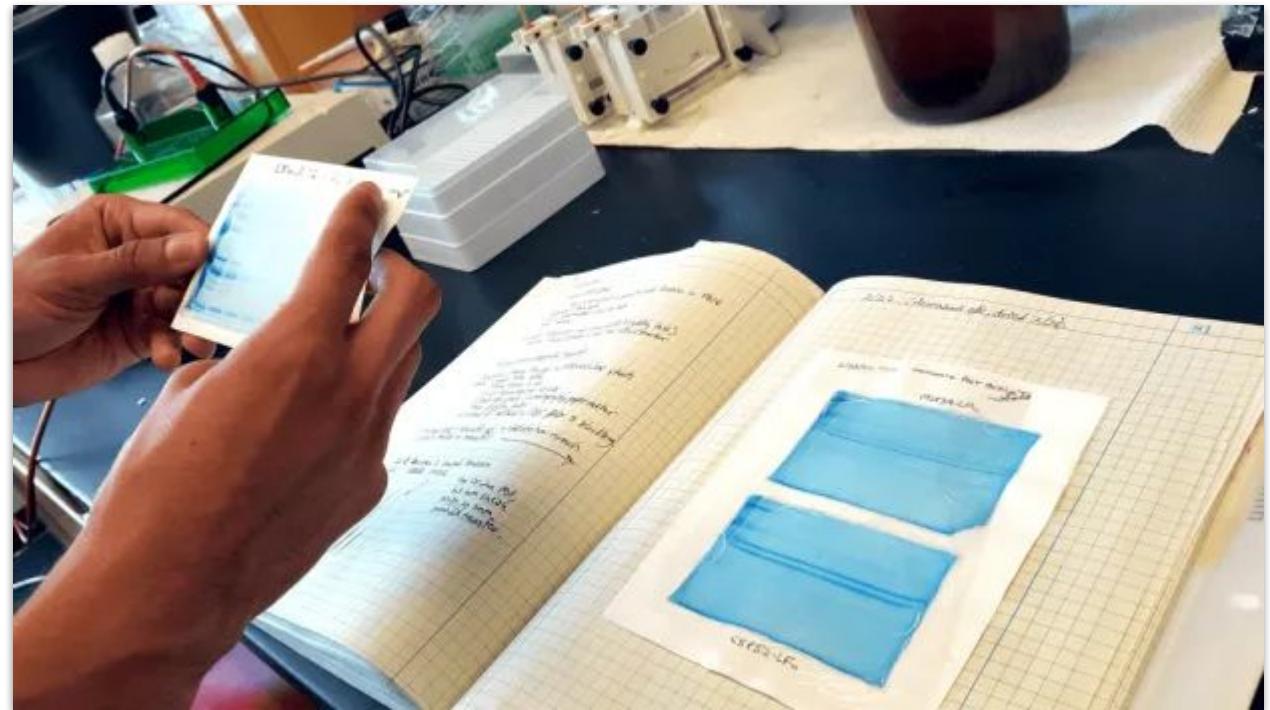
A data management plan (DMP) is a written document that describes the data you expect to acquire or generate during the course of a research project, how you will manage, describe, analyze, and store those data, and what mechanisms you will use at the end of your project to share and preserve your data.



<https://researchdata.epfl.ch/plan-fund/dmp>

Record Keeping

- Complete record of procedures, reagents, data, and thoughts to pass on to other researchers
- Explanation of why experiments were initiated, how they were performed, and the results
- Legal document to prove patents and defend your data against accusations of fraud



A graduate student at the [Harvard AIDS Initiative](#) examines a gel used to study T cell responses in HIV-infected individuals. Consider the Lab Notebook (<https://aids.harvard.edu/consider-the-lab-notebook>)

Here's the scenario:

Your group leader comes to you and asks you to retrieve a particular experiment that he performed 5 years ago.

First, you have to find that particular notebook among the hundreds that are stored in the lab's several filing cabinets. Half an hour and one notebook later, you find it, but it is really hard to organize the information you have acquired.

A lot of details are missing, and you can't even decipher the handwriting in some cases.

You end up setting up most of the methodology from scratch, and of course you aren't surprised that your experiments show different results than those observed 5 years ago.



Marta Teperek
@martateperek

Follow

Trying to find information in colleagues' hard copy lab books is a task for a detective ...
@HyvonenGroup #elncam

Notebook no.	Date 13 January 2017	101
Title Work for Sherlock Holmes		Continued from page

Marko Hyvonen
mh256@cam.ac.uk

5:33 AM - 13 Jan 2017

Electronic Lab Notebooks

Benefits:

- Searchable
- Export data
- Easily shareable
- Easily accessible
- Use the mobile App
- Embed images, protocols, etc

Considerations:

- Cost
- Software
- Storage
- Security
- Interactivity
- Migration (in & out)

Electronic Lab Notebooks

Electronic Lab Notebooks at HMS

Lab notebooks are good for writing down procedures, observations, conclusions and for drawing flow charts and diagrams by hand. However, in order to accommodate the increase of digital data collected, researchers have taped instrumentation and computer printouts onto the pages of their notebooks, or cross-referenced larger data sets by recording file names and locations in the notebook.

An ELN (electronic lab notebook) is a software tool that in its most basic form is a digital version of a page in a paper lab notebook. In this electronic notebook you can record your observations, draw graphs, store data and images, and search for data using your computer or mobile device. This offers several advantages over a paper notebook:

The number of available ELN tools is increasing and the functions available in each tool may be confusing to evaluate all of the advantages and limitations of each tool for your project.

The Electronic Lab Notebook Matrix has been created to aid HMS researchers in evaluating the various ELN tools currently available. This matrix allows researchers to compare and contrast the numerous solutions available and select the best options in-depth.

Questions about Electronic Lab Notebooks at HMS? Contact us at datamanagement@hms.harvard.edu

Features	Specifications															
	Benchling	Biovia	Confluence	Doccollab	ECL	ELOG	Evernote	Exemplar	Findings	Hivebench	IDBS	LabArchives	LabCollector	LabWare	LabVantage	LabW
Interactivity																
Intuitive Interface Design	✓	No response received	*	*	No response received	*	No response received	No response received	*	*	*	✓	No response received	*	*	
Auto Metadata Harvest	*	No response received	✗	✓	No response received	✗	No response received	No response received	✗	✓	*	✓	No response received	✗	✓	
Search functions can search across file formats and beyond types	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
Ability to manipulate files and images	*	No response received	*	*	No response received	*	No response received	*	*	*	*	*	*	*	*	
Support for multiple open windows	✓	*	✓	✓	✓	No response received	✓	✓	*	✓	✓	✓	✓	*	*	
Ability to link out	✗	No response received	*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	No response received	✓	
Support for Researcher Documentation																
Hyperlink support	✓	No response received	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	No response received	✓	✓	
Metadata Creation Prompts	✓	No response received	✗	✓	✓	No response received	✗	No response received	✗	✓	✓	✓	No response received	✗	✓	
Rights Management (licensing)	*	No response received	*	✓	✓	No response received	*	No response received	✓	*	✓	✓	✓	✓	*	
Protocol Integration	✓	*	✓	✓	✓	No response received	✓	*	*	✓	✓	✓	✓	*	*	
Adaptability to Lab workflows																
Accounts/Permissions Levels	✓	No response received	*	✓	✓	✓	✓	✓	*	*	✓	✓	✓	✓	*	
Internal Data Sharing	✓	*	*	✓	✓	No response received	✓	✓	✓	✓	✓	✓	✓	✓	*	
Adaptable to a Variety of Workflows	*	No response received	*	*	*	No response received	*	No response received	*	*	*	*	*	*	*	
Compatibility with authoring tools	✓	No response received	*	✓	✓	No response received	✗	No response received	✗	*	*	*	*	No response received	✗	
Windows Compatible	✓	No response received	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	*	
Macintosh Compatible	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	*	
Linux Compatible	✓	✓	✗	✓	✓	✓	✓	No response received	✓	✓	✓	✓	✓	No response received	✓	
Android Compatible	✓	✓	✓	✓	✓	✓	✓	No response received	✓	✓	✓	✓	✓	✓	*	
iOS Compatible	✓	✓	✓	✓	✓	✓	✓	No response received	✓	✓	✓	✓	✓	✓	*	
Storage																
Cloud Storage	✓	No response received	✗	✓	✓	No response received	✓	No response received	✓	✓	✓	✓	✓	✓	*	
Local Storage	*	No response received	✓	✓	✗	No response received	✓	No response received	✓	✓	✓	✓	✓	No response received	✓	
Hybrid (cloud/local) Storage	*	No response received	✗	✗	No response received	✗	No response received	No response received	✓	✓	✗	✓	✓	No response received	✗	
Versioning	*	*	*	*	*	No response received	*	No response received	*	*	*	*	*	*	*	
File Redundancy	*	No response received	*	*	No response received	*	No response received	*	No response received	*	*	*	*	No response received	*	
Creates stable URLs or persistent identifiers for entries	✓	No response received	✓	✓	No response received	✓	No response received	✓	No response received	✓	✓	✓	✓	No response received	✓	
Can unregistered users access the data found at persistent links?	✓	No response received	✓	✗	No response received	✗	No response received	✓	No response received	*	*	*	*	No response received	✗	
Storage Capacity - Users	*	No response received	*	*	No response received	*	No response received	*	No response received	*	*	*	*	No response received	*	
Storage Capacity - Max File Size	*	No response received	*	*	No response received	*	No response received	*	No response received	*	*	*	*	No response received	*	

