

# Rich data management and publishing on websites



Amit Chourasia  
San Diego Supercomputer Center, UCSD

Share...

Manage...

Search...

Visualize...

***...data, as never before!***

## **Data creation**

By individuals has been increasing steadily  
By teams has been rising at a rapid pace

## **Data use/reuse**

By individuals has been increasing  
Within teams has been increasing  
Outside team has been increasing

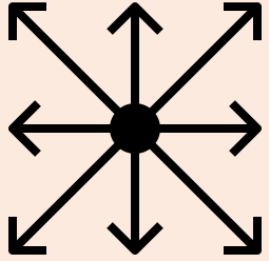
## **Research teams**

Much larger  
Heterogeneous with specialists  
Geographically spread  
Utilize diverse array of tools and platforms

## **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**



## Information dispersion

- Scattered data and its context leads to content fragmentation and poor information/knowledge retention
- Painful to maintain and enforce qualitative controls



## Weak presentation

- Data presented without context makes its consumption difficult (Imagine Youtube as list of files/folders)



## Dark data

- Lack of context makes data harder to search/find, its essentially lost from a practical standpoint
- Most people cannot find their own data

## Usual stumbling blocks

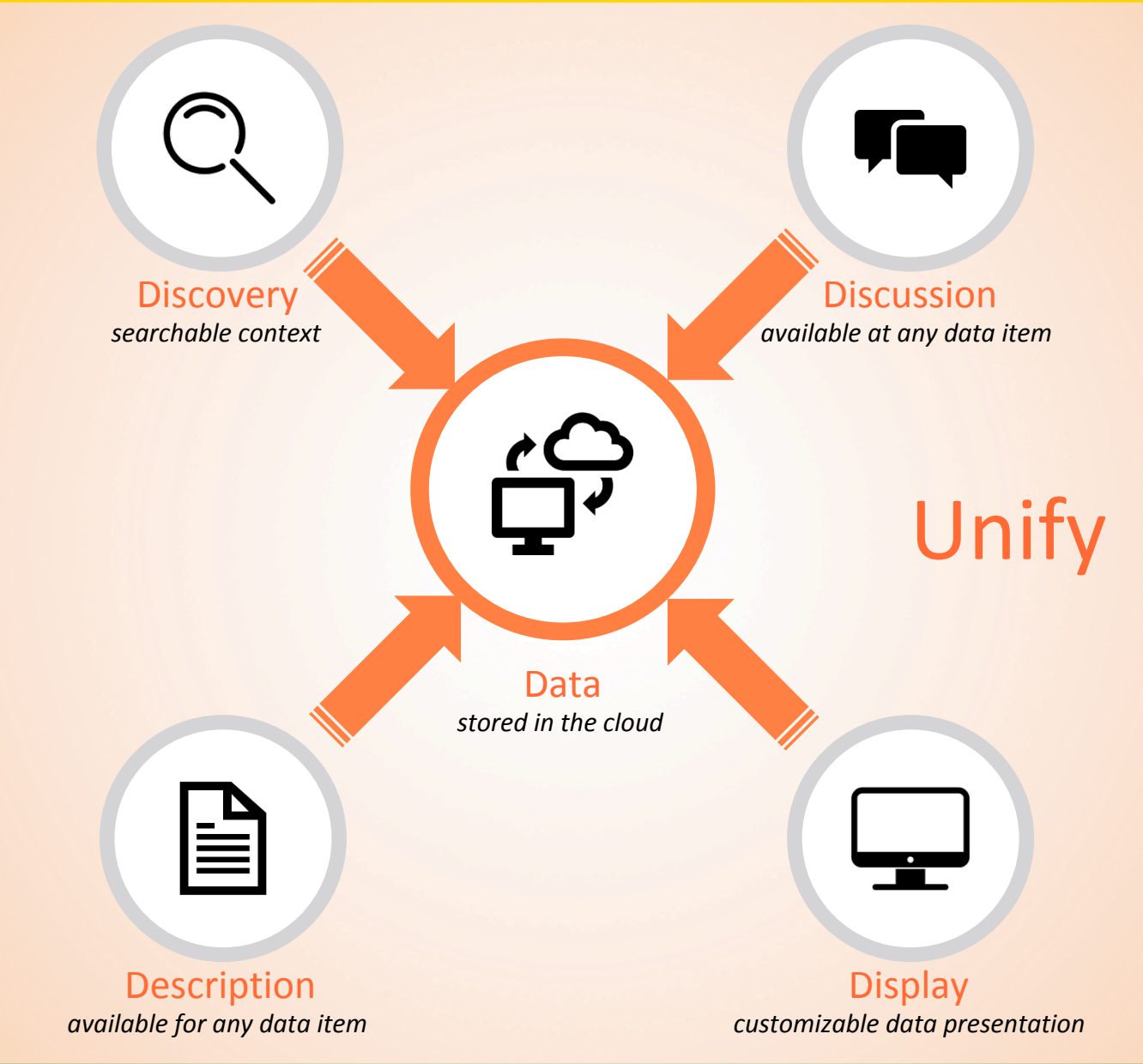
Access control

Collaboration

Storage

Transfer

Automation



*Provide rich data organization and management to overcome scattering of data, discussion, and description while presenting it with rich display in a consumable form.*

## SeedMe: Stream, Encode, Explore and Disseminate My Experiments

Share, manage, search and visualize data, as never before

### Powerful data handling

- New content centric file system
- Effective data organization, management & sharing

### Make data consumable

- Expressive presentation
  - Flexible layouts
  - Automatic visualizations to make data useful at a glance
- Rich collaboration
  - Share, describe, and discuss any data item
  - Enable easier discovery
- Versatile accessibility: Mobile to Desktop to Supercomputers

### Extensible platform

- Add custom metadata fields to the file system
- Add full website capability
- Scriptable automation



## **Designed for researchers**

Streamline and make your data more valuable. Accelerate your research with improved efficiency.

