Using the SeedMeLab Data Management System on Comet

Amit Chourasia, SDSC/UCSD

XSEDE Webinar: Mar 19, 2020

Comet data share website: hpcshare.sdsc.edu

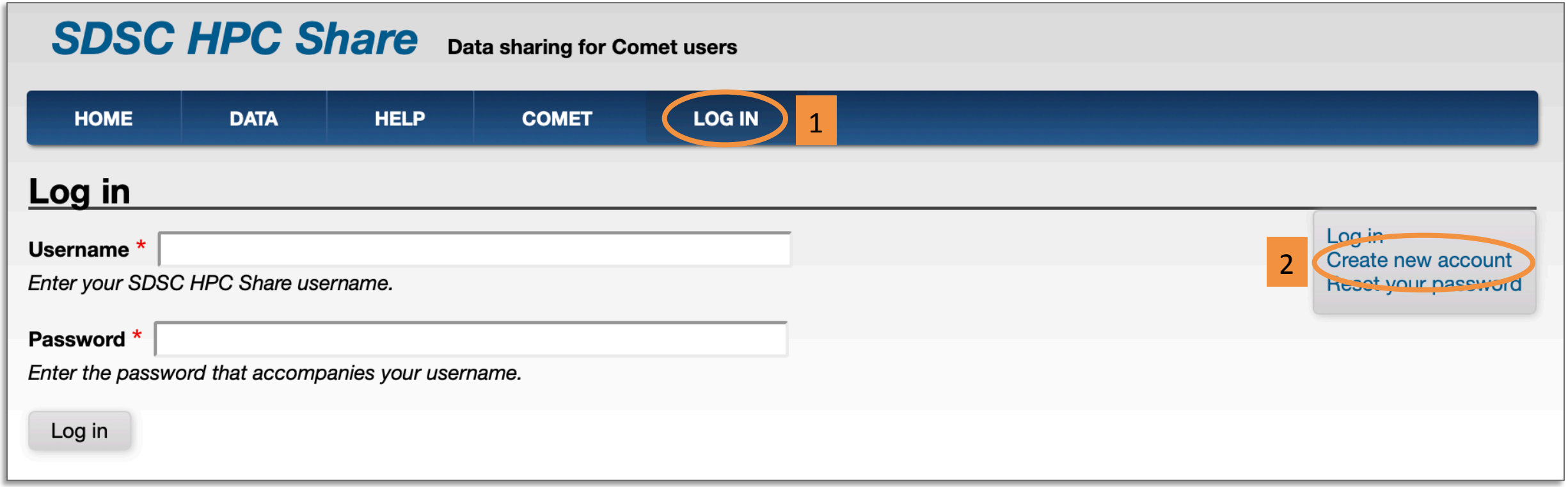
Slides: hpcshare.sdsc.edu/foldershare/96

SeedMeLab

1. Self register at hpcshare.sdsc.edu
2. SeedMeLab overview
    a. Problem & Motivation
    b. Key features
    c. Use cases
3. Usage from web browser
4. Usage from Comet
5. Preview: set up for command line usage from other computers
6. Restrictions
7. Discussion

Create your account

Visit hpcshare.sdsc.edu



To prevent spam self registration is restricted to emails ending in .edu or .gov or .unavco.org domains

Contact us for an invite if you don't have such email

Motivation

Data creation ^n

Data use/reuse ++

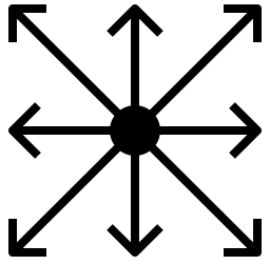
Research team size +

Trends

- Data doubling every year
- Sprawling data governance issues
- Open data and FAIR data mandates by Government & Sponsors