Electronic Lab Notebook Matrix

<https://datamanagement.hms.harvard.edu/electronic-lab-notebooks>

OpenLAB ELN

For Biologists/Genome Researchers



For those studying genomic variations, OpenLAB ELN safeguards data quality, documents all sample findings and information, and makes key information readily available and searchable.

- Quickly enter scientific data, from all types of life science experiments, whether it is structured (Excel sheets, instrument data, graphs, DNA sequences) or unstructured information (gel images, blots, photos, videos)
- Reuse protocols and methods
- Collaborate on shared experiments

EXAMPLE WORKFLOW



Plan



Hybridized Samples



Analysis



Results

- DNA extraction
- Sample preparation
- Digestion
- Labelling
- Hybridization
- Scan & analyze

Open Science Framework

The image shows the Open Science Framework (OSF) website. At the top, there's a dark header bar with the OSF logo, a search bar, support links, a donation button, and two buttons for "Sign Up" and "Sign In". Below the header is a large, dark blue background featuring a network graph of nodes and connections. In the center, the text "Open Science Framework" is displayed in a large, white, sans-serif font, with the subtitle "A scholarly commons to connect the entire research cycle" underneath. To the left, a screenshot of the OSF dashboard is shown, displaying a list of projects and a prominent orange play button icon. On the right, there's a sign-up form with fields for "Full name", "Contact email", and "Confirm email", followed by a checkbox for accepting terms and policies, and a "Sign up free" button.

OSF HOME ▾

Search Support Donate Sign Up Sign In

Open Science Framework

A scholarly commons to connect the entire research cycle

FREE AND OPEN SOURCE. START NOW.

Full name

Contact email

Confirm email

I have read and agree to the [Terms of Use](#) and [Privacy Policy](#).

Sign up free



Lab Example

Public

8

...

Contributors: Ian Sullivan, Courtney K. Soderberg, Jennifer Freeman Smith, Brandon Thorpe

Affiliated institutions: Center For Open Science

Date created: 2017-06-06 12:00 PM | Last Updated: 2018-03-22 03:41 PM

Category: Project

Description: This project demonstrates one way the OSF can be used to create a space for a lab to share materials and research.

License: CC0 1.0 Universal [i](#)

Wiki

This is an example project showing how the OSF might be used by a lab to create a shared lab space, share lab standards and resources, and collate the work that is being done by individuals/groups within the lab.

Getting started

Welcome to the lab! This wiki contains the steps to take at key moments in the life of a lab like when someone joins the lab, when you start a new experiment or when the p...

[Read More](#)

Files

[Filter](#)



Citation

Components

New Experiment Template

Sullivan, Soderberg, Smith & 1 more

Start all new experiments by forking this section.

Lab Documents

Sullivan, Soderberg, Smith & 1 more

Internal documentation and lab reference material should go here.

Lab Meetings

Sullivan, Soderberg, Smith & 1 more

Discussion notes, articles, and other internal musings.

Example 1: OSF Example Project



OSFHOME ▾

Search

Support

Donate

Sign Up

Sign In

JHU Electronic Lab Notebook Template

Files

Wiki

Analytics

Registrations



JHU Electronic Lab Notebook Template

Public

P 6

...

Contributors: [Reid Boehm](#), [Chen Chiu](#), [Dave Fearon](#)

Affiliated institutions: [Johns Hopkins University](#)

Date created: 2017-03-15 03:36 PM | Last Updated: 2018-04-13 12:09 PM

Category: Project

Description: A template for researchers to use to structure an OSF project as an Electronic Lab Notebook

License: CCO 1.0 Universal

Wiki



JHU Electronic Lab Notebook Template

by

[JHU Data Management Services](#)

JHU DMS created this template to get you started using OSF as a Lab Notebook. Each component within uses the wiki space to provide a brief explanation about suggested content and in some cases links to other helpful resources.

Unlike some electronic lab notebooks, this design works mainly as a "binder" for gathering and acces...
[Read More](#)

Citation



Components

Data Management Plan

[Boehm, Chiu & Fearon](#)

A space to include your data management plan

Literature

[Boehm, Chiu & Fearon](#)

A space to collect associated literature that relates to your experiment.

Protocols

[Boehm, Chiu & Fearon](#)

A space to include the documentation of the governing regulations and processes of the experiment/project

Example 2: JHU ELN Template

Activity! ELN Scorecard

Form groups!

Each group will evaluate one ELN option for one of the scorecard categories

Use the ELN Matrix to score
the ELN (or go to the product
webpage for more information)

Add “additional considerations” to think about when evaluating ELNs!

Final Thoughts on ELNs

Paper and electronic lab notebooks can work together

It may take time to find a solution

One size does not fit all!

Really think about:

- What happens when you leave?
- How can people access the correct version of your data?
- Is it even easily accessible for you?

Keep track of things

Preregister important study design and analysis information. Free tools to help you make your first registration include AsPredicted, Open Science Framework, and Registered Reports. Clinical trials use Clinicaltrials.gov.

Track changes to your files using version control.

Document everything done by hand in a README file and data dictionary.

Version Control

Records changes

Keeps track of change history

Illustrates changes between versions

Lets you share more easily

Lets you collaborate more easily



THIS IS GIT. IT TRACKS COLLABORATIVE WORK
ON PROJECTS THROUGH A BEAUTIFUL
DISTRIBUTED GRAPH THEORY TREE MODEL.

COOL. HOW DO WE USE IT?

NO IDEA. JUST MEMORIZIZE THESE SHELL
COMMANDS AND TYPE THEM TO SYNC UP.
IF YOU GET ERRORS, SAVE YOUR WORK
ELSEWHERE, DELETE THE PROJECT,
AND DOWNLOAD A FRESH COPY.