## **Organize and Share data**

Organize data as familiar folders/files and easily share it privately with collaborators or make it public

## **Deep context**

Add context to any data item and discuss it via a simple commenting system, which serves as excellent resource when writing research articles.

## **Rich metadata**

Add relevant fields along with your data for quick review, filtering and search

## **Visualize and present**

Present your data in a consumable form via automatic visualizations, rich formatted text that supports scientific notation

## **Extend your brand**

Add distinction to your brand and data by including a custom domain name, logo, layout and even use it a website

## **Integrate**

Automate your data flow with our API

## **Versatile access**

Access your data on phone, desktop or an a supercomputer

## **Open source**

SeedMe is open source and built on the popular Drupal content management system.

## Side-by-Side Comparison

### SeedMe

#### Relativistic Laser-Plasma Simulation Group

[DATA](#) [MY ACCOUNT](#) [LOG OUT](#)

Home » Personal files

Custom branding & domain

#### Coulomb Implosion

Owned by you

Name	Modified	Owner	Size
<a href="#">Total electric field</a> Snapshots of the total electric field $E = \sqrt{E_x^2 + E_y^2}$ . The color-coding is	Mon, 01/28/2019 - 14:10	John Doe	25.65 MB
<a href="#">s_n_e</a> Snapshots of the electron density. The color-coding is $n_e$ in 1/m3.	Mon, 01/28/2019 - 14:26	John Doe	6.90 MB
<a href="#">OpenGL mesh memory use.csv</a> Bench mark of OpenGL performance	Mon, 01/28/2019 - 14:10	Carl Smith	1.30 KB
<a href="#">input.deck</a> Input parameters used in this simulation	Mon, 01/28/2019 - 14:10	Carl Smith	4.35 KB
<a href="#">E_tot_0011.png</a> Snapshot of final total energy	Mon, 01/28/2019 - 14:10	Carl Smith	719.40 KB

5 items | [Image slideshow](#)

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

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

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

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

$R = 60 \text{ nm}$  - 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  $\Delta t = 20 \text{ fs}$  long, with the field envelope given by  $(1 - \cos(2\pi t/\Delta t))/2$ .

