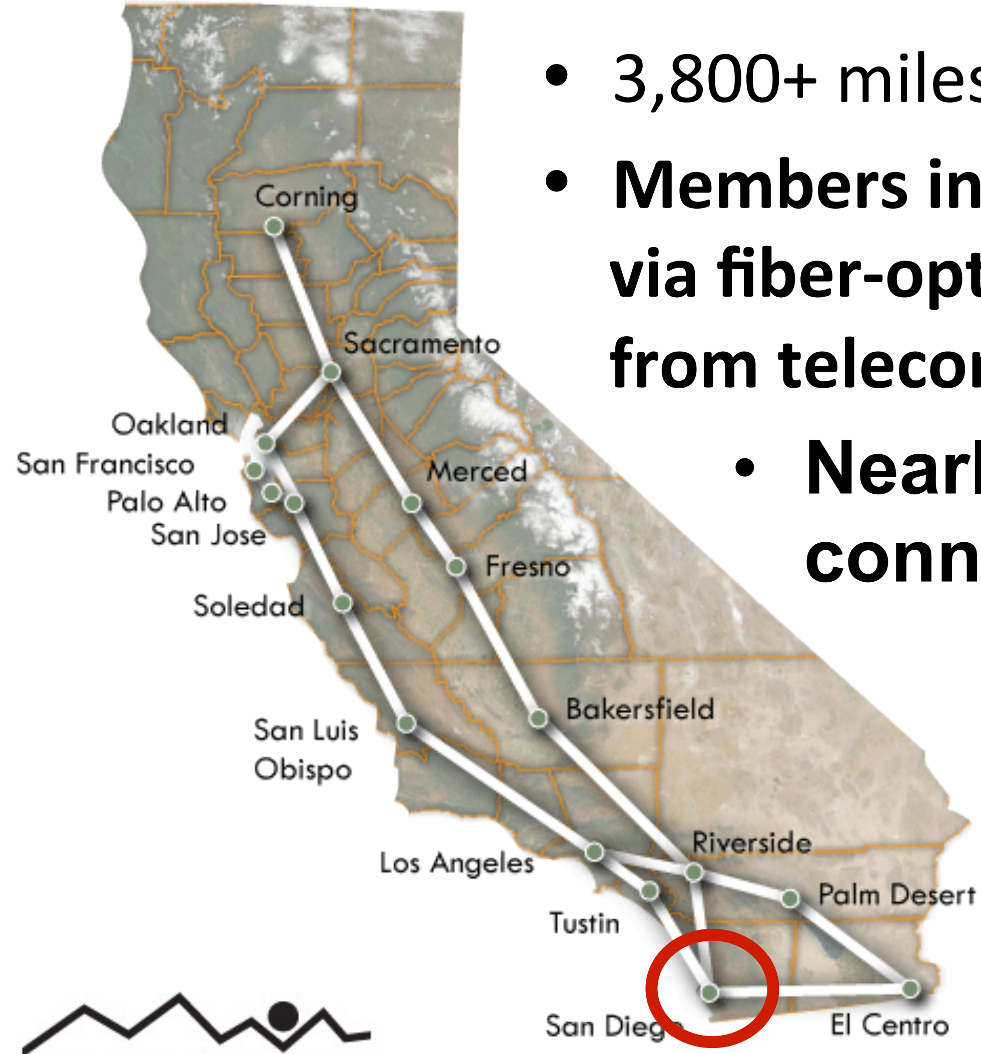
Big Data: Storing, Moving, Mining, Visualizing

Things HiPACC Could be Leveraging

Topics

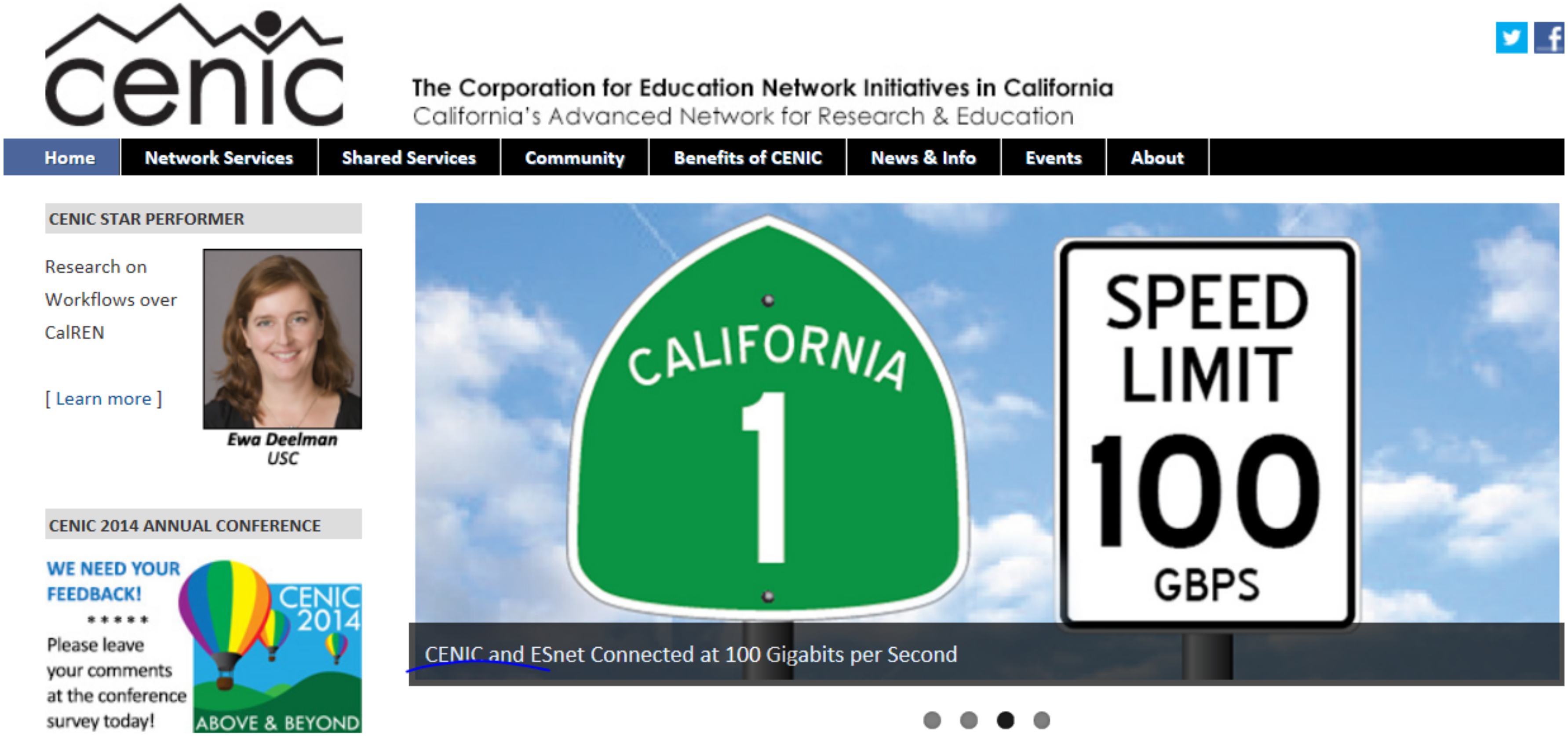
- Moving: CENIC @100G
- Storing: Data Oasis
- Mining: Gordon, Comet
- Visualizing: yt
- Sharing: SeedMe.org

