SeedMe: Data Sharing Building Blocks

Amit Chourasia, David Nadeau and Michael Norman
San Diego Supercomputer Center, University of California San Diego

✓ Open source
✓ Web based
✓ Cross-platform

What is SeedMe?

SeedMe = Stream Encode Explore and Disseminate My Experiments

SeedMe is a platform that enables easy sharing of transient and preliminary data for a broad research computing community by offering cyberinfrastructure as a service and a modular software stack that could be customized. SeedMe is based on Drupal content management system as a set of building blocks with additional PHP modules and web services clients.

In this poster we present our progress on implementing a web based modular data sharing platform that collocates shared data, along with the data's context, including descriptions, discussion, light-weight visualizations, and support files. This project is an evolution of the earlier SeedMe project, which created prototype data sharing tools and garnered user feedback from real-world use. The new SeedMe platform is developing modular components for data sharing, lightweight visualization, collaboration, DOI registration, video encoding and playback, REST APIs, command-line data import/export tools, and more. These modules may be added to any web site based upon the widely used open-source Drupal content management system. The new SeedMe modules allow extensive customization enabling the sites to select and enhance functionality to provide features specific to a research community or a project.

SeedMe building blocks

Authentication + Authorization	Virtual file system	Access control
Sharing	Search / Index	Small data formats
Field formatters	REST API	Clients (Java, Python) + Command line
Light visualization	Rich text description	Discussion

Project website: dibbs.seedme.org

Why build a platform?

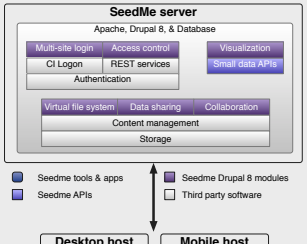
Data sharing is essential and pervasive in scientific research. The requirements for data sharing vary as research projects mature and iterate through early designs and prototypes with a small number of collaborators, and develop into publishable results and larger collaborator teams. Along the way, preliminary and transient results often need to be shared, discussed, and visualized with a quick turn-around time in order to guide the next steps of the project. Data sharing throughout this process requires that the data itself be shared, along with essential context, such as descriptions, provenance, scripts, visualizations, and threaded discussions. However, current consumer-oriented data sharing solutions mainly rely on local or cloud file systems or web-based drop boxes. These mechanisms are rather basic and are largely focused on data storage for individual use, rather than data collaboration. Using them for scientific data sharing is cumbersome.

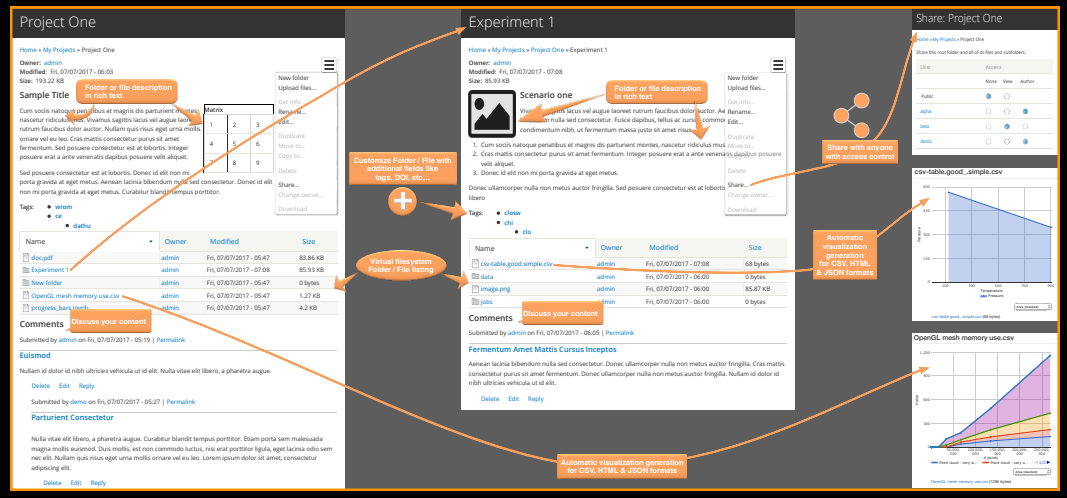
For whom?

- **Researchers**
  Use as collaboration hub or personal dashboard
- **Developers**
  Integrate scientific applications
- **Project repositories**
  Project specific customization
- **Gateways**
  Service for data sharing, data publishing, data escrow
- **CI providers**
  Offer SeedMe platform to your user base/communities.

Use how?

- **As a cloud service**
  demo.seedme.org
  www.seedme.org
- **DIY - Run own instance**
  On your own hardware
  Condo hardware
- **Provider runs an instance**
  Your institution
  National centers
  Commercial vendors
- ~~**No lock in**~~





- Virtual file system with colocated data, description and discussion
- Integrated lightweight visualization tools for quick analysis
- Secure access, sharing, and access controls
- Cross-platform tools, APIs, and Drupal modules
- Post & query data from HPC jobs, workflows, apps, browsers, and command lines

Acknowledgements: This work is 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 NSF."