### Google Drive

Help, Settings, App View, Notifications

My Drive > Coulomb Implosion

Name	Owner	Last modified
<a href="#">Total electric field</a>	me	2:24 PM me
<a href="#">s_n_e</a>	me	2:27 PM me
<a href="#">OpenGL mesh memory use.csv</a>	me	2:24 PM me
<a href="#">input.deck</a>	me	2:24 PM me
<a href="#">E_tot_0011.png</a>	me	2:24 PM me

## Side-by-Side Comparison

### SeedMe

#### Relativistic Laser-Plasma Simulation Group

- DATA
- MY ACCOUNT
- LOG OUT

Home » Personal files

#### Coulomb Implosion

Owned by you

Name	Modified	Owner	Size
Total electric field Snapshots of the total electric field $E = \sqrt{E_x^2 + E_y^2}$ . The color-coding is	Mon, 01/28/2019 - 14:10	John Doe	25.65 MB
s_n_e Snapshots of the electron density. The color-coding is $n_e$ in 1/m3.	Mon, 01/28/2019 - 14:26	John Doe	6.90 MB
OpenGL mesh memory use.csv Bench mark of OpenGL performance	Mon, 01/28/2019 - 14:10	Carl Smith	1.30 KB
input.deck Input parameters used in this simulation	Mon, 01/28/2019 - 14:10	Carl Smith	4.35 KB
E_tot_0011.png Snapshot of final total energy	Mon, 01/28/2019 - 14:10	Carl Smith	719.40 KB

5 items | [Image slideshow](#)

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

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

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

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

$R = 60 \text{ nm}$  - 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  $\Delta t = 20 \text{ fs}$  long, with the field envelope given by  $(1 - \cos(2\pi t/\Delta t))/2$ .

Description on any item

### Google Drive

- ?
- ⚙
- ☰
- 🔔

My Drive > Coulomb Implosion

Name	Owner	Last modified
Total electric field	me	2:24 PM me
s_n_e	me	2:27 PM me
OpenGL mesh memory use.csv	me	2:24 PM me
input.deck	me	2:24 PM me
E_tot_0011.png	me	2:24 PM me



## Side-by-Side Comparison

### SeedMe

#### Relativistic Laser-Plasma Simulation Group

- DATA
- MY ACCOUNT
- LOG OUT

Home » Personal files

#### Coulomb Implosion

Owned by you

Name	Modified	Owner	Size
<b>Total electric field</b> Snapshots of the total electric field $E = \sqrt{E_x^2 + E_y^2}$ . The color-coding is	Mon, 01/28/2019 - 14:10	John Doe	25.65 MB
<b>s_n_e</b> Snapshots of the electron density. The color-coding is $n_e$ in 1/m3.	Mon, 01/28/2019 - 14:26	John Doe	6.90 MB
<b>OpenGL mesh memory use.csv</b> Bench mark of OpenGL performance	Mon, 01/28/2019 - 14:10	Carl Smith	1.30 KB
<b>input.deck</b> Input parameters used in this simulation	Mon, 01/28/2019 - 14:10	Carl Smith	4.35 KB
<b>E_tot_0011.png</b> Snapshot of final total energy	Mon, 01/28/2019 - 14:10	Carl Smith	719.40 KB

PIC simulations of a  
 $\lambda = 1 \mu\text{m}$  - laser wa  
 $I = 5 \times 10^{18}$  watt/c  
 $n_e = n_i = 30n_{crit}$  -  
 $R = 60 \text{ nm}$  - radius  
 cell size is 2 nm by 2  
 Laser electric field is  
 Each pulse is  $\Delta t =$

#### FolderShare comments

Submitted by Carl Smith on Mon, 01/28/2019 - 14:13

[Permalink](#)

#### OpenGL benchmark

I have added the results from OpenGL benchmarks

[Edit](#) [Reply](#)

*Discussion on any item*

### Google Drive

Help Settings Grid Notifications

My Drive > Coulomb Implosion

Name	Owner	Last modified
<b>Total electric field</b>	me	2:24 PM me
<b>s_n_e</b>	me	2:27 PM me
<b>OpenGL mesh memory use.csv</b>	me	2:24 PM me
<b>input.deck</b>	me	2:24 PM me
<b>E_tot_0011.png</b>	me	2:24 PM me