California Has the Most Advanced Big Data Optical Network in the Nation



- 3,800+ miles of optical fiber
- **Members in all 58 counties connect via fiber-optic cable or leased circuits from telecom carriers**
    - **Nearly *10,000* sites connect to CENIC**
        - *10,000,000+* Californians use CENIC each day
        - Governed by members on the segmental level

CENIC, Internet2, ESNet now connected at 100G



The Corporation for Education Network Initiatives in California

California's Advanced Network for Research & Education

CENIC STAR PERFORMER

Research on Workflows over CalREN

[Learn more]

Ewa Deelman
USC

CENIC 2014 ANNUAL CONFERENCE

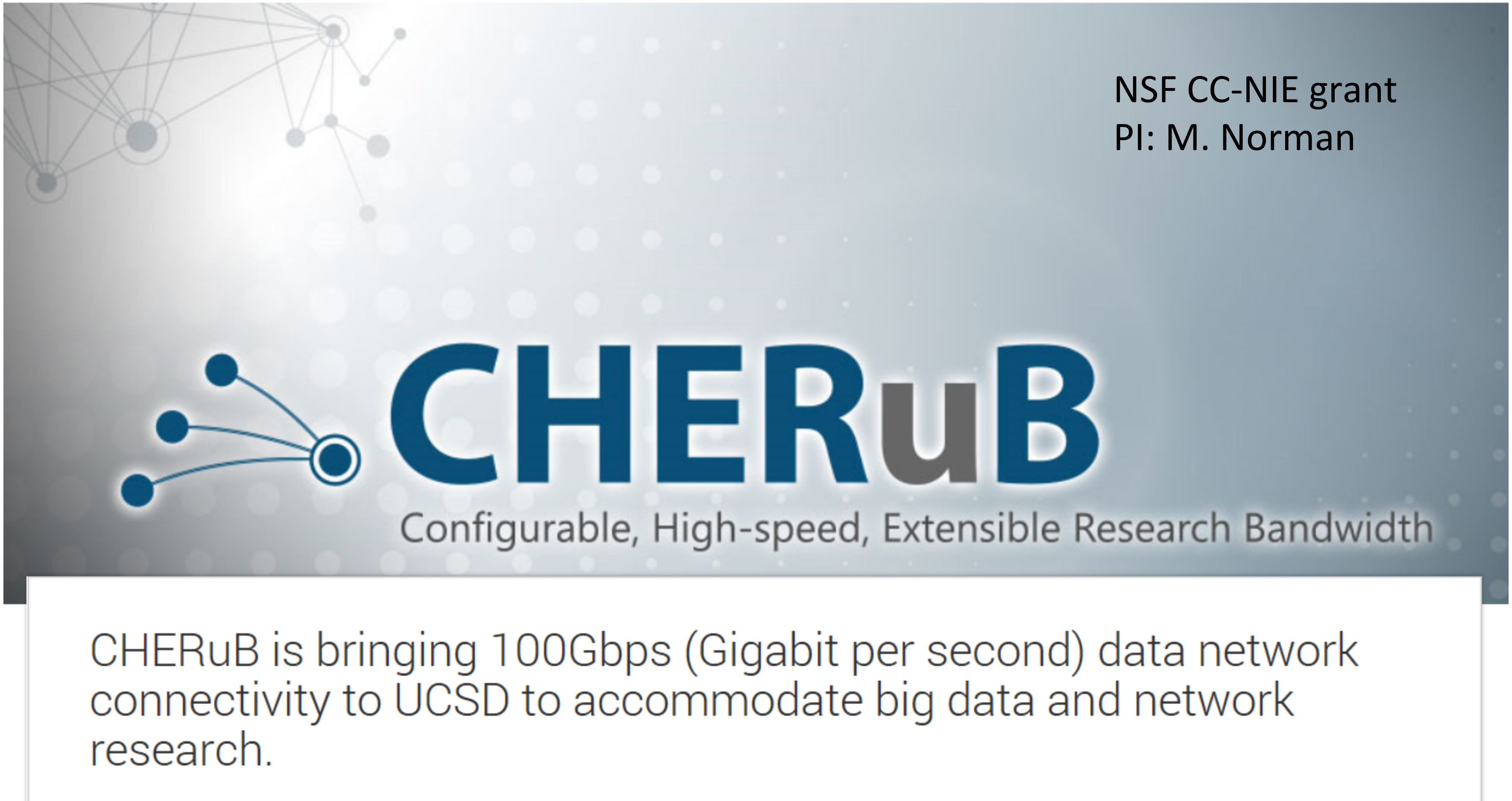
WE NEED YOUR FEEDBACK!

