



# Web-based data-sharing building blocks for researchers and research projects

SeedMe2 project website: [dibbs.seedme.org](http://dibbs.seedme.org)

Amit Chourasia

San Diego Supercomputer Center, UC San Diego

SeedMe2 Team: David Nadeau, Mike Norman & me

# Presentation overview

- Background
- SeedMe2 positioning with other services
- SeedMe2 building blocks
  - Architecture
  - Modules
  - Value proposition
  - Use cases
- Demonstration

# Keywords

**SeedMe** = Stream encode explore and disseminate **My** experiments

**SeedMe2** = Second generation full rewrite

## **Cloud drives**

Cloud hosted solutions such as dropbox, onedrive, google drive, etc.

## **Consumable data**

- Useful at a glance
- Can be handled by stock web browser
- Displayable on many devices (Phone – PC)

## **Hot data**

High value data at a given time

# Data (static)

Rapidly growing and no longer easily consumable

## **Data sources**

- Sensors (Satellites, Telescopes, Colliders, ...)
- Simulations (Weather, Engineering, ...)
- Derivation (Refinement, visualization ...)
- Transformation
- Aggregation

## Stumbling blocks

Transfer

Storage

Access  
control

Collaboration

Automation

## Issues due to content dispersion

Data, Description, Discussion

Description in mind

Data in cloud

Discussion on emails

## Complexity due to context

Display and Discovery

Display data with  
context

Discover data with  
context

# 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

<b>Features</b>	<b>SeedMe2</b>	<b>Cloud drives (Dropbox, One drive, etc.)</b>
<b>Focus</b>	Arbitrary data, context & metadata	Arbitrary data
<b>Sharing scheme</b>	Read or write	Read or write
<b>File system hierarchy</b>	Tree	Tree
<b>Extensible metadata</b>	Yes	No
<b>Add description on file/folder</b>	Yes	Varies
<b>Add comments on file/folder</b>	Yes	For text documents only
<b>Command line client</b>	Yes	No (from third-party clients)
<b>Visualization</b>	Automatic visualization for CSV files	No
<b>Branding and layout</b>	Customizable and personable	Fixed
<b>Extensible as a website</b>	Yes	No
<b>Hosting options</b>	On-premise or on any cloud platform	Central hosting
<b>Open source</b>	Yes	No

# SeedMe2 architecture

Drupal 8

## SeedMe2 modules

1. FolderShare
2. Chart Suite
3. Formatter Suite

Database

(MySQL variants)

PHP

+

Webserver

(Apache,  
NGINX,  
Etc.)

Users

Web browser

Command line

REST client

Structured  
Data API  
(PHP library)