## Side-by-Side Comparison

### SeedMe

#### Relativistic Laser-Plasma Simulation Group

- DATA
- MY ACCOUNT
- LOG OUT

Home » Personal files

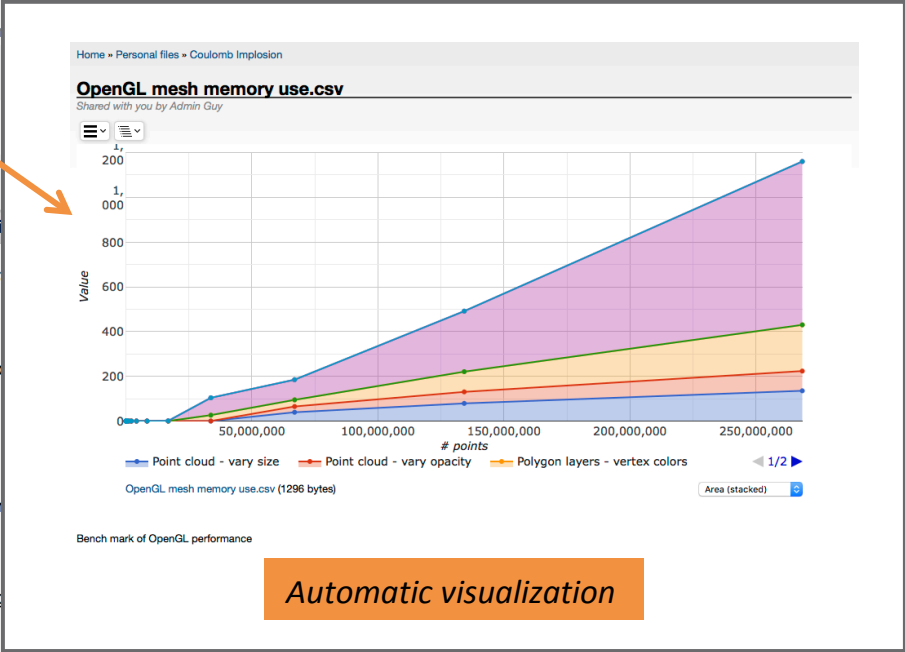
#### Coulomb Implosion

Owned by you

Name	Modified	Owner	Size
<ul style="list-style-type: none"> <li>Total electric field</li> <li>s_n_e</li> <li>OpenGL mesh memory use.csv</li> <li>input.deck</li> <li>E_tot_0011.png</li> </ul>	Mon, 01/28/2019 - 14:10	John Doe	25.65 MB

PIC simulations of a rectangular plasma with a cavity irradiated by a laser pulse with the following parameters:

- $\lambda = 1 \mu\text{m}$  - laser wavelength;
- $I = 5 \times 10^{18} \text{ watt/cm}^2$  - peak intensity of each laser pulse;
- $n_e = n_i = 30n_{crit}$  - initial electron and ion densities;
- $R = 60 \text{ nm}$  - radius of the cavity;
- cell size is 2 nm by 2 nm; 20 electron macro-particles are used per cell;
- Laser electric field is in the plane of the simulation;
- Each pulse is  $\Delta t = 20 \text{ fs}$  long, with the field envelope given by