*****

Please leave your comments at the conference survey today!

CENIC and ESnet Connected at 100 Gigabits per Second

UCSD will be connected to CENIC and ESNet at 100G by end of 2014



NSF CC-NIE grant
PI: M. Norman

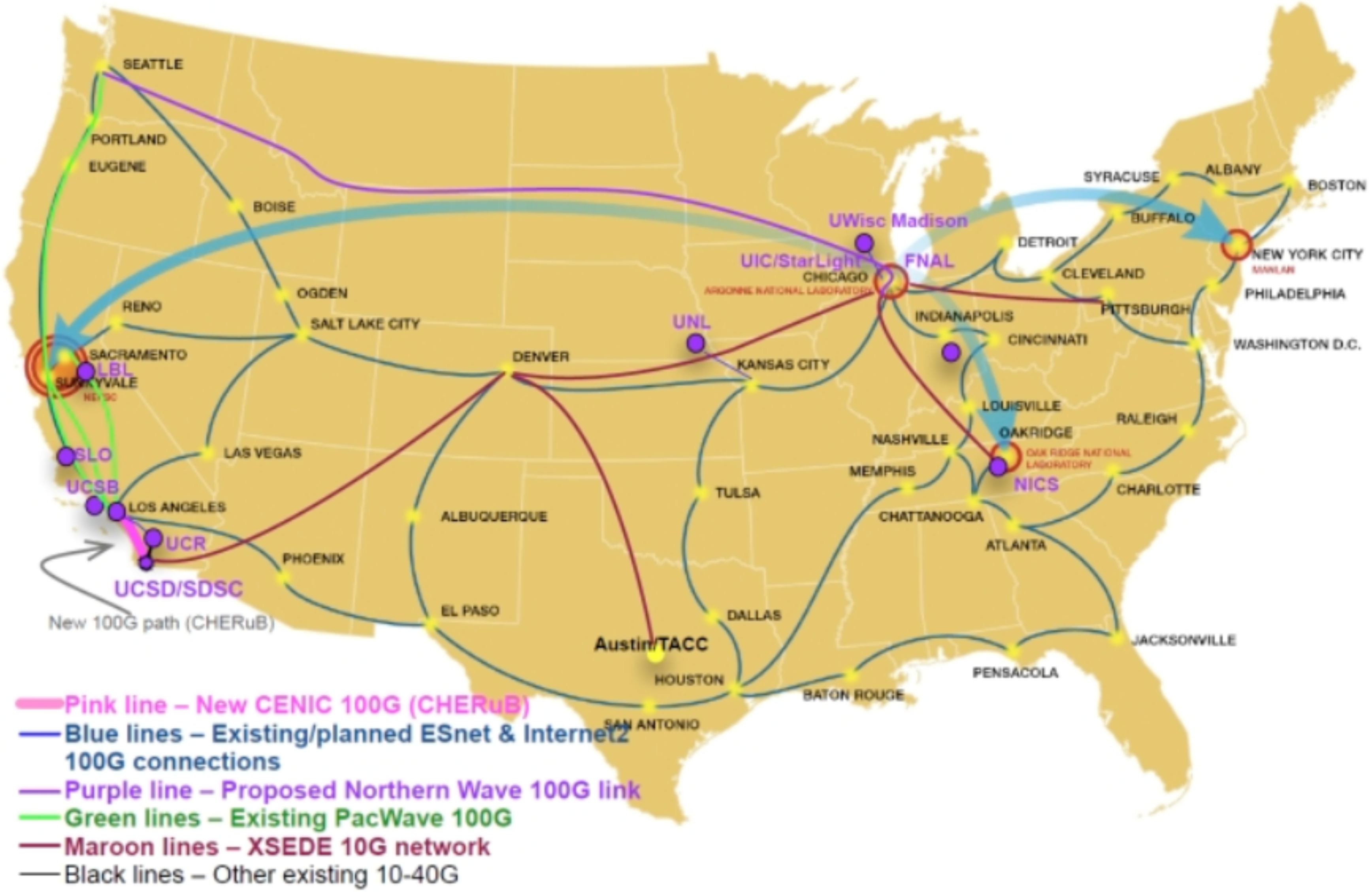
CHERuB

Configurable, High-speed, Extensible Research Bandwidth

CHERuB is bringing 100Gbps (Gigabit per second) data network connectivity to UCSD to accommodate big data and network research.