Lack of powerful and rich data management solutions

Problems



Information fragmentation

Scattered data is harder to manage and use well



Weak presentation

Data presentation with mere listing lacks depth



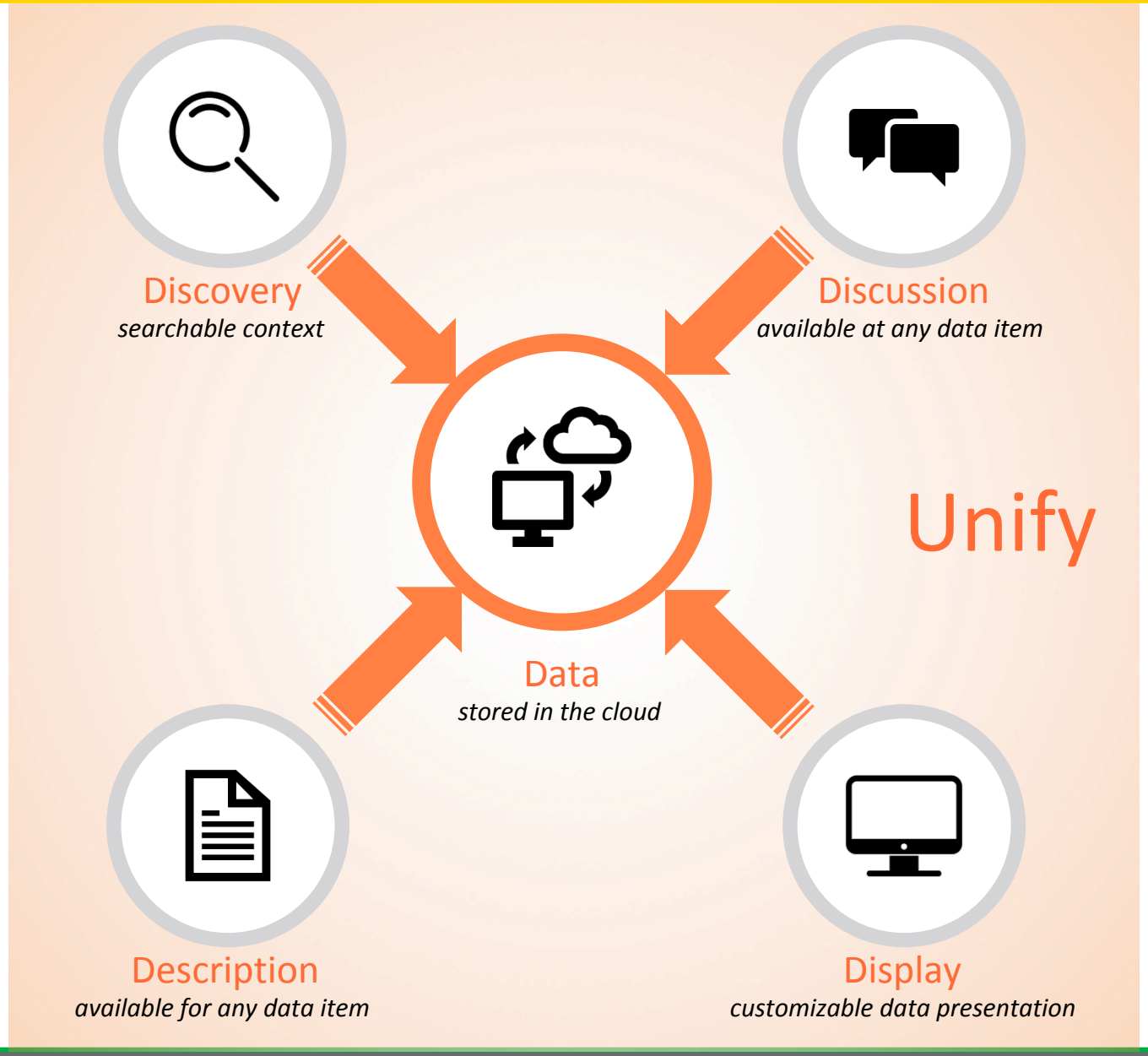
Dark data

Without context, data is harder to find, use and reuse

Stumbling blocks

- Storage
- Transfer
- Automation
- Access control
- Collaboration

Goal



A standalone system with full control for research team or application.