Automatic visualization

### Google Drive

- ?
- ⚙
- ☰
- 🔔

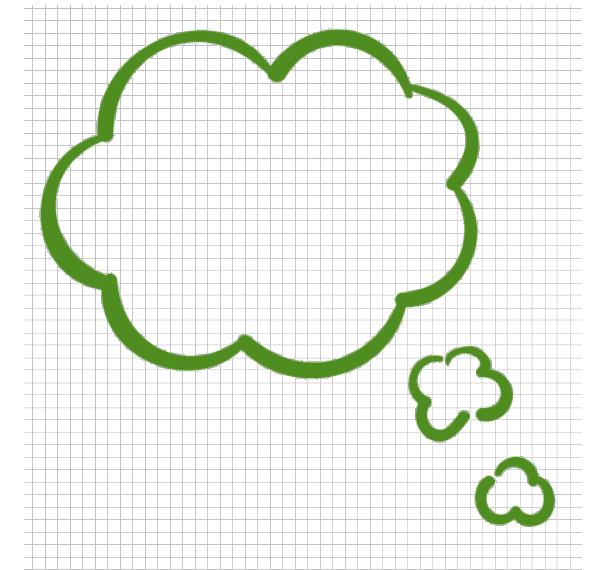
My Drive > Coulomb Implosion

Name	Owner	Last modified
Total electric field	me	2:24 PM me
s_n_e	me	2:27 PM me
OpenGL mesh memory use.csv	me	2:24 PM me
input.deck	me	2:24 PM me
E_tot_0011.png	me	2:24 PM me

*“My inbox is decongested (100 less emails per week), I am able to provide swift and actionable feedback to my students and Post-docs rapidly. All data from my group is in one place, making it easy when writing research papers.” - Professor at UCSD*

*“With SeedMe we save approximately 30 minutes daily by easily sharing data with my Professor and research group. With faster feedback my research productivity pace has significantly increased.” - Graduate Student and Postdoctoral Researcher at UCSD*

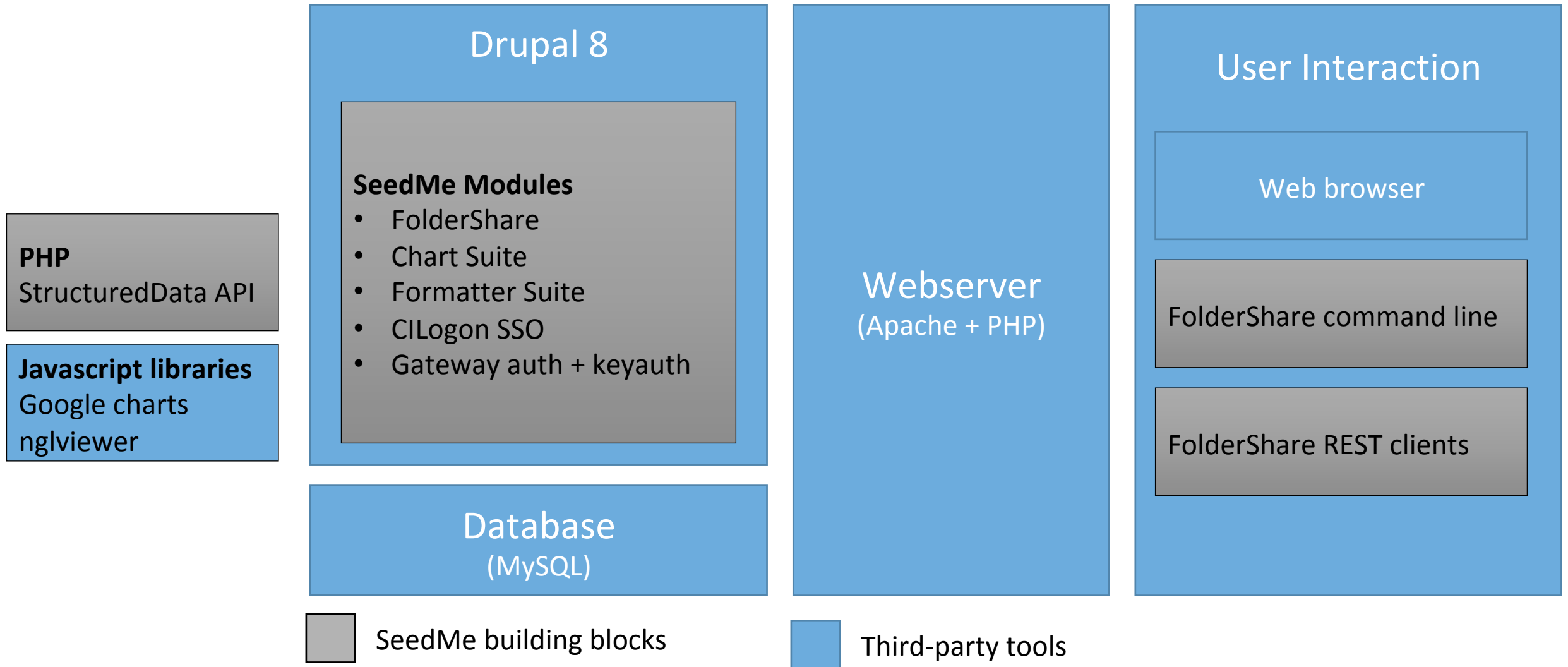
*“Storing metadata will provide future context to research results” – Graduate Student at UCSD*



