


Researcher Perspectives – NIH Generalist Repository Ecosystem Initiative (GREI) Workshop

Jason Williams
Cold Spring Harbor Laboratory, DNA Learning Center
 [@JasonWilliamsNY](https://twitter.com/JasonWilliamsNY)

Generalist Repository Ecosystem Initiative (GREI) Workshop, Online
January 2023

DOWNLOAD SLIDES



bit.ly/grei-jan-2023



Cold Spring Harbor Laboratory

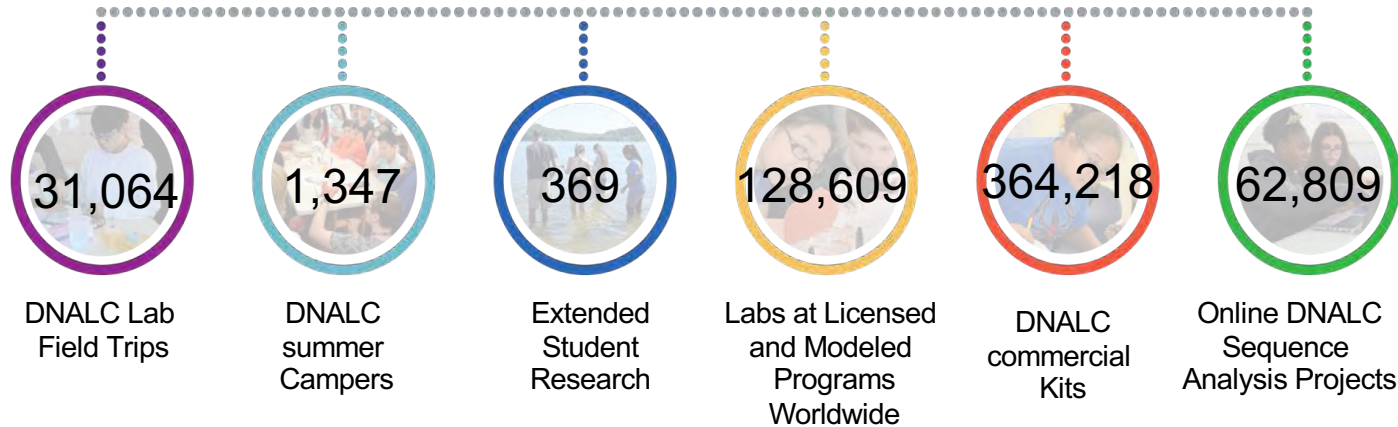


**“Prepare students and families
to thrive in the gene age...”**



CSHL DNA Learning Center

588,416 Annual Exposures





An **open data publishing platform**
& community committed to the
open availability and routine re-
use of all research data

AN OPEN DATA PUBLISHING PLATFORM & COMMUNITY

Vision: **All research data is openly available
and routinely re-used**

Mission: **Enable and promote** the re-use of research data



Make the most of your research data

Dryad is a community-owned resource
Learn more about our [organizational memberships](#)

Submit Now

49,600+ data publications
193,300+ researchers
69,100+ international institutions
1,270+ academic journals

How it works



Login

Use your ORCID. If your institution is a [Dryad member](#), connect to your existing credentials.



Submit

Whether or not your data are related to an article, [upload](#) your data files and receive a citable DOI.



Review

Our [curators](#) will check through your submission to ensure the data are usable. They may contact you with advice or questions.



Cite

[Cite](#) and promote your data publication!

I went to school for...