http://SeedMeLab.org

Share…

Manage…

Search…

Visualize…

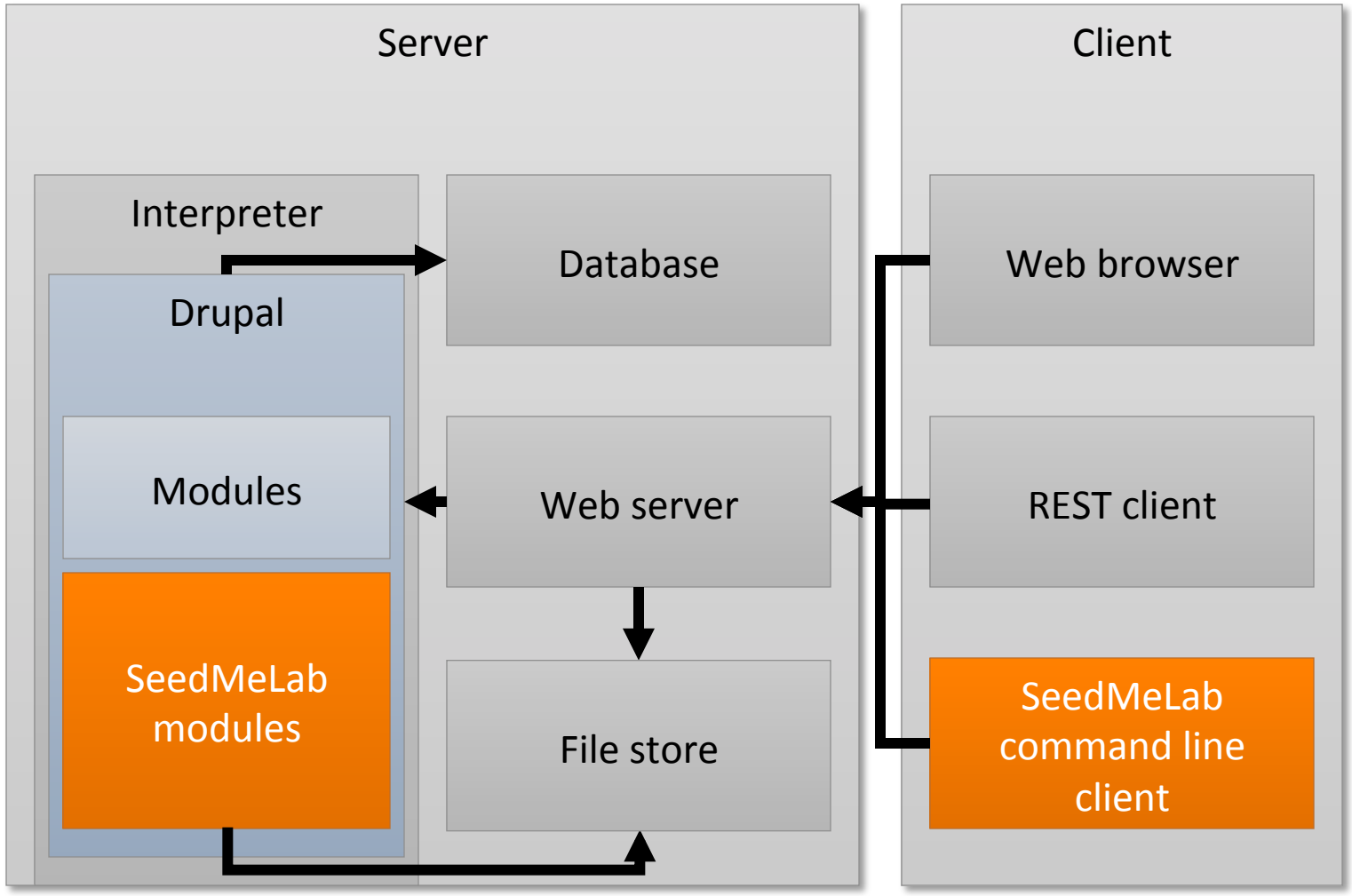
…data, *as never before!*

SeedMe : Stream Encode Explore and Disseminate My Experiments

SeedMeLab is a web based data management solution
that enables research teams or cyberinfrastructure system users
struggling with data and its (access/metadata/description & discussion/presentation)
to accelerate research by boosting knowledge transfer/retention.

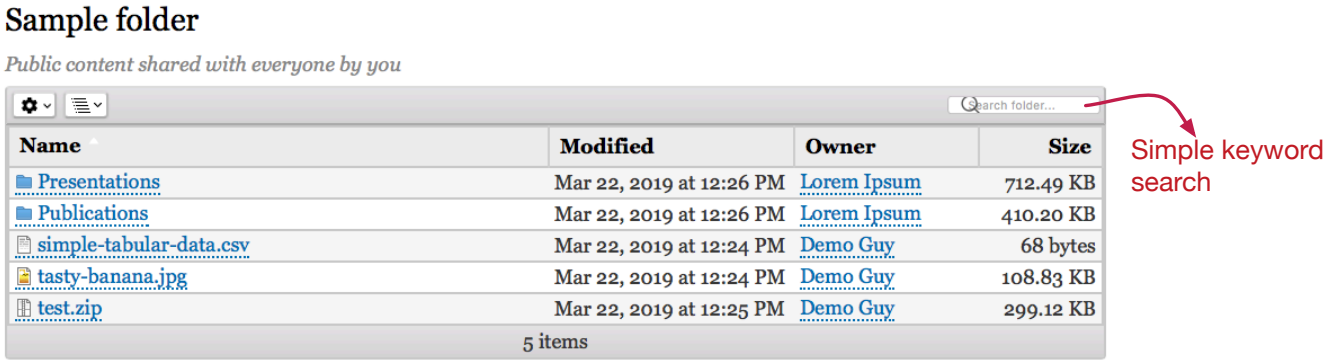
SeedMeLab

- Specialized modules for Drupal
  - Shared files and folders
  - Web services for REST clients
  - Light visualization
  - Single sign-on
- Specialized REST client
  - Command-line access to remote shared files
- Open Source



Preview SeedMeLab user interface and features

Data organization and search



Data description

Sample folder

Public content shared with everyone by you

Name	Modified	Owner	Size
Presentations	Mar 22, 2019 at 12:26 PM	Lorem Ipsum	712.49 KB
Publications	Mar 22, 2019 at 12:26 PM	Lorem Ipsum	410.20 KB
simple-tabular-data.csv	Mar 22, 2019 at 12:24 PM	Demo Guy	68 bytes
tasty-banana.jpg	Mar 22, 2019 at 12:24 PM	Demo Guy	108.83 KB
test.zip	Mar 22, 2019 at 12:25 PM	Demo Guy	299.12 KB
5 items			

Add description to any file/folder via '*Edit description*' menu option.