Metadata

Data documentation provides the information necessary to fully understand and interpret the data

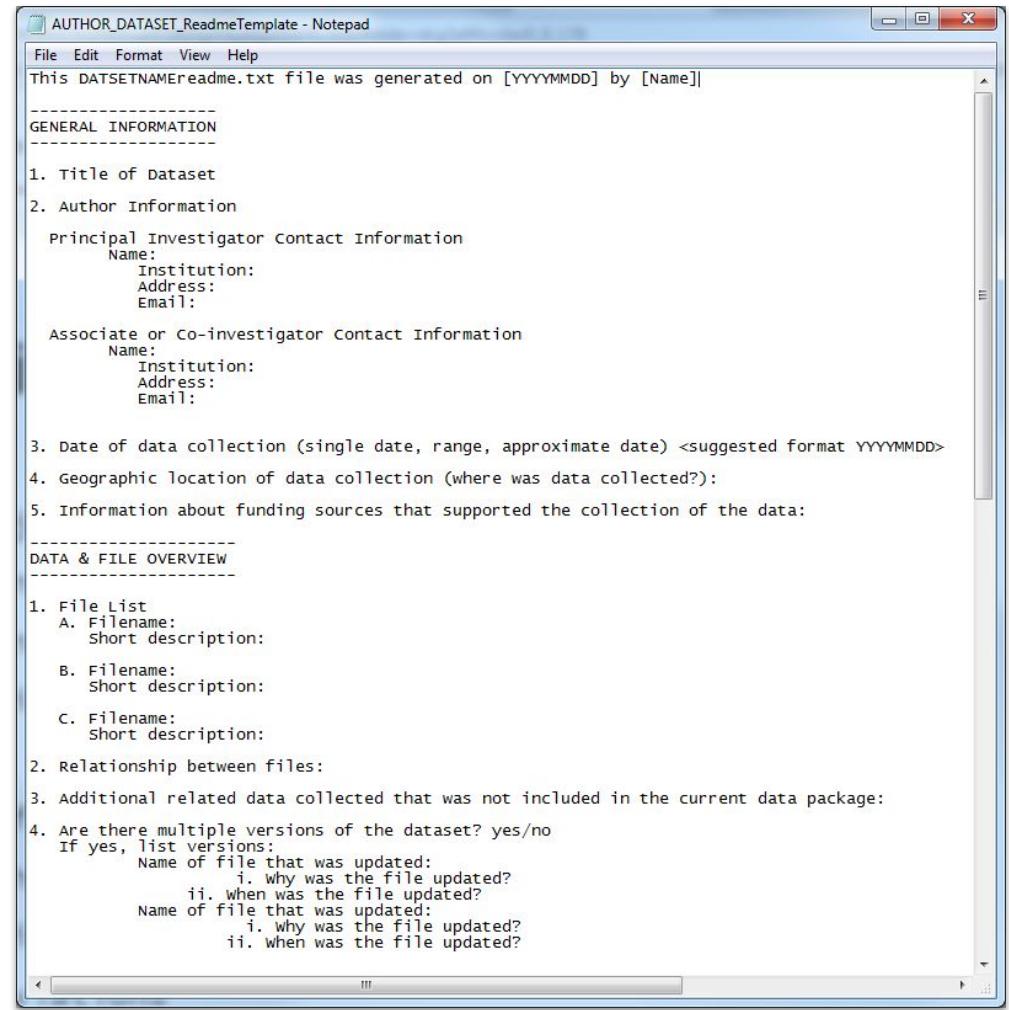
Metadata should be standardized, consistent and interoperable, and facilitates discovery, preservation and archiving of data



Andy Warhol, *Big Torn Campbell's Soup Can (Pepper Pot)*, 1962 The Andy Warhol Museum, Pittsburgh Founding Collection, Contribution The Andy Warhol Foundation for the Visual Arts, Inc.

README File

- Basic project information
- Title, Contributions, Grant Info
- Contact information
- All locations of where data live, including backups
- Useful information about the files and how they are organized
- Explain file naming conventions and abbreviations



Report your research transparently

Collect, organize, cite, and share research with reference management tools Mendeley, Refworks or Zotero. **Collaborate on manuscripts** using collaborative writing and publishing systems including Authorea and Overleaf.

Share your protocols and interventions explicitly and transparently.

Write a transparent report. Guidelines from the Equator Network or processes like Registered Reports can help.

Methodology

Simply sharing data and code is not enough to allow for scientific reproducibility, resulting in a rising focus on the importance of reporting detailed methods

Common but unhelpful:

“contact author for details”

“we used a slightly modified version of the protocol reported in paper XYZ”



Morgan Halane
@themorgantrail

Follow

Looking for protocol in 1997 paper: "as described in (x) et al '96". Finds '96 paper: "as described in (x) '87." Finds '87 paper: Paywall.



9:20 PM - 1 Nov 2017 from Pohang-si, Republic of Korea

35 Retweets 83 Likes



protocols.io

protocols > Running VirSorter in the CyVerse Discovery Environment

protocols groups researchers news

Description Guidelines Steps

▶ STEP 1 DESCRIPTION

Register an account with CyVerse.

▼

EXTERNAL LINK

<https://user.cyverse.org/register/>

▶ STEP 2 DESCRIPTION

EDITING STEP 1

Search components

Amount

Annotation

Command

Dataset

Duration / Timer

Expected result

External Link

Reagent / Mixture

Section

Software package

The screenshot shows a web-based protocol editor on the protocols.io platform. At the top, there's a navigation bar with links for 'protocols', 'groups', 'researchers', and 'news'. Below the navigation is a dark header bar with tabs for 'Description', 'Guidelines', and 'Steps'. The 'Steps' tab is active, indicated by a blue underline. On the left, a vertical sidebar has icons for navigating between steps (forward, back, first, last), a search bar, and a trash bin. Step 1 is titled 'STEP 1' and has a 'DESCRIPTION' section containing a text input field with a toolbar and a note to 'Register an account with CyVerse.' Step 2 is titled 'STEP 2' and has an empty 'DESCRIPTION' section. Below the steps, there's an 'EXTERNAL LINK' section with a button to copy the URL 'https://user.cyverse.org/register/'. To the right of the steps is a sidebar titled 'EDITING STEP 1' which lists various components that can be added to a step, each in its own blue-bordered box: 'Amount', 'Annotation', 'Command', 'Dataset', 'Duration / Timer', 'Expected result', 'External Link', 'Reagent / Mixture', 'Section', and 'Software package'.



High quality DNA from Fungi for long read sequencing e.g. PacBio, Nanopore MinION Version 2

Benjamin Schwessinger and Megan McDonald¹

¹ Australian National University

23 Extraction I
Pour off ethanol and remove remaining ethanol with pipette. Spin down briefly and remove any remaining ethanol with pipette.

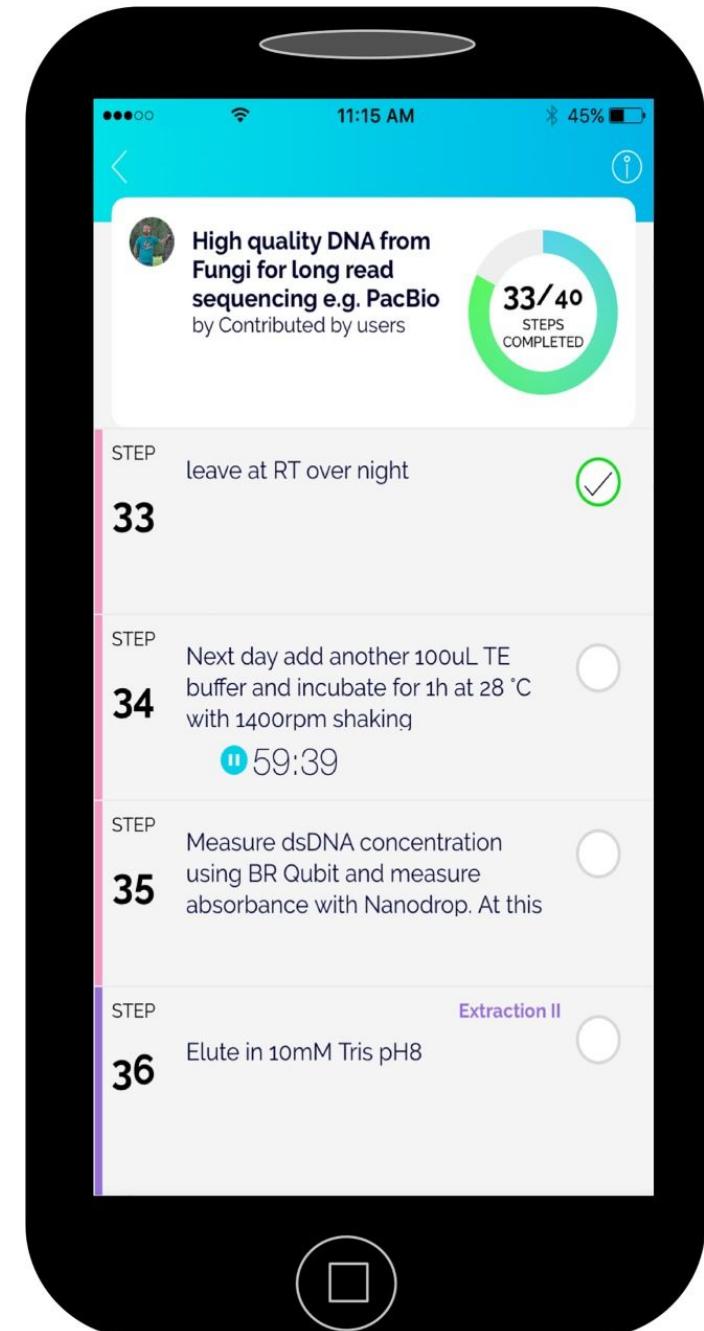
24 Extraction I
Air-dry pellet for 2-2 mins

TIMER
00:07:00

NOTES

Jet Beekwilder September 14, 2017 01:20:07
Hi Megan, what do you mean by 2-2 minutes? The timer says 7 minutes, did you mean 2-7 minutes?

