Rich Data Sharing for HPC Users

Amit Chourasia, San Diego Supercomputer Center, UC San Diego

http://SeedMeLab.org

Share...
Manage...
Search...
Visualize...

...data, as never before!
Presentation slides: https://hpcshare.sdsc.edu/foldershare/96

Schedule zoom conversation
https://calendly.com/achourasia

Contact
help@seedmelab.org

Video overview of SeedMeLab
1. What is SeedMeLab? (90 sec)
2. Managing data on SeedMeLab for users (3 mins)
3. Managing data via command line REST client for automation and power users (5 mins)
4. Customizing SeedMeLab for Principal Investigator/Project manager (4 mins)

Resources
HPCShare help documentation
SeedMelab quick trial from web browser
SeedMeLab Dockerized recipe for technical adopters/evaluators

Upcoming event
Webinar: Implementing Research Data Management for Labs & Grants
Apr 28, 2021 | Event details
1. Self register at [hpcshare.sdsc.edu](http://hpcshare.sdsc.edu)
2. SeedMeLab overview (software that powers hpcshare)
   a. Problem & Motivation
   b. Use cases
3. Using HPC Share
   a. Key Features
   b. Web browser use demonstration
   c. Command line use - Hands-on from Expanse cluster
   d. Review: Set up for command line usage from other computers
4. Discussion
Create your account at hpcshare.sdsc.edu

Visit [hpcshare.sdsc.edu](http://hpcshare.sdsc.edu)

Self registration is restricted to emails ending in .edu or .gov or .unavco.org domains.

If you don’t have such an email address, send a chat message to Jeff Sale to get an invite.
Acknowledgements

Users and integrators for their support & feedback.
Masters, undergraduate & high school interns

Science Gateways Community Institute
Focus Week + Additional engagements including Usability, Marketing & Value proposition

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

“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.”

San Diego Supercomputer Center, UC San Diego
Data stumbling blocks

Stumbling blocks

- Access control
- Collaboration
- Storage
- Transfer
- Automation
Problems for stakeholders

Information dispersion
- Missing context – emails, meeting notes, protocols, parameters, etc
- Scattered data is harder to manage and use well
- Lack of control on where data is stored shared

Weak presentation
- Data presentation requires context to give it depth and meaning

Dark data
- Without context, data is harder to find and use
Data shall be
1. Accessible on the web via Browser, API & Command line
2. Share-able with access control
3. Annotate-able with context & metadata
4. Presentable in a rich format

Built on framework (Drupal) that is
1. Easily usable, customizable and extensible
2. Deployable in turn-key fashion
3. Mature & sustainable
4. Supported by a large ecosystem
5. Open source & distributable ([Docker’ized recipe](#))
HPCShare key features

• Upload/download (<2 GB/per file) from Expanse cluster or elsewhere to Web
  Via command line, API or web browser.

• Annotate data with rich text
  Such as arbitrary context, metadata, equations, etc.

• Create preview visualizations of small tabular data
  CSV and JSON formats.

• Collaborate
  Via data sharing and comments on data.
Watch video: SeedMeLab overview
Data Management

Organize and manage data
SeedMeLab File Manager
Builds upon status quo
Standard tree hierarchy for files and folders
Standard file operations

Upload files via drag & drop
Data Expressiveness

Describe & Discuss Data
Add custom metadata fields filesystem wide
e.g. Identifiers, dates, keywords, taxonomy, etc.

Add rich description to any file/folder with
* Formatted text
* Link, list & table
* Images & videos
* Equations
Sample folder

Public content shared with everyone by you

<table>
<thead>
<tr>
<th>Name</th>
<th>Modified</th>
<th>Size</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Images</td>
<td>11/26/20 1:07 AM</td>
<td>1.14 MB</td>
<td>Site Admin</td>
</tr>
<tr>
<td>Vis Showcase</td>
<td>11/26/20 1:07 AM</td>
<td>2.87 MB</td>
<td>Site Admin</td>
</tr>
<tr>
<td>Visible spectrum.png</td>
<td>12/13/20 10:52 PM</td>
<td>607 bytes</td>
<td>Site Admin</td>
</tr>
<tr>
<td>mandelbrot.rom4</td>
<td>11/26/20 1:07 AM</td>
<td>744.01 KB</td>
<td>Site Admin</td>
</tr>
<tr>
<td>seedmelab-overview.pdf</td>
<td>11/26/20 1:08 AM</td>
<td>1.87 MB</td>
<td>Site Admin</td>
</tr>
</tbody>
</table>

The visible spectrum is the portion of the electromagnetic spectrum that is visible to the human eye.