Drupal 8  
Contributed  
Modules

e.g.  
Federated  
Authentication  
via OIDC module



Project contributions



# SeedMe2 Platform Ecosystem

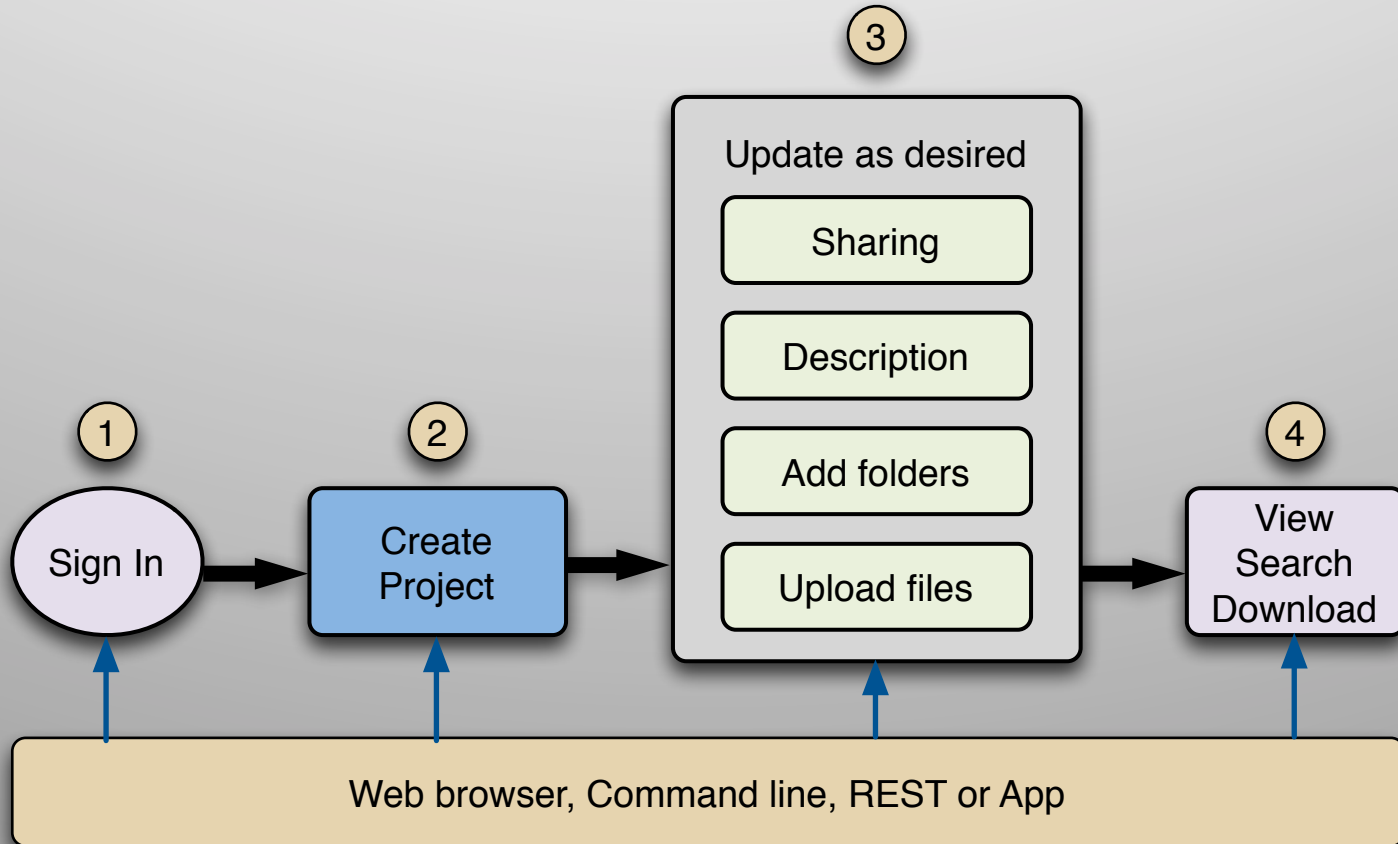
## **Drupal (Content Management System)**

- Widely used in industry, academia and government  
(third most popular Open CMS on web after Wordpress & Joomla)
- Modular architecture with large ecosystem  
(over 1,000 contributed modules)
- Large developer & support community  
(4,000 contributors to core + thousands more)
- Security advisory and updates every month
- Versatile deployment options  
(personal hosting, institutional hosting, cloud hosting)

# FOLDERSHARE module

- **Virtual file system (fieldable)**
  - Entity data model & access controls
  - File management & security
  - Display using Views
  - UI & Command plugins
  - Search
  - Web services
- **Configurable by sites**
  - Add new fields e.g. Keywords, comments, flags, DOIs
  - Customize data display
- **Extensible by developers**
  - Develop custom plugins
  - Explore API at <http://api.try.seedme.org>

# Workflow



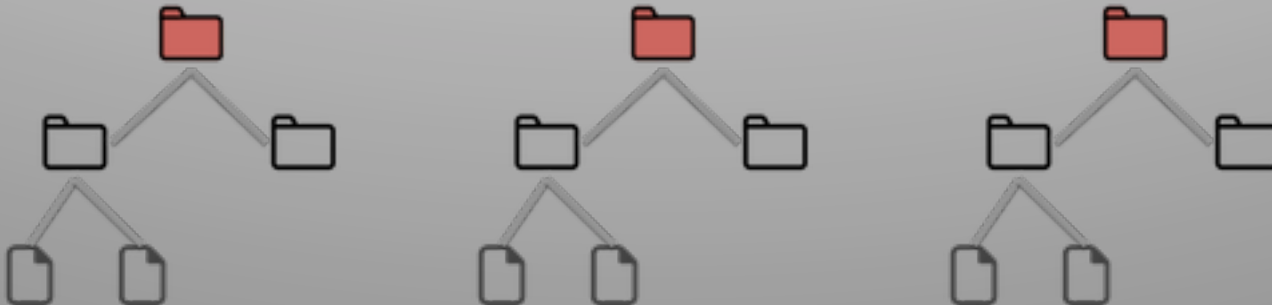


# FolderShare: Data sharing

User may grant following privileges

- None (default)
- View (read only)
- Author (read, write and delete)
- Anonymous aka Public (readable by anyone visiting the site)