# Target users/Use cases (Academic)

- **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)
- **Developers**
  - Integrate scientific applications to share demonstration data, test datasets and highlighted content
  - Make it available for user contribution
- **Science gateways**
  - Provide shared data repositories related to computation
  - Data publishing
- **CyberInfrastructure providers**
  - Offer SeedMe building blocks to your user base

# SeedMe Architecture



# Drupal modules

**Chart Suite**  
(CSV, json)

Light  
Visualization

**Formatter Suite**  
Text presentation

**FolderShare: Virtual file system**

Access control      Sharing      Search / Index

**Web services**

REST API      Clients      Command line

**Gateway**

Cross authentication  
Privileged access

**CILogon Auth**  
Federated  
Authentication &  
Authorization

# FOLDERSHARE module

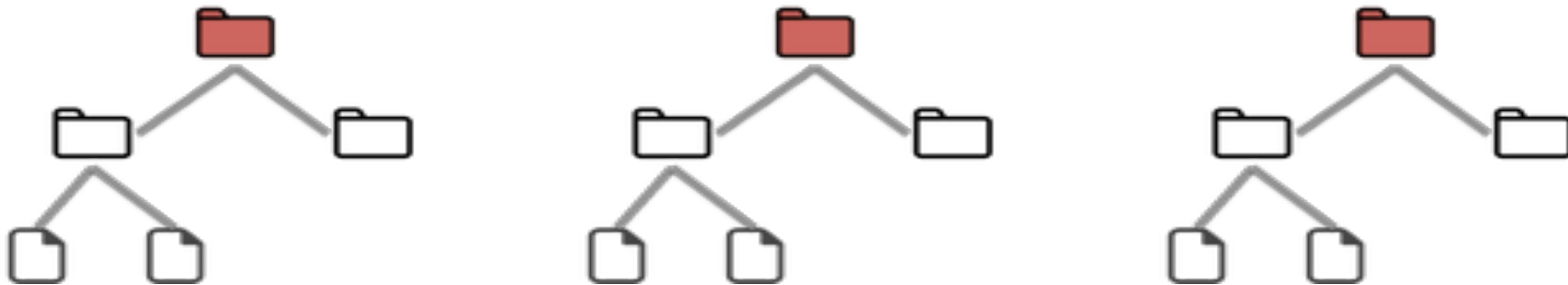
- **Virtual file system (fieldable)**
  - Entity data model & access controls
  - File management & security
  - Display using Views
  - UI & Command plugins
  - Search
  - Web services
- **Configurable by sites**
  - Add new fields e.g. Keywords, comments, flags, DOIs
  - Customize data display
- **Extensible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

# FolderShare: Data sharing

User may grant following privileges

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

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





# FolderShare: Trivia

**Lines of code** (excluding comments and blank lines)

Drupal 8.6.13 core (831k) + vendor (165k)

## **Top Drupal-8.6.13 core modules**

1. Drupal migrate (106k)
2. Views (73k)
3. System (49k)
4. **FolderShare (32k) + Client (3k) -> 6 major, 16 minor versions during development**
5. Node (22k)

# 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 with embedded data URL
- ...

# CILogon Auth module

Single SignOn (SSO) from academic institutions (2,500 + world wide)

Uses CILogon service that provides free authentication

- ~~Avoids hassle with Shibboleth, SAML, etc~~
- Override site wide registration/activation issues
- Set Role for users logging in via SSO
- Generate better usernames
- Store organization info for user
- Integrates with User Restriction module for further customization

# Demonstration

Culmination of 4.5 years of R&D

Try at <http://dibbs.seedme.org>

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