



Share, collaborate & automate

# SeedMe2: Data Sharing Building Blocks

Amit Chourasia

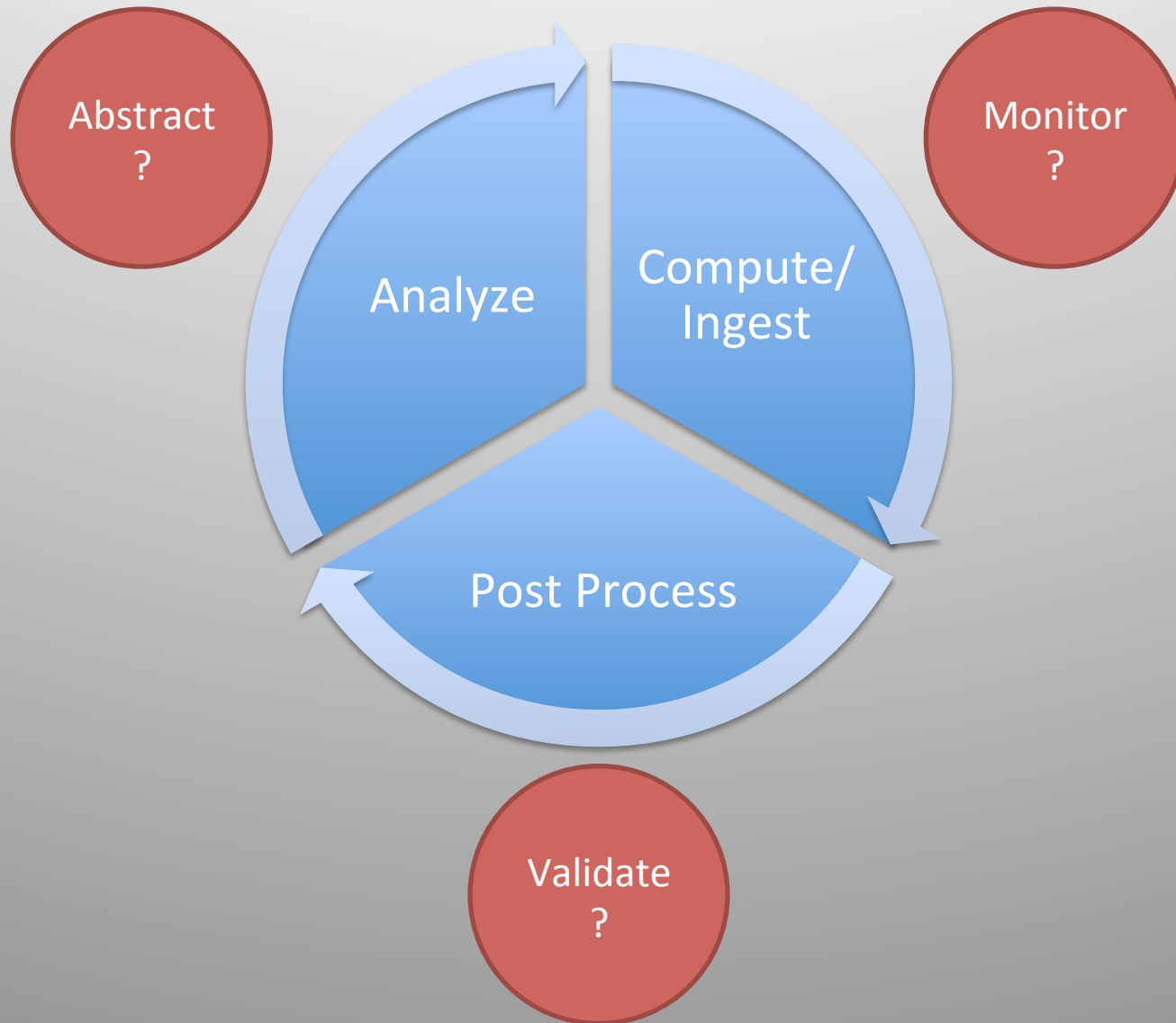
