Louisiana's Cyberinfrastructure
to Support Higher Education Research & Education

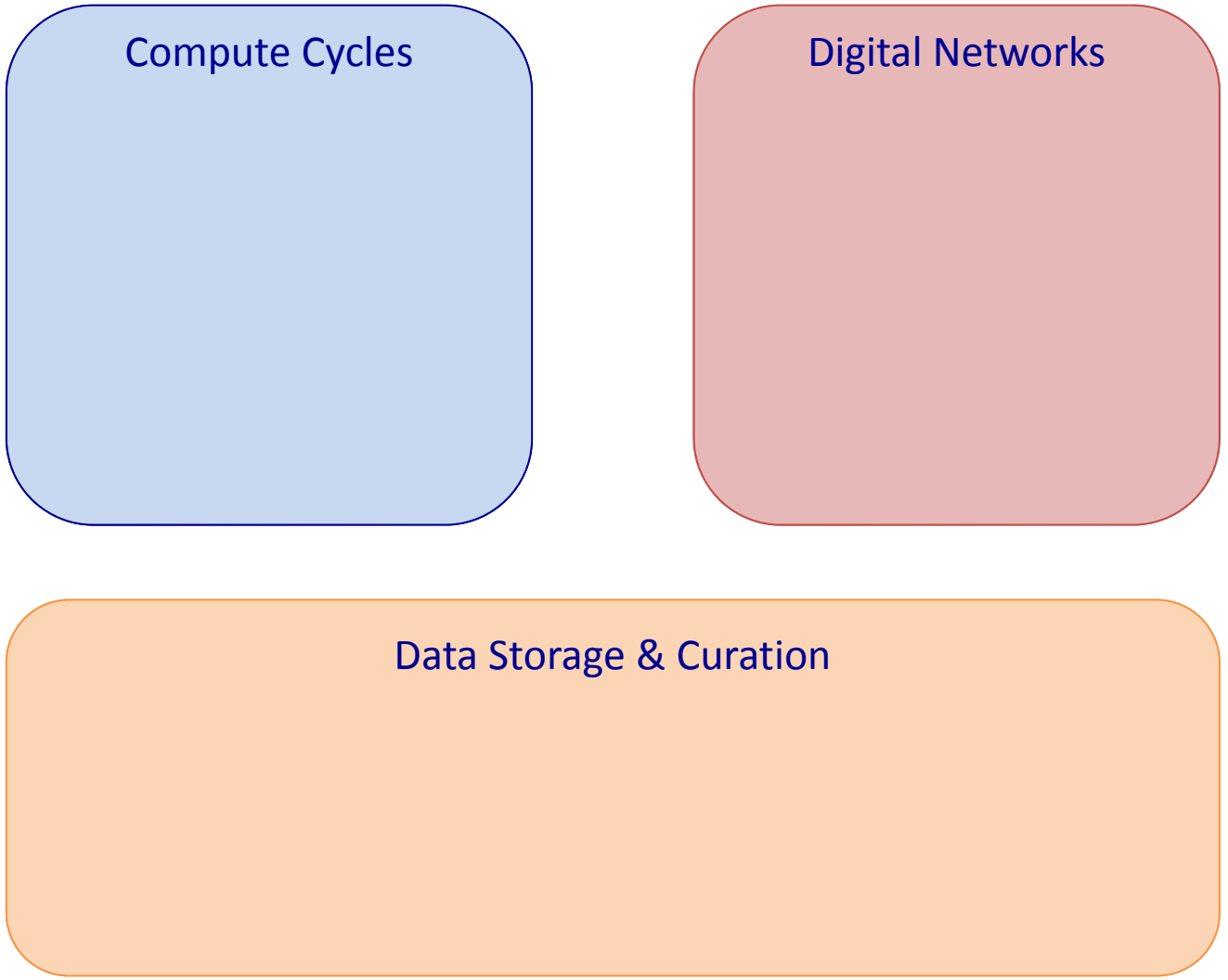
Joel E. Tohline

Chair, LONI Management Council

Director, LSU Center for Computation & Technology

8 October 2013

CyberInfrastructure of Vibrant Research Universities



Louisiana's CyberInfrastructure (status in October, 2013)

