Data sharing building blocks

A SeedMe project

FolderShare: Building a data sharing cloud on Drupal 8 for researchers

Amit Chourasia, **David Nadeau** & Michael Norman

San Diego Supercomputer Center, UC San Diego

Project code: dibbs.seedme.org/downloads

or

drupal.org/projects/foldershare

drupal.org/projects/smalldata

Trial website: sandbox.seedme.org

Project website: dibbs.seedme.org

Presentation Overview

1. Background & motivation
2. Architecture
3. SeedMe2 platform
   - **FolderShare: Virtual file system for Drupal**
     - » Entity data model & access control
     - » File management & security
     - » Views integration
     - » UI & Command plugins
     - » File formatters
     - » Web services
   - **Small data module and API**
     - – Visualization
4. Target users/Use cases
5. Screenshots
6. **Demo**

SeedMe Projects

SeedMe 1

a.k.a. Stream **e**ncode **e**xplore and **d**isseminate **M**y **e**xperiments

- **Video encoding from visualization images was the main focus**
- Based on Drupal 7
- In production as Platform as a Service (PaaS)

SeedMe 2 : Data sharing building blocks

Evolution of the original SeedMe project (Complete rewrite)

- **Data organization, sharing and data management is the main focus**
- Incorporates user feedback from original SeedMe project
- Based on Drupal 8
- Built for distribution (open source) and extension

SeedMe 2's focus

- **Enable rapid access to data to consumable data over the web**

  Must be handled by stock web browser (Upload < 2GB per file)

  Must be displayable on many devices (Phone to PC)

- **Rich presentation**

  Support customization/extension

- **Discoverability**

  Index and search files, descriptions, etc..

- **Three way interaction**

  Web browser

  Command line

  API

Data management stumbling blocks

- Transfer
- Storage
- Access control
- Collaboration
- Automation

But what about Presentation and Discovery?

Data management stumbling blocks

- Transfer
- Storage
- Access control
- Collaboration
- Automation

Three D's : Data, Description, Discussion

Issues due to content dispersion

- Data in the Cloud
- Description in someone's mind
- Discussion on emails

Related solutions

Filesystem based solutions

Tools
SCP
FTP

Middleware
Globus
IRods
NEWT

Software repositories
GitHub
SVN
CVS

File hosting
Cloud drives
(Dropbox, etc)
WebDAV

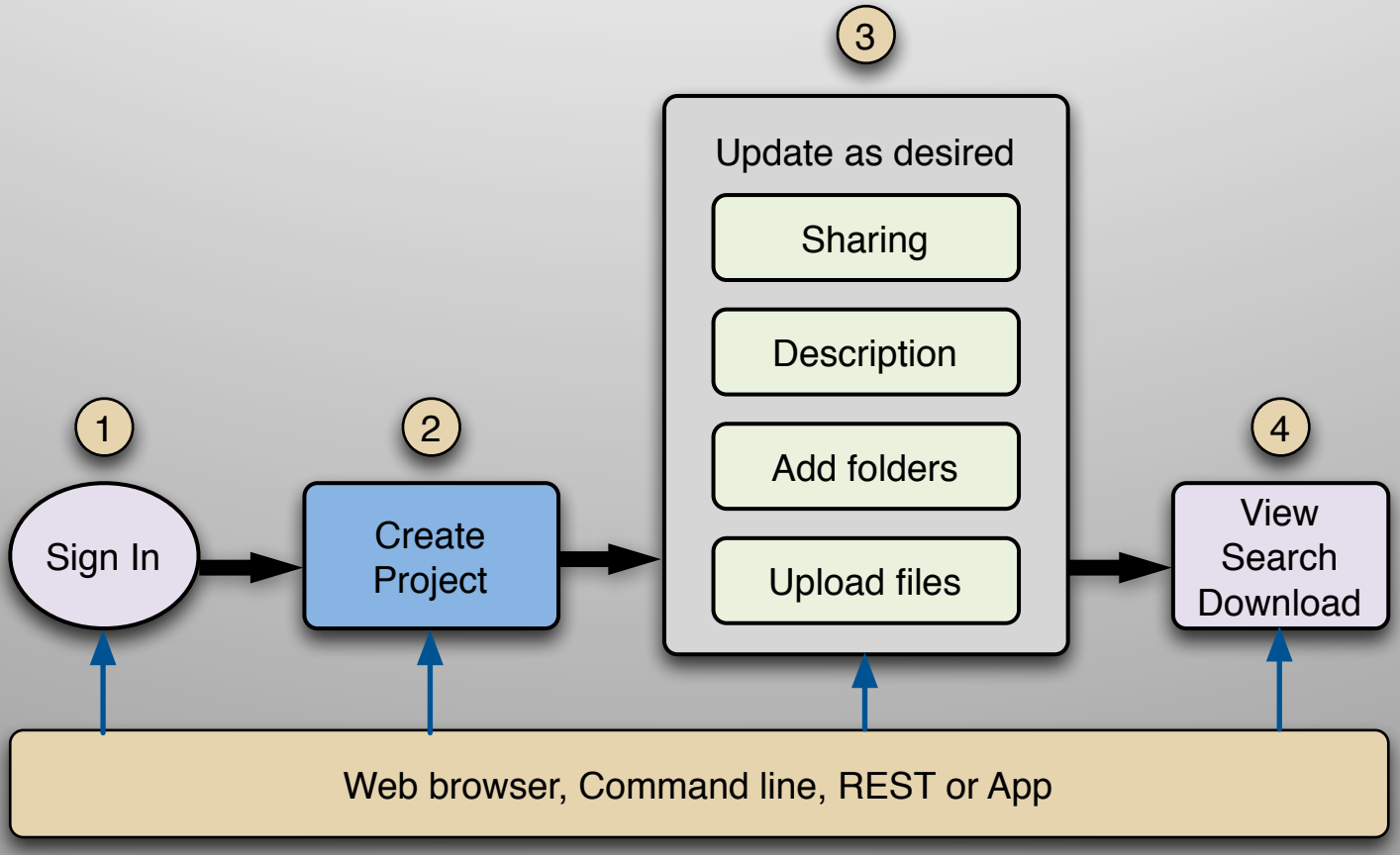
Content management system based solutions

HubZero
FigShare

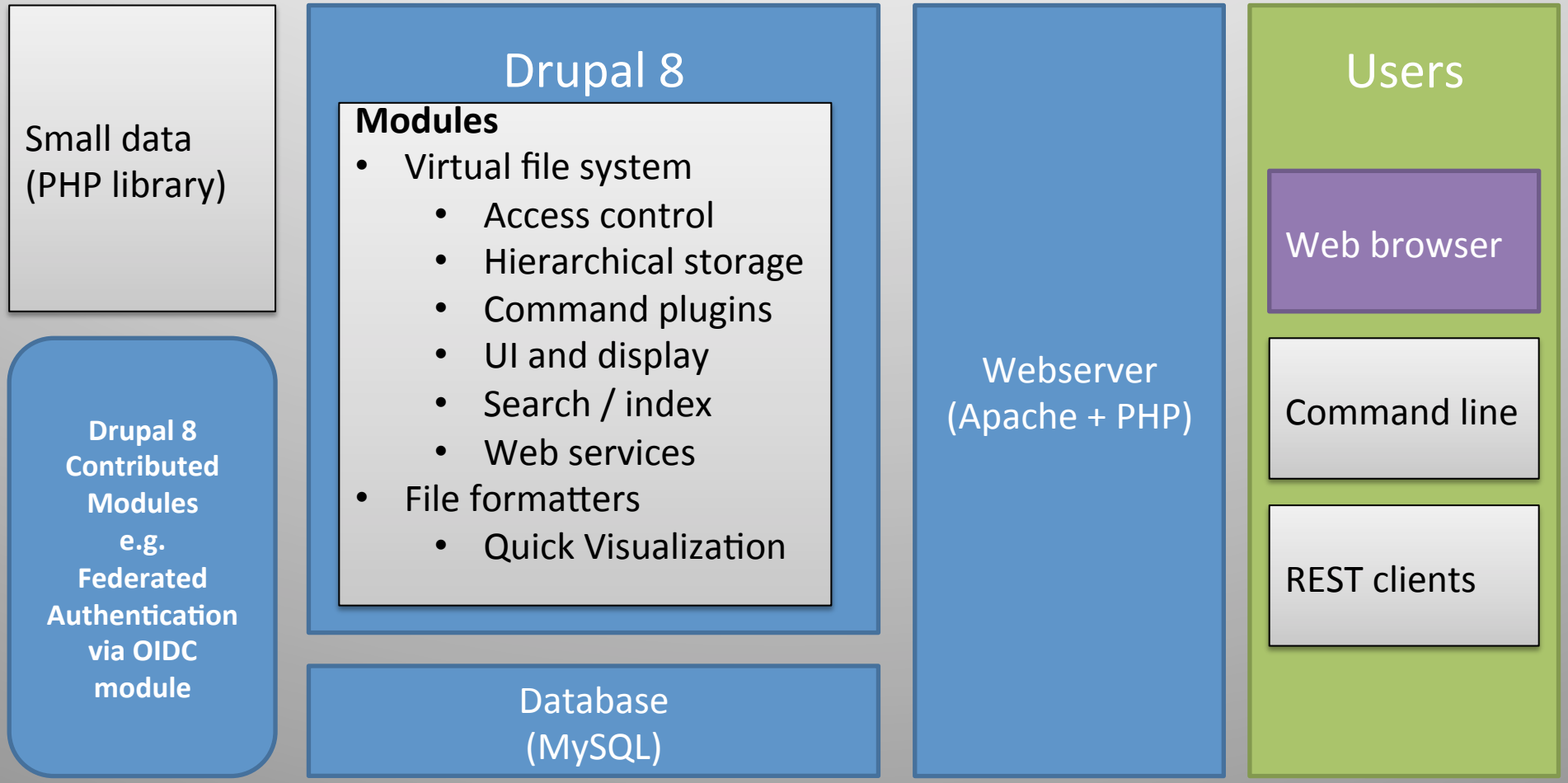
Limitations of existing solutions

Lack extensibility
Lack support for rich content (description, discussion, etc…)
Lack independent developer support
Lack 3 way interaction via web browser & command line & API
Resource restricted

Workflow


3
Update as desired
Sharing
Description
Add folders
Upload files
1
Sign In
2
Create Project
4
View Search Download
Web browser, Command line, REST or App

Architecture



Drupal (Content Management System)

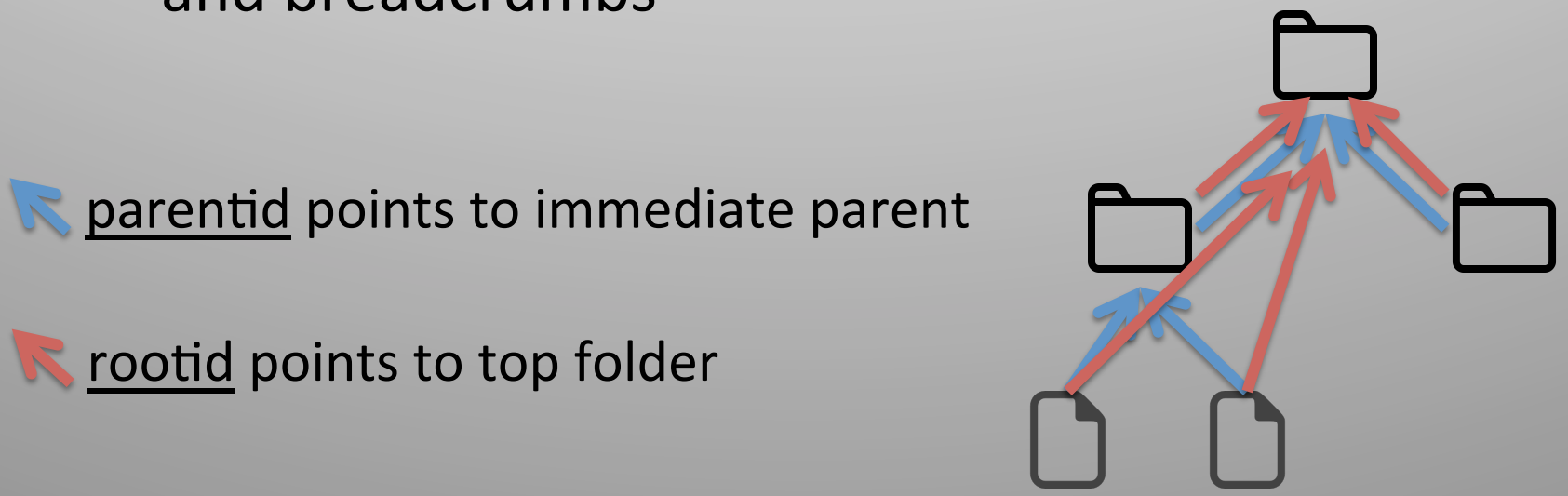
- Modular and extensible architecture
  (1,000+ community contributed modules for version 8)
- Widely used in industry, academia and government
  (third most popular CMS on web)
- Large developer & support community
  (4,000 contributors for Drupal 8 core + thousands more)
- Security advisory and updates for core and stable contributed modules every month
- Versatile deployment options
  (personal hosting, institutional hosting, cloud hosting)

FolderShare module

- Virtual file system (fieldable):
  - Entity type & API
  - Access controls
  - Usage tracking
  - Views, displays, breadcrumbs, forms
  - Plugins for field formatters, search, views, actions, and queue workers
- Configurable by sites
  - e.g. Keywords, comments, flags, DOIs
- Extensible by developers

Files & folders

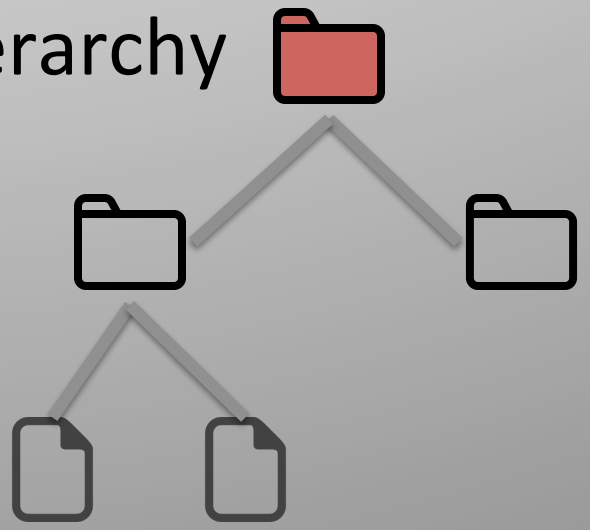
- Children point to parents
  - Parent IDs enable fast queries for all children of a folder
  - Root IDs enable fast queries for access controls and breadcrumbs



Access controls

- Drupal account-based
- Permissions + access control list on top folders
  - List of users that can view and author

- Top folder controls entire hierarchy
  - Simpler than desktop OSes
  - Similar to file sharing services
  - Fast to check access



File storage

- Folders only exist in database
- Files described in database & stored on disk

- Disk directory !== folder hierarchy
  - Better for security and load balancing
  - Files have generated names
    - Avoids name collisions
    - Avoids character set and name length limits

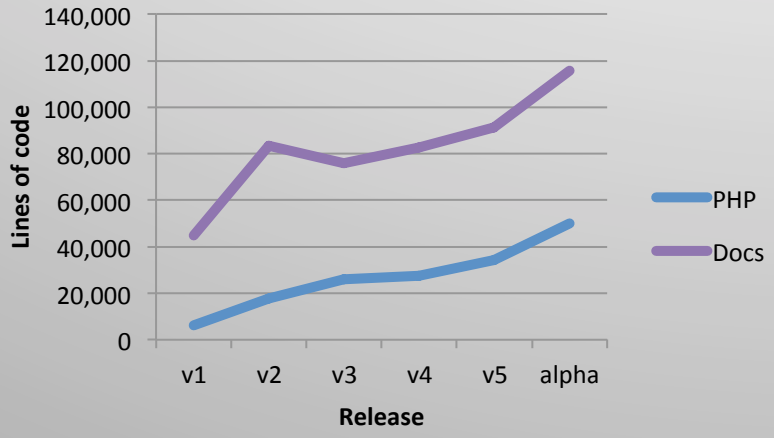
Views

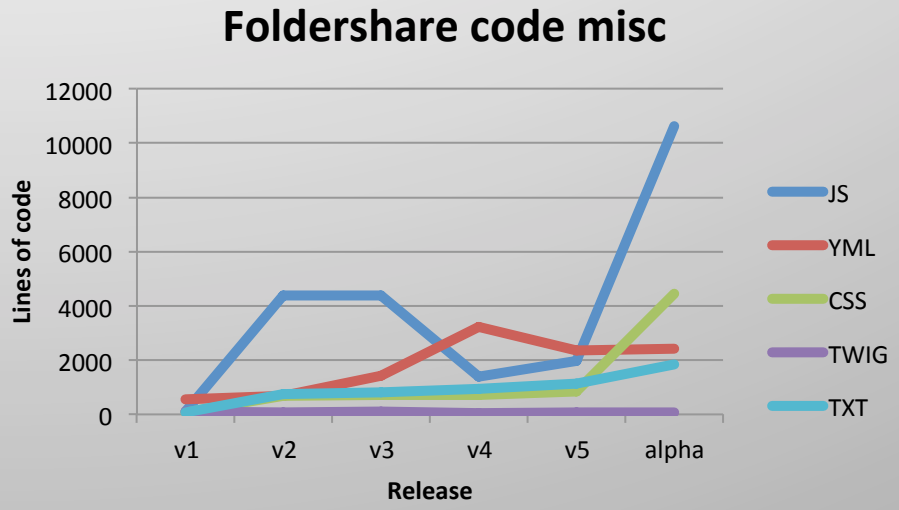
- List personal, public, and shared files & folders
  - Pages & embedded views in folder pages
- Integrated desktop-like UI
  - Select files and folders
  - Then choose menu command
- Three UI variants:
  - No scripting
  - Scripting but no AJAX
  - Scripting with AJAX

Plugins

- Field formatters
  - Folder names, entity references, MIME-type icons
- Search
  - Index and present results
- Queue worker
  - Update folder hierarchy sizes in background
- Actions & custom commands
  - Menu UI items to add, delete, etc

Code trivia



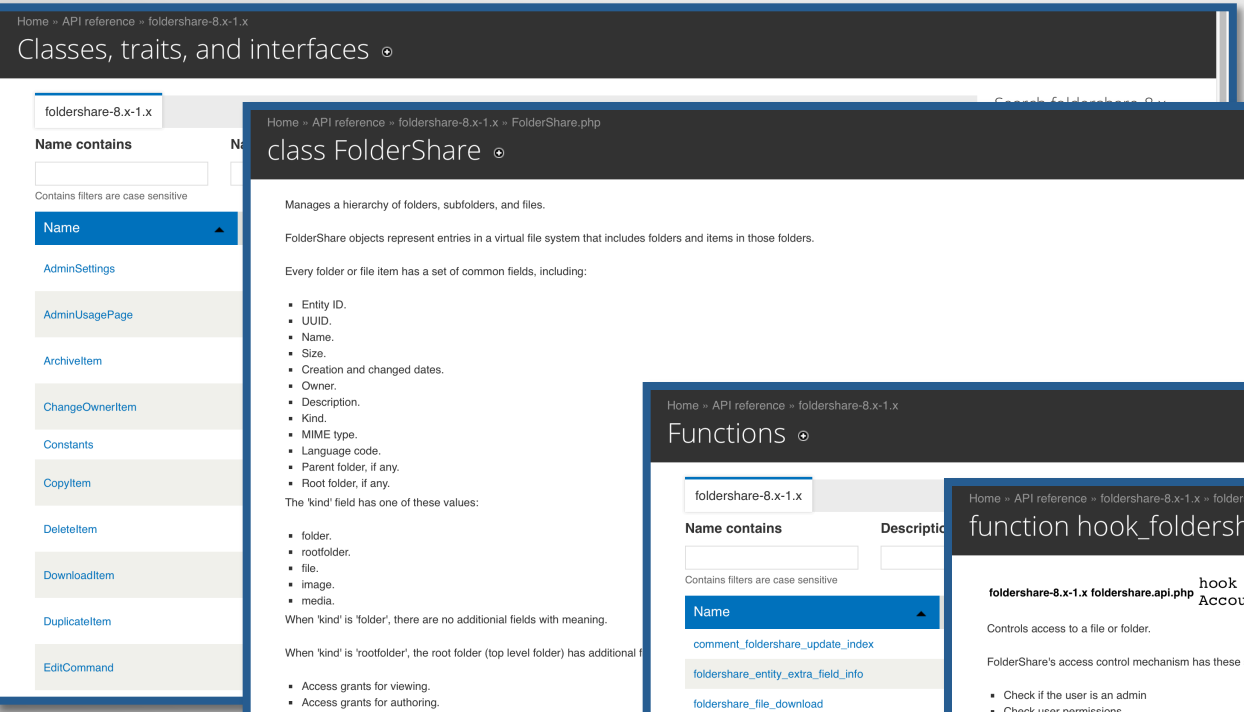


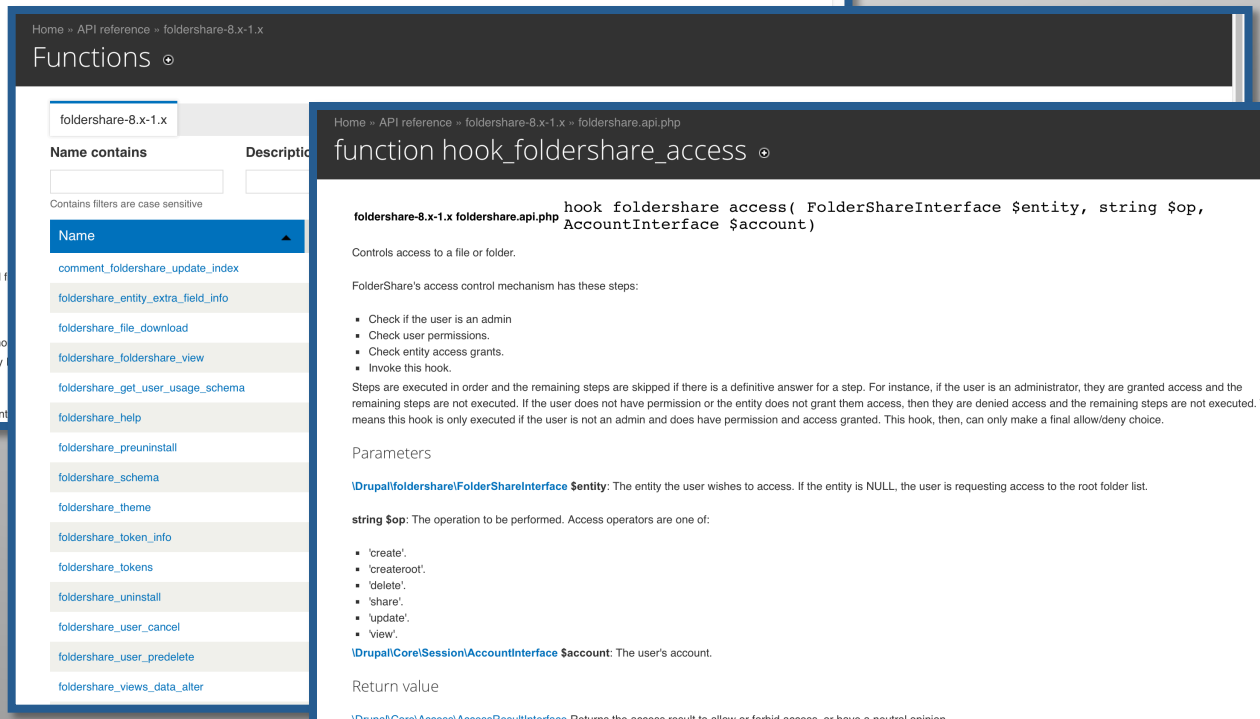
Total lines of php code

Node : 25,639 (Drupal core 8.5.0)

Foldershare: 50,156 (Alpha version)

Foldershare API Documentation





SmallData API & Module

SmallData

- Structured data parsers & writers
  - Tables, trees, and graphs
  - JSON, CSV, TSV, TXT, etc.
- Field formatters
  - Light-weight visualization
  - Line plots, bar charts, pie charts, etc.