I need to know this for my research...

```
import urllib2
eutils = 'http://www.ncbi.nlm.nih.gov/entrez/eutils/'
efetch = 'efetch.fcgi?'
s = eutils + efetch

targets = ['J04243', 'M60064']
idString = 'id=' + ','.join(targets)
s += idString + '&db=nucleotide&rettype=fasta'
fileObject = urllib2.urlopen(s)
data = fileObject.read().strip()

entries = data.split('\n\n')
title, sequence = entries[0].split('\n', 1)
print title.split(' ', 1)[0]

# prints:
# >gi|154102|gb|J04243.1|STYHEMAPRE
```

1932



2012



Gaps multiply and perpetuate

@ClinPsychDavid

REALITY	EQUALITY	EQUITY	JUSTICE	INCLUSION
<p>One gets more than is needed, while the other gets less than is needed. Thus, a huge disparity is created.</p>	<p>The assumption is that everyone benefits from the same supports. This is considered to be equal treatment.</p>	<p>Everyone gets the support they need, which produces equity.</p>	<p>All 3 can see the game without supports or accommodations because the cause(s) of the inequity was addressed. The systemic barrier has been removed.</p>	<p>Everyone is INCLUDED in the game. No one is left on the outside; we <u>didn't</u> only remove the barriers keeping people out, we made sure they were valued & involved.</p>

Credit: Saskatoon Health Service; Revised version - David Murphy



Sequence-a-genome camp

2021

- **Species:** Duckweed (*S. polyrhiza* ~150 mb)
- **Relevance:** Biofuels and climate change
- **Partners:** CSHL/HudsonAlpha
- **Results:** ~7Gb DNA sequence and partial assembly



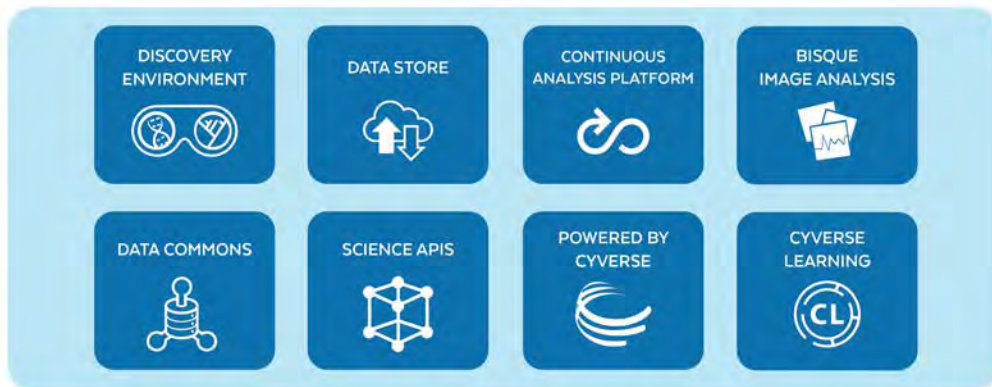
Two species of duckweeds (*Wolffia globosa* and *Spirodela polyrhiza*) taken in Waimanalo, Hawai'i by Eric Guinther

2022

- **Species:** Jamaican Broom (*C. glandulosa* v. *mirabilis* ~unknown mb)
- **Relevance:** Endemic/endangered
- **Partners:** U. Puerto Rico
- **Results:** +9GB and counting; in-progress



