Sharing computational results via SeedMe platform

Amit Chourasia, San Diego Supercomputer Center, UC San Diego

SeedMe: Stream Encode Explore and Disseminate My Experiments
SeedMe name inspired by Seed: proliferate and grow
Presentation Overview

- Situate context & define problem
- Introduce SeedMe
- Sample use cases
- Sample user interaction
## Accessibility Irony

<table>
<thead>
<tr>
<th>Year</th>
<th>Event/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>“The web was originally conceived and developed to meet the demand for automatic information-sharing between scientists in universities and institutes around the world.”</td>
</tr>
<tr>
<td>1993</td>
<td>“NCSA Mosaic, or simply Mosaic, is the web browser credited with popularizing the World Wide Web.”</td>
</tr>
</tbody>
</table>

**Mission Dashboard at NASA – JPL**

Voyager space crafts have been sharing information since **1977**

**2015**

Computation workloads cannot easily share information with us

- No place to push information (no infrastructure)
- Lack of easy tools

**2007**

Mobile devices can share content instantly
Computation Cycle

- Analyze
- Compute/Ingest
- Post Process

- Abstract?
- Monitor?
- Validate?
Not All Members Have Access
----
Need Exact Location + Permissions

Many customers

Consumable Products
(Preliminary Results)
(Transient Content)

Direct Access

Direct Web Sharing

Inflexible Security Policy

SeedMe.org
Sharing Consumable Products

(Many customers)

Email
- Download + Send
- Scattered results
- Can’t share larger content
- Manual

Webpage
- Download + Upload
- Handle Privacy
- Significant time commitment
- Manual

Cloud Drive
- Download + Upload
- Cannot describe content
- Manual

SeedMe.org
Compute

Anticipate

React

Waterworld (1995)
Consumable Content Dominates

71% of 83 million files transferred were less than one megabyte on Kraken in 2013

Pitfalls in sharing derived content

- Download
- Upload
- Download
- (Round Trip + 1)

- Video Encoding
- Complexity

- Missing Easy
- Automation

- How To
- Describe &
- Discuss
- Content

- Replication
- &
- Scalability

SeedMe.org
## Why not use existing tools?

<table>
<thead>
<tr>
<th>Features</th>
<th>SeedMe</th>
<th>Figshare</th>
<th>Dropbox &amp; GDrive</th>
<th>YouTube &amp; Vimeo</th>
<th>Flickr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>✔ File, Image, Video</td>
<td>✔ File, Image, Video</td>
<td>✔ File, Image, Video</td>
<td>❌ Video only</td>
<td>❌ Image, Video</td>
</tr>
<tr>
<td>Sharing</td>
<td>✔ Public, Group, Private</td>
<td>❓ Public, Group, Private</td>
<td>✔ Public, Group, Private</td>
<td>❓ Public, Private</td>
<td>❌ Public</td>
</tr>
<tr>
<td>Describe All Content</td>
<td>✔</td>
<td>❌</td>
<td>❌ Video only</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Text Tickers</td>
<td>✔</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Metadata Capability</td>
<td>✔</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Periodic Updates</td>
<td>✔</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Upload Method</td>
<td>✔ Command Line, API, Web</td>
<td>❓ API, Web</td>
<td>❓ API, Web Browser</td>
<td>❓ API, Web</td>
<td>❌ Web</td>
</tr>
<tr>
<td>Provide Upload Tools</td>
<td>✔</td>
<td>❌</td>
<td>❌ Dropbox only</td>
<td>❌</td>
<td>❌</td>
</tr>
<tr>
<td>Video Resolution</td>
<td>✔ Arbitrary</td>
<td>❌</td>
<td>❌ Native</td>
<td>❌ Up to 4K</td>
<td>❌ Up to HD</td>
</tr>
<tr>
<td>Image Sequence to Video</td>
<td>✔</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
<td>❌</td>
</tr>
</tbody>
</table>

- ✔ Desirable
- ❓ Limiting
- ❌ Crippling
SeedMe

Share, collaborate, & automate
Scientific data sharing made easy!

Share easily
- Share data, images, & videos with selected colleagues.
- Access from any computer, phone, or tablet.

Collaborate securely
- Discuss preliminary & published results.
- Control who can view and comment on your content.

Automate quickly
- Post data from HPC jobs.
- Create videos from image sequences.

SeedMe.org
SeedMe Collection

Ticker
(Text Message)
Ephemeral

Files
(Classified into three categories)

Images
(plots)

Videos
(videos)

Files
(catch all)

Sequences
(Image set)

Privacy + Collaborators
(Public, Group, Private) + (emails)

Meta Data
(Title, Description, Tags, Key Value Pairs, Credits, License)
SeedMe: How it Works

1. Sign In
2. Create Collection
3. Set Privacy & Sharing
4. Set Metadata
5. Upload Content

Options:
- Web Browser
- Command Line or REST (Automate scripts, workflow or app)
- Update as desired
- Phone or Tablet or Computer (Web Browser)
- View or Download on Web
## SeedMe User Interactions