Other UC campuses with 100G connectivity

Campus	Connectivity	Date
UCLA	100G	Now
UCSC	100G	2014
UCD	100G	2015
UCB	100G	2015



Pink line – New CENIC 100G (CHERuB)

Blue lines – Existing/planned ESnet & Internet2 100G connections

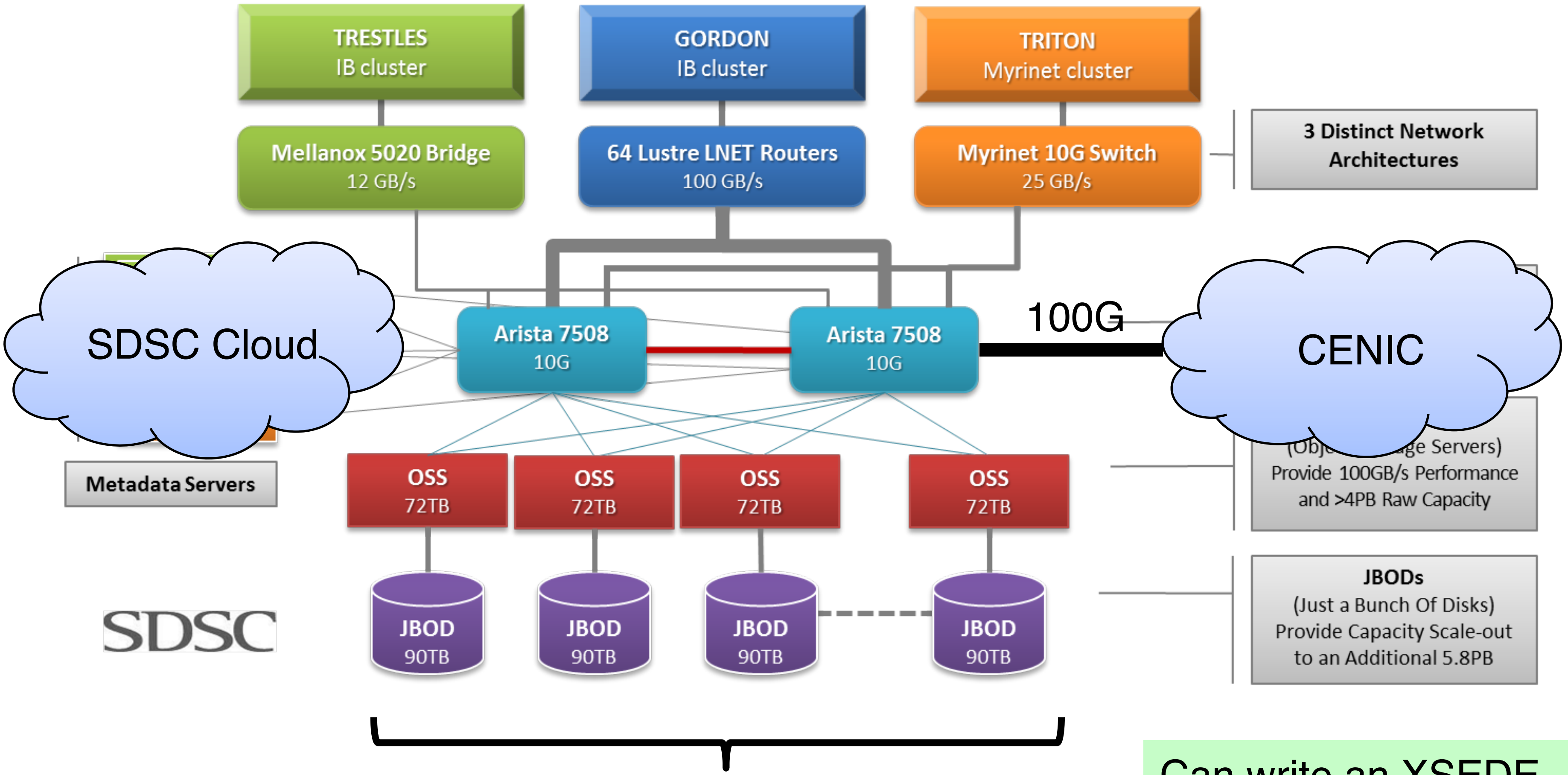
Purple line – Proposed Northern Wave 100G link