Compute Cycles

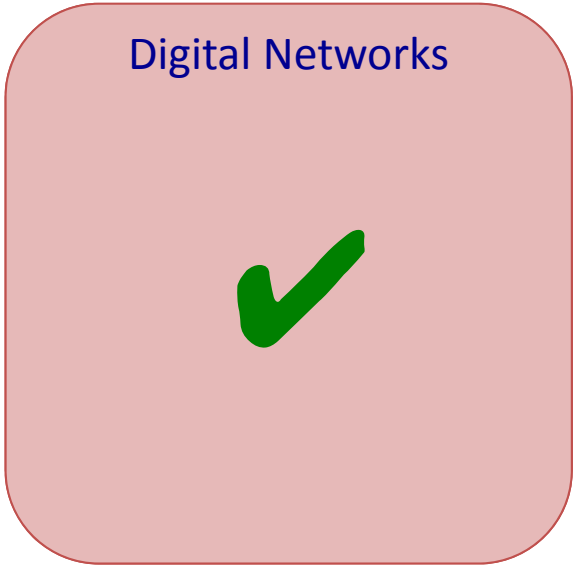
Digital Networks

Data Storage & Curation

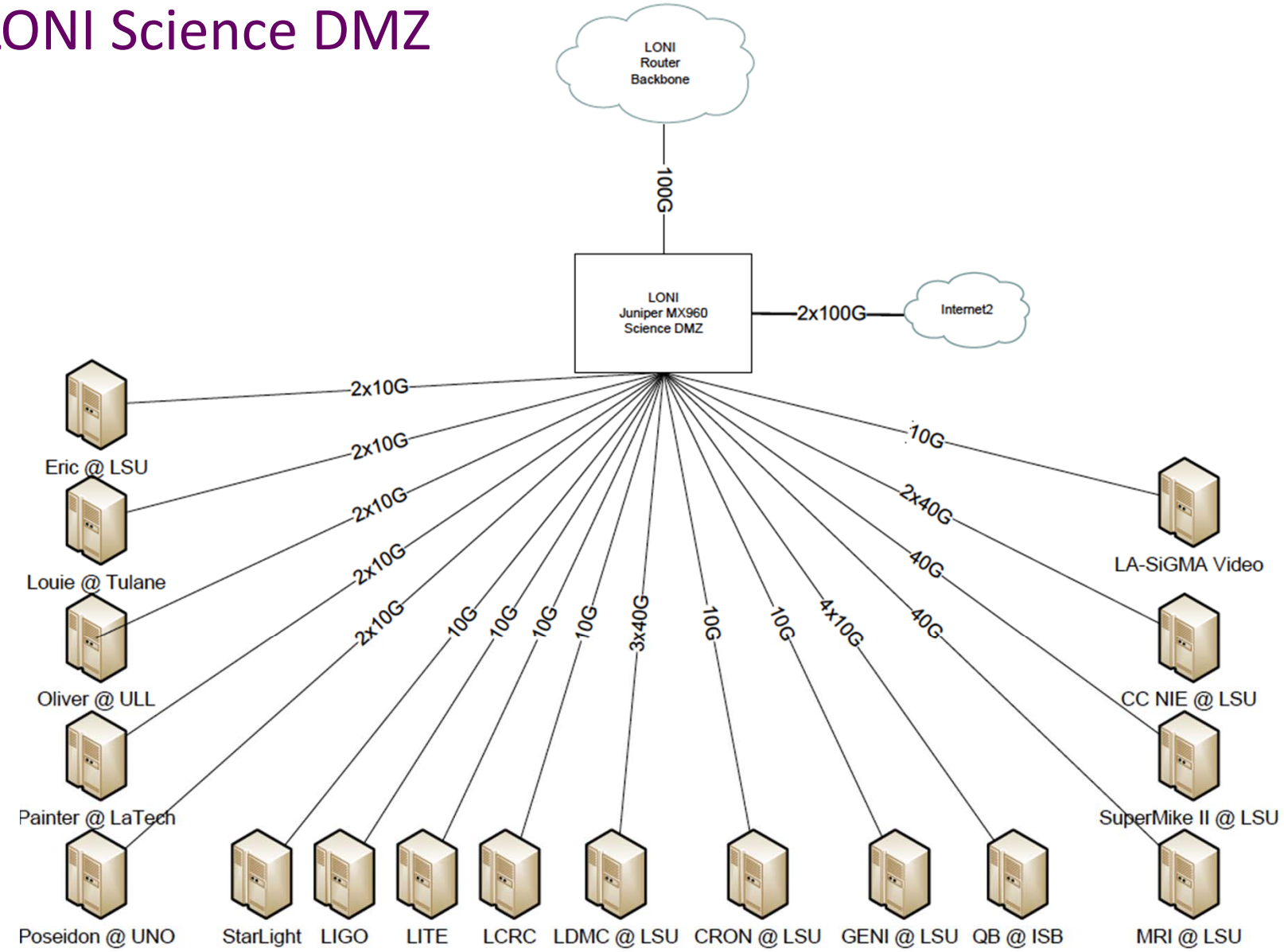
under construction

Louisiana's CyberInfrastructure (status in October, 2013)

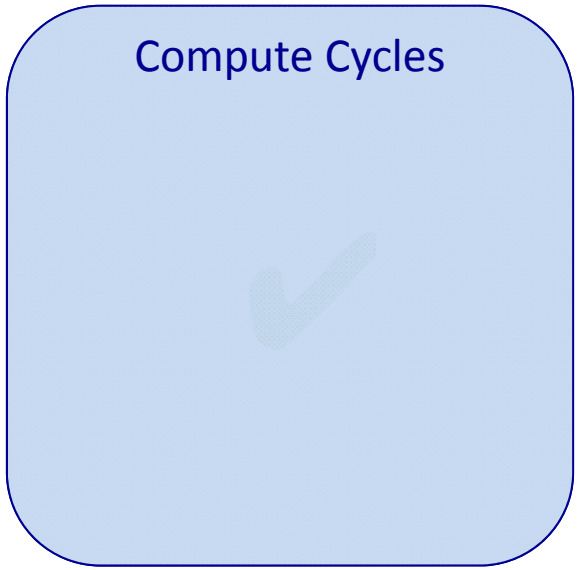
- ✓ LONI (Louisiana Optical Network Initiative): Higher education's 2004-2014 state-funded 40-Gbps optical network
- ✓ Xavier: $1.2M in 2010 NSF/RII-C2 funding; 10-Gbps connectivity into Xavier University
- ✓ CADIS: $0.5M in 2012 NSF/CC-NIE funding; provides 100-Gbps connectivity between HPC systems and new DMC building
- ✓ Big Data: $0.95M in 2013 NSF/CC-NIE funding; modest-scale, distributed GENI-rack environment
- ✓ LONI Science DMZ implementation



LONI Science DMZ



Louisiana's CyberInfrastructure (status in October 2013)



- ✓ SuperMike-II: $2.6M in LSU funding; installed in fall 2012
- ✓ Shelob: $0.54M in 2012 NSF/CNS funding; a GPU-loaded, heterogeneous, computing platform
