



The Open Science Workspace for Collaborative Data-driven Discovery

## Practical Help to Manage Your Data Store

Jason Williams

Cold Spring Harbor Laboratory, DNA Learning Center / CyVerse

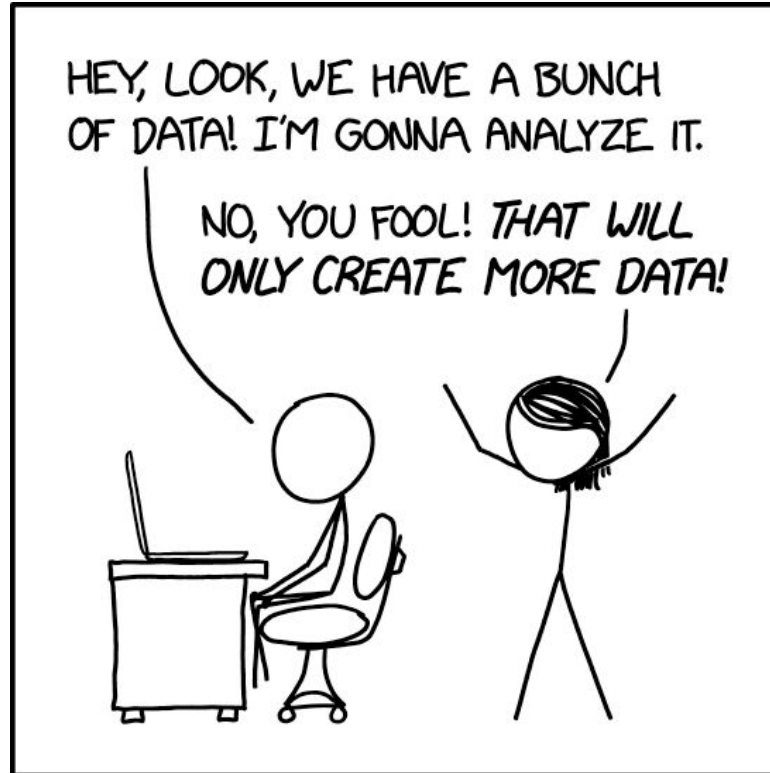


@JasonWilliamsNY



CyVerse is supported by the National Science Foundation under Grant Nos. DBI-0735191, DBI-1265383, and DBI-1743442

