Sharing computational results via SeedMe platform

Amit Chourasia, San Diego Supercomputer Center, UC San Diego

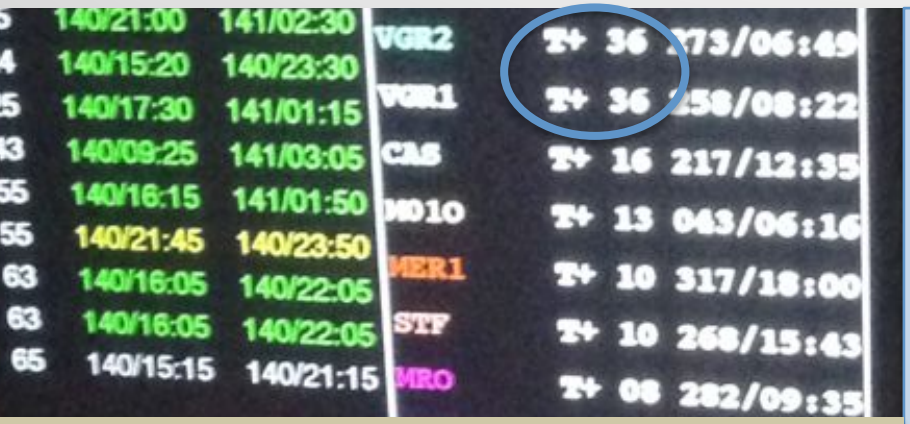
SeedMe: **S**tream **E**ncode **E**xplore and **D**isseminate **M**y **E**xperiments

SeedMe name inspired by Seed : proliferate and grow

Presentation Overview

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

Accessibility Irony



Mission Dashboard at NASA – JPL
Voyager space crafts have been sharing information since **1977**

1990

"The web was originally conceived and developed to meet the demand for automatic information-sharing between scientists in universities and institutes around the world."

The birth of the web. CERN
http://home.web.cern.ch/topics/birth-web

1993

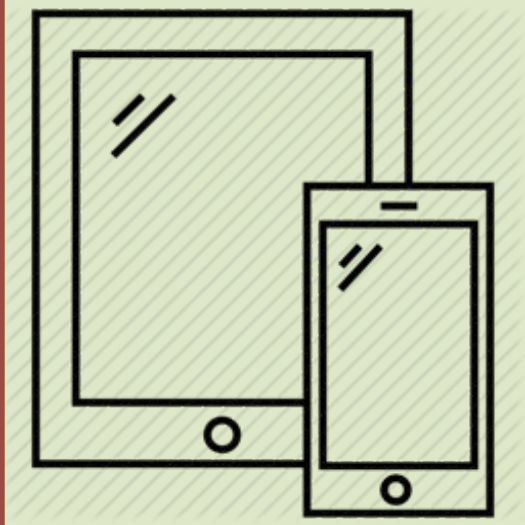
"NCSA Mosaic, or simply Mosaic, is the web browser credited with popularizing the World Wide Web."

Mosaic (Web browser), Wikipedia
http://en.wikipedia.org/wiki/Mosaic_(web_browser)