Green lines – Existing PacWave 100G

Maroon lines – XSEDE 10G network

Black lines – Other existing 10-40G

Data Oasis Heterogeneous Architecture



4.5 PB (/scratch and /project)

Can write an XSEDE proposal for a storage allocation; e.g., /oasis/proj/hipacc

SDSC Cloud
http://cloud.sdsc.edu

OpenStack components

- SWIFT object store
- NOVA compute

openstack™
CLOUD SOFTWARE

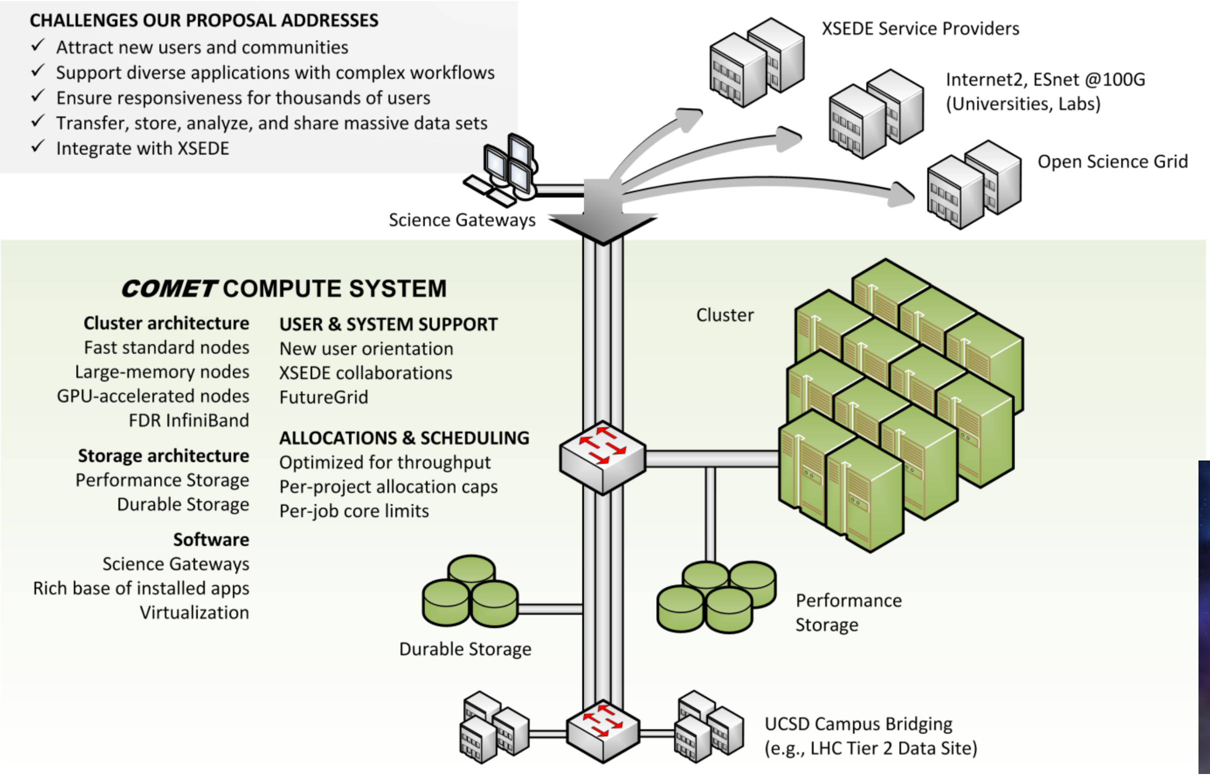
Functionality

- Easily locating research data (web accessible)
- Storage facilities that can hold huge datasets
- New web, compute servers with few clicks
- Publishing and sharing data

Data Mining: Gordon Data-Intensive Supercomputer (XSEDE resource)

- First HPC system with flash SSD (300 TB)
- All I/O channels “maxed out” to accommodate Big Data movement
- Can “spin up” a SSD based Hadoop clust from the batch queu individual users
- Fore-runner to Catal system at LLNL

Comet supercomputer (2015) will support Big Data workflows for all fields of scholarship



CHALLENGES OUR PROPOSAL ADDRESSES

- ✓ Attract new users and communities
- ✓ Support diverse applications with complex workflows
- ✓ Ensure responsiveness for thousands of users
- ✓ Transfer, store, analyze, and share massive data sets
- ✓ Integrate with XSEDE

COMET COMPUTE SYSTEM

Cluster architecture
Fast standard nodes
Large-memory nodes
GPU-accelerated nodes
FDR InfiniBand

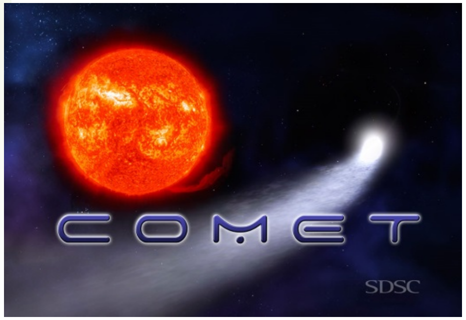
Storage architecture
Performance Storage
Durable Storage

Software
Science Gateways
Rich base of installed apps
Virtualization

USER & SYSTEM SUPPORT
New user orientation
XSEDE collaborations
FutureGrid

ALLOCATIONS & SCHEDULING
Optimized for throughput
Per-project allocation caps
Per-job core limits