[Reply](#)



Manuscript Preparation

Collaborative Writing Tools

- Code-writing: GitHub, Bitbucket, GitLab, CodePlex
- Collaborative real-time editors: Etherpad, Hackpad, GoogleDocs, Microsoft Office, Authorea
- Wikis: Wikipedia, Wikia

LaTeX

High-quality typesetting system

Features designed for the production of technical and scientific documentation

LaTeX is the de facto standard for the communication and publication of scientific documents

[Free online introduction to LaTeX](#)

Overleaf

Overleaf v2—Launch Announcement

SIGN UP SIGN IN

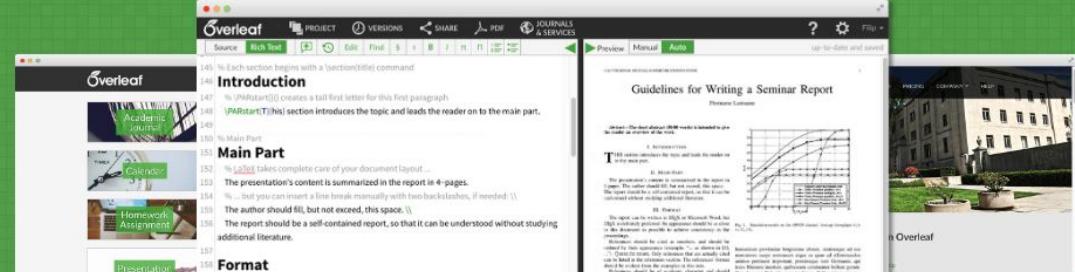
Overleaf

FEATURES & BENEFITS ▾ TEMPLATES PRICING COMPANY ▾ HELP

Collaborative Writing and Publishing

Join 2,000,000 authors enjoying the easiest way to create, collaborate and publish online

Start writing now! CREATE A NEW PAPER



Over 17 million projects, authors from 3600 institutions and more than 2400 templates

Overleaf is an online LaTeX and Rich Text collaborative writing and publishing tool that makes the whole process of writing, editing and publishing scientific documents much quicker and easier.

Source

Rich Text



Edit

Find



B

I



Preview Manual Auto

refresh preview

Template for preparing your research report submission to PNAS using Overleaf

1

51 \dropcap{T}his PNAS journal template is provided to help you write your work in the correct journal format. Instructions for use are provided below.

52

53 Note: please start your introduction without including the word ``Introduction'' as a section heading (except for math articles in the Physical Sciences section); this heading is implied in the first paragraphs.

54

55 Guide to using this template on Overleaf

56

57 Please note that whilst this template provides a preview of the typeset manuscript for submission, to help in this preparation, it will not necessarily be the final publication layout. For more detailed information please see the \href{http://www.pnas.org/site/authors/format.xhtml}{[PNAS Information for Authors]}.

58

59 If you have a question while using this template on Overleaf, please use the help menu ('?') on the top bar to search for \href{https://www.overleaf.com/help}{[help and tutorials]}. You can also \href{https://www.overleaf.com/contact}{[contact the Overleaf support team]} at any time with specific questions about your manuscript or feedback on the template.

60

61 Author Affiliations

62

63 Include department, institution, and complete address, with the ZIP/postal code, for each author. Use lower case letters to match authors with institutions, as shown in the example. Authors with an ORCID ID may supply this information at submission.

64

65 Submitting Manuscripts

66

67 All authors must submit their articles at \href{http://www.pnascentral.org/cgi-bin/main.plex}{[PNAS Central]}

Template for preparing your research report submission to PNAS using Overleaf

Author One^{a,b,1}, Author Two^{b,1,2}, and Author Three^c

^aAffiliation One, ^bAffiliation Two, ^cAffiliation Three

This manuscript was compiled on February 23, 2016

Please provide an abstract of no more than 250 words in a single paragraph. Abstracts should explain to the general reader the major contributions of the article. References in the abstract must be cited in full within the abstract itself and cited in the text.

Keyword 1 | Keyword 2 | Keyword 3 | ...

This PNAS journal template is provided to help you write your work in the correct journal format. Instructions for use are provided below.

Note: please start your introduction without including the word "Introduction" as a section heading (except for math articles in the Physical Sciences section); this heading is implied in the first paragraphs.

Guide to using this template on Overleaf

Please note that whilst this template provides a preview of the typeset manuscript for submission, to help in this preparation, it will not necessarily be the final publication layout. For more detailed information please see the PNAS Information for Authors.

If you have a question while using this template on Overleaf, please use the help menu ('?') on the top bar to search for [help and tutorials]. You can also [contact the Overleaf support team] at any time with specific questions about your manuscript or feedback on the template.

Author Affiliations. Include department, institution, and complete address, with the ZIP/postal code, for each author. Use lower case letters to match authors with institutions, as shown in the example. Authors with an ORCID ID may supply this information at submission.

Submitting Manuscripts. All authors must submit their articles at PNASCentral. If you are using Overleaf to write your article, you can use the "Submit to PNAS" option in the top bar of the editor window.

Format. Many authors find it useful to organize their manuscripts with the following order of sections: Title, Author Affiliation, Keywords, Abstract, Significance Statement, Results, Discussion, Materials and methods, Acknowledgments, and References. Other orders and headings are permitted.

Manuscript Length. PNAS generally uses a two-column format averaging 67 characters, including spaces, per line. The maximum length of a Direct Submission research article is six pages and a PNAS PLUS research article is ten pages including all text, spaces, and the number of characters displaced by figures, tables, and equations. When submitting tables, figures, and/or equations in addition to text, keep the text for your manuscript under 39,000 characters (including spaces) for

Direct Submissions and 72,000 characters (including spaces) for PNAS PLUS.

References. References should be cited in numerical order as they appear in text; this will be done automatically via bibtex, e.g. [1] and [2, 3]. All references, including for the SI, should be included in the main manuscript file. References appearing in both sections should not be duplicated. SI references included in tables should be included with the main reference section.

Data Archival. PNAS must be able to archive the data essential to a published article. Where such archiving is not possible, deposition of data in public databases, such as GenBank, ArrayExpress, Protein Data Bank, Unidata, and others outlined in the Information for Authors, is acceptable.

Language-Editing Services. Prior to submission, authors who believe their manuscripts would benefit from professional editing are encouraged to use a language-editing service (see list at www.pnas.org/site/authors/language-editing.xhtml). PNAS does not take responsibility for or endorse these services, and their use has no bearing on acceptance of a manuscript for publication.

Digital Figures. Only TIFF, EPS, and high-resolution PDF for Mac or PC are allowed for figures that will appear in the main text, and images must be final size. Authors may submit USD or PRC files for 3D images; these must be accompanied by 2D representations in TIFF, EPS, or high-resolution PDF format. Color images must be in RGB (red, green, blue) mode. Include the font files for any text.

Figures and Tables should be labelled and referenced in the standard way using the \label{} and \ref{} commands.

Figure 1 shows an example of how to insert a column-wide figure. To insert a figure wider than one column, please use

Significance Statement

Authors must submit a 120-word maximum statement about the significance of their research paper written at a level understandable to an undergraduate educated scientist outside their field of specialty. The primary goal of the Significance Statement is to explain the relevance of the work in broad context to a broad readership. The Significance Statement appears in the paper itself and is required for all research papers.

Please provide details of author contributions here.

Please declare any conflict of interest here.

¹A.O. (Author One) and A.T. (Author Two) contributed equally to this work (remove if not applicable).

