COUNTWAY LIBRARY

Countway Library, @HMSCountway

eLIFE

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?g=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)

Code

Github https://github.com/ (code repository; free for public repos)
Jupyter Notebooks http://jupyter.org/ (open source web-app for creating & sharing live code, equations, and more)
Code Ocean https://codeocean.com/ (computational reproducibility platform; free to upload, share & publish executable code with DOI; pay for more computing time over freemium limit)
Conda and BioConda https://conda.io/docs/ and https://bioconda.github.io/ (A operating system independent package environment manager for the command line)
Docker and Biocontainers https://docs.docker.com/ and http://biocontainers.pro (A container ecosystem to package code and data on the command line.

Binder https://mybinder.org/ (A tool to make your github repository an online docker image run in the cloud)
Galaxy https://usegalaxy.org/ (A web and graphic interface based bioinformatics platform. Needs local set-up for larger data handling.)

Reagents

Addgene https://www.addgene.org/ (nonprofit plasmid repository)
CiteAb https://www.citeab.com/ (antibody search engine with results sorted by citations)
Quartzy https://www.quartzy.com/ (manage lab inventory)
Resource Identification Portal (RRIDs) https://scicrunch.org/resources (unique identifired for antibodies, model organisms, and tools used in PubMed or PMC)

Methods

Bio-Protocol https://bio-protocol.org/ (A peer-reviewed protocol journal; free to read & publish)
protocols.io http://protocols.io/ (an open access repository of science methods; free to read & publish)

Data

DataDryad http://datadryad.org/ (curated digital repository; free to access, $120 to publish dataset up to 20GB)
Figshare http://datadryad.org/ (free digital repository, 5GB per file limit)
Zenodo https://zenodo.org/ (free digital repository; 50GB per dataset limit)

Authoring

Zotero https://www.zotero.org/ (free and open-source reference management software to manage bibliographic data and related research materials)
Mendeley https://www.mendeley.com/ (free desktop & web program reference manager)
RefWorks https://refworks.proquest.com/ (web-based commercial reference management software package)
Authorea https://www.authorea.com/ (free online real-time collaborative writing & publishing system with unlimited collaborators)
Overleaf https://www.overleaf.com/ (free online collaborative writing & publishing system)
Manuscripts https://www.manuscriptsapp.com/ (offline Mac app authoring tool; free unlimited usage)

Practical tips for reproducibility

1. Plan for reproducibility before you start
    a. **Write a study plan or protocol** and track new versions.
    b. **Set-up a reproducible project** using an electronic lab notebook to organize and track your work. Avoid saving proprietary file formats.

2. Keep track of things
    a. **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.
    b. **Track changes** to your files using version control.
    c. **Document** everything done by hand in a README file and data dictionary. **Karl Broman's Data Organization**: http://kbroman.org/dataorg/pages/dictionary.html

3. Report your research transparently
    a. **Share your protocols and interventions** explicitly and transparently.
    b. **Write a transparent report**. Guidelines from the Equator Network or processes like Registered Reports can help.
    c. **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.

4. Archive & share your materials
    a. Share and licence your research
        i. Data
            1. Avoid supplementary files, licence, and share your data using a repository. **How to License Research Data**: http://www.dcc.ac.uk/resources/how-guides/license-research-data.
        ii. Materials & reagents
            1. Licence your published materials so they can be reused. **Creative Commons License Picker**: https://creativecommons.org/choose/
            2. Deposit reagents with repositories like Addgene, The Bloomington Drosophila Stock Center, and ATCC to make them easily accessible to other researchers.
        iii. Software
            1. Licence your code using Code Ocean or Github. **Open Source Licences**: https://opensource.org/licenses.

5. Further reading:

- **Ten Simple Rules for Reproducible Computational Research**: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003285
- **Reproducibility in Science**: http://ropensci.github.io/reproducibility-guide/
- **Open Science MOOC**: https://opensciencemooc.eu/ **and** https://opensciencemooc.github.io/site/Resources/#three
- **Tools and Resources for Reproducibility Series at protocols.io:** goo.gl/r7GKMA
- **Managing Laboratory Notebooks** http://colinpurrington.com/tips/lab-notebooks
- **General File and Folder Organization** https://zapier.com/blog/organize-files-folders/
- **File Naming Conventions** http://www.exadox.com/en/articles/file-naming-convention-ten-rules-best-practice

6. Example studies:

- **Gene family innovation, conservation and loss on the animal stem lineage**
  - Paper: https://doi.org/10.7554/eLife.34226
  - Protocols: dx.doi.org/10.17504/protocols.io.kwscxees
  - Data: https://doi.org/10.6084/m9.figshare.5686984.v2
- **A robust method for transfection in choanoflagellates illuminates their cell biology and the ancestry of animal septins**
  - Paper: https://doi.org/10.1101/343111
  - Protocols: http://www.protocols.io/groups/king-lab
  - Constructs: http://www.addgene.org/Nicole_King
- **Implicating candidate genes at GWAS signals by leveraging topologically associating domains**
  - Paper: https://dx.doi.org/10.1038/ejhg.2017.108
  - Code: https://zenodo.org/record/163950#.W0hqTdJKjIU
  - Docker workflow: https://zenodo.org/record/166556#.W0hqc9JKjIU
- **mcSCRB-seq: sensitive and powerful single-cell RNA sequencing**
  - Protocol: dx.doi.org/10.17504/protocols.io.p9kdr4w
  - Paper: https://doi.org/10.1101/188367
  - Code: https://github.com/cziegenhain/Bagnoli_2017
- **TransRate: reference-free quality assessment of de novo transcriptome assemblies**
  - Paper: https://dx.doi.org/10.1101%2Fgr.196469.115
  - Code: https://github.com/Blahah/transrate
  - Tutorial: http://hibberdlab.com/transrate/
- **Genomic insights into members of the candidate phylum Hyd24-12 common in mesophilic anaerobic digesters**
  - Paper: https://doi.org/10.1038/ismej.2016.43

    - Code: https://github.com/Kirk3gaard/Publications
- **Experimenting with Reproducibility: a case study of Robustness in Bioinformatics**
    - Paper: https://doi.org/10.1093/gigascience/giy077
    - Code: https://github.com/sje30/waverepo

- **A Bayesian Mixture Modelling Approach For Spatial Proteomics**
    - Paper: https://doi.org/10.1101/282269
    - Code: https://github.com/lgatto/2018-tagm-paper

Tailored from: https://tinyurl.com/plantbio-repo

addgene
The nonprofit plasmid repository

@Addgene

protocols.io

@ProtocolsIO

CODE OCEAN
BETA

@CodeOceanHQ