2 PF compute
7 PB Lustre PFS
4.5 PB 2nd copy PFS
(old Data Oasis)

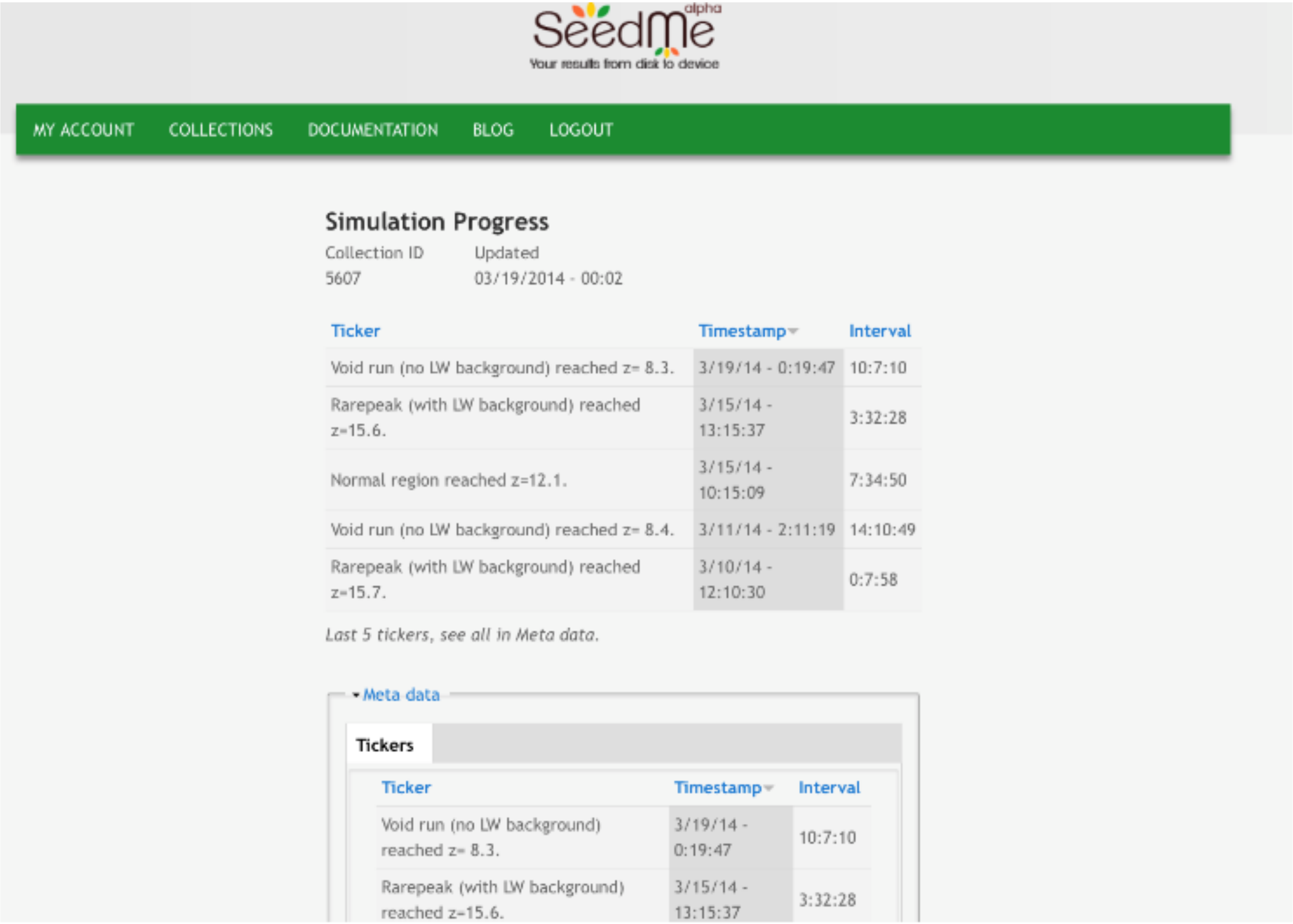


Sharing: SeedMe.org

SeedMe alpha

Your results from disk to device

- A new SaaS web service under development at SDSC
- Think of it as Flickr+YouTube for scientists