2015

Computation workloads cannot easily share information with us

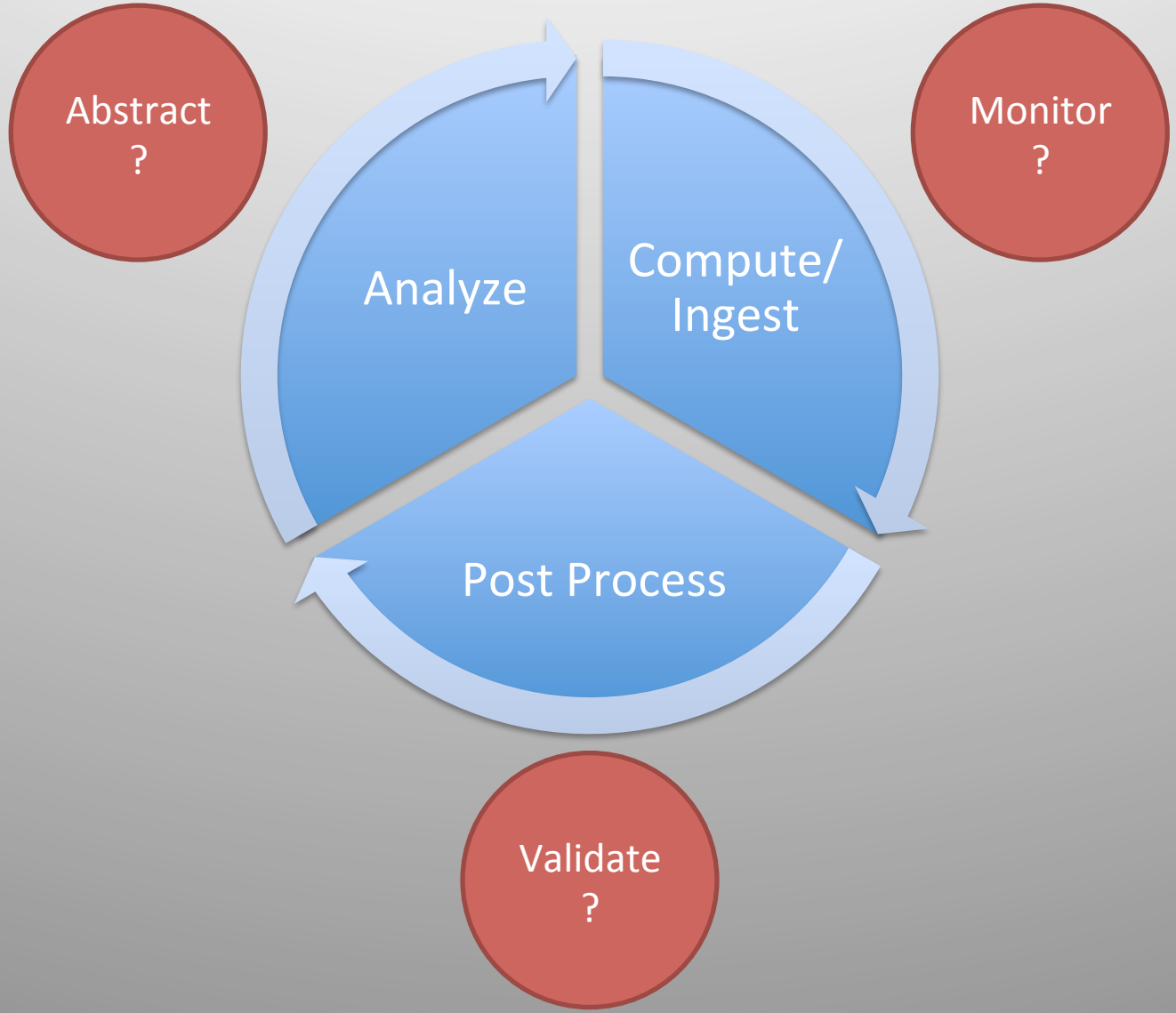
- **No place to push information (no infrastructure)**
- **Lack of easy tools**