# The cause of and solution to all our problems...



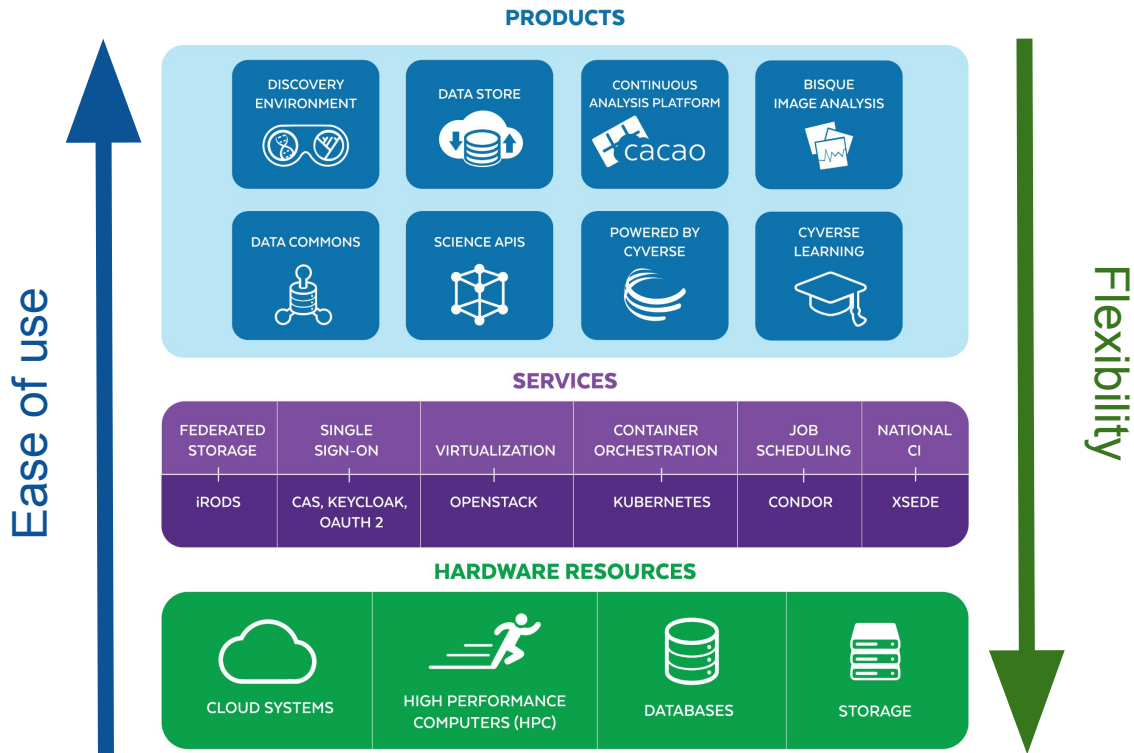
CREDIT: <https://xkcd.com/2582/>

# Big data, bigger problems

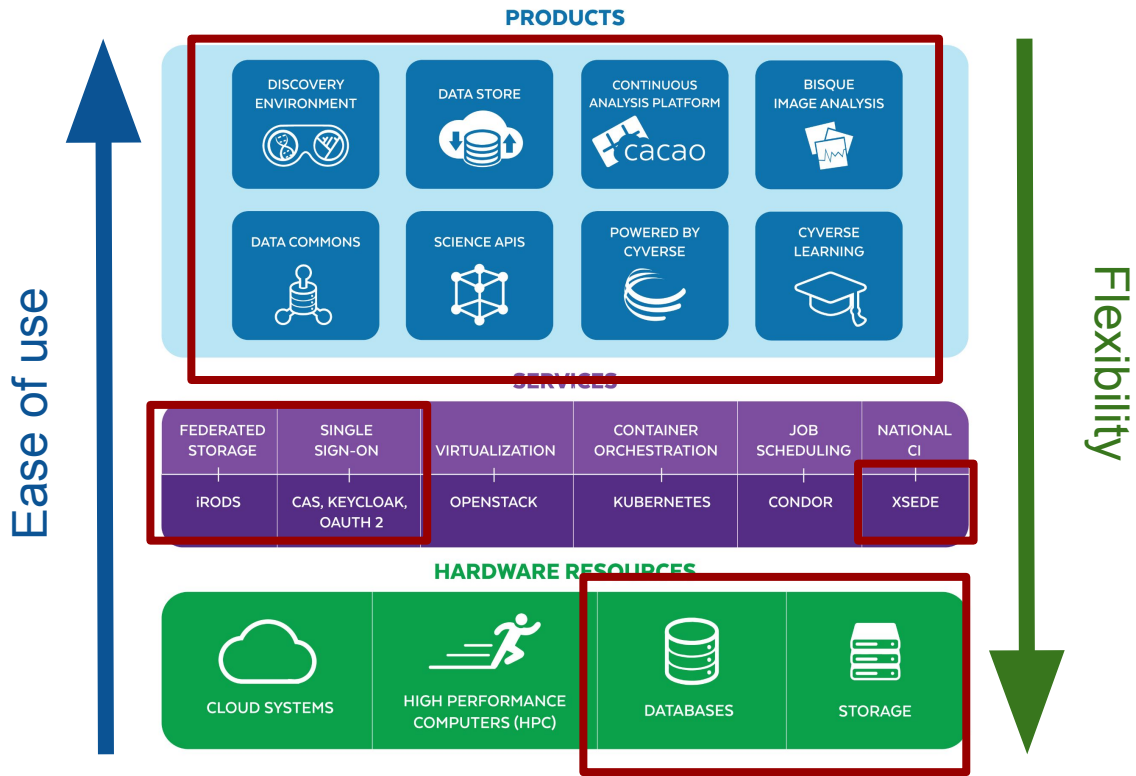


- Difficult / slow transfers
- Expense for storage / backup
- Difficult to share and publish
- Analysis
- Metadata (What Is metadata?)