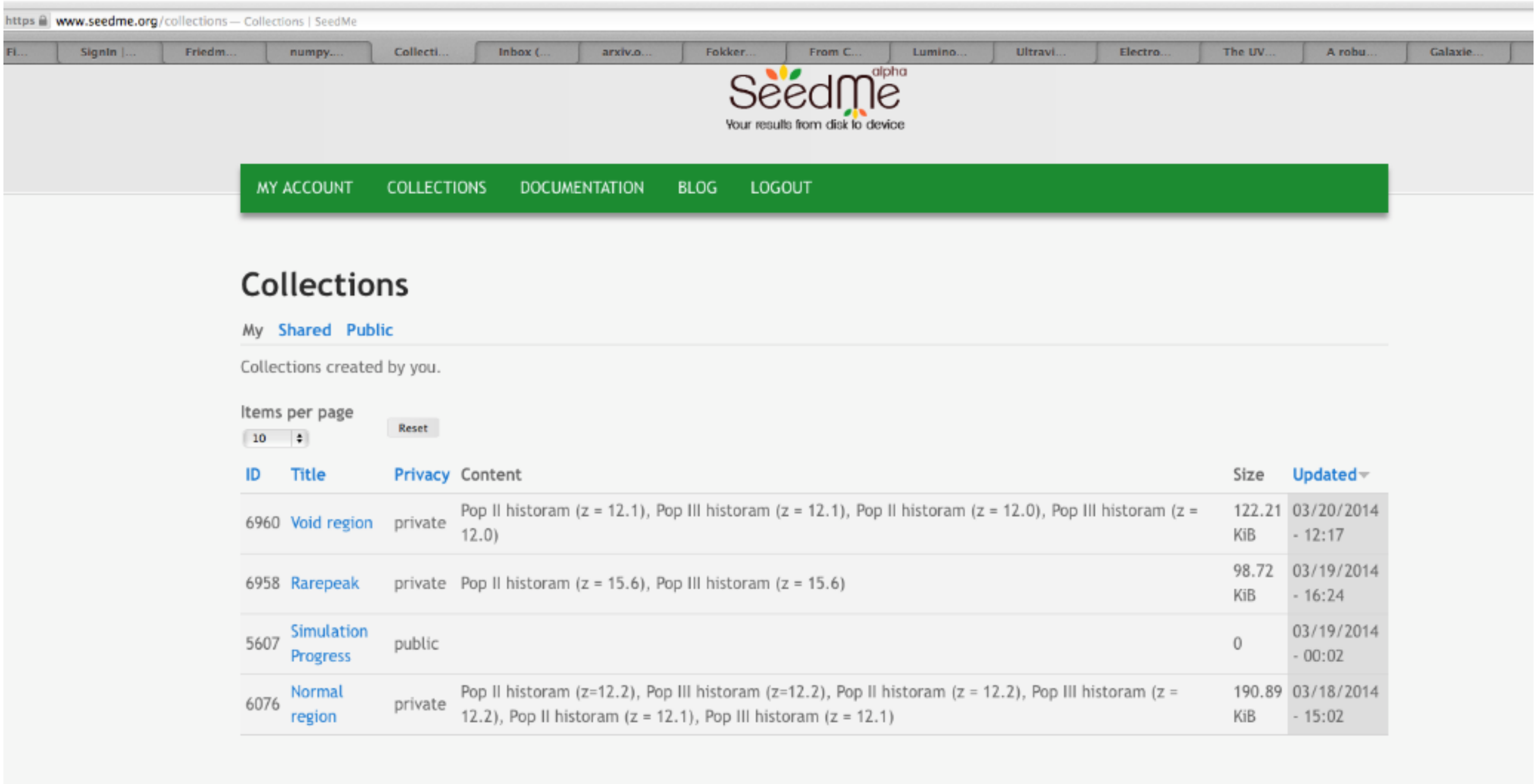
Simulation Progress

Collection ID	Updated
5607	03/19/2014 - 00:02

Ticker	Timestamp	Interval
Void run (no LW background) reached z= 8.3.	3/19/14 - 0:19:47	10:7:10
Rarepeak (with LW background) reached z=15.6.	3/15/14 - 13:15:37	3:32:28
Normal region reached z=12.1.	3/15/14 - 10:15:09	7:34:50
Void run (no LW background) reached z= 8.4.	3/11/14 - 2:11:19	14:10:49
Rarepeak (with LW background) reached z=15.7.	3/10/14 - 12:10:30	0:7:58

Last 5 tickers, see all in Meta data.

Ticker	Timestamp	Interval
Void run (no LW background) reached z= 8.3.	3/19/14 - 0:19:47	10:7:10
Rarepeak (with LW background) reached z=15.6.	3/15/14 - 13:15:37	3:32:28



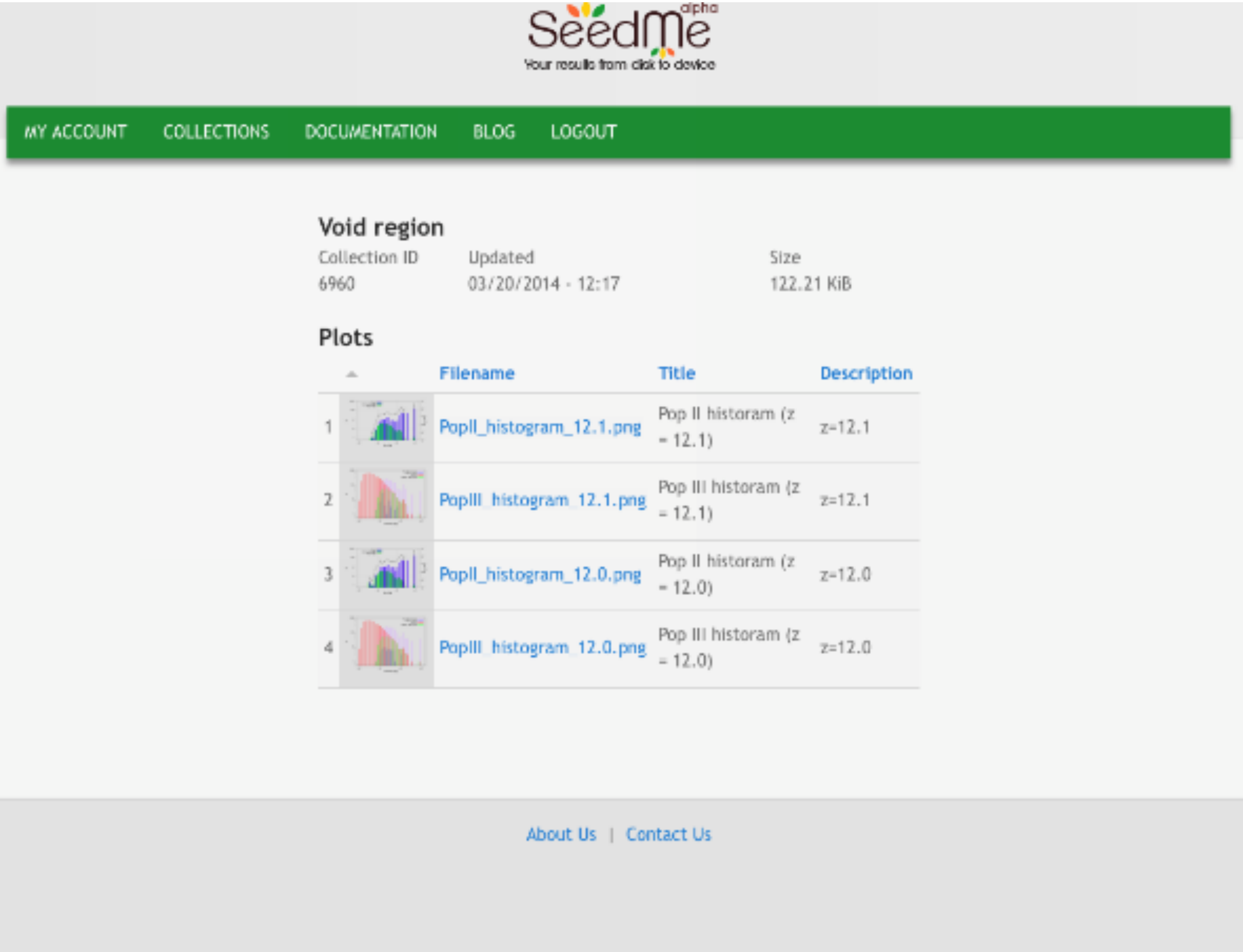
Collections

My Shared Public

Collections created by you.