Features	More features
Organize and Share data Organize data as familiar folders/files and easily share it privately with collaborators or make it public	**Automate** Automate your data flow with provide REST client and command line utility
Rich metadata Add new fields to your data for quick review, filtering and search	**Customize** Operations that are available to different roles

Add description to any file/folder

Equations (Use Latex syntax)

When $a \neq 0$, there are two solutions to $ax^2 + bx + c = 0$ and they are

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Data Discussion

Sample folder

Public content shared with everyone by you

Search folder...

Name	Modified	Owner	Size
Presentations	Mar 22, 2019 at 12:26 PM	Lorem Ipsum	712.49 KB
Publications	Mar 22, 2019 at 12:26 PM	Lorem Ipsum	410.20 KB
simple-tabular-data.csv	Mar 22, 2019 at 12:24 PM	Demo Guy	68 bytes
tasty-banana.jpg	Mar 22, 2019 at 12:24 PM	Demo Guy	108.83 KB
test.zip	Mar 22, 2019 at 12:25 PM	Demo Guy	299.12 KB
5 items			

Add description to any file/folder via '*Edit description*' menu option.

Features	More features
Organize and Share data Organize data as familiar folders/files and easily share it privately with collaborators or make it public	**Automate** Automate your data flow with provide REST client and command line utility
Rich metadata Add new fields to your data for quick review, filtering and search	**Customize** Operations that are available to different roles

FolderShare comments

Demo Guy
Mon, 03/25/2019 - 17:28
Permalink

All files/folders can...

All files/folders can include comments.

Edit Reply

Lorem Ipsum
Mon, 03/25/2019 - 17:30
Permalink

Thats fantastic! Now our...

Thats fantastic! Now our discussion and data are in one place.

Reply

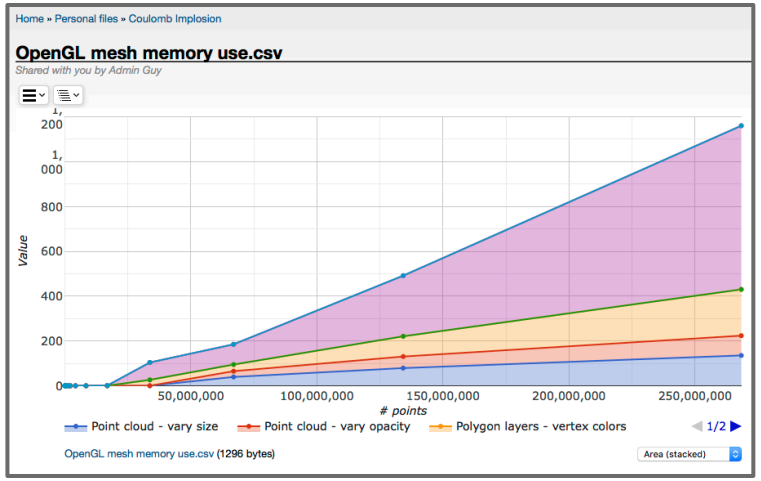
FolderShare entity is fieldable, add comments, etc.