PRODUCTS



SERVICES



HARDWARE RESOURCES



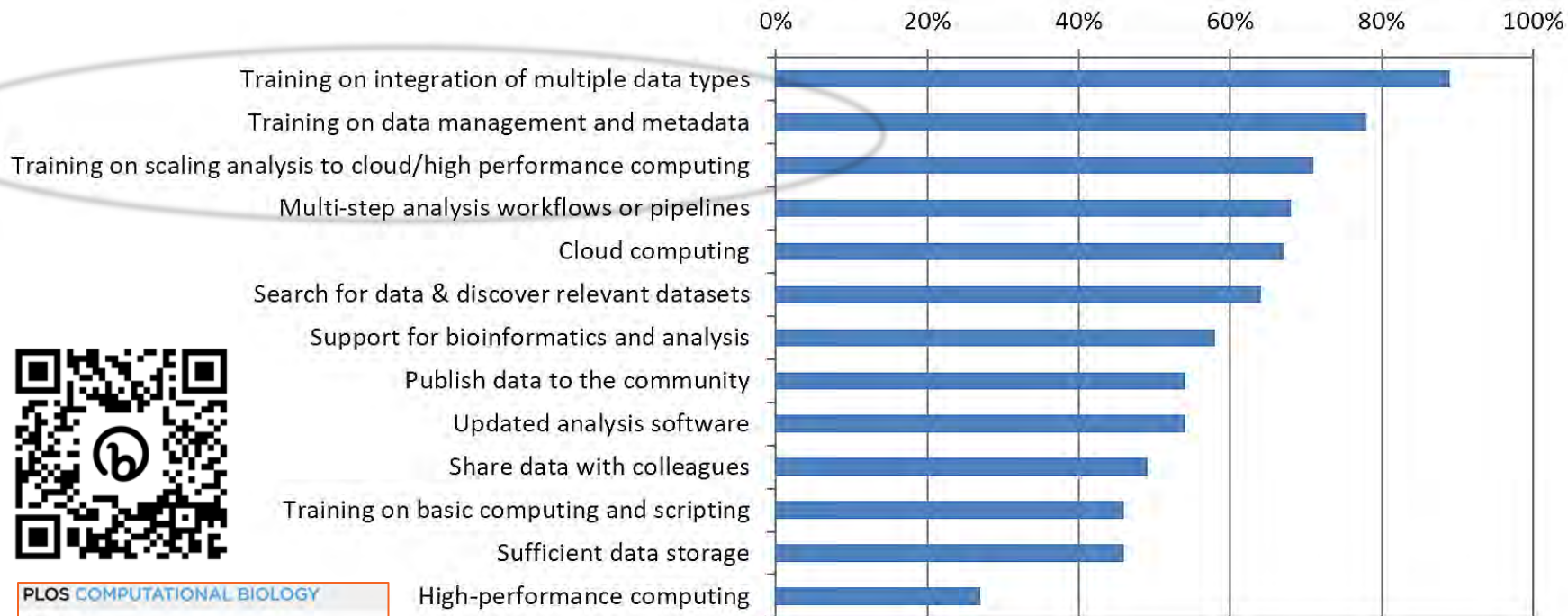
- NSF-Subsidized cyberinfrastructure resources
- Easy-to-use interfaces
- Facilitates national/international collaboration (CyVerse UK, CyVerse Austria)



DBI-0735191, DBI-1265383, DBI-1743442



Training is the biggest need



PLOS COMPUTATIONAL BIOLOGY
















bioRxiv preprint doi: <https://doi.org/10.1101/2018.08.14.244811>; this version posted August 14, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Unmet needs for analyzing biological big data: A survey of 704 NSF principal investigators

Anthony Barone, Jason Williams, David Huxley



Barriers to integration of bioinformatics into undergraduate life sciences education: A national study of US life sciences faculty uncover significant barriers to integrating bioinformatics into undergraduate instruction

Jason J. Williams , Jennifer C. Drew , Sebastian Galindo-Gonzalez , Srebrenka Robic , Elizabeth Dinsdale , William R. Morgan , Eric W. Triplett , James M. Burnette III , Samuel S. Donovan , Edison R. Fowlks , Anya L. Goodman , Nealy F. Grandgenett , Carlos C. Goller , [...], Mark A. Pauley   [view all]

Published: November 18, 2019 • <https://doi.org/10.1371/journal.pone.0224288>



Effective and Inclusive Career-spanning Training



The Bicycle Principles
v1.0.3 - October 2022

Search

Introduction to the Recommenda... Next >

The Bicycle Principles for Effective, Inclusive, and Career-spanning Short-format Training

Improving Professional Development in the Life Sciences and Beyond

Announcement
November 2022

We're collecting feedback: Let us know what you think about the recommendations to improve short-format training. We will be conducting surveys and focus groups from now through February 2023. **Participants will be compensated for their time.**

Home

- The need for a community-driven principle-based framework
- The Bicycle Principles for short-format training
- The Principles and this website
- Banbury Working Group
- Citations and publications
- Funding