²To whom correspondence should be addressed. E-mail: author@email.com

Archive & share your materials

Data: Avoid supplementary files, licence, and share your data using a repository.

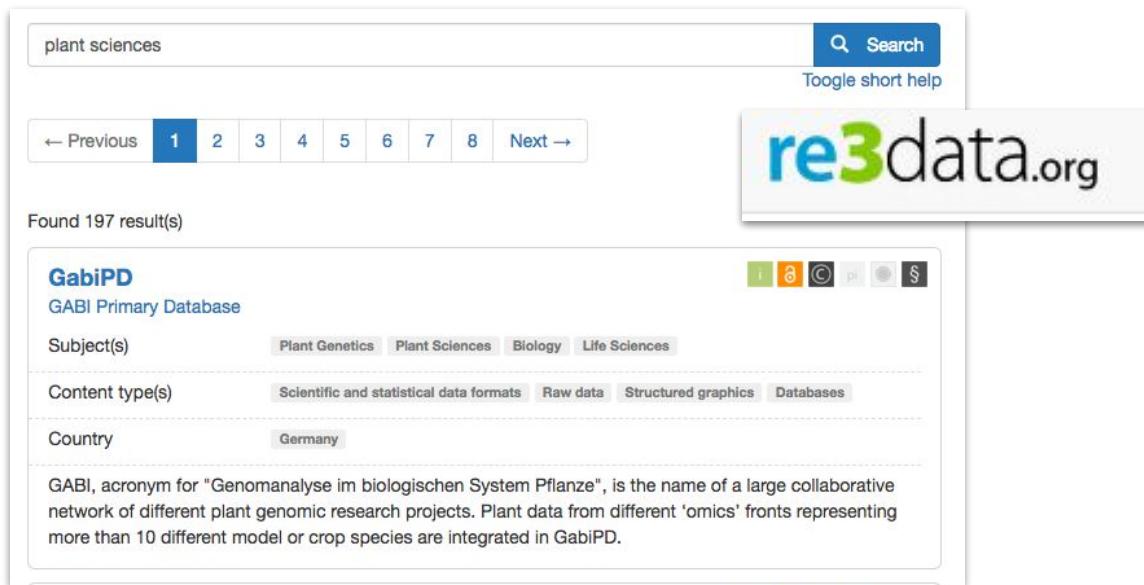
Materials & reagents: Licence your published materials so they can be reused. Deposit reagents with repositories like Addgene, The Bloomington Drosophila Stock Center, and ATCC to make them easily accessible to other researchers.

Software: Licence your code using Code Ocean or Github.

Data Sharing

- Funder specified repository
- Institutionally specified data repository
- Domain or discipline-specific data repository
- Repository of Research Data Repositories

<https://www.re3data.org>



The screenshot shows the re3data.org search interface. In the top left, there's a search bar with "plant sciences" typed in. To its right is a blue "Search" button and a "Toggle short help" link. Below the search bar is a navigation menu with links for "← Previous", page numbers 1 through 8, and "Next →". A large "re3data.org" logo is centered below the search bar. Underneath the logo, the text "Found 197 result(s)" is displayed. The main content area features a card for "GabiPD" (GABI Primary Database). The card includes sections for "Subject(s)" (Plant Genetics, Plant Sciences, Biology, Life Sciences), "Content type(s)" (Scientific and statistical data formats, Raw data, Structured graphics, Databases), and "Country" (Germany). A descriptive text at the bottom states: "GABI, acronym for 'Genomanalyse im biologischen System Pflanze', is the name of a large collaborative network of different plant genomic research projects. Plant data from different 'omics' fronts representing more than 10 different model or crop species are integrated in GabiPD."

In addition to a specified data repository, you can make a deposit to a general purpose repository:

- DataDryad <http://datadryad.org>
- Figshare <https://figshare.com>
- Zenodo <https://zenodo.org>



REPOSITORIES

Dryad
figshare
GigaScience
Harvard Dataverse
NIH and NCBI Repositories
Scientific Data
Zenodo
Additional Resources

HOME / BEST PRACTICES /

Repositories

The number of available resources for data sharing and data publication has increased substantially over the past few years, making it difficult for individual researchers to evaluate the advantages and limitations of the various options they search for the right solution to address their needs.

Here, we compare and contrast several of the general data repositories and data publication platforms available for biomedical science researchers. Click on the matrix below to see detailed descriptions of each resource.

Requirement		Dataverse	Dryad	figshare	Zenodo	GigaScience	Scientific Data
Data Size and Format							
Hosting of common file formats (e.g. csv, tsv, xls, xlsx, doc, pdf)		✓	✓	✓	✓	✓	✗
Hosting of proprietary file formats (e.g. raw image files)		✓	✓	✓	✓	✗	✗
Unlimited size per file		✗	✓	✗	✗	✓	✗
Unlimited total dataset size		✓	✓	✓	✓	✓	✗
Data Licensing							
CC0 waiver1	recommended		required		recommended	available	required
Data Attribution and Citation Tools							
Assignment of dataset DOIs		✓	✓	✓	✓	✓	✗
User Access Controls							
Tiered access (e.g. administrator-level, collaborator-level, curator-level)		✓	✗	✓	✗	✓	✗
Journal-integrated, anonymous access (for peer review pre-publication)		✗	✓	✓	✗	✓	✗
Optional embargo to data release following publication		✗	✓	✓	✓	✓	✗
Data Access Tools							
Comprehensive data and metadata search tools		✓	✗	✗	✗	✗	✗
Data access via direct download		✓	✓	✓	✓	✓	✗
Data downloading via API		✓	✓	✓	✓	✓	✗
Built-in tools for reading proprietary file formats		✗	✗	✓	✗	✗	✗
Integrated data analysis tools		✓	✗	✗	✗	✓	✗
Cost							
Data deposition fees	none	tiered	none	none	none	none	✗
Data maintenance fees	none	none	none	none	none	none	✗

Requirement	Dataverse	Dryad	figshare	Zenodo	GigaScience	Scientific Data
Data Size and Format						
Hosting of common file formats (e.g. csv, tsv, xls, xlsx, doc, pdf)	✓	✓	✓	✓	✓	✗
Hosting of proprietary file formats (e.g. raw image files)	✓	✓	✓	✓	✗	✗
Unlimited size per file	✗	✓	✗	✗	✓	✗
Unlimited total dataset size	✓	✓	✓	✓	✓	✗
Data Licensing						
CC0 waiver1	recommended	required	recommended	available	required	✗
Data Attribution and Citation Tools						
Assignment of dataset DOIs	✓	✓	✓	✓	✓	✗
User Access Controls						
Tiered access (e.g. administrator-level, collaborator-level, curator-level)	✓	✗	✓	✗	✗	✗
Journal-integrated, anonymous access (for peer review pre-publication)	✗	✓	✓	✗	✓	✗
Optional embargo to data release following publication	✗	✓	✓	✓	✓	✗
Data Access Tools						
Comprehensive data and metadata search tools	✓	✗	✗	✗	✗	✓
Data access via direct download	✓	✓	✓	✓	✓	✗
Data downloading via API	✓	✓	✓	✓	✓	✗
Built-in tools for reading proprietary file formats	✗	✗	✓	✗	✗	✗
Integrated data analysis tools	✓	✗	✗	✗	✓	✗
Cost						
Data deposition fees	none	tiered	none	none	none	✗
Data maintenance fees	none	none	none	none	none	✗

Data Repository Comparison Matrix

<https://datamanagement.hms.harvard.edu/repositories>

Reagent Management & Sharing

- Wasted time, money, resources when reagents are recreated
- Mistakes in recreation can lead to spurious results
- Individual labs don't usually have the resources to:
 - Keep track of all reagents created in lab
 - Consistently validate all reagents in the lab
 - Properly label and store all reagents
 - (Legally) distribute all reagents to interested researchers
- Reagents repositories are part of the solution!

Addgene

 **addgene**
The nonprofit plasmid repository

[Login](#) | [Create Account](#) | [Track Order](#)



[Find Plasmids](#) ▾ [Deposit Plasmids](#) ▾ [Find Virus](#) ▾ [How to Order](#) ▾ [Resources](#) ▾ [About Addgene](#) ▾

This website uses cookies to ensure you get the best experience. By continuing to use this site, you agree to the use of cookies. [Close](#)



Better. Faster. CRISPR.