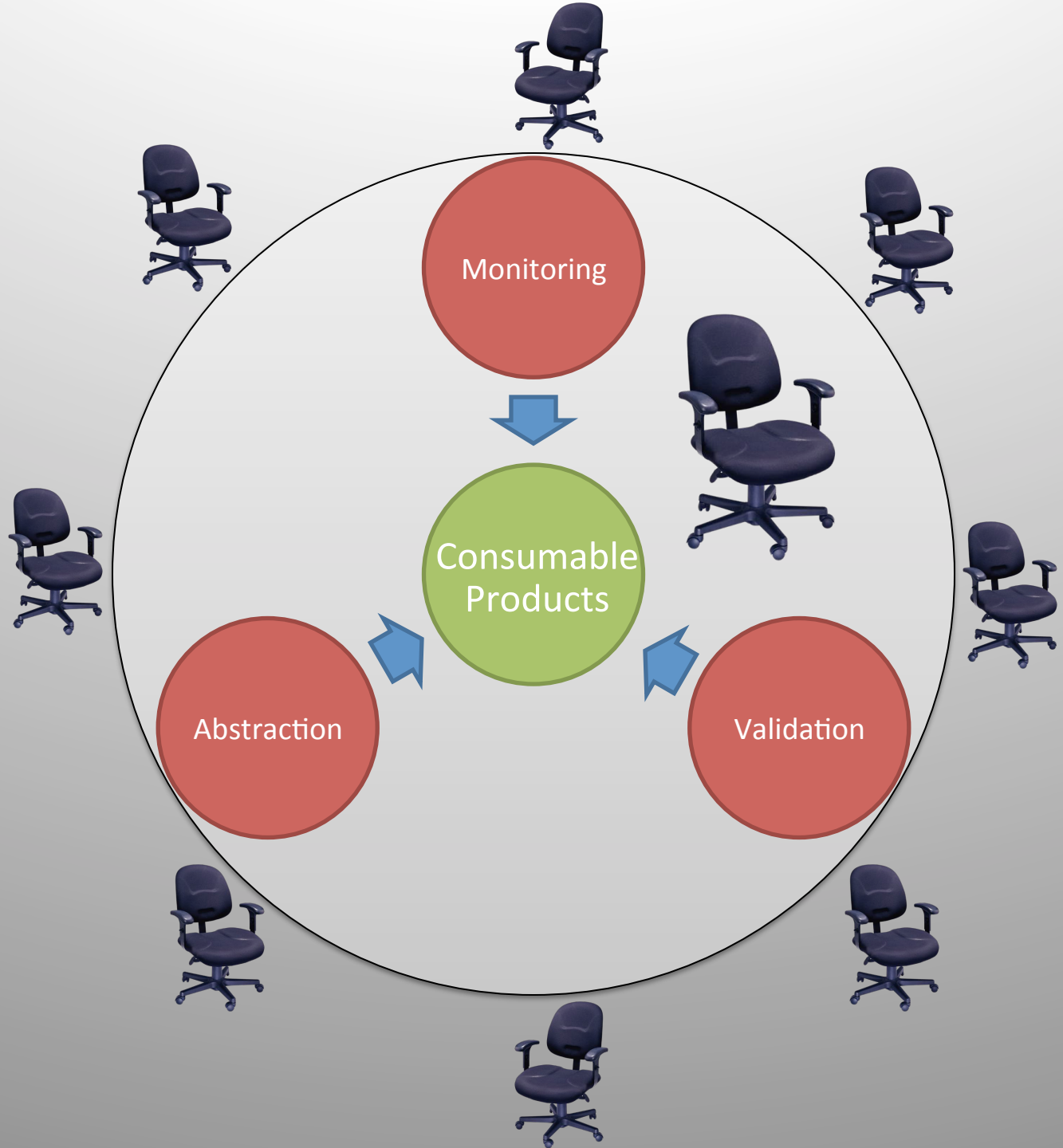


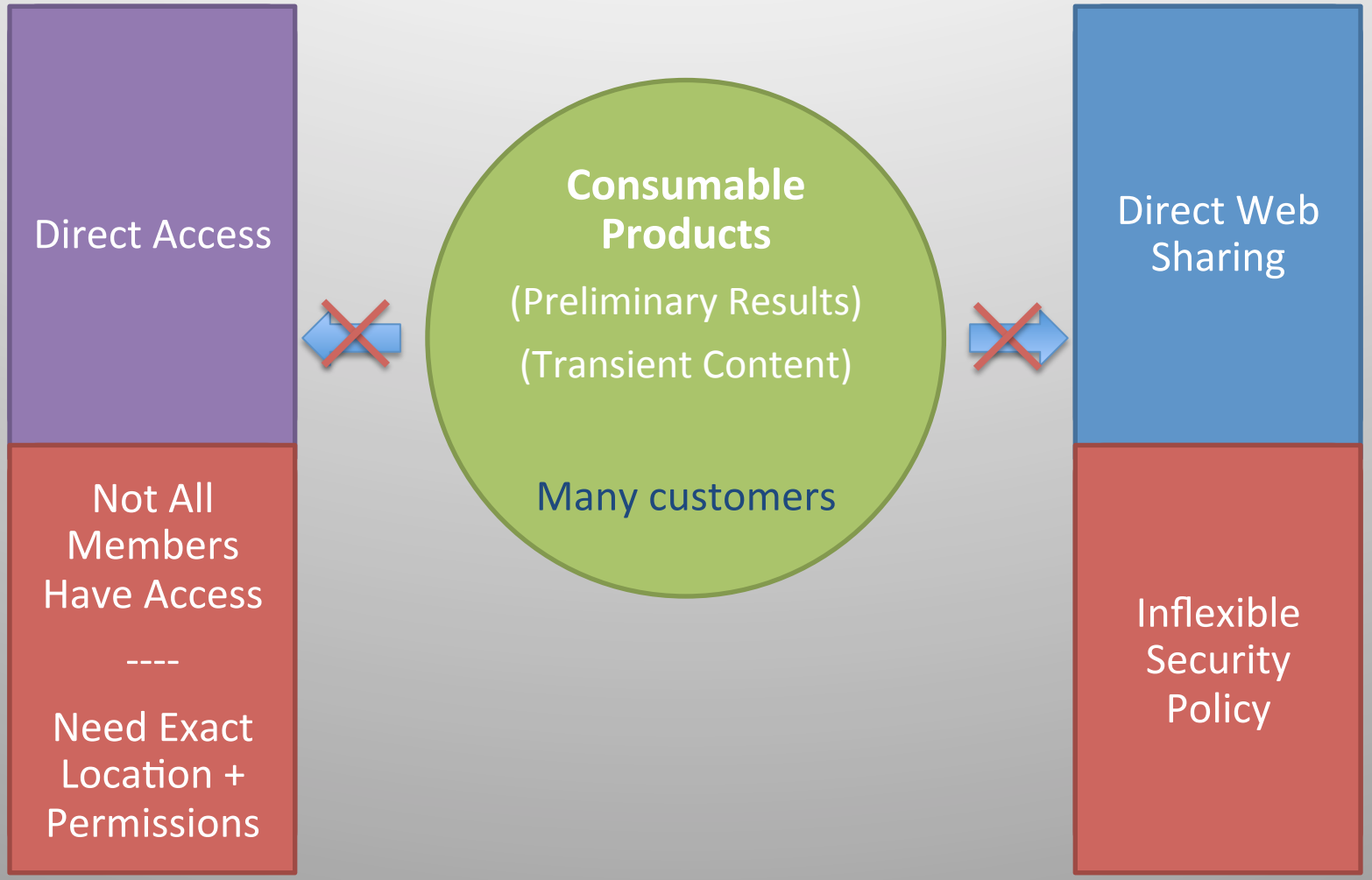
2007

Mobile devices can share content instantly

Computation Cycle






Direct Access
Not All Members Have Access

Need Exact Location + Permissions
Consumable Products
(Preliminary Results)
(Transient Content)
Many customers
Direct Web Sharing
Inflexible Security Policy

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

Compute

Anticipate

React

Waterworld (1995)

Consumable Content Dominates

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

S. Srinivasan, V. Hazlewood, and G. D. Peterson. 2014. Descriptive Data Analysis of File Transfer Data. In Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment (XSEDE '14). ACM, New York, NY, USA, Article 37 , 8 pages.