► Recommendations and Surveys

Glossary and Definitions

Community Feedback and Next Steps



bikeprinciples.org

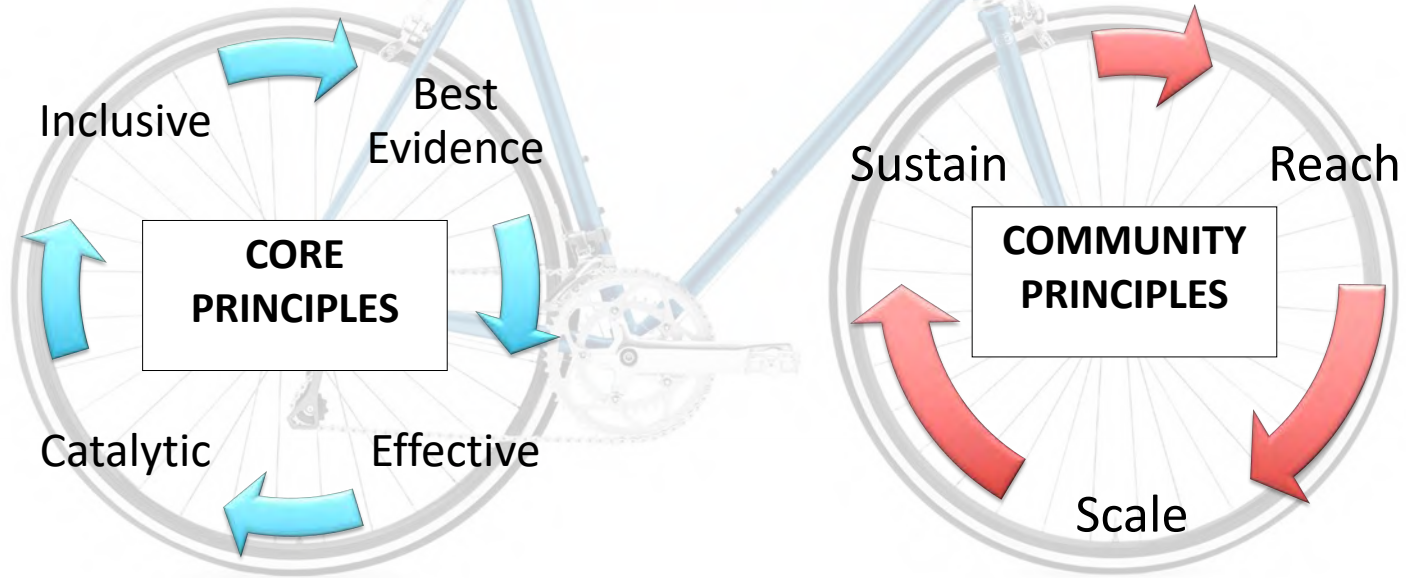


This material is based upon work supported by the National Science Foundation under [DRL/EHR:2027025](https://www.nsf.gov/awardsearch/showAward?AWDNO=DRL/EHR:2027025).



Cold Spring Harbor Laboratory

“The bicycle” – good for going far



A. Professionalize the training of short-format training instructors and instructional designers



Doctor studying a textbook

E. Deploy short-format training to counter inequity



Judge in robe fighting for justice

Building a community of practice in training



LIFESCITRAINERS.ORG

ACCELERATING SCIENTISTS

Why short-format training?

In many areas of the life sciences new technologies and approaches (especially, but not only computational ones) are changing rapidly. It's not possible for formal training (undergraduate/graduate) to keep pace, but short-format training can fill these gaps. Short-format training comes with its own set of challenges, and this community works together to address them.



How I Teach Life Scientists...by Using Reproducible and Scalable Learning Environments

May 11, 2022 @ No Comments

The combination of Docker + cloud computing service enables a teacher to create a highly scalable and flexible learning environment

[Continue reading >](#)



How I Teach Life Scientists...to Build Reproducible, Scalable Workflows with Nextflow

April 26, 2022 @ No Comments

The term "reproducible research" has been used to describe the idea that a scientific publication should be distributed along with all the raw data and metadata used in the study, all the code and/or computational notebooks needed to produce results...

[Continue reading >](#)



March 2022: Community Discussion – The Return to In-Person Training

March 14, 2022 @ No Comments

Many instructors have or shortly will be running their first in-person events since the pandemic, what will change?