Sr. Visualization Scientist

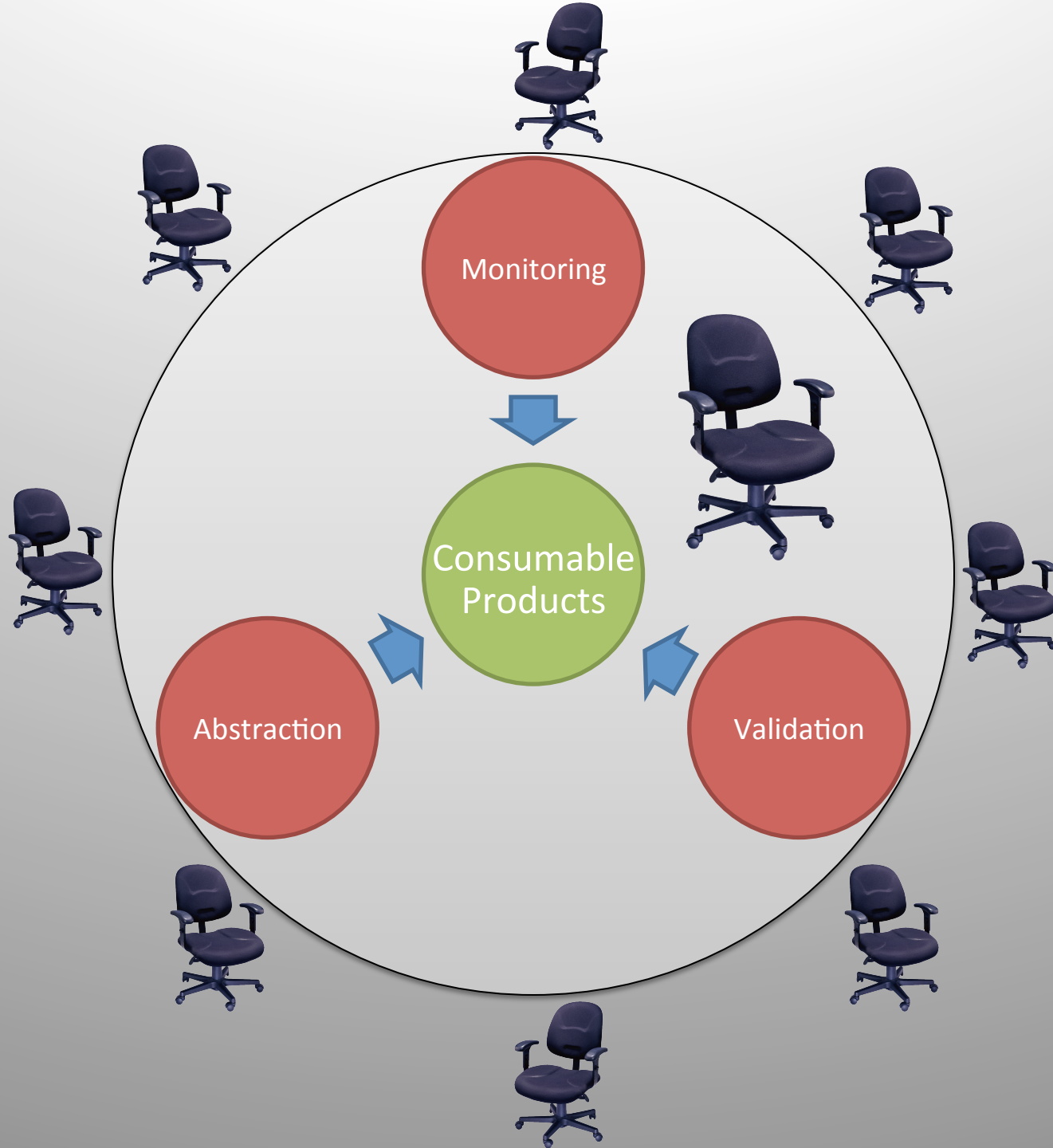
San Diego Supercomputer Center, UC San Diego