Pitfalls in sharing derived content

Download
Upload
Download

(Round Trip + 1)

Video
Encoding
Complexity

Missing Easy
Automation

How To
Describe &
Discuss
Content

Replication
&
Scalability

Why not use existing tools?

Features	SeedMe	Figshare	Dropbox & GDrive	YouTube & Vimeo	Flickr
Content	✔ File, Image, Video	✔ File, Image, Video	✔ File, Image, Video	✘ Video only	✘ Image, Video
Sharing	✔ Public, Group, Private	? Public, Group, Private (Limited of users)	✔ Public, Group, Private	? Public, Private	✘ Public
Describe All Content	✔	✘	✘	✘ Video only	✘
Text Tickers	✔	✘	✘	✘	✘
Metadata Capability	✔	✘	✘	✘	✘
Periodic Updates	✔	✘	✘	✘	✘
Upload Method	✔ Command Line, API, Web Browser	? API, Web Browser	? API, Web Browser	? API, Web Browser	✘ Web Browser
Provide Upload Tools	✔	✘	? Dropbox only	✘	✘
Video Resolution	✔ Arbitrary	✘	✘ Native	✘ Up to 4K	✘ Up to HD
Image Sequence to Video	✔	✘	✘	✘	✘

✔ Desirable ? Limiting ✘ Crippling

Share, collaborate, & automate

Scientific data sharing made easy!