# CyVerse Technology Stack



# CyVerse Technology Stack



# Data Store



A secure data storage and management service accessible from anywhere in the CyVerse ecosystem

- Securely manage your data across its entire lifecycle

# Data Store



A secure data storage and management service accessible from anywhere in the CyVerse ecosystem

- Securely manage your data across its entire lifecycle
- Automatic backup

# Data Store



A secure data storage and management service accessible from anywhere in the CyVerse ecosystem

- Securely manage your data across its entire lifecycle
- Automatic backup
- Numerous ways to upload/download/share



# Data Store



A secure data storage and management service accessible from anywhere in the CyVerse ecosystem

- Securely manage your data across its entire lifecycle
- Automatic backup
- Numerous ways to upload/download/share
- Based on open source IRODS technology

# Data Storage Tiers

	<b>Basic (free)</b>	<b>Regular</b>	<b>Pro</b>	<b>Commercial</b>
Data storage limit	5GB	50GB	3TB	5TB

[cyverse.org/subscribe](https://cyverse.org/subscribe)

# Data Storage FAQs

**Q: How do I know how much data I am using?**

# Data Storage FAQs

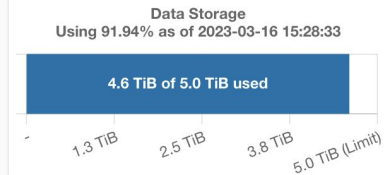
**Q: How do know how much data I am using?**

**A: Login to the Discovery Environment Dashboard**

## Resource Usage

Your current subscription tier is [Commercial](#).

Effective Period: 2023-01-10 17:12:18 - 2024-01-10 17:12:18



# Data Storage FAQs

**Q: What happens if I hit my storage limit?**

# Data Storage FAQs

**Q: What happens if I hit my storage limit?**

**A: You won't be able to add new data**

# Data Storage FAQs

**Q: What happens if I hit my storage limit?**

**Options:**

- 1. Delete some data**
- 2. Upgrade your data tier**

# Data Storage FAQs

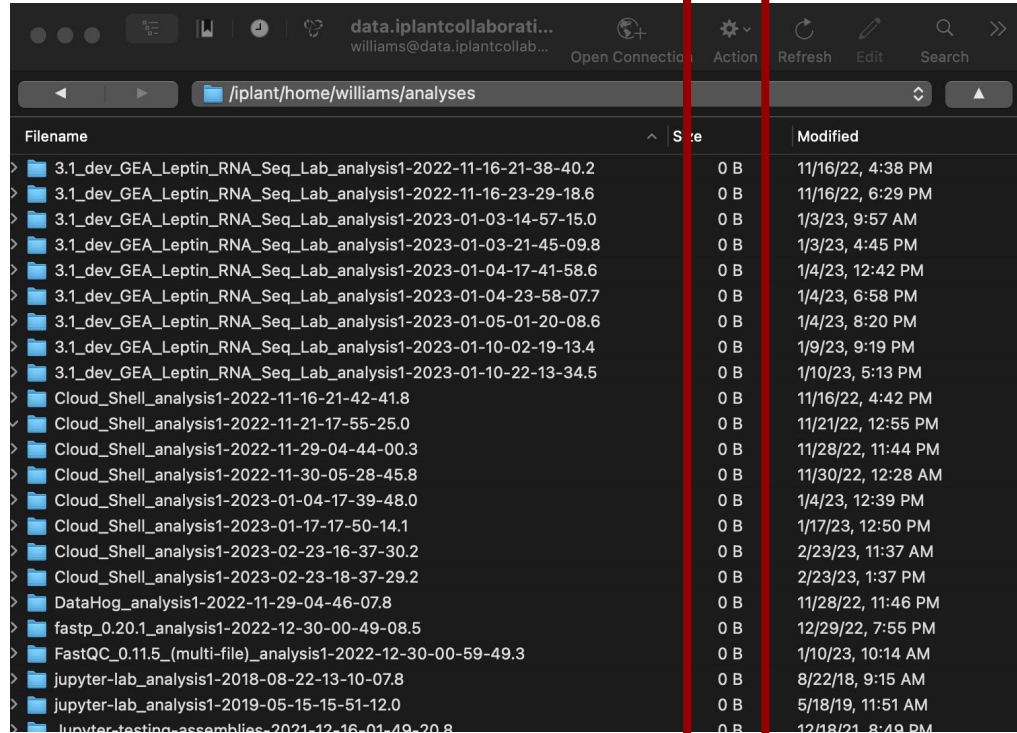
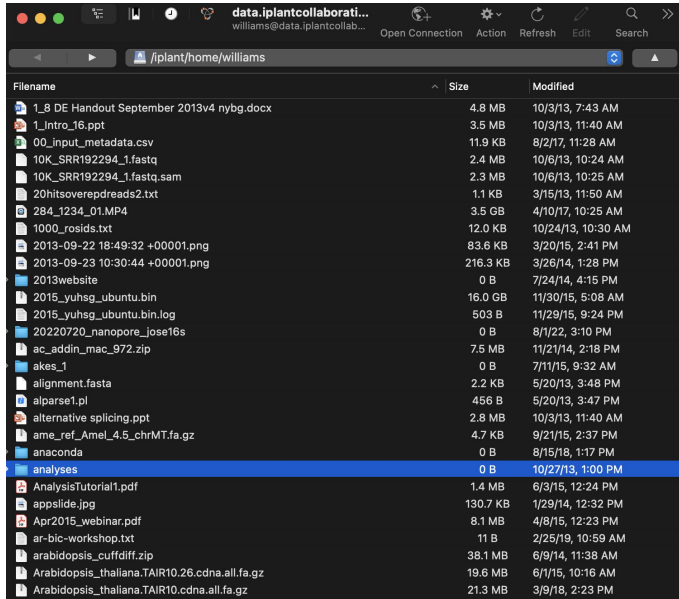
**Q: How can I tell how much data is in a folder?**