<table>
<thead>
<tr>
<th>Action</th>
<th>Command Line</th>
<th>API</th>
<th>Web Browser</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create Collection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Update Collection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Query Collection</td>
<td>✓</td>
<td>✓</td>
<td>Under Dev</td>
</tr>
<tr>
<td>Download Collection</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>View Collection</td>
<td>NA</td>
<td>NA</td>
<td>✓</td>
</tr>
<tr>
<td>Delete Collection</td>
<td>Under Dev</td>
<td>Under Dev</td>
<td>✓</td>
</tr>
</tbody>
</table>
SeedMe Processing

Ingestion (Atomic or Recurring) → Instant Processing (Text, Images) → Notification → Batch Processing (Videos) → Responsive Content Access
Use Case: Monitor

Computation state
- Progress %
- Progress parameters
e.g. Last step completed

Requirements
- Managed Sharing
- Universal access
- Automation

```
seedme.py -t "Progress Tracking" \  # Create new Collection

seedme.py -update 29643 \   # Update collection ID
  -tic "step 1" \  # Ticker Text
  -tic "step 2" \  # Ticker Text
  -fp "sample/files/doc.pdf"  # File Path
```

Command line example, but you can upload content simply via Web Browser
## Use Case: Create Dashboards

### Results from simulations, instruments, analysis

- Files
- Plots
- Sequences
- Videos

### Requirements

- Managed Sharing
- Universal access
- Automation

---

**seedme.py**

```bash
t "Quick Start" \
-privacy group \
-email alpha@sdsc.edu \
-notify \
-fp "sample/plots/node.png" \
-sp "sample/sequences/steam"
```

# Collection Title
# Privacy
# Share
# Notify (not automatic)
# File Path
# Seq Path

---

Command line example, but you can upload content simply via Web Browser
Use case: Sequence to Video

Sequences of images are generated by

- Visualizations
- Confocal scans
- Time lapse recording

Requirements

- Frame rate
- High Quality Encoding
- Managed Sharing
- Universal playback
- Automation

```
seedme.py -t "Seq Collection" \ 
-sp "sample/sequences/steam" \ 
-st "sequence title" \ 
-sd "Desc of sequence" \ 
-sr "5" \ 
-se
```

# Collection Title
# Sequence Path
# Sequence Title
# Sequence Description
# Sequence frame Rate
# Trigger Encode to create video from Sequence

Command line example, but you can upload content simply via Web Browser
Use Case: Share & Reuse

<table>
<thead>
<tr>
<th>Disseminate</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Results</td>
</tr>
<tr>
<td>• Data</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Share &amp; discover reusable content</th>
</tr>
</thead>
<tbody>
<tr>
<td>• IPython notebook</td>
</tr>
<tr>
<td>• Session/State files from softwares</td>
</tr>
</tbody>
</table>
Application Integration

Scientific apps shipping with SeedMe
   Kepler Workflow
   VisIt software

Under evaluation
   Paraview, Vapor, YT software

Integrate SeedMe Python/Java Client or write your own REST client
#! /bin/bash

# Create a place holder collection
output=$(./seedme.cmd -title "Place holder collection");

# Extract collection_id as place holder
id=$(echo $output | sed -e 's/^.*"collection_id":\(["\^]*\).*$\1/');

# Run your computation
./SCIENCE_APPLICATION

# Upload end results after computation
./seedme.cmd –up $id –fp output.txt

Further Guidance available on the website
https://www.seedme.org/documentation/integration
https://www.seedme.org/documentation/extract-collection-id
Demo: Web Browser

Collections category: My    Shared    Public
https://www.seedme.org/collections

You may Add, Edit collections and Notify collaborators

Add a new collection
https://www.seedme.org/collections/add

Examples:
• Result sharing https://www.seedme.org/node/5458
• Reusable content https://www.seedme.org/node/25907
Demo: Command Line Interaction

One Time Setup
1) Download - APIKey File (Move to your Home directory) (Requires sign-in)
2) Download - SeedMe Client (Python or Standalone)
3) Download - Sample data for testing (optional)
Demo: Command Line Interaction

Create a new collection

```bash
seedme.py -title "SeedMe Quick Start"
```
Getting Started

**Interaction Tools**

- Standalone executable (command line)
- Python client/module (API + command line)
- Java client
- Web Browser
- Curl command line utility (not recommended)

**Information**

- Quick start guide
- Use cases
- Documentation
- Tips for C, Fortran codes
- Blog
Upcoming
• Federated Login

Limitations
• No folders, only files
• Owner write only

Policy
• 100 mb limit per file
• Storage quota not enforced

Future planned work
Rewrite SeedMe2.0
• Support folders
• Support group write
• REST client in additional languages
• Support simple visualization like charts from text files
• To DOI or not to DOI?
• Open Source SeedMe
Acknowledgements

All early users for their valuable feedback

Kristen Levy, Mahidhar Tatineni, Michael Dwyer, Tom Hutton, Doug Weimer, SDSC

Apple Inc.: Provided test hardware/software on loan during conceptualization phase

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."
Invite you to visit & try SeedMe.org

Contact
amit @ sdsc.edu
https://www.seedme.org/contact

Twitter @SeedmeOrg