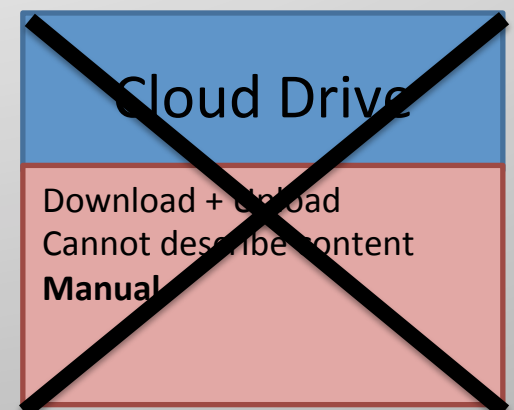
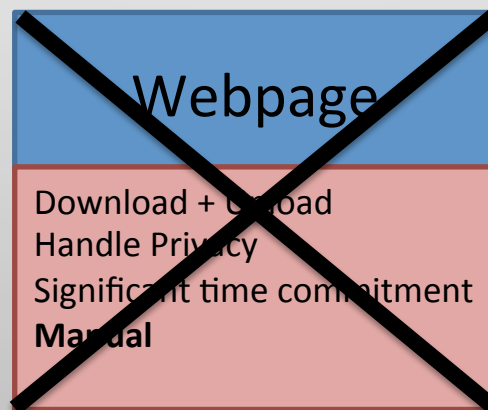
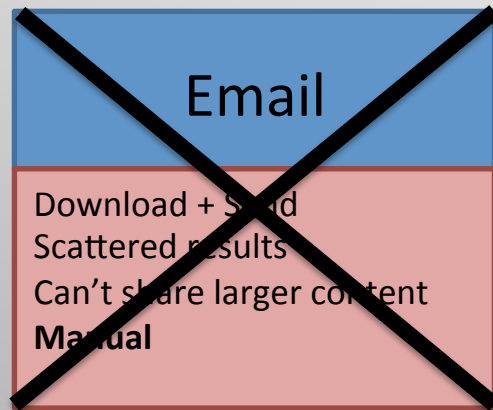
Best Practices in Data Infrastructure Workshop, Pittsburg, PA

May 17-18, 2016

# Computation Cycle







Consumer/Home grown Infrastructure is inadequate

## Gaps in compute infrastructure

- Collaborators need account to compute resource (non trivial/slow)
- Exact path to content
- Permissions to access content
- Context and content description away from data

## Stumbling blocks

Transfer

Store

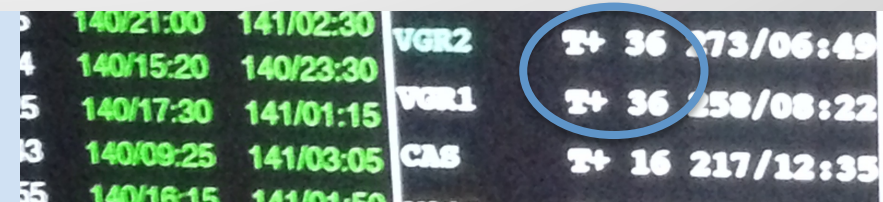
Access  
control

Collaboration

# Accessibility Irony

**1977 onwards** Voyager space crafts have been sharing information. ✓

Mission dashboard at NASA's JPL



A mission dashboard from NASA's JPL showing flight data for Voyager 1 and 2. The dashboard is a black screen with green and white text. It lists various parameters such as time, distance, and speed for both spacecraft. A blue circle highlights the 'T+ 36' value for Voyager 1, which is circled in blue.

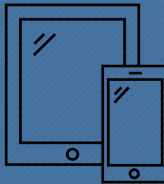
Parameter	VGR2	VGR1	CAS
T+ 36	273/06:49	258/08:22	217/12:35

**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.” ✓

**2007** Mobile devices can share content instantly. ✓



**2016 Computation workloads cannot easily share data / information X**

- Can't provide direct access
- No place to push information
- Lack of ready to use tools