[Find CRISPR Plasmids](#)

CRISPR Plasmids & Lentiviral Preps
Addgene has assembled the reagents and information you need to use CRISPR/Cas9 technology in your own lab.
CRISPR/Cas9 libraries, lentiviral vectors, and more.

Special Collections [View All »](#)

 [Viral Service](#)

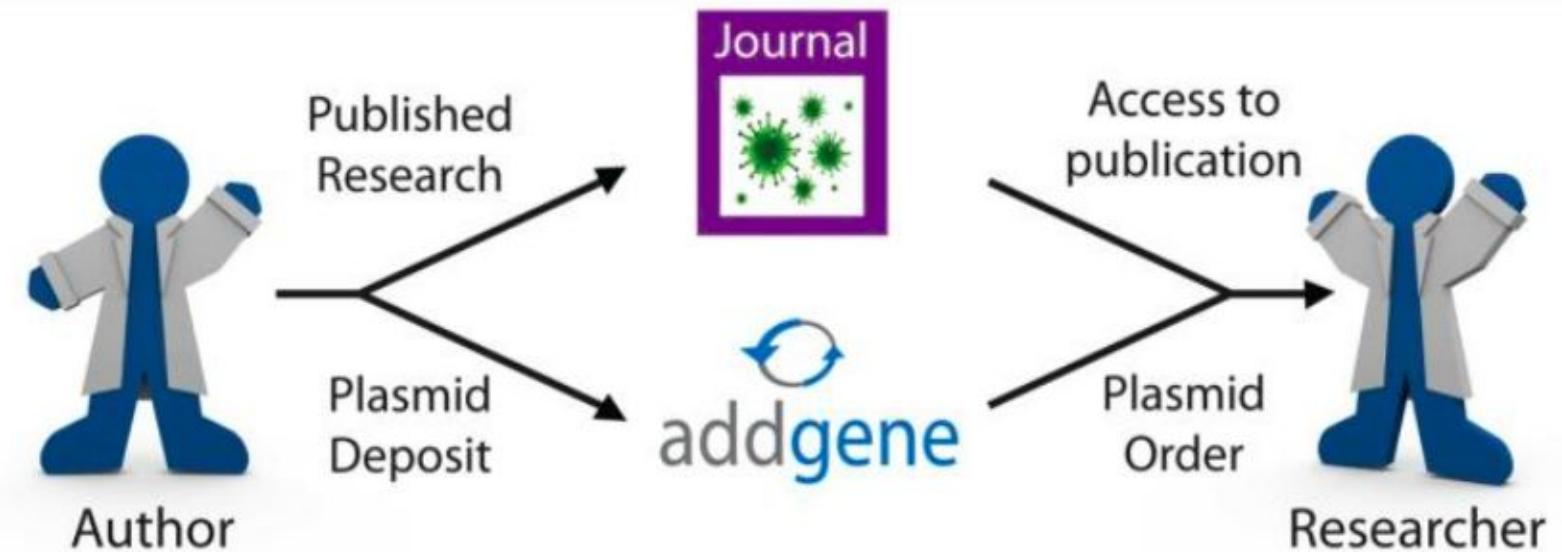
 [CRISPR Tools](#)

 [Fluorescent Proteins](#)

 [Cancer Resources](#)

Journals Recommend Addgene!

Plant Physiology
Journal of Endocrinology
eLife
Developmental Biology
PNAS
Nature
Scientific Reports
PeerJ
Microbiology Open
Society
.and more!



Software Environments

Keep track of analysis

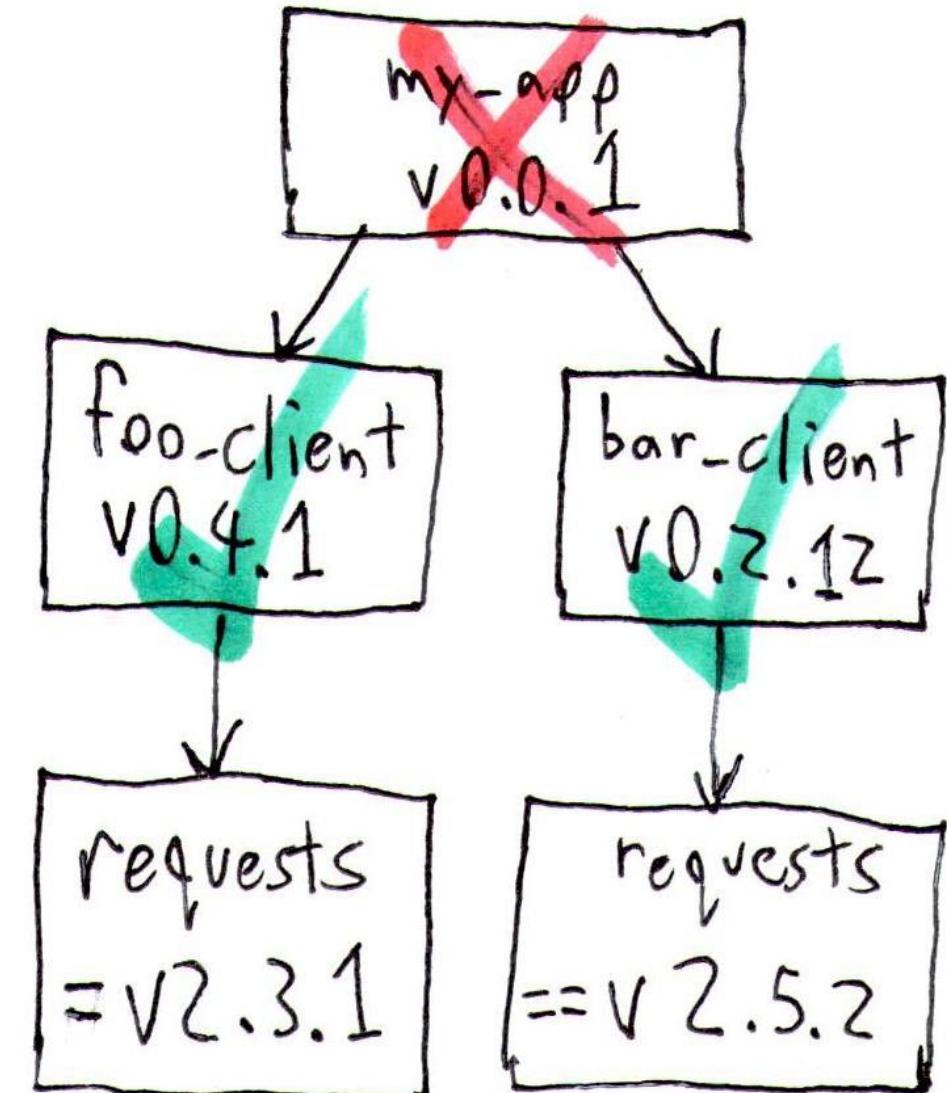
Interactive coding

Interactive data exploration

Imbedded visualization

Easy access to docstrings

Mix of code and documentation



Project Jupyter

The image shows two windows of the Jupyter Notebook application. The left window is a 'Welcome' screen with a 'WARNING' message about relying on the server. The right window is an open notebook titled 'Lorenz Differential Equations'.

Lorenz Differential Equations Notebook Content:

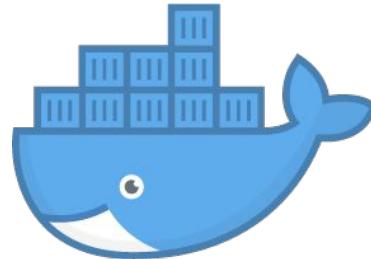
- Title:** Exploring the Lorenz System
- Description:** In this Notebook we explore the [Lorenz system](#) of differential equations:
- $$\dot{x} = \sigma(y - x)$$
- $$\dot{y} = \rho x - y - xz$$
- $$\dot{z} = -\beta z + xy$$
- Text:** This is one of the classic systems in non-linear differential equations. It exhibits a range of complex behaviors as the parameters (σ , β , ρ) are varied, including what are known as *chaotic solutions*. The system was originally developed as a simplified mathematical model for atmospheric convection in 1963.
- In [7]:** interact(Lorenz, N=fixed(10), angle=(0.,360.),
 σ=(0.0,50.0),β=(0.,5), ρ=(0.0,50.0))
- Interactive Parameters:**
 - angle: 308.2
 - max_time: 12
 - σ: 10
 - β: 2.6
 - ρ: 28
- Figure:** A 3D plot showing the Lorenz attractor, a complex, chaotic trajectory in three dimensions.