Items per page 10 Reset

ID	Title	Privacy	Content	Size	Updated
6960	Void region	private	Pop II historam (z = 12.1), Pop III historam (z = 12.1), Pop II historam (z = 12.0), Pop III historam (z = 12.0)	122.21 KiB	03/20/2014 - 12:17
6958	Rarepeak	private	Pop II historam (z = 15.6), Pop III historam (z = 15.6)	98.72 KiB	03/19/2014 - 16:24
5607	Simulation Progress	public		0	03/19/2014 - 00:02
6076	Normal region	private	Pop II historam (z=12.2), Pop III historam (z=12.2), Pop II historam (z = 12.2), Pop III historam (z = 12.2), Pop II historam (z = 12.1), Pop III historam (z = 12.1)	190.89 KiB	03/18/2014 - 15:02

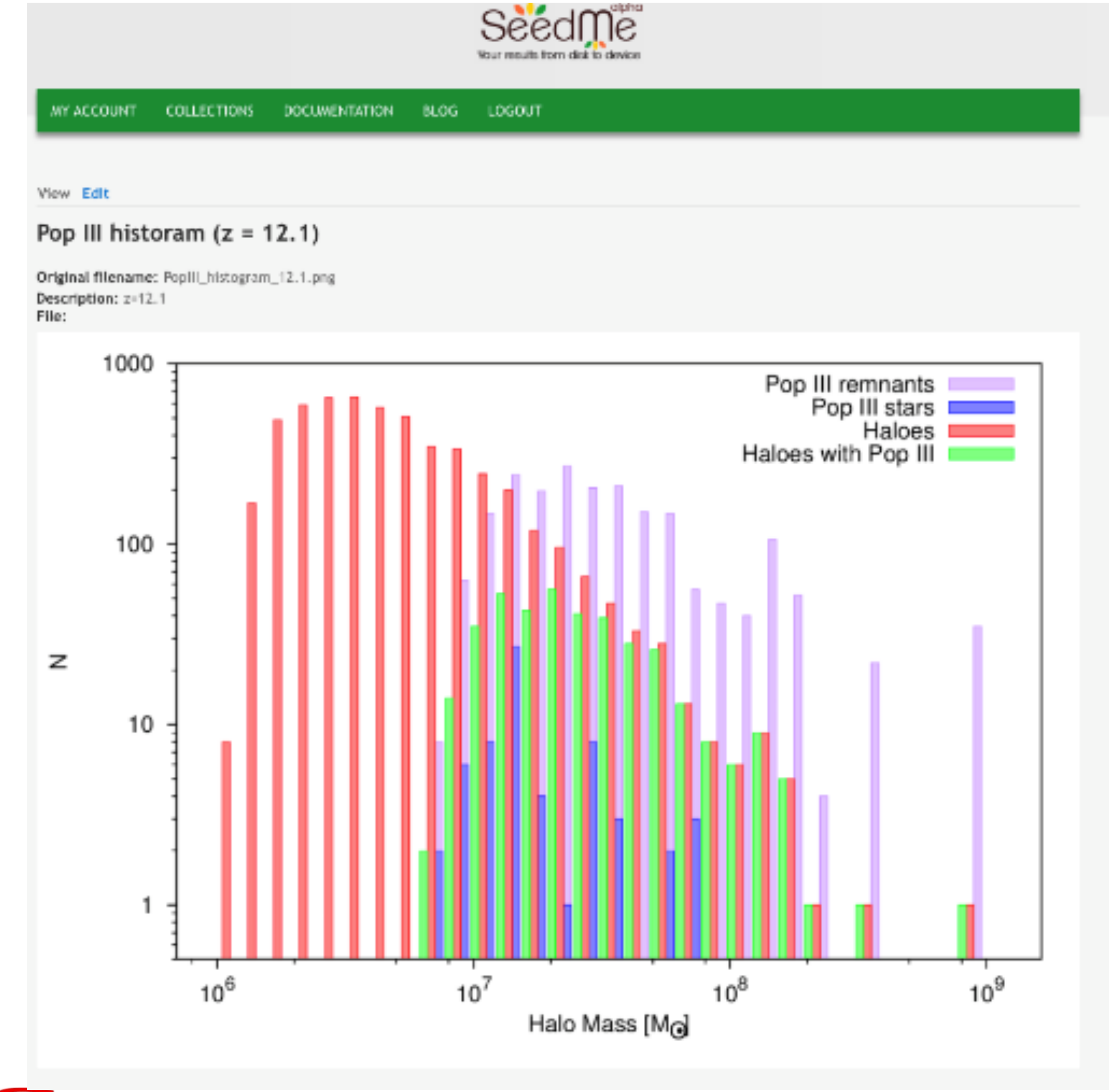


Void region

Collection ID	Updated	Size
6960	03/20/2014 - 12:17	122.21 KiB

Plots

		Filename	Title	Description
1		PopII_histogram_12.1.png	Pop II historam (z = 12.1)	z=12.1
2		PopIII_histogram_12.1.png	Pop III historam (z = 12.1)	z=12.1
3		PopII_histogram_12.0.png	Pop II historam (z = 12.0)	z=12.0
4		PopIII_histogram_12.0.png	Pop III historam (z = 12.0)	z=12.0



Pop III historam (z = 12.1)

Original filename: PopIII_histogram_12.1.png
Description: z=12.1
File: