



CYVERSE™

Transforming Science Through Data-driven Discovery

Don't let your data drag you down

Focus Forum Webinar

January 31, 2020

Ramona Walls – Assistant Research Professor

Bio5 Institute, University of Arizona

rwalls@email.arizona.edu  @RamonaWalls



Cold Spring Harbor Laboratory



Webinar Content/Goals

By the end of the webinar, you should have the resources to:

- Create a shared project on CyVerse
- Use metadata to manage your data
- Make your data publicly available

For additional topics, go to <https://www.cyverse.org/ffw> or <https://learning.cyverse.org/>.



Why CyVerse?

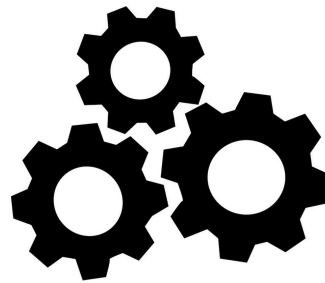


F
Findable

A
Accessible

I
Interoperable

R
Reusable



FAIR data: <https://www.nature.com/articles/sdata201618>

Making your data FAIR webinar: <https://www.youtube.com/watch?v=5aeBYTXaWqA>



Create a shared project on CyVerse

Prerequisites – all project members must:

- Create an account at user.cyverse.org
- Know how to upload, download, and move data on CyVerse

<https://learning.cyverse.org/projects/cyverse-group-project-quickstart>



First step – make a plan

- Information on Data Management Plans (DMP)
 - https://learning.cyverse.org/projects/foss-2020/en/latest/Data_management/dmp.html
- Key first steps:
 - Decide on roles for project member, including who has access to what data
 - Create folder hierarchy and file naming conventions
 - <https://guides.library.upenn.edu/datamgmt/fileorg>
 - Figure out what metadata are needed.
 - Write down the plan!



Follow through

- Assign a data manager to create folder hierarchy and assign permissions in the Data Store
 - https://learning.cyverse.org/projects/data_store_guide/en/latest/step4.html
- Conduct regular quality assurance activities
 - Are data clean and meaningful?
 - Are metadata in place?
 - Is the DMP being followed?
 - If it needs to be adjusted, write down the changes.



Hot off the presses

- See this paper that just came out:
- "Eleven tips for working with large datasets"
 - <https://www.nature.com/articles/d41586-020-00062-z>
- Challenge: versioning



Use metadata to manage your data

- Why use metadata:

- Make data FAIR
- Improve search
- Track provenance
- Sort data based on need, not just folder name

- CyVerse metadata flavors:

- System metadata – added automatically
- User metadata – customizable



Use metadata to manage your data

Live demo:

- Metadata:

- Add metadata in the DE
- Bulk metadata upload
- Search based on metadata
- Metadata templates
- User guide at

<https://wiki.cyverse.org/wiki/display/DEmanual/Using+Metadata+in+the+DE>

- Sharing data in the DE

- collaborators
- teams
- basic sharing
- *ichmod* for sharing via iCommands.



Publish your data

- **Public data: anyone can see them**
 - Community Released folders in the Data Commons
 - Not necessarily FAIR
 - There may still be restrictions on using them
- **Published data**
 - In a repository – CyVerse Curated data
 - Permanent identifier (DOI)
 - License
 - Standard metadata
- **To be FAIR, data must be published**



Publishing data through CyVerse

- Get a DOI through the Data Commons
 - tutorial: <https://cyverse-doi-request-quickstart.readthedocs-hosted.com/en/latest/index.html>
- Submit sequence data to SRA or WGS:
 - https://learning.cyverse.org/projects/sra_submission_quickstart/en/latest/index.html
 - <https://wiki.cyverse.org/wiki/pages/viewpage.action?pageId=34834057>



<http://datacommons.cyverse.org/>





CyVerse is supported by the National Science Foundation under Grants No. DBI-0735191, DBI-1265383 and DBI-1743442.



Cold
Spring
Harbor
Laboratory