Share easily

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

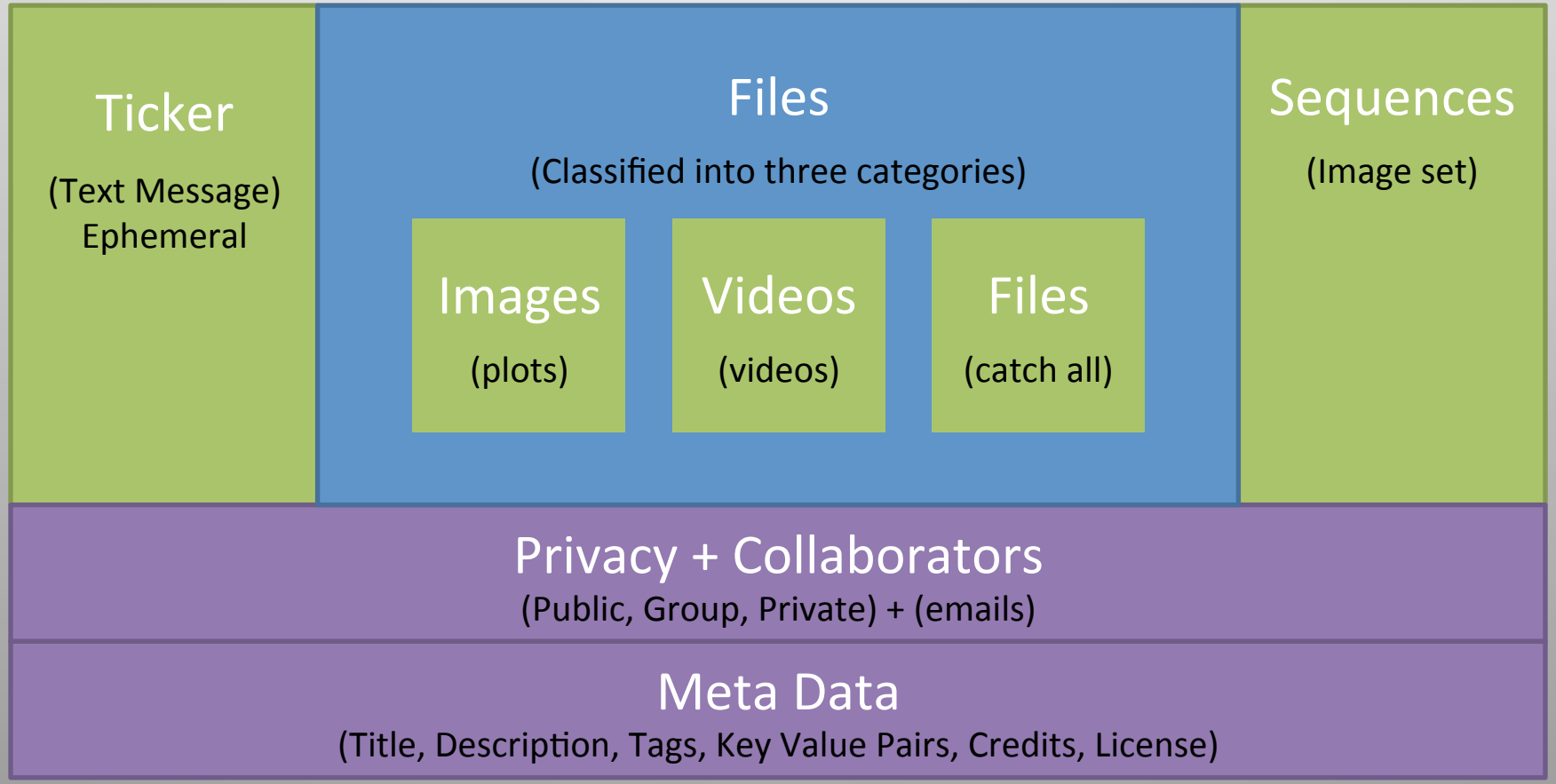
Collaborate securely

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

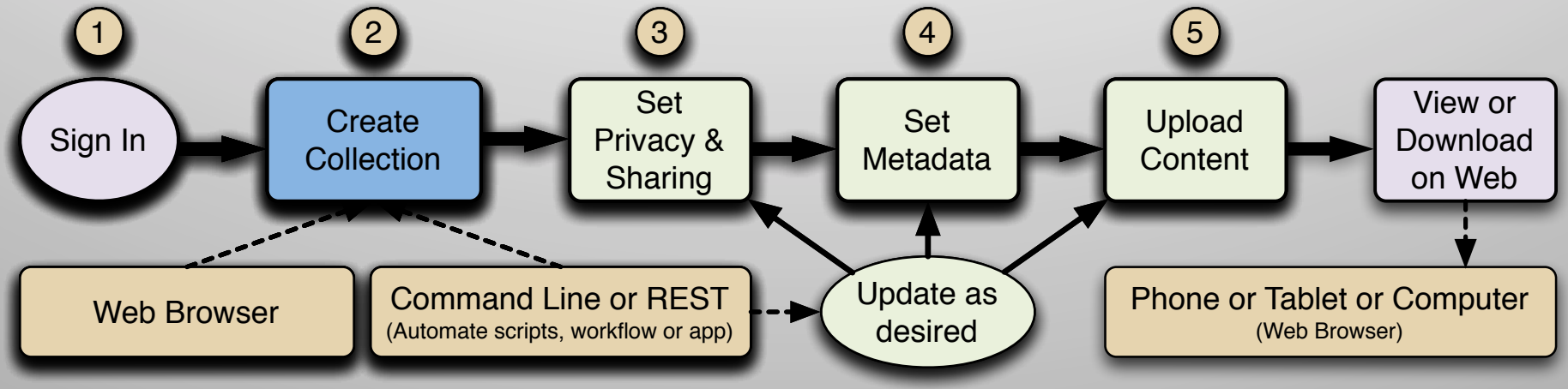
Automate quickly

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

SeedMe Collection



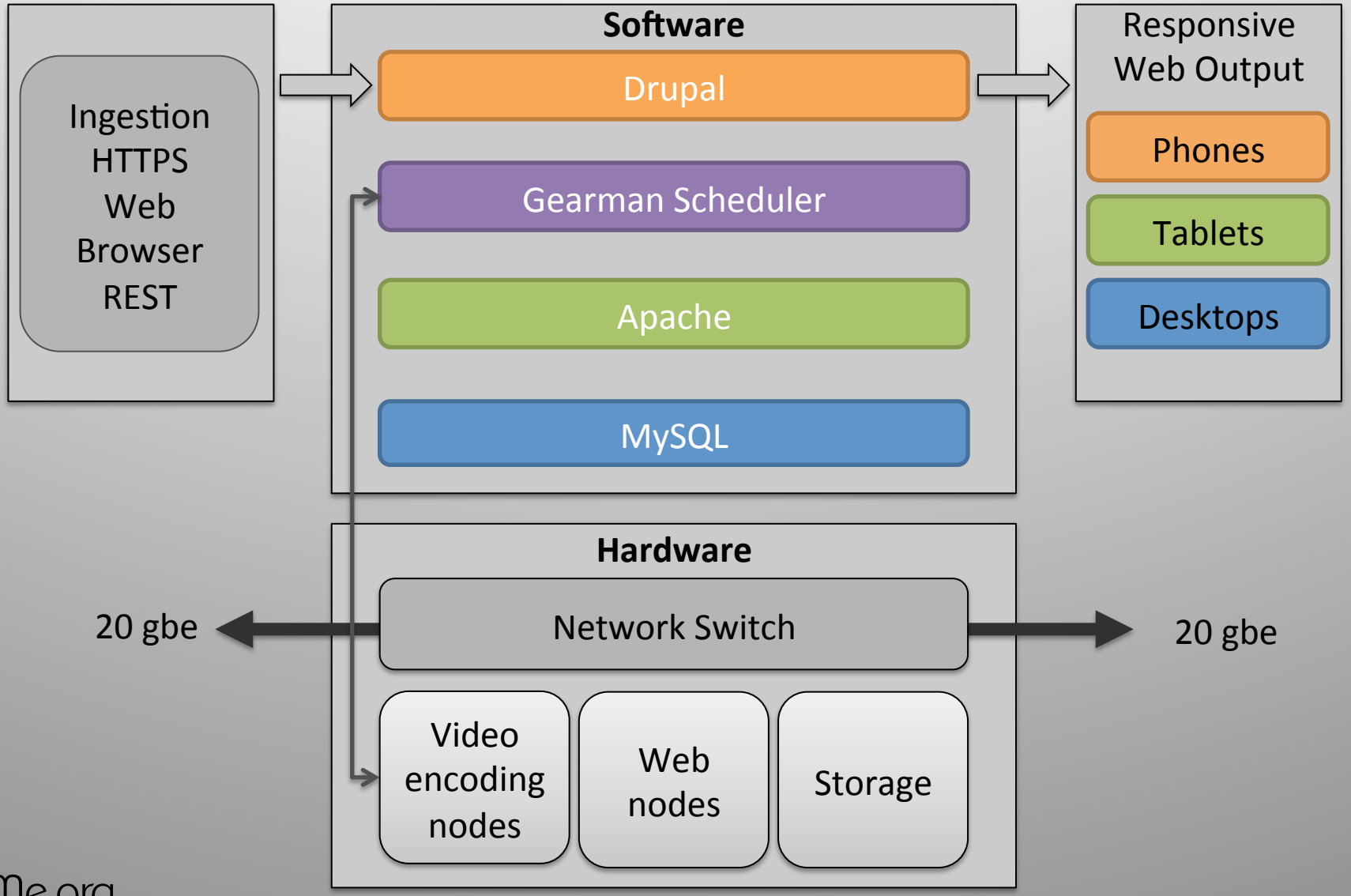
SeedMe: How it Works



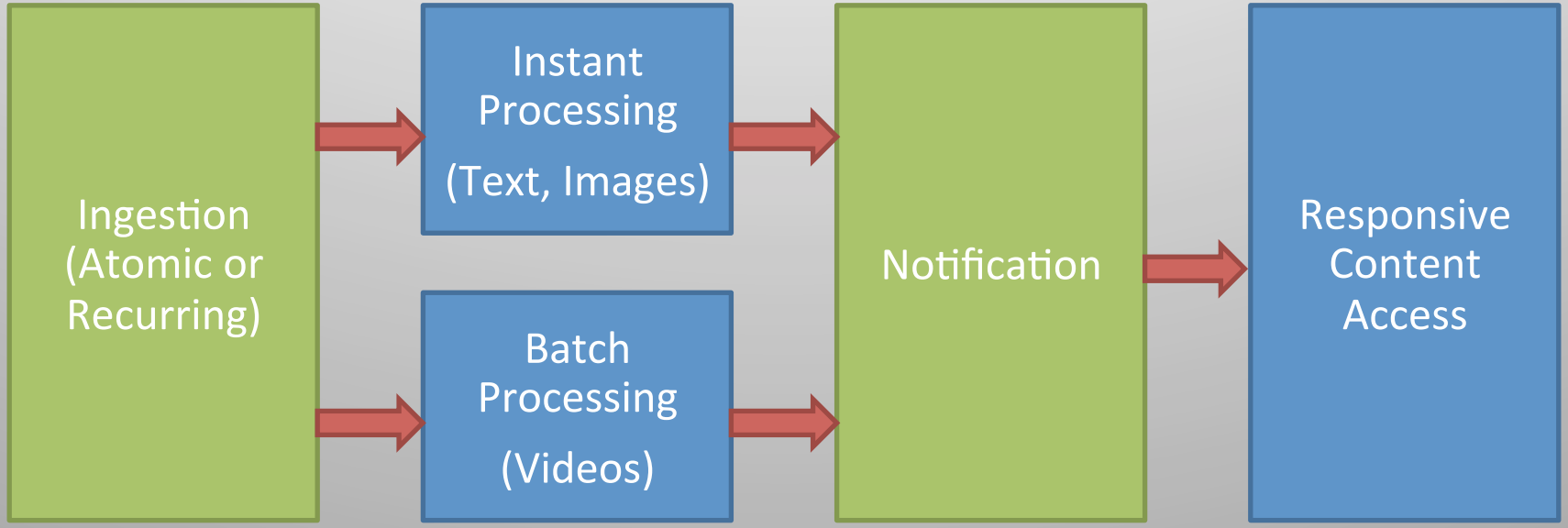
SeedMe User Interactions

Action	Command Line	API	Web Browser
Create Collection	✔	✔	✔
Update Collection	✔	✔	✔
Query Collection	✔	✔	Under Dev
Download Collection	✔	✔	✔
View Collection	NA	NA	✔
Delete Collection	Under Dev	Under Dev	✔

SeedMe Architecture



SeedMe Processing



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
```

SeedMe.org

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     -t “Quick Start” \                  # Collection Title
              -privacy group \                    # Privacy
              -email alpha@sdsc.edu \             # Share
              -notify \                           # Notify (not automatic)
              -fp "sample/plots/node.png" \       # File Path
              -sp "sample/sequences/steam”        # Seq Path
```

SeedMe.org

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” \                  # Collection Title
              -sp "sample/sequences/steam" \         # Sequence Path
              -st "sequence title" \                 # Sequence Title
              -sd "Desc of sequence" \               # Sequence Description
              -sr "5" \                              # Sequence frame Rate
              -se                                    # Trigger Encode to
                                                       create video from Sequence
```

SeedMe.org

Command line example, but you can upload content simply via Web Browser

Use Case: Share & Reuse

Disseminate

- Results
- Data

Share & discover reusable content

- IPython notebook
- Session/State files from softwares

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
```

Dummy Automation Example

```
#! /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:

- Dashboard https://www.seedme.org/node/5607 , https://www.seedme.org/node/15868
- 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

```
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**

SeedMe Team

Amit Chourasia, Mona Wong, David Nadeau

Andrew Ferbert, Michael Norman

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

Contact

amit @ sdsc.edu

https://www.seedme.org/contact

Twitter @SeedmeOrg