**Data sharing is supported from top-level folders**



# CHART SUITE module

Automatically generates visualizations of text data in CSV & JSON formats

- Configurable with FolderShare or as a standalone Drupal module

## **Uses**

- Structured Data API (parses and writes table, trees & graphs)
- Google Charts JavaScript library

# FORMATTER SUITE module

Provides 17 additional field/file formatters for Drupal 8 core

- Numbers in scientific notation
- Image conversion to base64 encoding
- ...

# SeedMe2 – Academic use cases

- **Researchers**
  - Use as collaboration hub
  - Use for data management plan
- **Project repositories**
  - Project specific customization, store, document, discuss and search contributed content
  - Use for data management plan (Combine website and data sharing)
- **Application cloud**
  - Integrate applications to share demonstration & test datasets
  - Open data cloud for user contribution
  - Data publishing
  - Extend Science Gateways with rich data sharing capabilities

## Current status

- In active use by friendly groups
- Alpha version available
- Evaluating offering managed service early 2019 onwards



# SeedMe2: Value proposition

SeedMe2 enables your team to effectively manage, share, search and present data in a readily consumable form on a secure, branded and fully customizable website that you own and control.

- Unify data, its description and its discussion in one spot
- Extend virtual file system to map your meta data attributes
- Generate visualization of CSV and JSON files automatically
- Add your branding and customize presentation style
- Optionally, use it as your project/group website
- Hosting: On-premise or on public cloud
- Leverage open source software & Drupal's large developer community

## **Also relevant for**

- Data management plans
- Open data and FAIR guidelines

# Summary

SeedMe2 provides extensible data sharing website for teams

## Acknowledgements

All early users and collaborators for their valuable feedback, SDSC staff for IT support. This material is based upon work supported by the National Science Foundation under Grant No. 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."

# **Explorers welcome (web browser needed)**

Project website: [dibbs.seedme.org](http://dibbs.seedme.org)

# **Integrators & power users invited**

Play with command line & PHP REST clients

# **Developers solicited**

Extend with plugins, preview API: [api.try.seedme.org](http://api.try.seedme.org)

# Real usage by a physics group

**Coulomb Implosion**

Shared By John Doe

A. Menu

C. Keyword search

Path hierarchy

Q Search folder...

Name	Modified	Owner	Size
E_tot_0011.png Snapshot of final total energy	Wed, 09/05/2018 - 15:13	Carl Smith	702.54 KB
input.deck Input parameters used in this simulation	Wed, 09/05/2018 - 15:10	John Doe	4.25 KB
n_e Snapshots of the electron density. The color-coding is $n_e$ in 1/m3.	Wed, 09/05/2018 - 15:08	Eva Graham	6.58 MB
Total electric field Snapshots of the total electric field $E = \sqrt{E_x^2 + E_y^2}$ . The color-coding is $E$ in...	Wed, 09/05/2018 - 15:08	Joe Debona	5.65 MB

4 items

D. Description snippets on file/folder

E. Full description on currently opened folder 'Coulomb Implosion' displaying simulation parameters

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

$\lambda = 1 \mu\text{m}$  - laser wavelength;

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

$n_e = n_i = 30n_{crit}$  - 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  $(1 - \cos(2\pi t/\Delta t))/2$ .

B. Share with others

Fig. 1. Sample screenshot showing a folder listing from usage by a physics group (people's names are changed). (A) A menu showing various operations. (B) A menu item for sharing, (C) A search field to search for items by their names, descriptions, and file content. (D) Trimmed description summaries associated with files/folders. (E) A full description of the currently open folder, including text and equations.

# Demonstration

- **Usage**
  - Web browser
  - Command line?
- **Administration**
  - Configuration
  - Field extension



SeedMe2 project website: [dibbs.seedme.org](http://dibbs.seedme.org)