Containers

Docker

Runs images as containers that are self contained with all code, programs, libraries included. No subsequent installation required.

- Isolated
- Portable
- Lightweight



Code Ocean

Configuration, preservation, & reuse of executable code using containers.

- **Preservation:** Code works today, tomorrow, next week, next year
- **Impact:** Enable easy reuse of code
- **Advanced Tech:** Tools following reproducibility best practices
- **Collaboration:** Code is easy to share and discover

Code pane: the list of code files in your project.

Editor pane: text editor for modifying/editing code files.

The screenshot shows the Code Ocean interface with four main panes:

- Code pane:** On the left, a sidebar titled "Code" lists project files: createpng.c, fttest.c, main.c, main.sh, makename.cc, pcg_basic.c, pcg_basic.h, pdfcode.c (selected), and random.c. A file named amiri-bold.ttf is listed under "Data".
- Editor pane:** The central area displays the "pdfcode.c" file content. The code includes various #include directives, defines for constants like COLOR_SCALE and MAX_CIRCLE_RADIUS, and several extern function declarations for random number generation and PDF creation.
- Data pane:** A large green callout at the bottom left points to the "Data" section of the sidebar, which contains the file amiri-bold.ttf.
- Results pane:** On the right, a sidebar titled "Results" shows the output of a run (Run 5302609) with a run time of 0h 00m 02s on Sep 13, 2017, at 12:36. It lists generated files: a.png (13.49 KB), e.png (12.99 KB), f.png (11.34 KB), l.png (6.96 KB), O.png (23.39 KB), output.pdf (4.74 MB), r.png (10.08 KB), sample.png (117.87 KB), and v.png (14.4 KB).

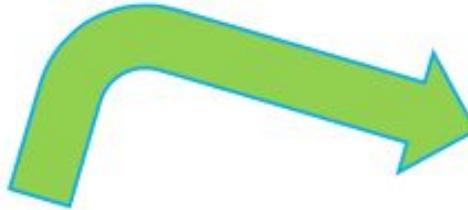
Data pane: the list of data resource files required by your project.

Results pane: the list of files produced by your project and available for download by users.

These digital tools can work together



Provides environment for scientists share and run published computer code used in data analysis without requiring expertise in a particular programming language. This can help assess results for selecting materials in future experiments.

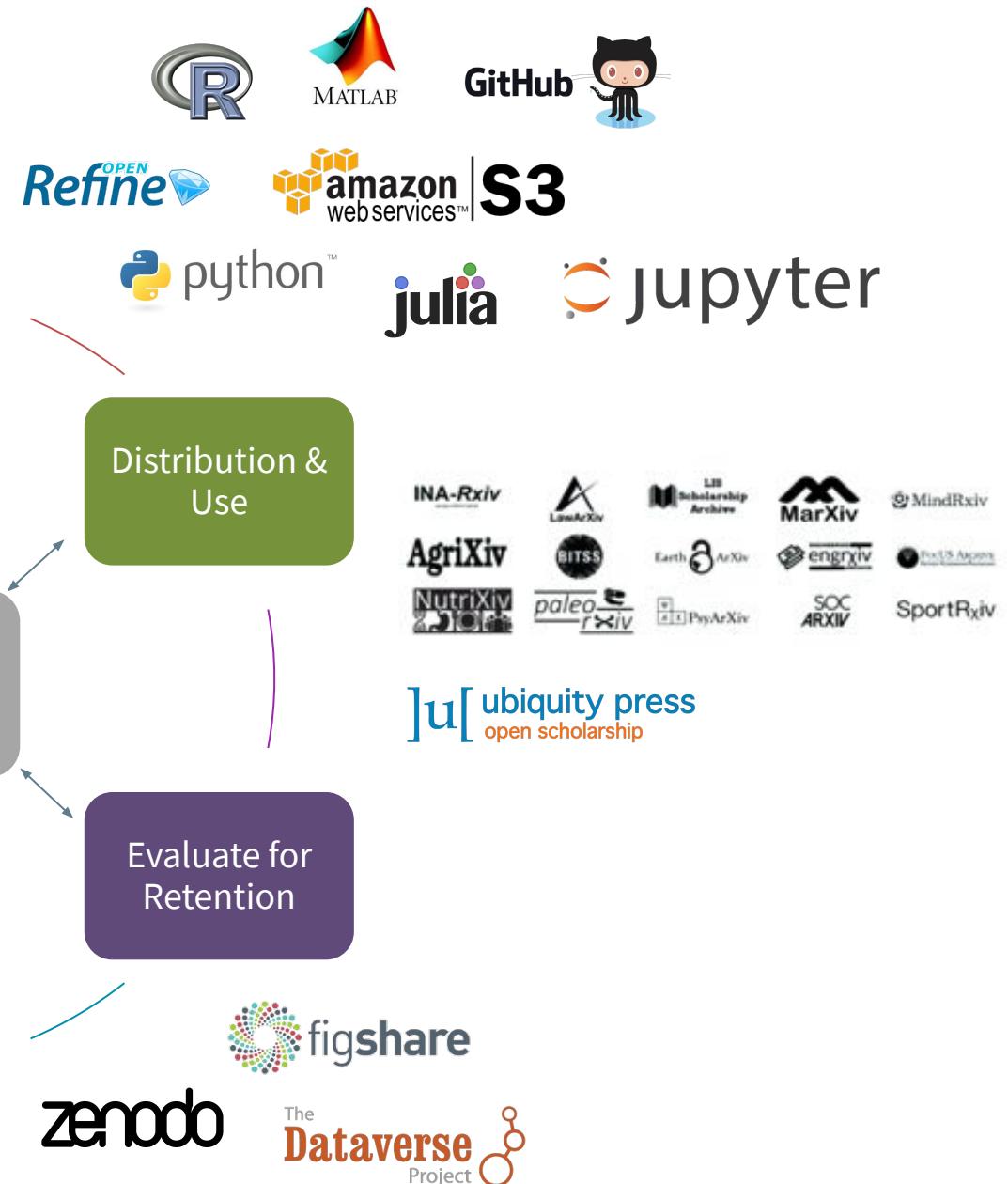
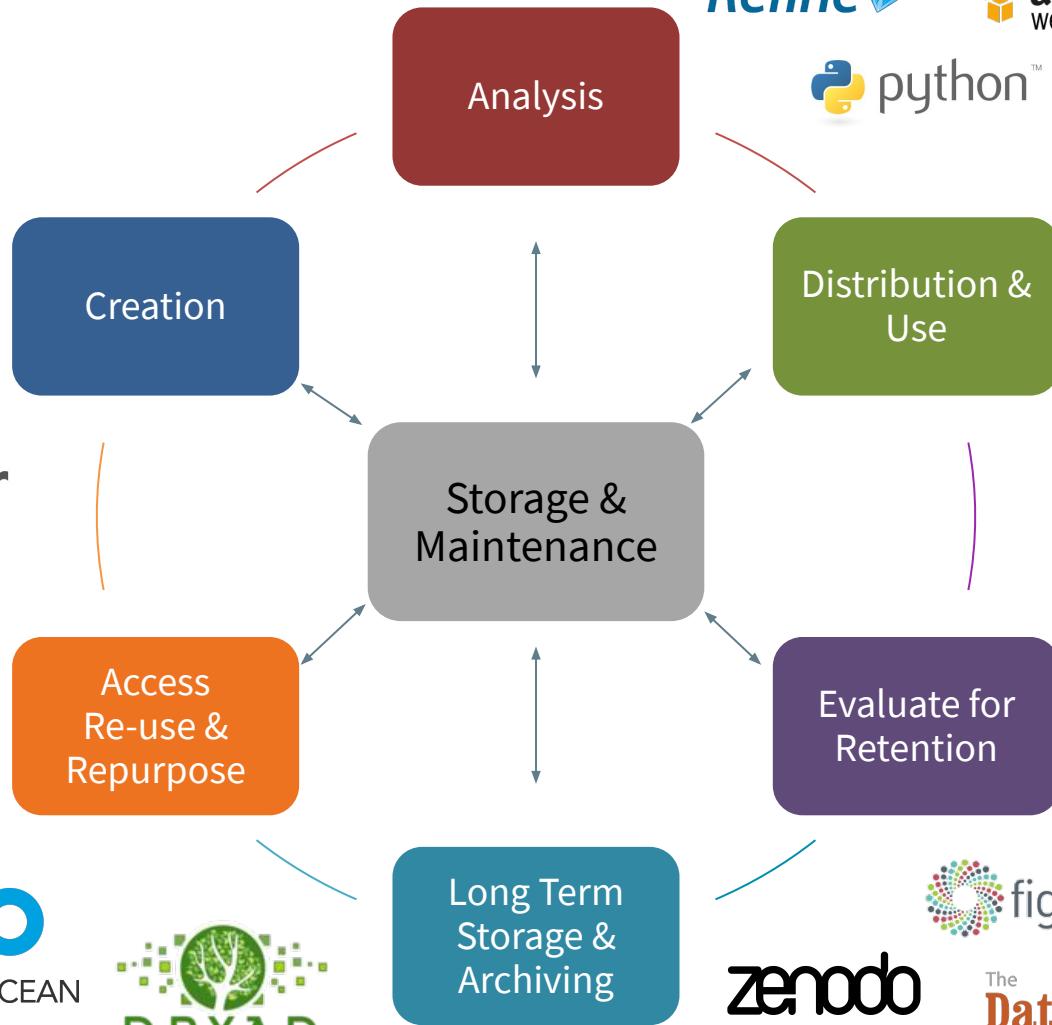