**A: See Cyberduck**





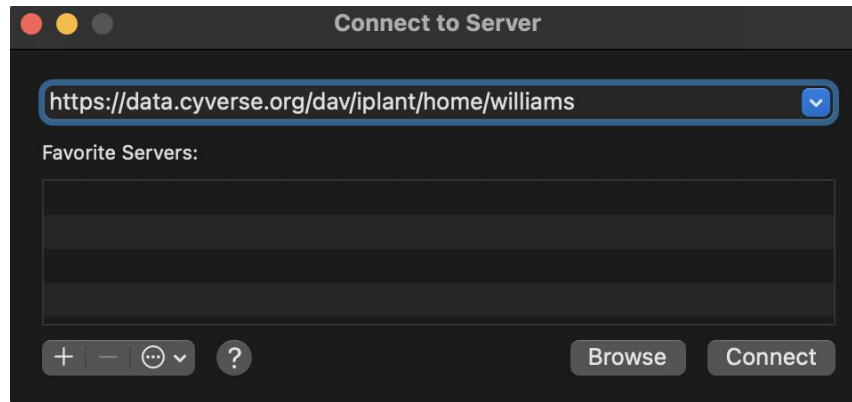
# Cyberduck



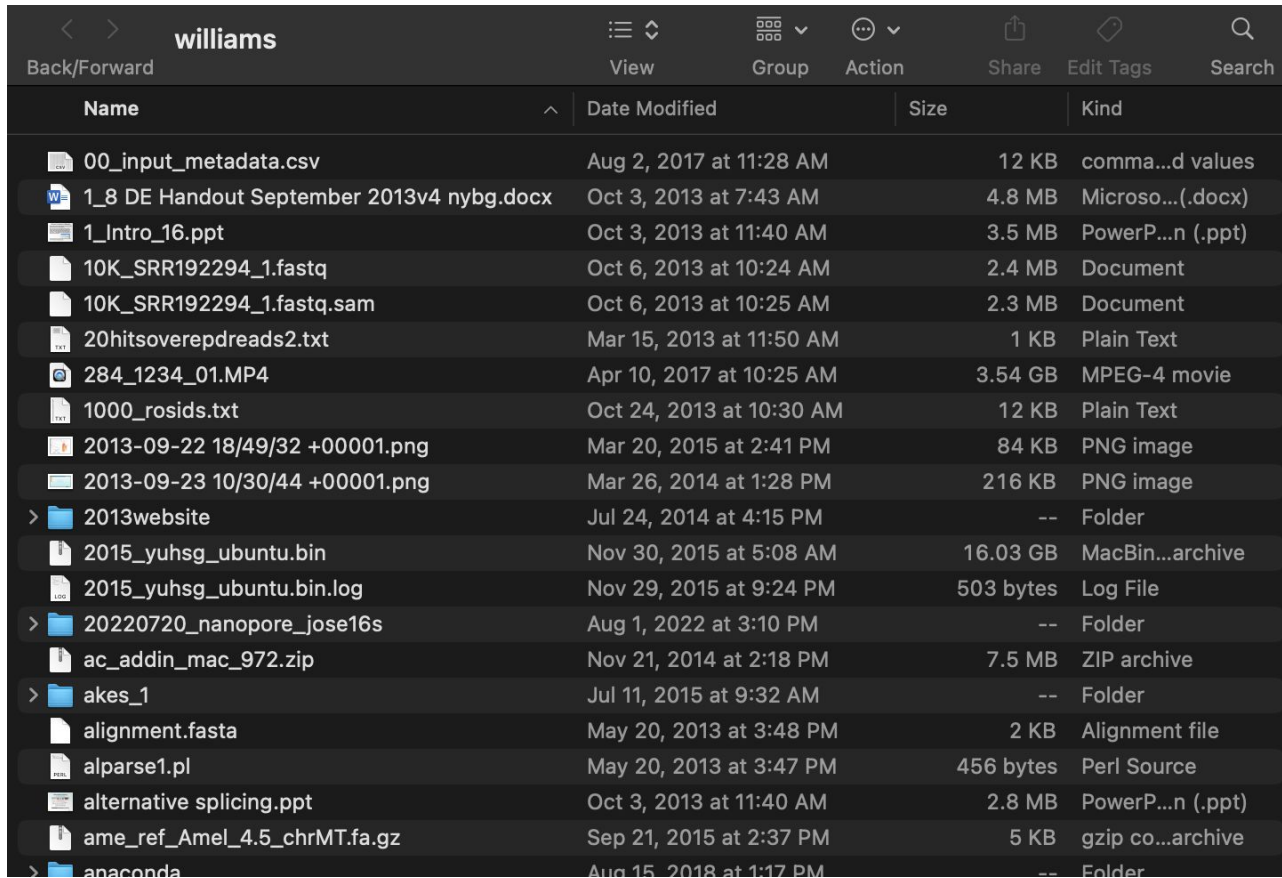
# Data Storage FAQs

**Q: How can I tell how much data is in a folder?**

**A: See WebDAV**



# WebDAV



Name	Date Modified	Size	Kind
00_input_metadata.csv	Aug 2, 2017 at 11:28 AM	12 KB	comma...d values
1_8 DE Handout September 2013v4 nybg.docx	Oct 3, 2013 at 7:43 AM	4.8 MB	Microso...( .docx)
1_Intro_16.ppt	Oct 3, 2013 at 11:40 AM	3.5 MB	PowerP...n (.ppt)
10K_SRR192294_1.fastq	Oct 6, 2013 at 10:24 AM	2.4 MB	Document
10K_SRR192294_1.fastq.sam	Oct 6, 2013 at 10:25 AM	2.3 MB	Document
20hitsoverepreads2.txt	Mar 15, 2013 at 11:50 AM	1 KB	Plain Text
284_1234_01.MP4	Apr 10, 2017 at 10:25 AM	3.54 GB	MPEG-4 movie
1000_rosids.txt	Oct 24, 2013 at 10:30 AM	12 KB	Plain Text