Light visualization

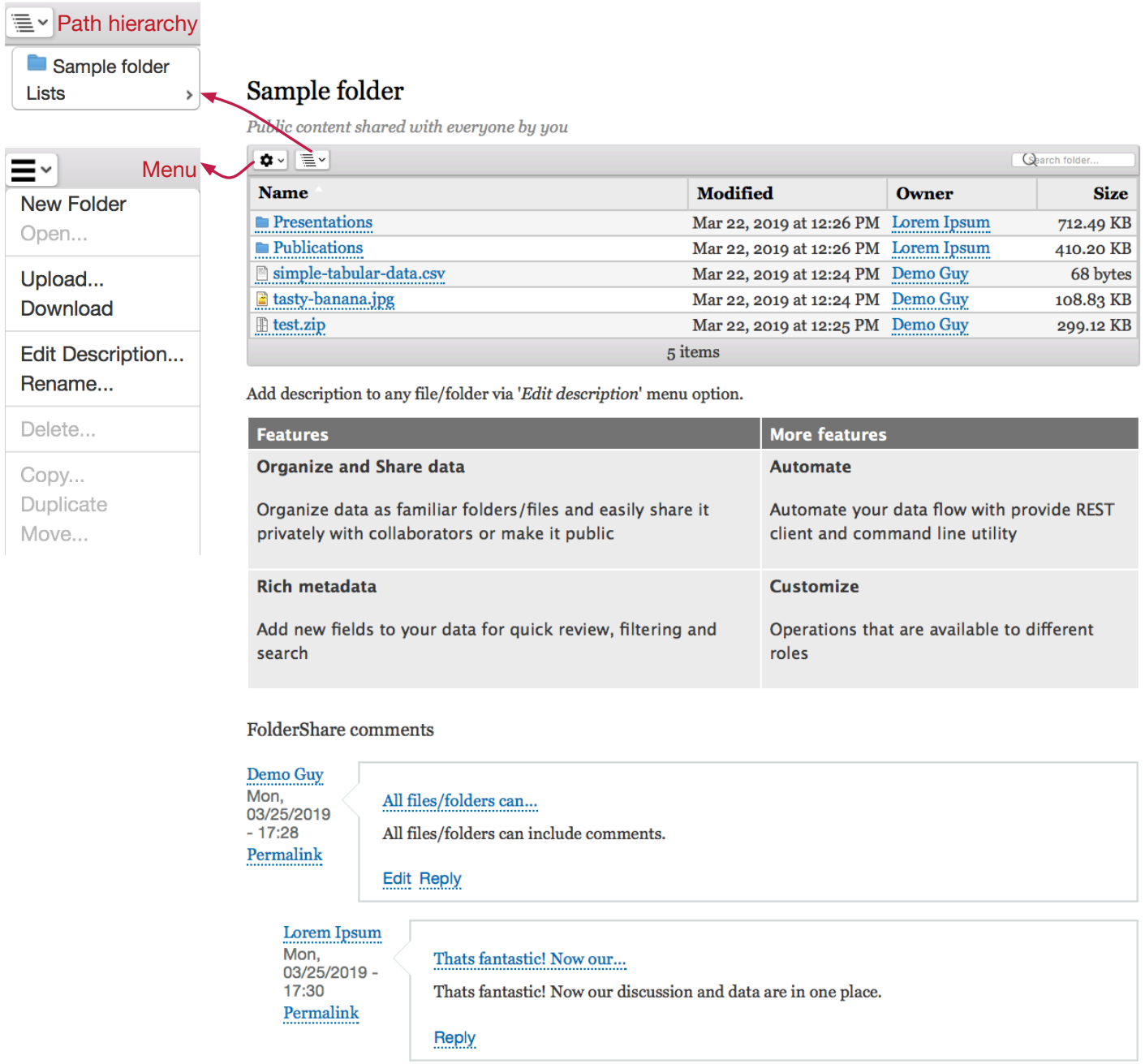
Automatic visualization for

- Simple tabular data in csv, json formats
- PDB files
- Newick files

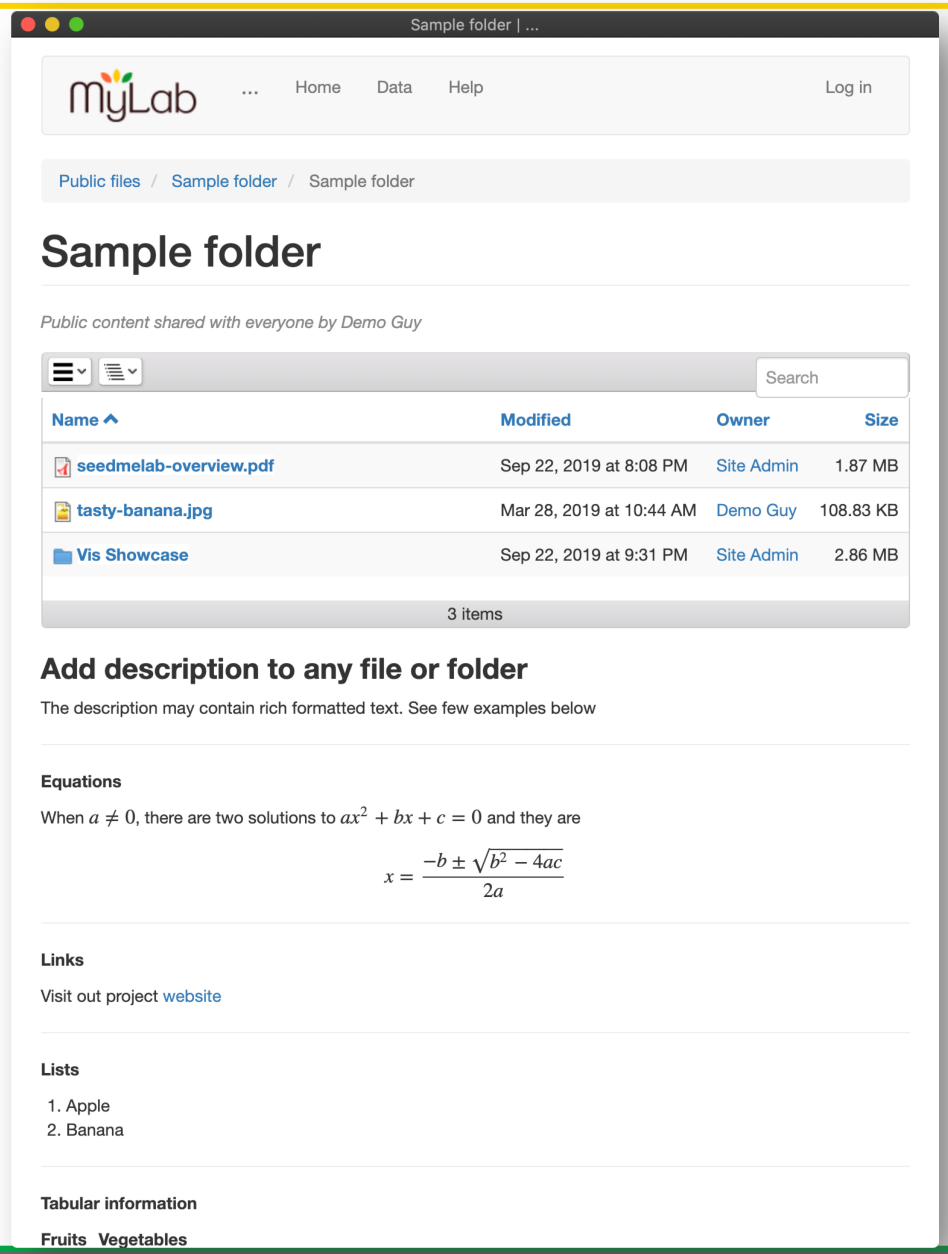