[Continue reading >](#)



“The illiterate of the 21st century will not be those who cannot read and write, but those who cannot learn, unlearn, and relearn”

– A. Toffler



<https://bit.ly/grei-jan-2023>

Thanks!



Cold Spring Harbor Laboratory
DNA LEARNING CENTER



CYVERSE™

DBI-0735191, DBI-1265383, and DBI-1743442

Bicycle Principles: This material is based upon work supported by the National Science Foundation under [DBL/EHR-2027025](#). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.



[@JasonWilliamsNY](#)



williams@cshl.edu

BACKUP SLIDES

OUR INSTITUTIONAL MEMBERS

California State University, East Bay, University Libraries	New Mexico State University, Research Cores Program	University of California, Merced (UCM)
Chan Zuckerberg Initiative	North Carolina State University, University Libraries	University of California, Riverside (UCR)
Claremont Graduate University	Ohio State University, University Libraries	University of California, San Diego (UCSD)
Claremont McKenna College	Pitzer College	University of California, San Francisco (UCSF)
Columbia University Libraries	Pomona College	University of California, Santa Barbara (UCSB)
Desert Research Institute	Scripps College	University of California, Santa Cruz (UCSC)
Harvey Mudd College	Stanford University, School of Medicine, Lane Medical Library	University of Illinois, University Library
Indian Institute of Technology Dhanbad	Texas Tech University, IT Technology Support	University of Minnesota Libraries
Keck Graduate Institute	The University of New South Wales Library	University of New Mexico,
Kyoto University	University of California, Berkeley (UCB)	
Lawrence Berkeley National Laboratory (LBNL)	Univer	
Macquarie University Library	Califor	
Michigan State University Libraries	Univer	
Montana State University Library	(UCD)	
Netherlands Institute of Ecology (NIOO-KNAW), Library and Information Services	Univer (UCI)	
	Univer Angel	

OUR PUBLISHER MEMBERS

American Academy for the Advancement of Science	Hindawi
American Genetic Association	John Wiley & Sons, Inc
American Society of Naturalists	Nordic Society Oikos
American Society of Plant Taxonomists	Oxford University Press
Association for Tropical Biology and Conservation	The Paleontological Society
British Ecological Society	The Royal Society
Cambridge University Press	Society for the Study of Evolution
eLife	Society of Systematic Biologists
European Society for Evolutionary Biology	The University of Chicago Press

Researchers, institutions and publishers invest in Dryad for the curation, open sharing and reuse of research data in all fields.



Principles of Open Scholarly Infrastructure



DRYAD

Not-for-profit and online since 2008. Learn more: datadryad.org

Data and metadata are:

- **Curated by our expert team to ensure suitability for publishing and to facilitate discovery and reuse**
- **Permanently stored in our CoreTrustSeal repository**
- **Published under a Creative Commons Public Domain (CC0) license**
- **Accessible via our open API**



DRYAD

Not-for-profit and online since
2008. Learn more: datadryad.org

- **Serving all research domains**
- **Leader in research data**
- **Interconnected**



DRYAD

Not-for-profit and online since
2008. Learn more: datadryad.org

A close-up photograph of a flower's reproductive parts, showing the green ovary and surrounding structures. The ovary is green and has several small, reddish-brown structures (likely anthers or stamens) attached to it. The background is dark, and there are blurred red petals visible.

How it works

Image CC0 <https://doi.org/10.5061/dryad.xd2547dd5>

“Data should be of sufficient quality to validate and replicate research findings”

* https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html#_ftn8



DRYAD

Not-for-profit and online since
2008. Learn more: datadryad.org



Data publishing at Dryad:

- 1. Ensures metadata quality**
- 2. Increases discoverability of data**
- 3. Connects data with other research outputs**
- 4. Promotes data citation**
- 5. Makes data count**



Data curation at Dryad:

- 1. Ensures data is appropriate and licensed for sharing**
- 2. Verifies data is accessible and usable**
- 3. Supports authors**