2013-09-22 18/49/32 +00001.png	Mar 20, 2015 at 2:41 PM	84 KB	PNG image
2013-09-23 10/30/44 +00001.png	Mar 26, 2014 at 1:28 PM	216 KB	PNG image
> 2013website	Jul 24, 2014 at 4:15 PM	--	Folder
2015_yuhsg_ubuntu.bin	Nov 30, 2015 at 5:08 AM	16.03 GB	MacBin...archive
2015_yuhsg_ubuntu.bin.log	Nov 29, 2015 at 9:24 PM	503 bytes	Log File
> 20220720_nanopore_jose16s	Aug 1, 2022 at 3:10 PM	--	Folder
ac_addin_mac_972.zip	Nov 21, 2014 at 2:18 PM	7.5 MB	ZIP archive
> akes_1	Jul 11, 2015 at 9:32 AM	--	Folder
alignment.fasta	May 20, 2013 at 3:48 PM	2 KB	Alignment file
alparse1.pl	May 20, 2013 at 3:47 PM	456 bytes	Perl Source
alternative splicing.ppt	Oct 3, 2013 at 11:40 AM	2.8 MB	PowerP...n (.ppt)
ame_ref_Amel_4.5_chrMT.fa.gz	Sep 21, 2015 at 2:37 PM	5 KB	gzip co...archive
> anaconda	Aug 15, 2018 at 1:17 PM	--	Folder



# Data Storage FAQs

**Q: How can I tell how much data is in a folder?**

**A: See iCommands**

```
(base) etacarinaeimac:profile williams$ ils -L /iplant/home/williams/284_1234_01.MP4|head -n1  
williams      0 CyVerseRes;ds14Broker;ds14 3540205382 2017-04-10.10:25 & 284_1234_01.MP4
```

# Data Storage FAQs

**Q: How can I tell how much data is in a folder?**

**A: See DataHog**



# Data Hog



DataHog

Summary

Browse Files

Duplicated Files

Manage File Sources

My iRODS Files (iRODS)

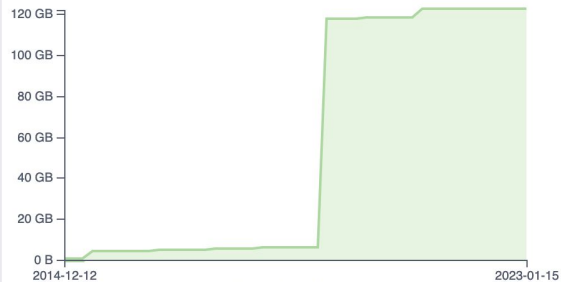
5837 files

122.14 GB

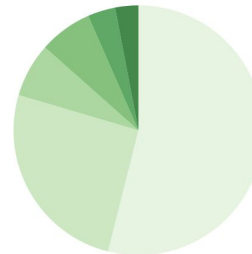
382 folders

Scanned 2023-03-16  
20:55

Total Space Occupied



File Type Breakdown



dat	65.86 GB
ovb	31.31 GB
WORKING	8.55 GB
zip	8.38 GB
fastq	4.38 GB

1-5



# Data Storage FAQs

**Q: How can I delete data?**

# Data Storage FAQs

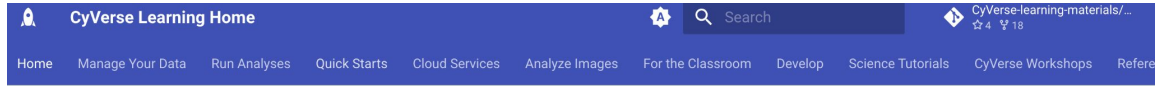
**Q: How can I delete data?**

**A: Your choice...**

- 1. Discovery Environment**
- 2. iCommands, WebDav, Cyberduck**



# Important resources



## Home

- [What is CyVerse?](#)
- [Self-Guided Tour](#)
- [Set Up an Account](#)
- [Getting Help](#)