FolderShare configuration

Admin menu

Structure > FolderShare

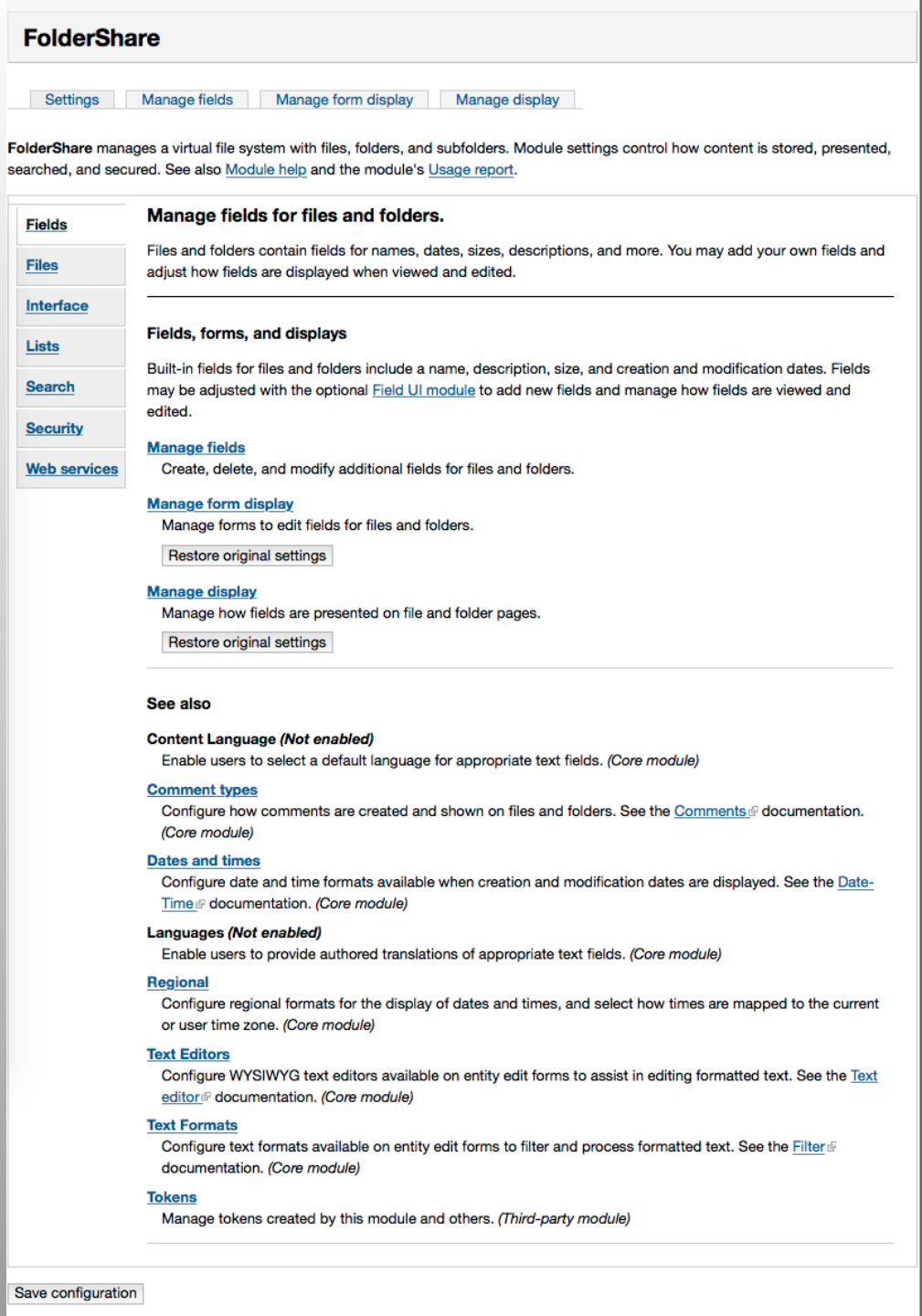
FOR ADMINISTRATORS

FolderShare configuration

Fields

Manage fields, forms & display

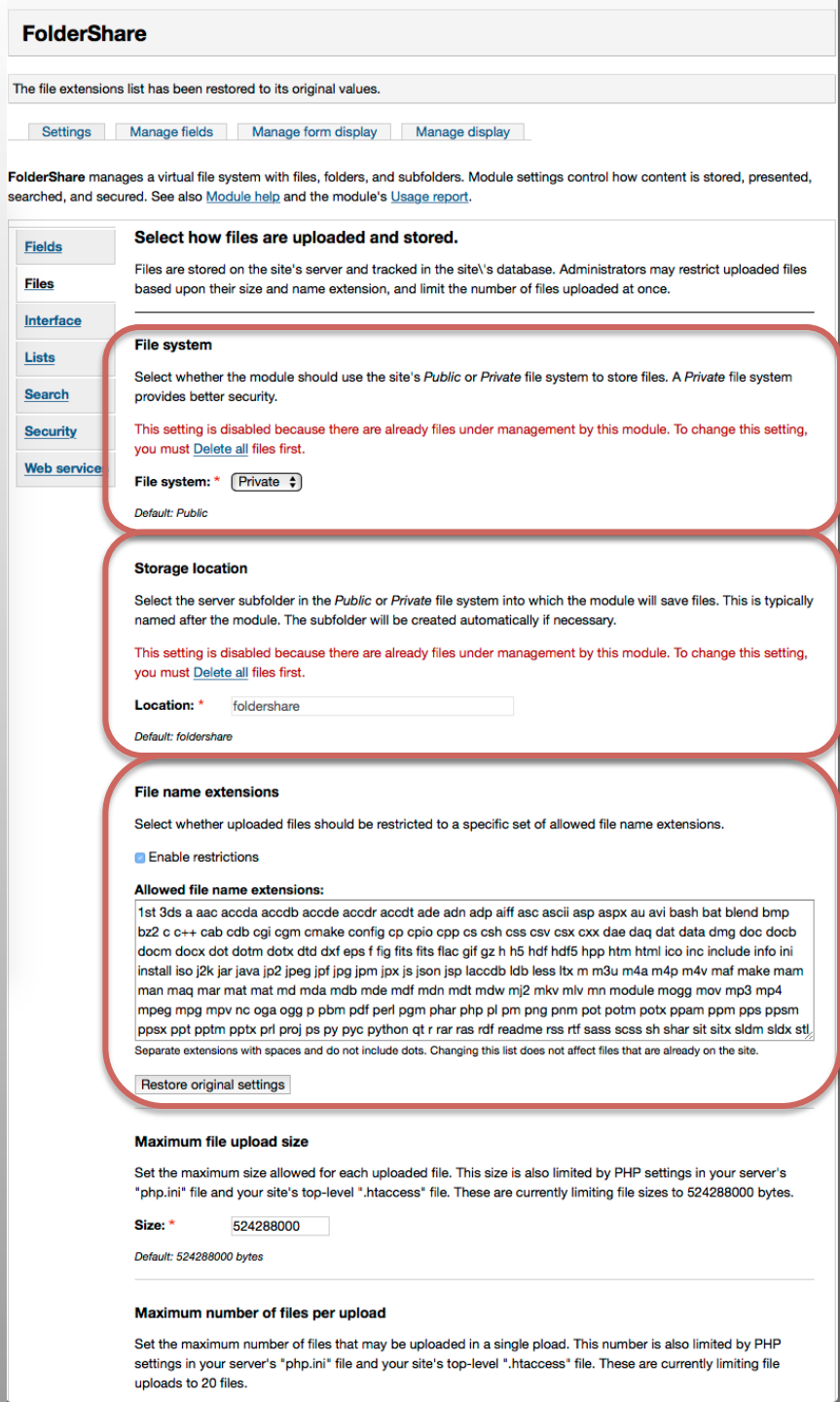
Configuration located in admin menu
Structure > FolderShare