### Properties of few colors in visible spectrum

<table>
<thead>
<tr>
<th>Color</th>
<th>Wavelength</th>
<th>Frequency</th>
<th>Photon energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Violet</td>
<td>380-450 nm</td>
<td>670-790 THz</td>
<td>2.75-3.26 eV</td>
</tr>
<tr>
<td>Blue</td>
<td>450-485 nm</td>
<td>620-670 THz</td>
<td>2.56-2.75 eV</td>
</tr>
</tbody>
</table>

Roots of quadratic equations
When \( a \neq 0 \), there are two solutions to \( ax^2 + bx + c = 0 \) and they are

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

Comments
Submitted by Site Admin on Fri, 12/27/2019 - 14:58

Comments may be posted to discuss any file or folder.
Present, brand and visualize data
Add custom metadata fields filesystem wide
e.g. Identifiers, dates, keywords, taxonomy, etc.
Search data and context
The **visible spectrum** is the portion of the electromagnetic spectrum that is visible to the human eye.

<table>
<thead>
<tr>
<th></th>
<th>V</th>
<th>450</th>
<th>B</th>
<th>480</th>
<th>G</th>
<th>530</th>
<th>Y</th>
<th>580</th>
<th>O</th>
<th>640</th>
<th>R</th>
<th>750</th>
</tr>
</thead>
</table>

Properties of few colors in visible spectrum

<table>
<thead>
<tr>
<th>Color</th>
<th>Wavelength</th>
<th>Frequency</th>
<th>Photon energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Violet</td>
<td>380-450 nm</td>
<td>670-790 THz</td>
<td>2.75-3.26 eV</td>
</tr>
<tr>
<td>Blue</td>
<td>450-485 nm</td>
<td>620-670 THz</td>
<td>2.56-2.75 eV</td>
</tr>
</tbody>
</table>

Roots of quadratic equations

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
Description of any data item
Discussion on any data item
Display with custom layout & visualization
Discover via searchable context
SeedMeLab: Use cases

**Research groups**
- Provide a branded data management system
  - Private collaboration hub
  - Public dissemination hub
  **Examples: Laser Plasma Lab, Structural Engineering Group, Material Science Group**

**Gateways and scientific applications**
- Integrate with existing systems
- Develop new applications
  **Examples: FlowGate portal, CIPRES Gateway, GenApp Gateway**

**Service Providers**
- Customize and offer SeedMeLab to their users
  **Examples: HPCShare @ SDSC**
Demonstration videos
Watch video: Managing data on SeedMeLab
Watch video: Customizing SeedMeLab
Watch video: Managing data via REST client on SeedMeLab
Auto generate and present visualization for any filetype (via plugins)

See Visualization Showcase Sample format for CSV & JSON files
1. On terminal logon to Expanse via SSH
2. **Open** *Usage from Expanse cluster* help page in your Web browser
3. Command line utility is named as foldershare

```bash
# Add location of module
export MODULEPATH=$MODULEPATH:/expanse/community/seedmelab/modules

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

# Create a shortcut/alias
fs="foldershare --host hpcshare.sdsc.edu --username YOUR_USERNAME --password YOUR_PASSWORD"

# Add the above to your ~/.bashrc file
```
# Get help for foldershare client
foldershare --help
Or
$fs --help

Follow along as demonstrated hands on
• FolderShare REST client is written in PHP
• Check if php already installed (Mac’s ship with php), else install it
• Download foldershare client
• Follow usage as demonstrate for Expanse
• See instructions on the help page
Help
https://hpcshare.sdsc.edu/help

Review REST client API on Expanse cluster at
/expanse/community/seedmelab/bin/foldershare_php_client.1.2/api

Upcoming REST client features
Using authentication from a json file
Python wrapper for foldershare client

Contact
help@seedme.org
Discussion

Contact
help@seedmelab.org

Schedule zoom conversation
https://calendly.com/achourasia

Website
SeedMeLab.org: Try - Adopt – Share

Resources
HPCShare help documentation
SeedMeLab quick trial from web browser
SeedMeLab Docker’ized recipe for technical adopters/evaluators

Upcoming event
Webinar: Implementing Research Data Management for Labs & Grants
Apr 28, 2021 | Event details