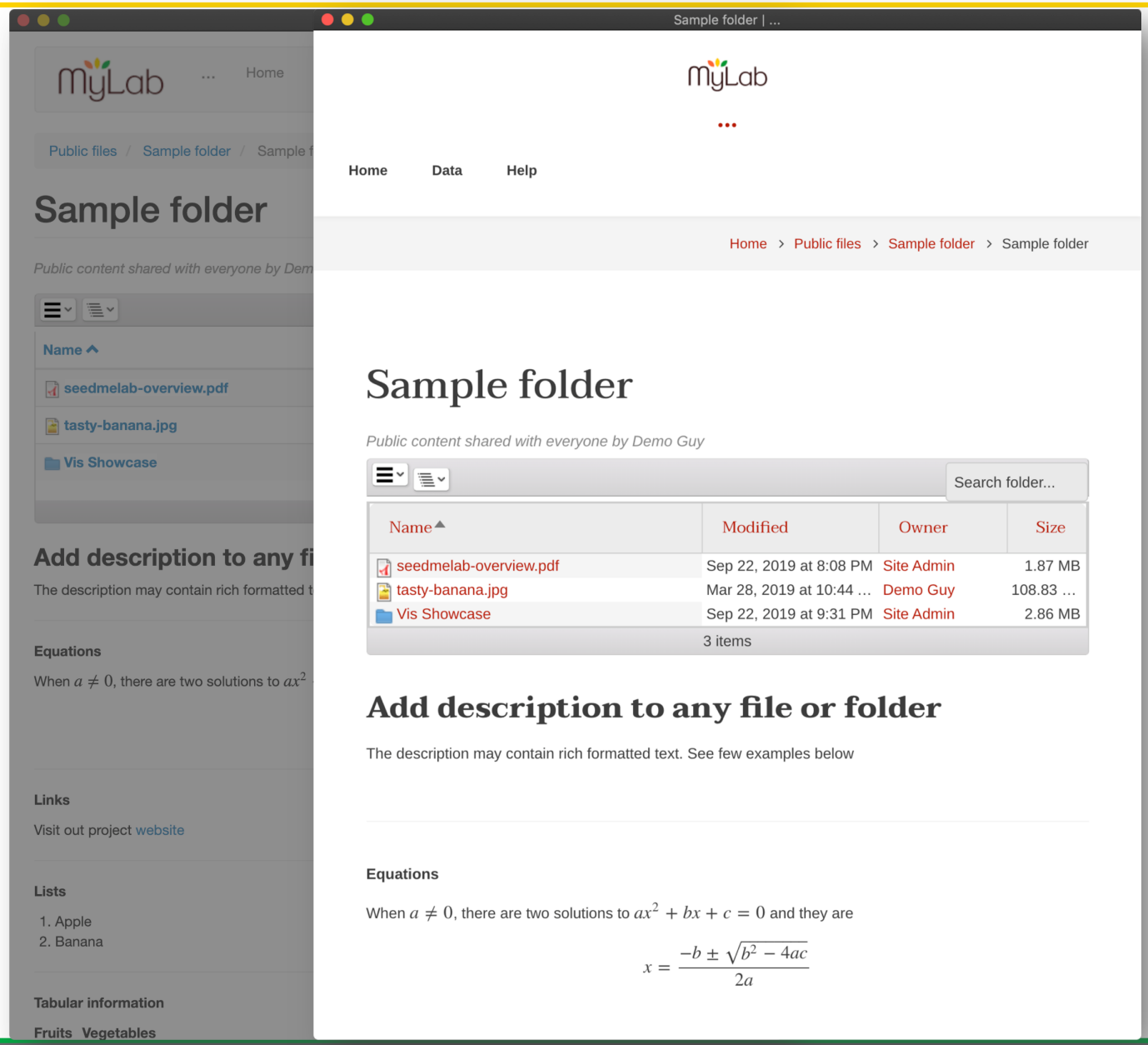
Data management



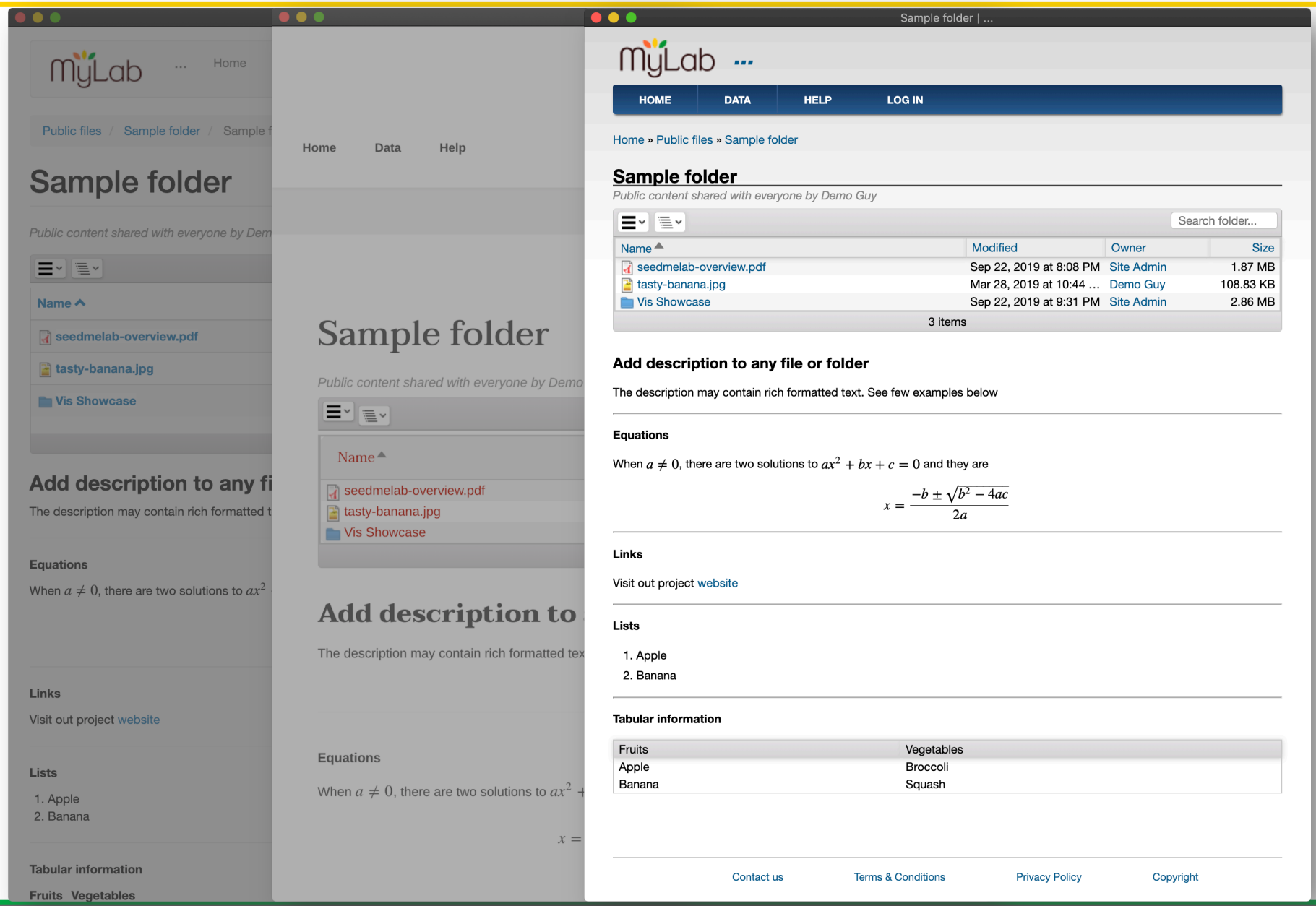
Personalized layout & branding



Personalized layout & branding



Personalized layout & branding



SeedMeLab has comprehensive administrative features such as

- Restrict upload file types and size
- Add new file system fields
- Change display layout
- Etc