# Rapid access to HOT consumable data\*

Transient data and preliminary results

- Statistics
- Benchmarks
- Visualizations
- Messages

\* Consumable data < 2GB per file

# Consumable data dominates transfers

**< 1 MB**

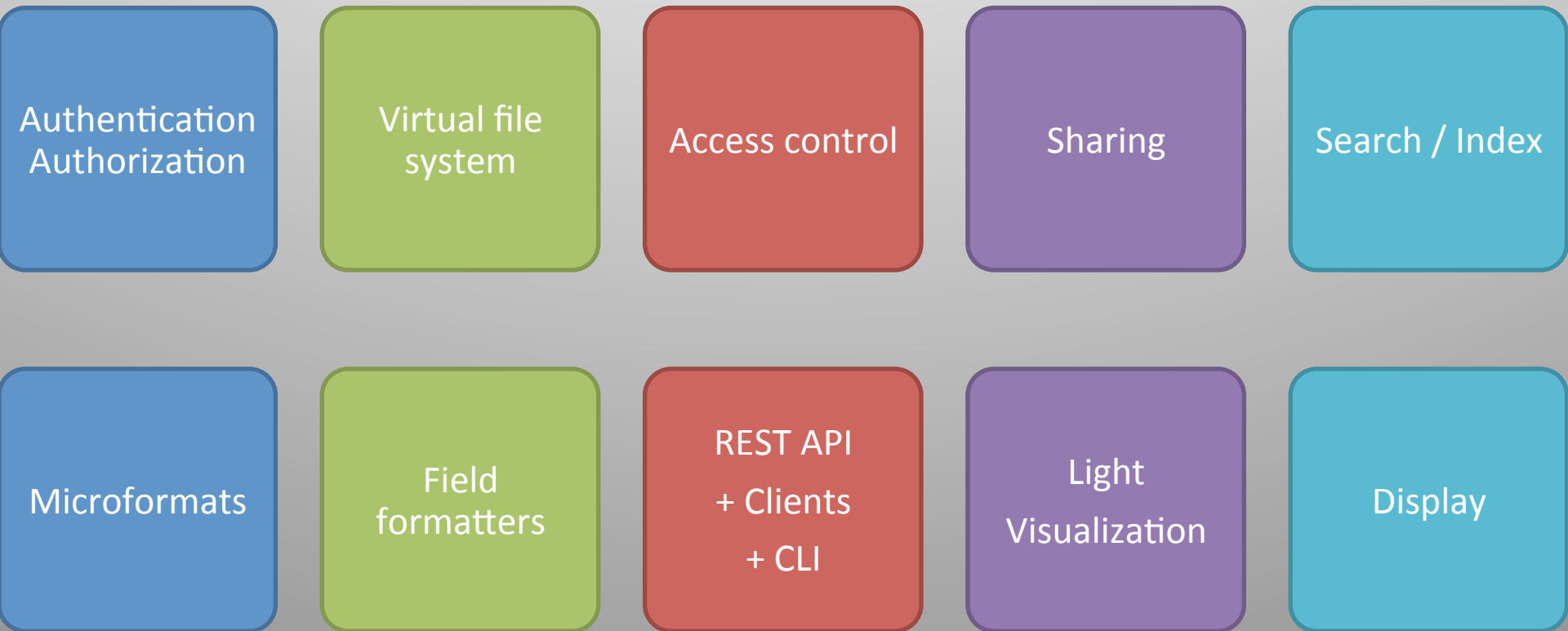
**71% of 83 million files transferred using Globus on Kraken in 2013**

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.

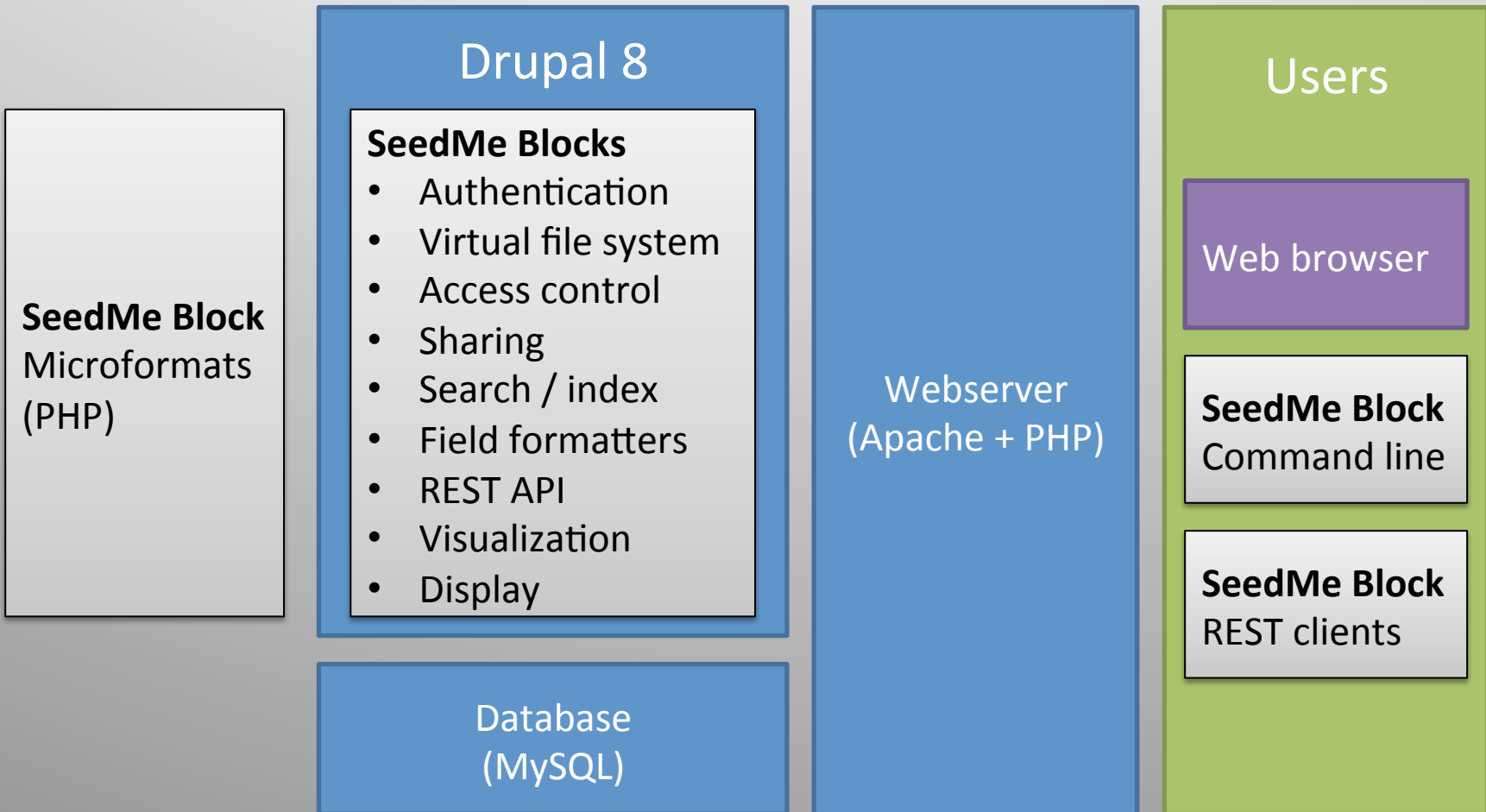


# SeedMe2 : Building blocks

SeedMe = Stream encode explore and disseminate My experiments



# SeedMe2 Architecture



# Target Users

- Batch computing users
  - Collaboration hub
  - Personal dashboard
- Data repositories
  - Plain - As provided
  - Custom : Project specific
  - Archive with DOI
- Gateways
  - Data escrow service/Sharing/publishing
- CI providers

# Use as you see fit

- Use as a service (Cloud)
- Customize and run your own instance
  - On your own hardware
  - Condo hardware
- Have a provider run an instance for you
  - Your institution
  - National centers
  - Commercial vendors

**~~No lock in~~**

# Best practices

- **Solid user experience**
  - User centric tools ( Graphical, Command line, API)
  - Robust and performant implementation
  - Extensive quality & assurance
- Pluggable/Modular architecture
  - Extension (customization)
  - Integration with scientific software
  - Interoperability with Gateways
- Leverage federated identity provider, but consider backup
- Leverage the Web
- Leverage broadly used community driven frameworks
- (Train, document, demonstrate)<sup>n</sup>

# Bad practices

- Unstable software
- Complex / Inconsistent functionality
- Mutating API
- Lack of performance
- Inadequate quality assurance

# Challenges

- Web browsers have quirks
- Provide service at high reliability
- Sustenance and persistence beyond grant
- Handling dependencies
  - Software maintenance
  - Infrastructure
  - Third-party services
- Feature creep
- Manage user expectations and demands
- Migration tools

# Findings

- Highly desired
  - Access control
  - Instant sharing
- Scholars very careful with opening the data
- Usage is in bursts
- Varied use cases
- Complex design is unattractive
- Web browser compatibility getting better, but still tough
- Providing high performance for logged-in users is tough





Amit Chourasia, Mona Wong, David Nadeau, John Moreland, Dmitry Mishin  
Andrew Ferbert, Michael Norman  
**SDSC**

## **Acknowledgements**

All users and application integrators for their valuable feedback

### **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 try



## Contact Us

help @ seedme.org