FolderShare configuration

Files

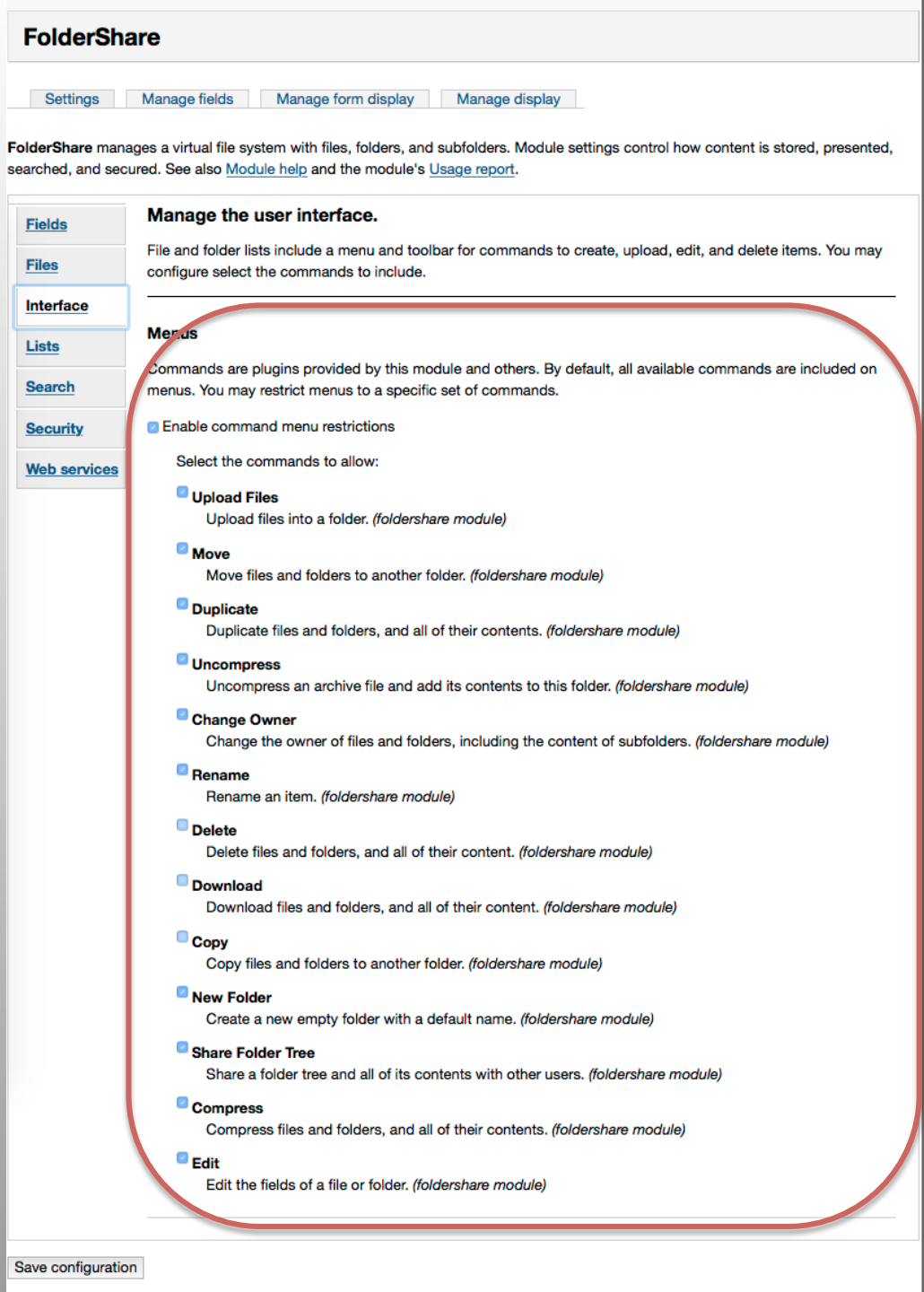
Storage location & upload restrictions



FolderShare configuration

Interface

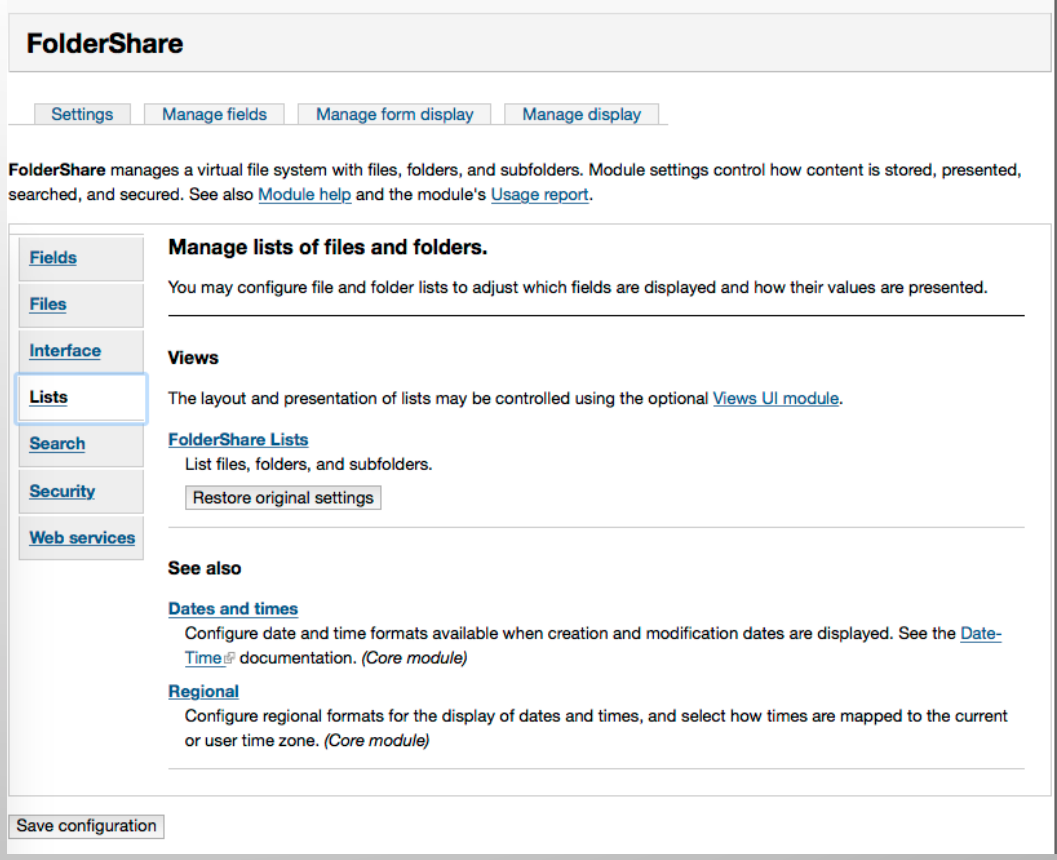
Command plugins



FolderShare configuration

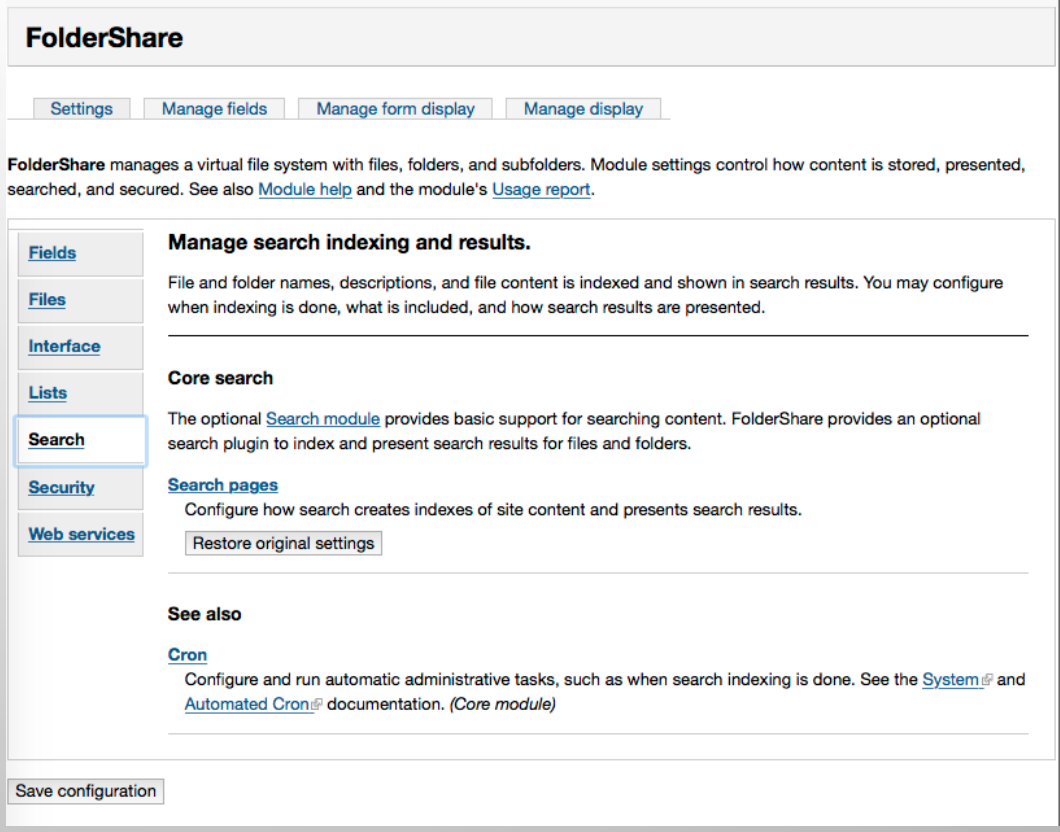
Lists

Manage listing of file and folders



FolderShare configuration

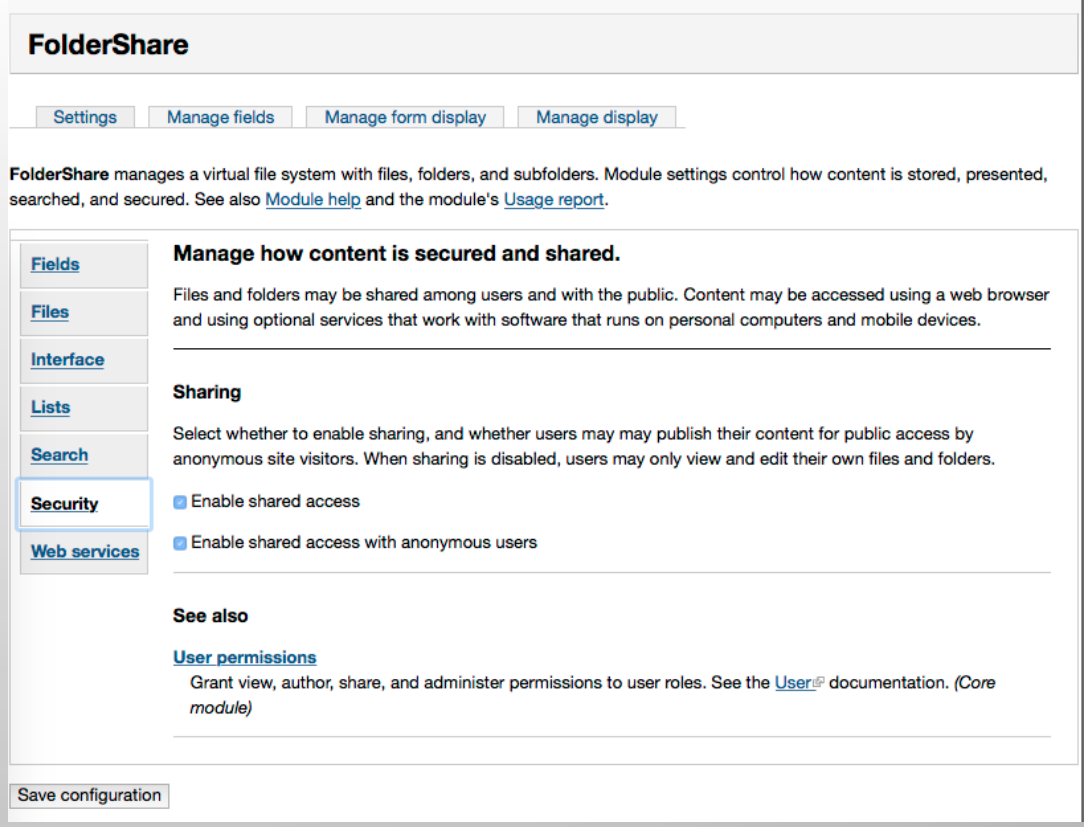
Search (optional)



FolderShare configuration

Security

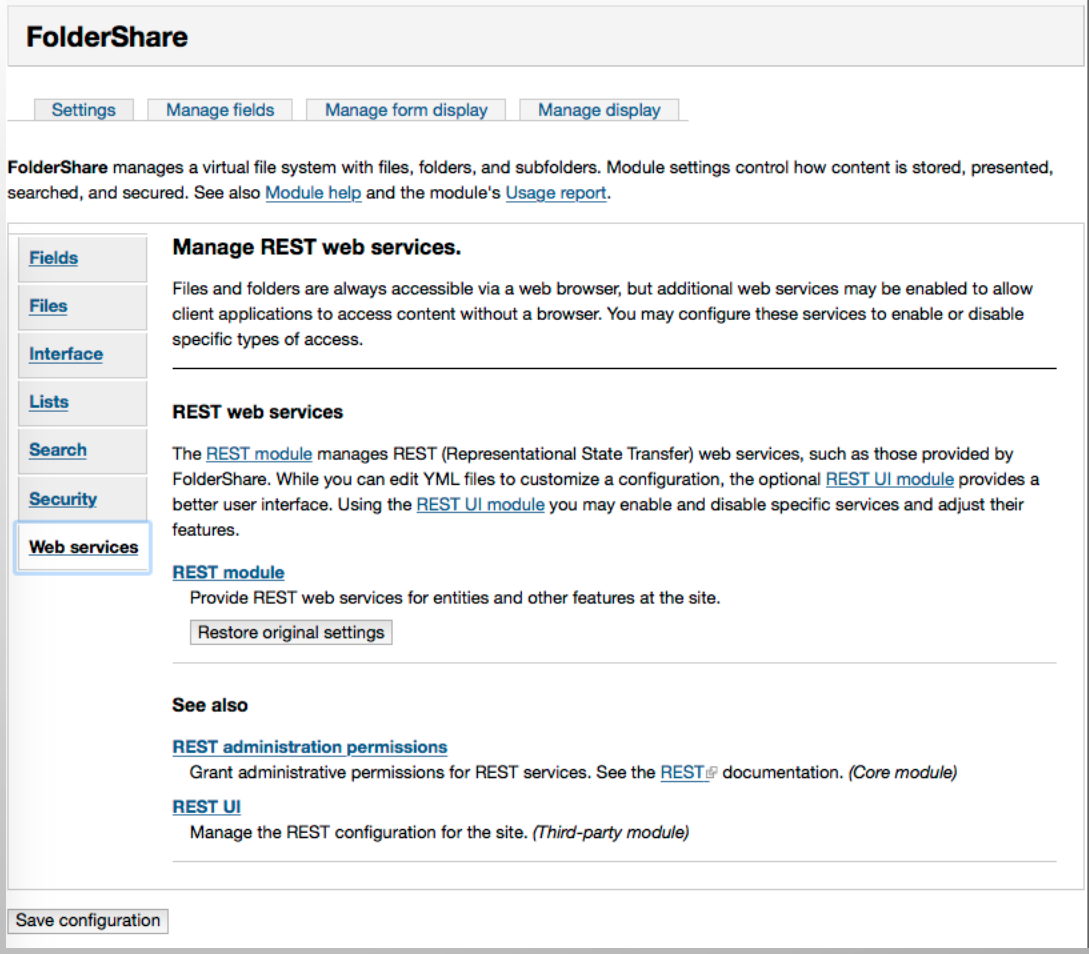
Manage sharing capabilities



FolderShare configuration

Web services (optional)

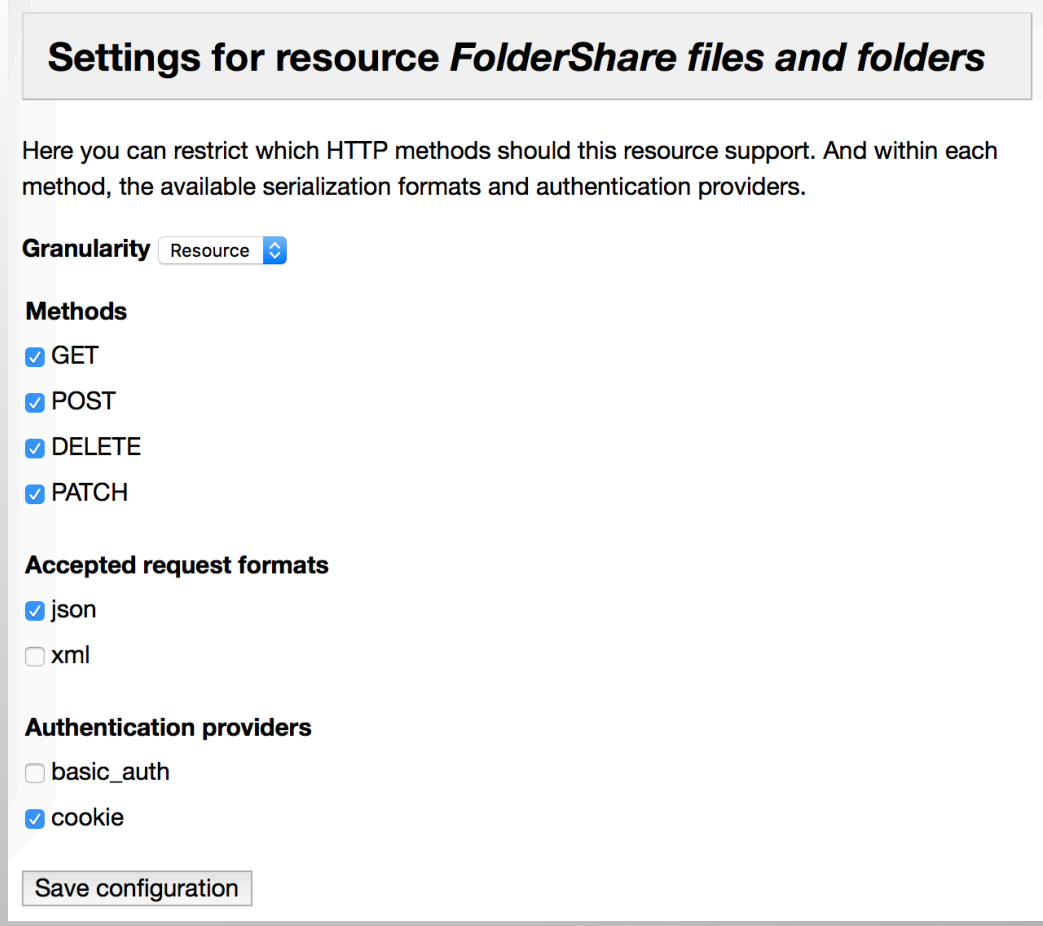
Manage REST capabilities



FolderShare REST settings

Requires REST UI contributed module

Manage REST operations



FolderShare Usage

Admin menu

Reports > FolderShare usage

FolderShare usage

Current usage by user

Users	Top-level folders	Folders	Files	Bytes
Anonymous	0	0	0	0 bytes
Dr. Aarush Ahuja	1	2	16	0 bytes
Dr. Buena Bonita	0	1	5	1.27 KB
Dr. Chandler Charles	1	2	20	1.54 KB
Site Administrator	0	0	0	120.68 MB
demo (Data manager)	0	0	0	0 bytes
Total				

FolderShare permissions

Permission	Anonymous user	Authenticated user	Data manager	Administrator
FolderShare				
Administer files and folders *Warning: Give to trusted roles only; this permission has security implications.* Administer public, private, and shared files and folders. Create, delete, view, download, edit, and change the ownership of content owned by any user.	☐	☐	☑	☑
Author files and folders Create, delete, view, upload, download, and edit files and folders owned by the user, and by other users that grant author access using folder sharing.	☐	☑	☑	☑
Create and share top-level folders Create and adjust settings for sharable top-level folders that group content that may be made public, kept private, or shared with specific users.	☐	☑	☑	☑
View files and folders View and download files and folders owned by the user, and by other users that grant view access using folder sharing.	☑	☑	☑	☑

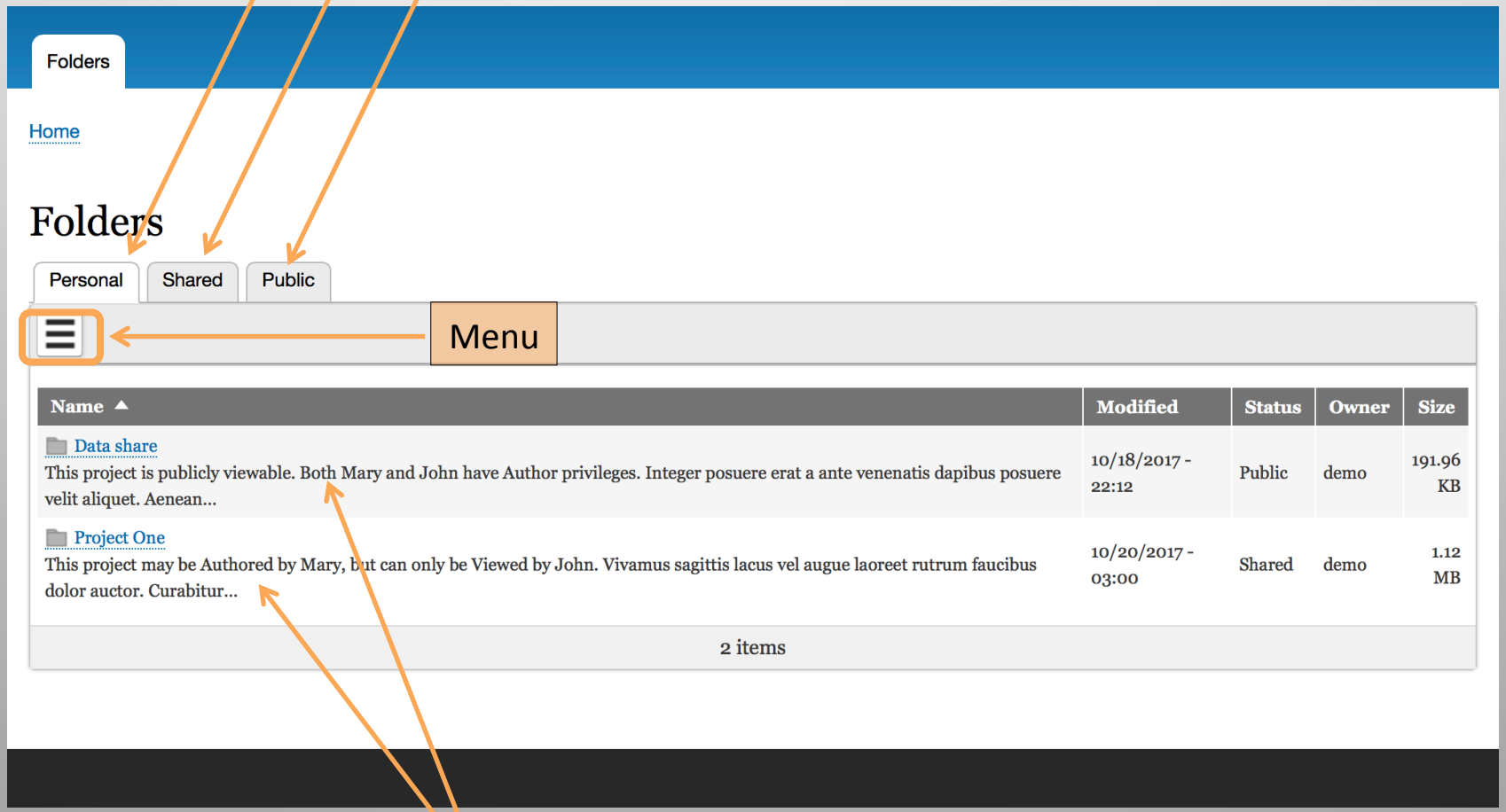
FOR USERS

Top folders owned by you

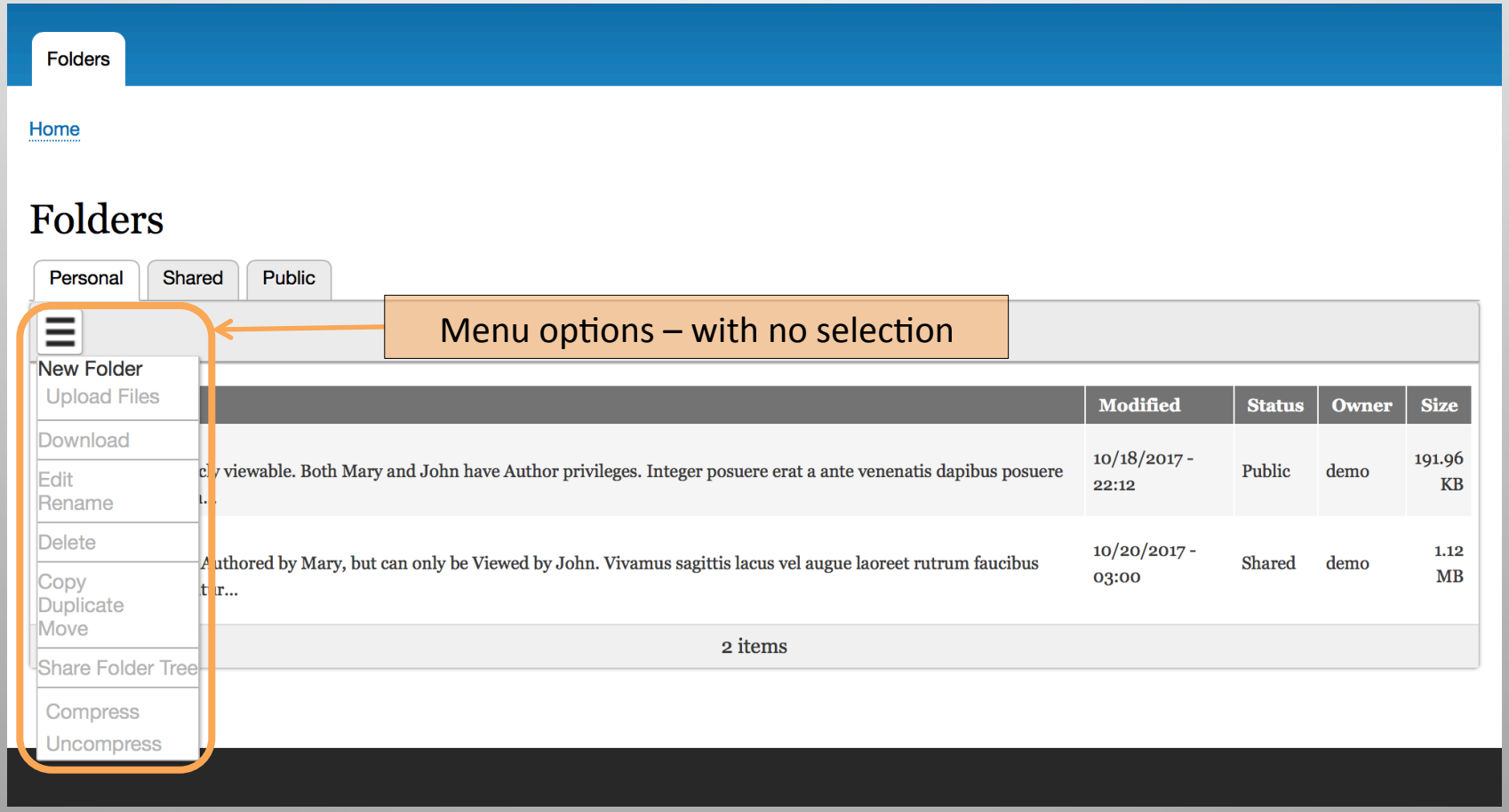
Top folders shared with you

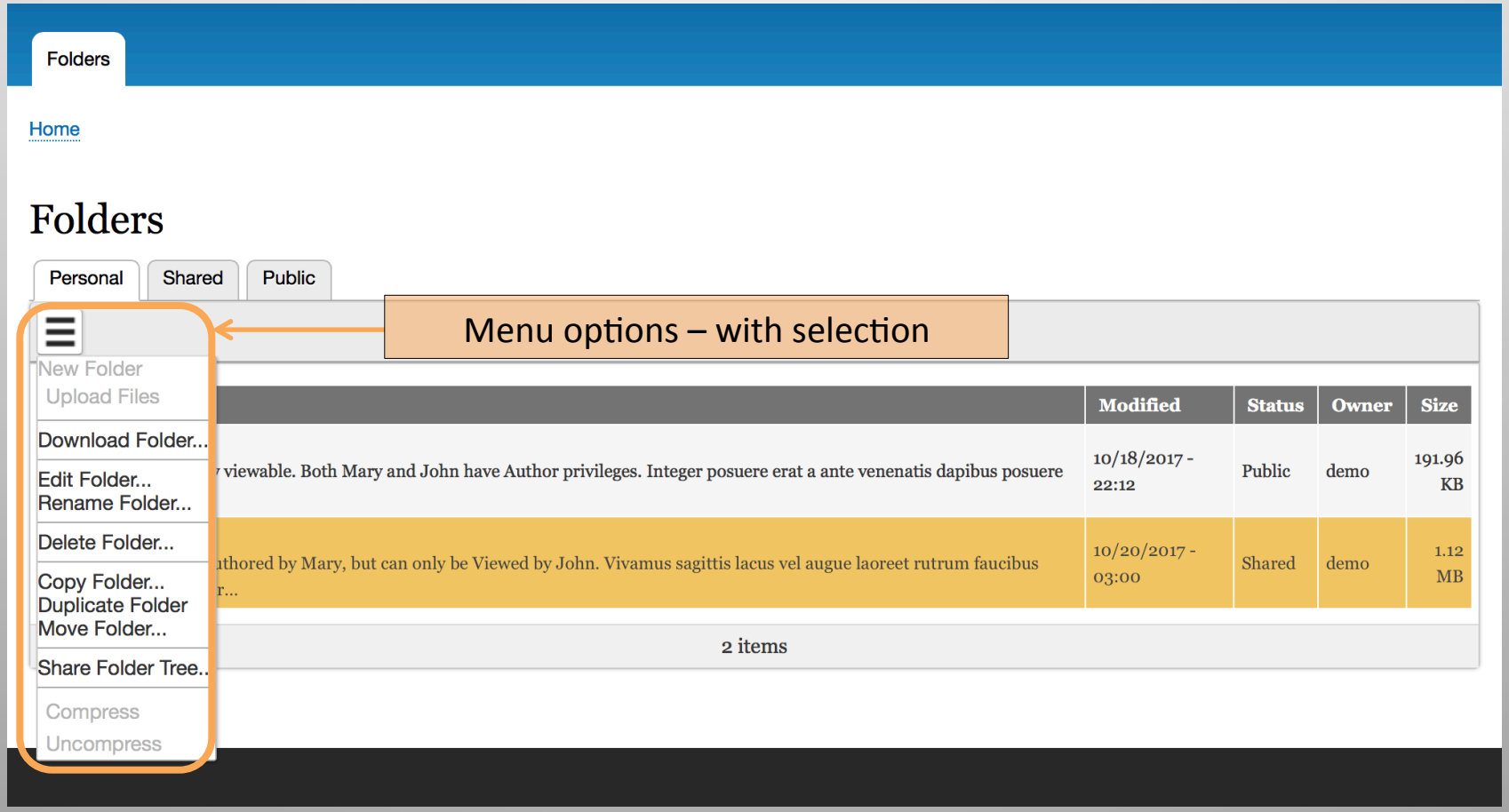
Public top folders

These lists display top level folders

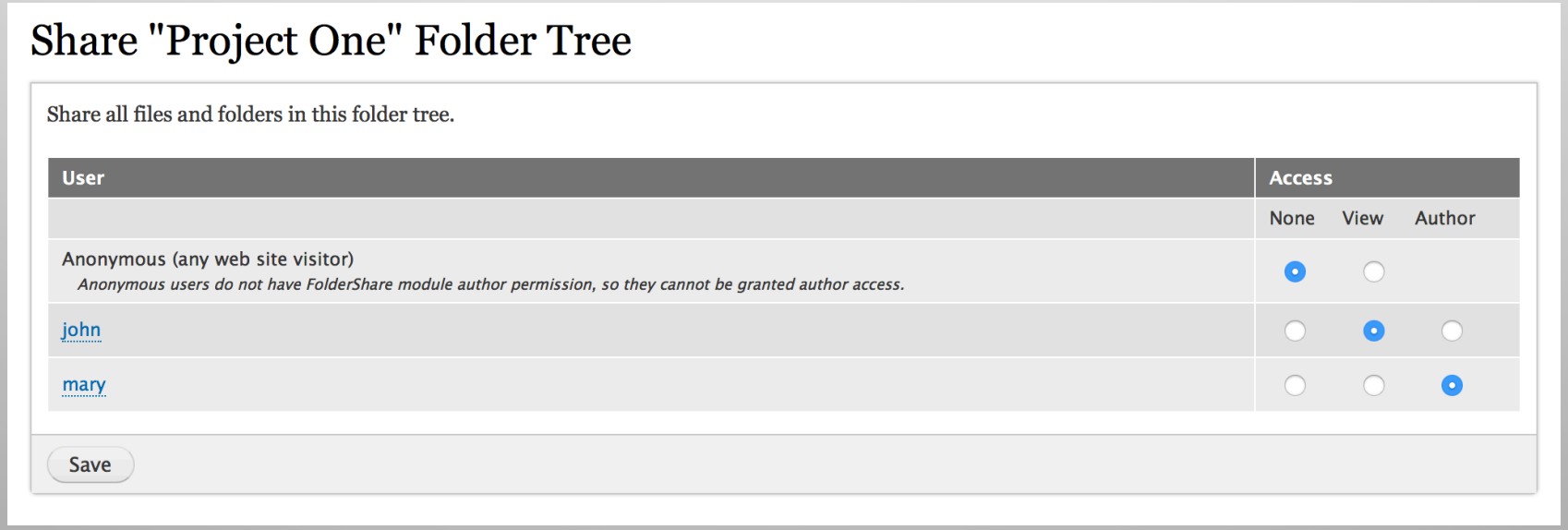


Folders may have a description


Folders
Home
Folders
Personal
Shared
Public
Menu options – with no selection
New Folder
Upload Files
Download
Edit
Rename
Delete
Copy
Duplicate
Move
Share Folder Tree
Compress
Uncompress
Modified
Status
Owner
Size
viewable. Both Mary and John have Author privileges. Integer posuere erat a ante venenatis dapibus posuere
10/18/2017 - 22:12
Public
demo
191.96 KB
uthored by Mary, but can only be Viewed by John. Vivamus sagittis lacus vel augue laoreet rutrum faucibus
r...
10/20/2017 - 03:00
Shared
demo
1.12 MB
2 items


Folders
Home
Folders
Personal
Shared
Public
Menu options – with selection
New Folder
Upload Files
Download Folder...
Edit Folder...
Rename Folder...
Delete Folder...
Copy Folder...
Duplicate Folder
Move Folder...
Share Folder Tree...
Compress
Uncompress
Modified
Status
Owner
Size
viewable. Both Mary and John have Author privileges. Integer posuere erat a ante venenatis dapibus posuere
10/18/2017 - 22:12
Public
demo
191.96 KB
thored by Mary, but can only be Viewed by John. Vivamus sagittis lacus vel augue laoreet rutrum faucibus
r...
10/20/2017 - 03:00
Shared
demo
1.12 MB
2 items

Sharing form to restrict access



Home » Folders

Project One

Name ▲	Modified	Status	Owner	Size
data Certain structured formats such as CSV, JSON and HTML table can be automatically visualized	10/20/2017 - 03:00	Shared	demo	6.92 KB
doc.pdf Every file may have a description.	10/20/2017 - 03:23	Shared	demo	83.86 KB
images Every folder may have a description in rich text.	10/20/2017 - 03:24	Shared	demo	1.03 MB
progress_bars.ipynb Every file may have a description in rich text.	10/20/2017 - 03:25	Shared	mary	4.2 KB

4 items

Sortable listing of files and folders.

Different users

Owner: demo
Modified: 10/20/2017 - 03:00
Description: This project may be *Authored* by Mary, but can only be *Viewed* by John.

Vivamus sagittis lacus vel augue laoreet rutrum faucibus dolor auctor. Curabitur blandit tempus porttitor. Donec sed odio dui. Maecenas sed diam eget risus varius blandit sit amet non magna.

- Nullam id dolor id nibh ultricies vehicula ut id elit. Etiam porta sem malesuada magna mollis euismod.
- Cras justo odio, dapibus ac facilisis in, egestas eget quam.
- Praesent commodo cursus magna, vel scelerisque nisl consectetur et.
- Duis mollis, est non commodo luctus, nisi erat porttitor ligula, eget lacinia odio sem nec elit.

Hello	World
Ace	Base
Uno	Dos

Integer posuere erat a ante venenatis dapibus posuere velit aliquet. Donec sed odio dui. Vestibulum id ligula porta felis euismod semper. Maecenas sed diam eget risus varius blandit sit amet non magna. Nullam quis risus eget urna mollis ornare vel eu leo. Vivamus sagittis lacus vel augue laoreet rutrum faucibus dolor auctor.

Home » Folders

Breadcrumbs shows path

Project One

Name ▲	Modified	Status	Owner	Size
data Certain structured formats such as CSV, JSON and HTML table can be automatically visualized	10/20/2017 - 03:00	Shared	demo	6.92 KB
doc.pdf Every file may have a description.	10/20/2017 - 03:23	Shared	demo	83.86 KB
images Every folder may have a description in rich text.	10/20/2017 - 03:24	Shared	demo	1.03 MB
progress_bars.ipynb Every file may have a description in rich text.	10/20/2017 - 03:25	Shared	mary	4.2 KB

4 items

Owner: demo
Modified: 10/20/2017 - 03:00
Description: This project may be *Authored* by Mary, but can only be *Viewed* by John.

Vivamus sagittis lacus vel augue laoreet rutrum faucibus dolor auctor. Curabitur blandit tempus porttitor. Donec sed odio dui. Maecenas sed diam eget risus varius blandit sit amet non magna.

- Nullam id dolor id nibh ultricies vehicula ut id elit. Etiam porta sem malesuada magna mollis euismod.
- Cras justo odio, dapibus ac facilisis in, egestas eget quam.
- Praesent commodo cursus magna, vel scelerisque nisl consectetur et.
- Duis mollis, est non commodo luctus, nisi erat porttitor ligula, eget lacinia odio sem nec elit.

Hello	World
Ace	Base
Uno	Dos

Integer posuere erat a ante venenatis dapibus posuere velit aliquet. Donec sed odio dui. Vestibulum id ligula porta felis euismod semper. Maecenas sed diam eget risus varius blandit sit amet non magna. Nullam quis risus eget urna mollis ornare vel eu leo. Vivamus sagittis lacus vel augue laoreet rutrum faucibus dolor auctor.

Home » Folders

Project One

Name ▲	Modified	Status	Owner	Size
data Certain structured formats such as CSV, JSON and HTML table can be automatically visualized	10/20/2017 - 03:00	Shared	demo	6.92 KB
doc.pdf Every file may have a description.	10/20/2017 - 03:23	Shared	demo	83.86 KB
images Every folder may have a description in rich text.	10/20/2017 - 03:24	Shared	demo	1.03 MB
progress_bars.ipynb Every file may have a description in rich text.	10/20/2017 - 03:25	Shared	mary	4.2 KB

4 items

Owner: demo
Modified: 10/20/2017 - 03:00
Description: This project may be *Authored* by Mary, but can only be *Viewed* by John.

Vivamus sagittis lacus vel augue laoreet rutrum faucibus dolor auctor. Curabitur blandit tempus porttitor. Donec sed odio dui. Maecenas sed diam eget risus varius blandit sit amet non magna.

- Nullam id dolor id nibh ultricies vehicula ut id elit. Etiam porta sem malesuada magna mollis euismod.
- Cras justo odio, dapibus ac facilisis in, egestas eget quam.
- Praesent commodo cursus magna, vel scelerisque nisl consectetur et.
- Duis mollis, est non commodo luctus, nisi erat porttitor ligula, eget lacinia odio sem nec elit.

Hello	World
Ace	Base
Uno	Dos

Integer posuere erat a ante venenatis dapibus posuere velit aliquet. Donec sed odio dui. Vestibulum id ligula porta felis euismod semper. Maecenas sed diam eget risus varius blandit sit amet non magna. Nullam quis risus eget urna mollis ornare vel eu leo. Vivamus sagittis lacus vel augue laoreet rutrum faucibus dolor auctor.

Every folder and file may add a description. (Formatted text field aka Body field in Drupal's Node)

Home » Folders

Project One

Name ▲	Modified	Status	Owner	Size
data Certain structured formats such as CSV, JSON and HTML table can be automatically visualized	10/20/2017 - 03:00	Shared	demo	6.92 KB
doc.pdf Every file may have a description.	10/20/2017 - 03:23	Shared	demo	83.86 KB
images Every folder may have a description in rich text.	10/20/2017 - 03:24	Shared	demo	1.03 MB
progress_bars.ipynb Every file may have a description in rich text.	10/20/2017 - 03:25	Shared	mary	4.2 KB

4 items

Owner: demo
Modified: 10/20/2017 - 03:00
Description: This project may be *Authored* by Mary, but can only be *Viewed* by John.

Vivamus sagittis lacus vel augue laoreet rutrum faucibus dolor auctor. Curabitur blandit tempus porttitor. Donec sed odio dui. Maecenas sed diam eget risus varius blandit sit amet non magna.

- Nullam id dolor id nibh ultricies vehicula ut id elit. Etiam porta sem malesuada magna mollis euismod.
- Cras justo odio, dapibus ac facilisis in, egestas eget quam.
- Praesent commodo cursus magna, vel scelerisque nisl consectetur et.
- Duis mollis, est non commodo luctus, nisi erat porttitor ligula, eget lacinia odio sem nec elit.

Hello	World
Ace	Base
Uno	Dos

Integer posuere erat a ante venenatis dapibus posuere velit aliquet. Donec sed odio dui. Vestibulum id ligula porta felis euismod semper. Maecenas sed diam eget risus varius blandit sit amet non magna. Nullam quis risus eget urna mollis ornare vel eu leo. Vivamus sagittis lacus vel augue laoreet rutrum faucibus dolor auctor.

FolderShare comments

demo
Wed, 10/18/2017 - 05:16
Permalink

Duis mollis, est non commodo...

Duis mollis, est non commodo luctus, nisi erat porttitor ligula, eget lacinia odio sem nec elit.

Edit Reply

mary
Wed, 10/18/2017 - 05:36
Permalink

Nullam quis risus eget urna...

Nullam quis risus eget urna mollis ornare vel eu leo. Nulla vitae elit libero, a pharetra augue. Aenean lacinia bibendum nulla sed consectetur. Etiam porta sem malesuada magna mollis euismod. Cras mattis consectetur purus sit amet fermentum.

Reply

Add new comment

Subject

Comment *

Source

About text formats

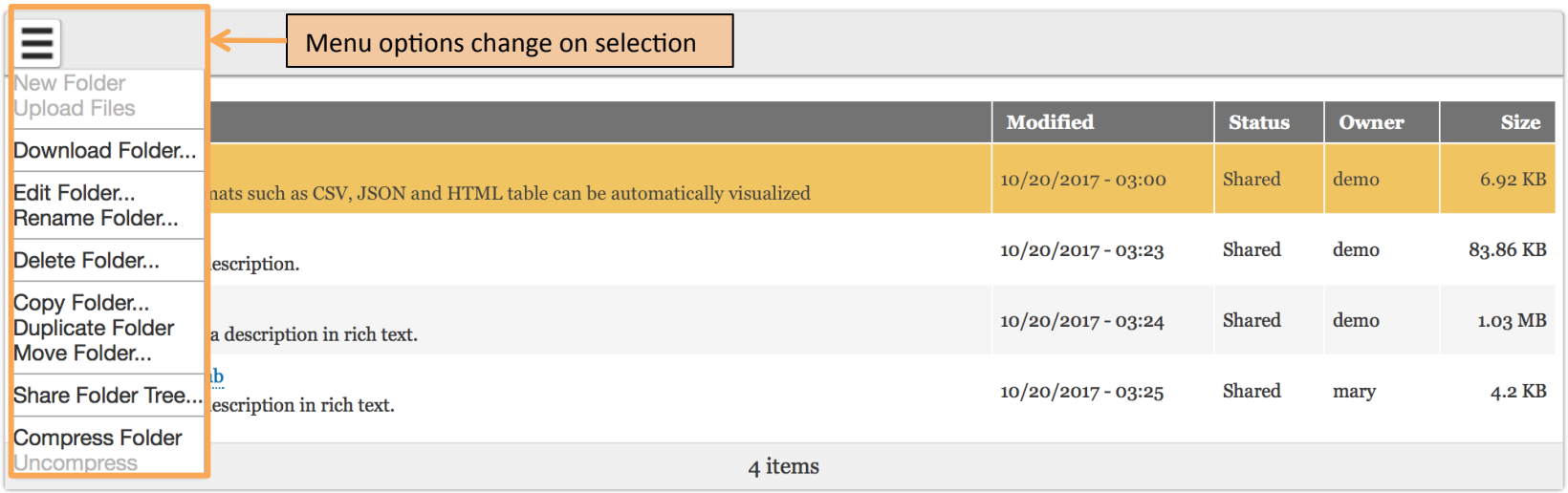
Save Preview

The FolderShare entity is fieldable.

Add customs fields such as Comments, Keywords/tags, taxonomy to FolderShare.

Home » Folders

Project One



Owner: demo
Modified: 10/20/2017 - 03:00
Description: This project may be *Authored* by Mary, but can only be *Viewed* by John.

Vivamus sagittis lacus vel augue laoreet rutrum faucibus dolor auctor. Curabitur blandit tempus porttitor. Donec sed odio dui. Maecenas sed diam eget risus varius blandit sit amet non magna.

- Nullam id dolor id nibh ultricies vehicula ut id elit. Etiam porta sem malesuada magna mollis euismod.
- Cras justo odio, dapibus ac facilisis in, egestas eget quam.
- Praesent commodo cursus magna, vel scelerisque nisl consectetur et.
- Duis mollis, est non commodo luctus, nisi erat porttitor ligula, eget lacinia odio sem nec elit.

Hello	World
Ace	Base
Uno	Dos

Integer posuere erat a ante venenatis dapibus posuere velit aliquet. Donec sed odio dui. Vestibulum id ligula porta felis euismod semper. Maecenas sed diam eget risus varius blandit sit amet non magna. Nullam quis risus eget urna mollis ornare vel eu leo. Vivamus sagittis lacus vel augue laoreet rutrum faucibus dolor auctor.

Home » Folders » Project One

View sub folder
Breadcrumbs shows path

data

Name ▲	Modified	Status	Owner	Size
Apple assets, 2008-2009.htm Every file may have a description in rich text.	10/20/2017 - 03:25	Shared	demo	1.04 KB
Benchmark of original Matlab image preprocessing times.json Every file may have a description in rich text.	10/20/2017 - 03:26	Shared	demo	1.82 KB
Benchmark of revised image preprocessing times.json Every file may have a description in rich text.	10/20/2017 - 03:26	Shared	demo	795 bytes
Cost breakdown of Google Nexus 7.json Every file may have a description in rich text.	10/20/2017 - 03:26	Shared	demo	798 bytes
csv-table.good.simple.csv Every file may have a description in rich text.	10/20/2017 - 03:26	Shared	demo	68 bytes
Image classification work breakdown schedule.json Every file may have a description in rich text.	10/20/2017 - 03:26	Shared	demo	1.18 KB
OpenGL mesh memory use.csv Every file may have a description in rich text.	10/20/2017 - 03:27	Shared	demo	1.27 KB

7 items

Owner: demo
Modified: 10/20/2017 - 03:00
Description: Certain structured formats such as CSV, JSON and HTML table can be automatically visualized

FolderShare comments

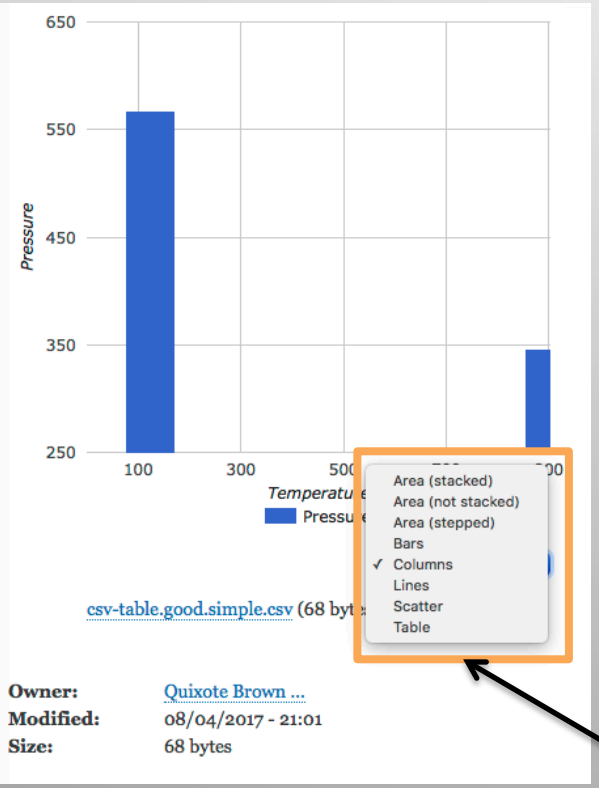
demo
Wed, 10/18/2017 - 22:50
Permalink

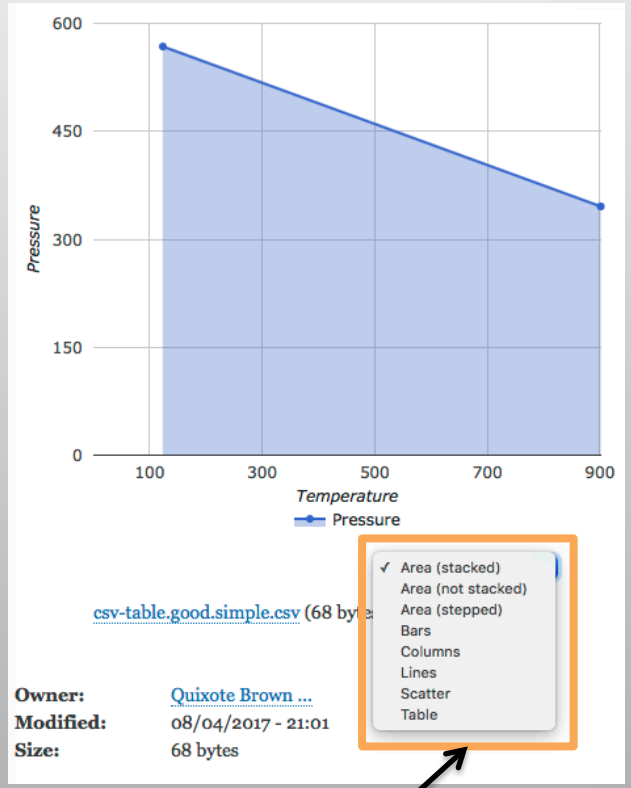
Donec id elit non mi porta...

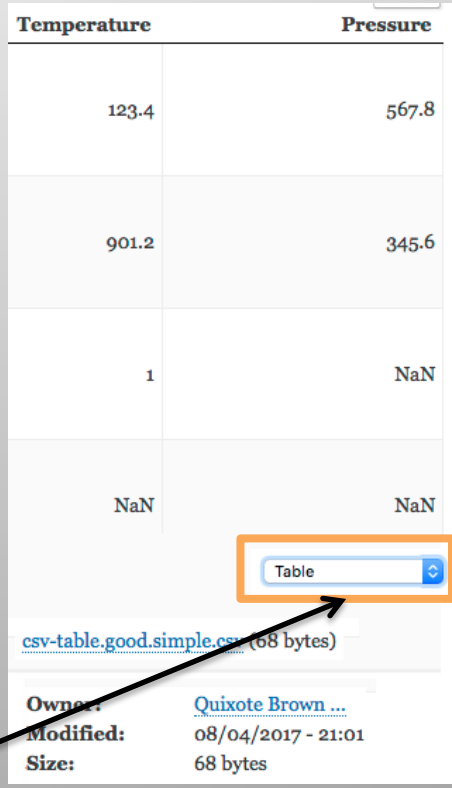
Donec id elit non mi porta gravida at eget metus. Donec sed odio dui.

Edit Reply

Quick visualization of CSV & JSON files







Visualizations can be switched interactively to different chart types

Sample comma

```
foldershare --help
foldershare --host http://demo.seedme.

help
```

```
$ foldershare --help
Usage is: foldershare [HOST_OPTIONS] [COMMAND] [OPTIONS]

Create, delete, change, and manage files and folders on a remote web site
that uses the FolderShare module.

Commands
--------
  config            Show the host administrator configuration.
  cp                Copy files and folders.
  get               Download files and folders.
  ls                List files and folders.
  mkdir             Make new folders.
  mv                Move files and folders.
  put               Upload files and folders.
  rm                Remove files or folders.
  rmdir             Remove empty folders.
  stat              Show the status of files and folders.
  update            Update fields on files and folders.
  usage             Show the user's storage use.
  version           Show version numbers.

Add "--help" after any command to show its description and flags.

Host and account
----------------
All commands require a remote host and account:

  --host HOSTNAME   Select the host (and optional port) for the site.
  --username NAME   Specify the user login name on the site.
  --password PASS   Specify the user password on the site.

Example:
  foldershare --host myhost -username me --password pw ls /

If the username and/or password are omitted, the command will prompt for them.

Additional options
------------------
  --help            Show this help information.
  --verbose         Be verbose while performing the operation.
  --version         Show application version numbers.
  --format type     Select how information is displayed:
                      auto     (default) Use an operation-specific default
                      full     Show full-detail nested-array output
                      keyvalue Show a table of key-value pairs
                      linux    Show a Linux-style output
                      text     Show key-value pairs as text

For Linux/macOS/BSD-style commands, the format type defaults to 'linux'
and output is formatted similar to those OS commands.

Example:
  foldershare --host myhost --format full stat /myfolder
```

Sample command line interaction

```
foldershare --help
foldershare --host http://demo.seedme.org --user dave --password 'cliRocks!'

help

ls --help
ls /
ls -l "/Classification Collection"
ls -l "/Classification Collection/Preliminary Results"

mkdir --help
mkdir "/test"
mkdir "/test/data"

upload --help
upload  "plots/villi.png"        "/test"
upload  "plots/composite.png"    "/test"

Upload "sample-small-data/OpenGL mesh memory use.csv"                    "/test/data"
Upload "sample-small-data/Image classification breakdown schedule.json" "/test/data"

download --help
download "/test" "/Users/amit/downloads"

rm --help
rm -rf "/test"
```

SeedMe2 – Target users/Use cases

- **Researchers**
  - Collaboration hub for a project or personal website
- **Custom project repositories**
  - Include project specific customization (e.g. taxonomy, keywords)
- **Developers**
  - Integrate your scientific application
- **Science gateways**
  - Data sharing
  - Data publishing
  - Data escrow service
- **CyberInfrastructure providers**
  - Offer SeedMe2 platform to your user base/communities

Real usage by physicists

Coulomb Implosion

View | Devel

Name ▲	Modified	Status	Owner	Size
EPOCH input deck	03/13/2018 - 18:05	Shared		4.25 KB
n_e Snapshots of the electron density. The color-coding is n_e in 1/m^3.	03/13/2018 - 18:07	Shared		6.58 MB
n_i Snapshots of the ion density. The color-coding is n_i in 1/m^3.	03/13/2018 - 18:07	Shared		6.33 MB
Total electric field Snapshots of the total electric field $E = \sqrt{E_x^2 + E_y^2}$. The color-coding is E in volt/m.	03/13/2018 - 18:08	Shared		5.65 MB
Total electric field [zoom] Snapshots of the total electric field $E = \sqrt{E_x^2 + E_y^2}$ in the central part of the target. The color-coding is E in vol	03/13/2018 - 18:10	Shared		1.2 MB

5 items | Image slideshow

Owner:
Modified: 03/13/2018 - 17:52

PIC simulations of a rectangular plasma with a cavity irradiated by 4 laser pulses.

$\lambda = 1\ \mu$m - laser wavelength;

$I = 5 \times 10^{18}$ watt/cm^2 - peak intensity of each laser pulse;

$n_e = n_i = 30 n_{crit}$ - initial electron and ion densities;

$R = 60$ nm - radius of the cavity;

cell size is 2 nm by 2 nm; 20 electron macro-particles and 20 ion macro-particles per cell;

Laser electric field is in the plane of the simulation;

Each pulse is $\Delta t = 20$ fs long, with the field envelope given by $(1 - \cos(2\pi t/\Delta t))/2$.

Deployment/service options

DIY - Run your own instance (Your branding + domain)

- On your own hardware
- Cloud hosted

Provider/vendor runs an instance

- Your institution
- CI centers or IT services
- Commercial vendors

Cloud service (Our domain & branding)

Try it yourself!

Explorers welcome (web browser needed)

Trial website: sandbox.seedme.org

Developers invited

Download code: https://dibbs.seedme.org/downloads

SeedMe Team

Amit Chourasia, David Nadeau, Dmitry Mishin & Michael Norman

San Diego Supercomputer Center | University of California San Diego

Acknowledgements

All users and application integrators for their valuable feedback

National Science Foundation

This material is based upon work supported by the National Science Foundation under Grant No. ACI-1235505 and ACI-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."

Talk to us

amit @ sdsc.edu

Keen to learn about potential uses cases in your work

Tutorial: Thursday afternoon

Building data centric website for research projects and research groups

Requirements

1. **Install docker**
2. **Download start/final images (~ 1.2GB)**