Welcome to the CyVerse Learning Center

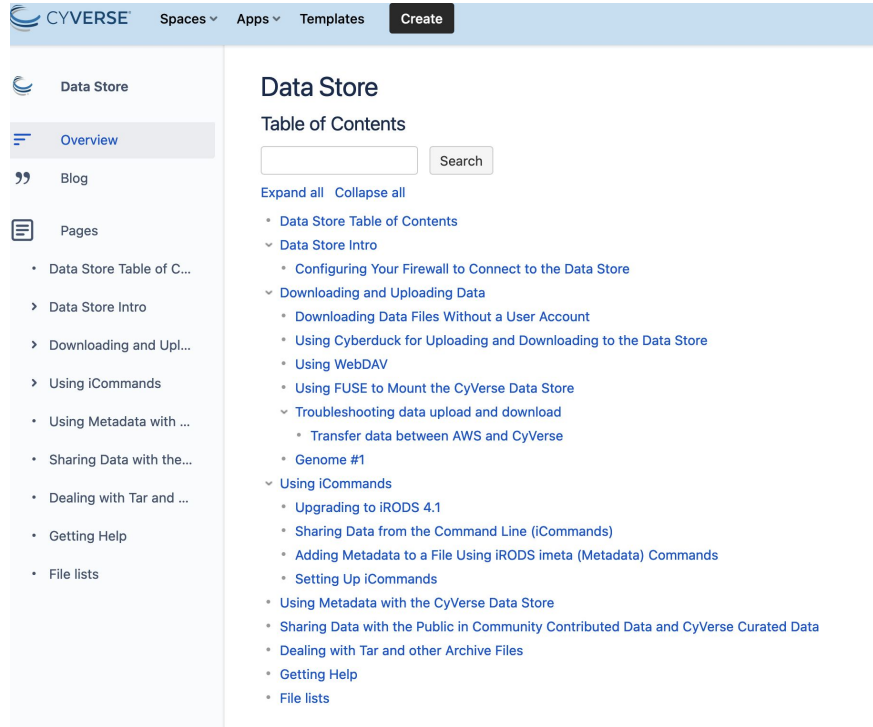
CyVerse provides life scientists with powerful computational infrastructure to handle huge datasets and complex analyses, thus enabling data-driven discovery.

In the Learning Center, you can find information about CyVerse platforms and documentation on how to use features and services, quick starts for data management and analysis tasks, tutorials, and more. Use the navigation bar on the left side of the screen or the search bar at the top to explore the learning materials.



**learning.cyverse.org**

# Important resources



The screenshot shows the CyVerse Data Store interface. At the top, there is a navigation bar with the CyVerse logo, 'Spaces', 'Apps', 'Templates', and a 'Create' button. On the left, a sidebar menu includes 'Data Store', 'Overview' (highlighted), 'Blog', and 'Pages'. The 'Pages' section lists various topics like 'Data Store Table of C...', 'Data Store Intro', 'Downloading and Uploading Data', 'Using iCommands', 'Using Metadata with ...', 'Sharing Data with the...', 'Dealing with Tar and ...', 'Getting Help', and 'File lists'. The main content area is titled 'Data Store' and 'Table of Contents'. It features a search input field with a 'Search' button and links for 'Expand all' and 'Collapse all'. Below this is a detailed list of links, including 'Data Store Table of Contents', 'Data Store Intro', 'Configuring Your Firewall to Connect to the Data Store', 'Downloading and Uploading Data', 'Downloading Data Files Without a User Account', 'Using Cyberduck for Uploading and Downloading to the Data Store', 'Using WebDAV', 'Using FUSE to Mount the CyVerse Data Store', 'Troubleshooting data upload and download', 'Transfer data between AWS and CyVerse', 'Genome #1', 'Using iCommands', 'Upgrading to iRODS 4.1', 'Sharing Data from the Command Line (iCommands)', 'Adding Metadata to a File Using iRODS imeta (Metadata) Commands', 'Setting Up iCommands', 'Using Metadata with the CyVerse Data Store', 'Sharing Data with the Public in Community Contributed Data and CyVerse Curated Data', 'Dealing with Tar and other Archive Files', 'Getting Help', and 'File lists'.

## Data Store manual



[support@cyverse.org](mailto:support@cyverse.org)



Thank you!



Start Using the Discovery Environment Now!

[de.cyverse.org](https://de.cyverse.org)

### DATA & ANALYSIS

- 5 GB free data storage
- 100s open-source scientific apps
- Containers and notebooks
- Visualize & interact with data

### YOUR WORKSPACE

- Manage and share data
- Perform analysis with your own or community datasets
- Do large-scale science from your web browser
- Build and customize apps

### COLLABORATION

- Secure, shared workspace for your team
- Reproducibility tools
- Manage the data lifecycle
- Make data more FAIR
- Open your science

### LEARN, TEACH & TRAIN

- Tutorials and documentation
- Webinars
- Workshops
- Teach using CyVerse

### COMMUNITY

- Join 95K+ users
- In-app chat support
- Find publicly available data
- Share data and analyses
- Deploy your own CyVerse



[Sign Up | Subscribe](#)  
[user.cyverse.org](https://user.cyverse.org)

[Learn More](#)  
[www.cyverse.org](https://www.cyverse.org)