If useful, follow up for a discussion

SeedMeLab - Use cases

Research groups

- Use as collaboration hub and/or group website
- Implement and realize data management plan for various projects
- Position research groups well to comply with **FAIR data mandate**
- **Example: Laser Plasma Lab @ UCSD**

Applications

- Enable applications with rich data sharing and data management
- Enable application users to contribute and share reusable content
- Data publishing
- **Example: CIPRES Gateway, GenApp Gateway, FlowGate & NEMAR data repositories**

Cyberinfrastructure services providers

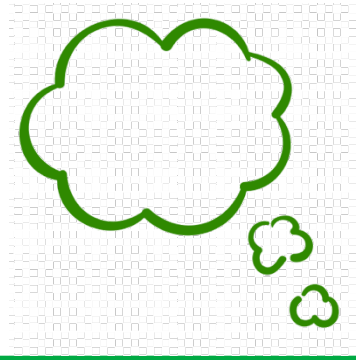
- Deploy instance of SeedMeLab for their user base
- **Example: HPC Share @ SDSC, UCSD**

Feedback from early adopters

“My inbox is decongested… swift feedback to team … easier to write research papers.”
- *Alexey Arefiev, UC San Diego*

“Fulfills user demand for data sharing … easy integration with little development effort that leaves us time left over to focus on other opportunities.”
- *Mark Miller*, CIPRES project, UC San Diego

“Provides advanced file system to GenApp users and applications … saving our team’s time to focus on other development priorities.”
- *Emre Brookes*, GenApp Framework, UT Health Science Center San Antonio



SeedMeLab

HPC Share: Hands on usage

1. Web browser
2. Command line on Comet
3. Command line setup on other computers

HPC Share – Usage via Web browser

Open: https://hpcshare.sdsc.edu
Visit: ***Data*** menu

Manage data
Use the gear icon
- Create New folder
- Upload files into “New folder”

Share data
Only possible from top level files/folders under *Personal files*
Select one file or folder > Chose Share or Invite action from gear icon
- Share with site members
- Invite non-site members

Visualize data
Visit https://hpcshare.sdsc.edu/foldershare/2
Open any file in “*Vis Showcase”* folder

Access controls

Share access controls can be only set at the top level files/folders from Personal files.
All descendants of top level folder have same access privileges

Following privileges are available
Everyone: Read only for anyone that can access this site aka Public
View: Read only for specified users
Author: Read and write privilege for specified users

Use Share action on top level file/folder to grant access to site user or make it public
Use Invite action on top level file/folder to grant access to non-site user

HPC Share – Comet usage documentation

Open: https://hpcshare.sdsc.edu

1. Visit: *Help* menu
2. Visit: *Usage from Comet cluster at SDSC*

Log in to Comet via SSH

```
# You may may add the following to your ~/.bashrc file
export MODULEPATH=$MODULEPATH:/software/community/seedmelab/modules

# Now load the foldershare module with version noted below as follows
module load foldershare/1.1.0

# Get help for foldershare client
foldershare --help
```

HPC Share – Comet interactive usage

Follow along as demonstrated hands on

Usage from Comet: Scriptable usage

Create full commands and add to bash script

SeedMeLab

Recipe

1. Install PHP
2. Download foldershare client
3. Usage similar to comet

Documentation: https://hpcshare.sdsc.edu

1. Visit: *Help* menu
2. Visit: *Command line setup on other computers*

Follow instructions in # 2 above

1. Maximum upload size for single file 2 GB
2. Access control supported from top level only (not like POSIX)

3. POSIX groups are not supported
4. Data versioning is not supported

Final remarks

1. hpcshare.sdsc.edu available for use by Comet users
2. Contact us for feedback/support
3. Available for one-on-one consultation to discuss your use case

Acknowledgements

All users and application integrators for their kind support & feedback.
SGCI (Focus Week + Usability study + additional engagements)

SeedMeLab Team: Amit Chourasia, David Nadeau, Michael Norman, Scott Sakai and RDS
Masters, Undergraduate and high school interns (17 students thus far)

National Science Foundation

This material is based upon work supported by the National Science Foundation under Grant No. OAC-1443083 & OAC-1235505

"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."