Provides quality-controlled genetic material and makes it easier to share materials generated during research.



Use to access up-to-date protocols, no matter what stage of research, and create protocols with necessary modifications that can then be shared with others.

Digital Workspaces



Summary

Reproducible research practices enable you to:

- Organize experiments productively
- Accurately analyze results
- Share results with future researchers
- Share techniques
- Share reagents with future researchers
- Accelerate science

The tools discussed here should provide you with the framework to make your research more reproducible and will save you time and resources in the long term!



Harvard Biomedical Data Management

Best practices & support services for research data lifecycles

[About ▾](#)[Best Practices ▾](#)[Plan ▾](#)[Store ▾](#)[Share ▾](#)[Resources & Support](#)[HOME / PLAN /](#)

Tips for Reproducibility

[Planning Overview](#)[Tips for Reproducibility](#)[Data Management Plan](#)[Metadata Overview](#)[Metadata Guides](#)

"A growing chorus of concern, from scientists and laypeople, contends that the complex system for ensuring the reproducibility of biomedical research is failing and is in need of restructuring"

[Policy: NIH plans to enhance reproducibility](#)

What is Reproducibility?

"refers to the ability of a researcher to duplicate the results of a prior study using the same materials and procedures as were used by the original investigator. So in an attempt to reproduce a published statistical analysis, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis to determine whether they yield the same results."

[Add New](#)🔍[Submit Questions and Feedback](#)[Upcoming Trainings & News](#)[Receive Data Management Updates](#)

HARVARD
MEDICAL SCHOOL



COUNTWAY
LIBRARY

Tips for Reproducibility

<https://datamanagement.hms.harvard.edu/tips-reproducibility>



Harvard Biomedical Data Management

Best practices & support services for biomedical data management

[About ▾](#)[Best Practices ▾](#)[Plan ▾](#)[Store](#)[HOME / PLAN /](#)

Tips for Reproducibility

"A growing chorus of concern, from scientists and research funders, has highlighted the need for a more transparent and complex system for ensuring the reproducibility of biomedical research. This is particularly true for studies in need of restructuring."

[Policy: NIH plans to enhance reproducibility](#)

What is Reproducibility?

"refers to the ability of a researcher to duplicate the results of a published study using the same materials and procedures as were used by the original authors. It requires that the researcher attempt to reproduce a published statistical analysis, using the same raw data to build the same analysis files and running the same statistical analysis to determine whether they yield the same results."

Tips for Reproducibility

<https://datamanagement.hms.harvard.edu/tips-reproducibility>



Countway Library, @HMSCountway



eLife Science, @eLife

Reproducibility Resources & Tools

Data management

Harvard University Data Management page <https://datamanagement.hms.harvard.edu>

Kbroman Lab <http://kbroman.org/dataorg> (Short primer on data storage and handling)

Purdue Library <http://guides.lib.purdue.edu/c.php?q=353013&p=2378292> (Short primer on data management and file naming conventions)

Data One Best Practices <https://www.dataone.org/best-practices> (Detailed resource on how to handle data throughout its life-cycle)

Best Practices for Biomedical Research Data Management MOOC

<http://bit.ly/HMS-RDM-MOOC> (Online course aimed at a broad audience on recommended practices for managing research data)

Electronic Lab Notebooks (ELN)

Harvard University ELN guide <https://tinyurl.com/Harvard-ELN> (Great summary about current ELNs and what they do)

Benchling <https://benchling.com/> (free)

Evernote <https://evernote.com/> (free and \$\$)

Labguru <https://www.labguru.com/> (\$)

sciNote <https://scinote.net/> (open source, free)

Open Science Framework <https://osf.io/> (free)

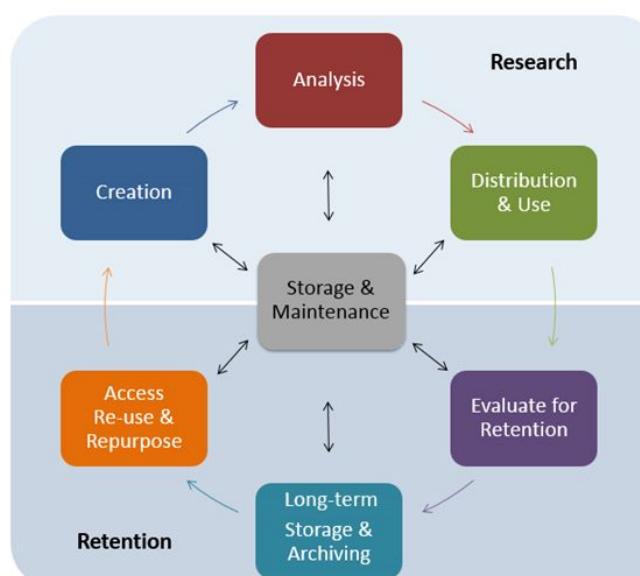
Questions?

Harvard Biomedical Data Management
Best practices & support services for research data lifecycles

About ▾ Best Practices ▾ Plan ▾ Store ▾ Share ▾ Resources & Support

Data Management
Data Management is the process of providing the appropriate labeling, storage, and access for data at all stages of a research project. We recognize that best practices for each of these aspects of data management can and often do change over time, and are different for different stages in the data lifecycle.

Early and attentive management at each step of the data lifecycle will ensure the discoverability and longevity of your research.



← December 2018 →

S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

🔍

FEATURED NEWS



DMWG Featured in Nature Article: How to pick an electronic laboratory notebook
Thursday, August 9, 2018

Submit Questions and Feedback
Upcoming Trainings & News
Receive Data Management Updates

Upcoming Workshops / Seminars

Data Skills: Planning for Research Success

Tuesday, December 11
1:00 – 2:00 pm
Cannon Room

Register: bit.ly/RDM-12-11

Metadata

January TBD

datamanagement.hms.harvard.edu

bit.ly/rdm-survey

Key Resources

Harvard Biomedical Data Management
datamanagement.hms.harvard.edu

Center for the History of Medicine | Archives and Records Management
www.countway.harvard.edu/chom/archives-and-records-management

Research Information Technology Solutions
rits.hms.harvard.edu

Office of the Vice Provost for Research | Research Data Security & Management
vpr.harvard.edu/pages/research-data-security-and-management

Harvard Catalyst | The Harvard Clinical and Translational Science Center
catalyst.harvard.edu

Office for Scholarly Communications
osc.hul.harvard.edu/policies