- ✓ Melete: $0.9M in NSF/CNS/MRI funding; an interaction-oriented, software-rich cluster w/ tangible interface support
- ✓ SuperMIC: $4M in 2013 NSF/ACI/MRI funding; 1 Pflops system shared with national XSEDE community
- ✓ Commitment of funding from Board of Regents to replace the flagship HPC system (Queen Bee) of LONI

Louisiana's
High-Performance Computing (HPC) Systems

According to Operating System (OS)

Linux Clusters

- LSU's HPC
  - SuperMIC (1050 TF)

    NEW! Acquisition in progress
  - SuperMike-II (220 TF)
  - Shelob (95 TF)
  - Tezpur (15.3 TF)
  - Philip (3.5 TF)
- LONI
  - Queen Bee (50.7 TF)
  - Five (@ 4.8 TF)

AIX Clusters

- LSU's HPC
  - Pandora (IBM P7; 6.8 TF)
  - Pelican (IBM P5+;1.9 TF)

    Decommissioned
- LONI
  - Five (IBM P5; @ 0.85 TF)

    Decommissioned

According to Operating System (OS)

Linux Clusters

- LSU's HPC
  - SuperMIC (1050 TF)

  NEW! Acquisition in progress
  - SuperMike-II (220 TF)
  - Shelob (95 TF)
  - Tezpur (15.3 TF)
  - Philip (3.5 TF)
- LONI
  - Queen Bee (50.7 TF)
  - Five (@ 4.8 TF)

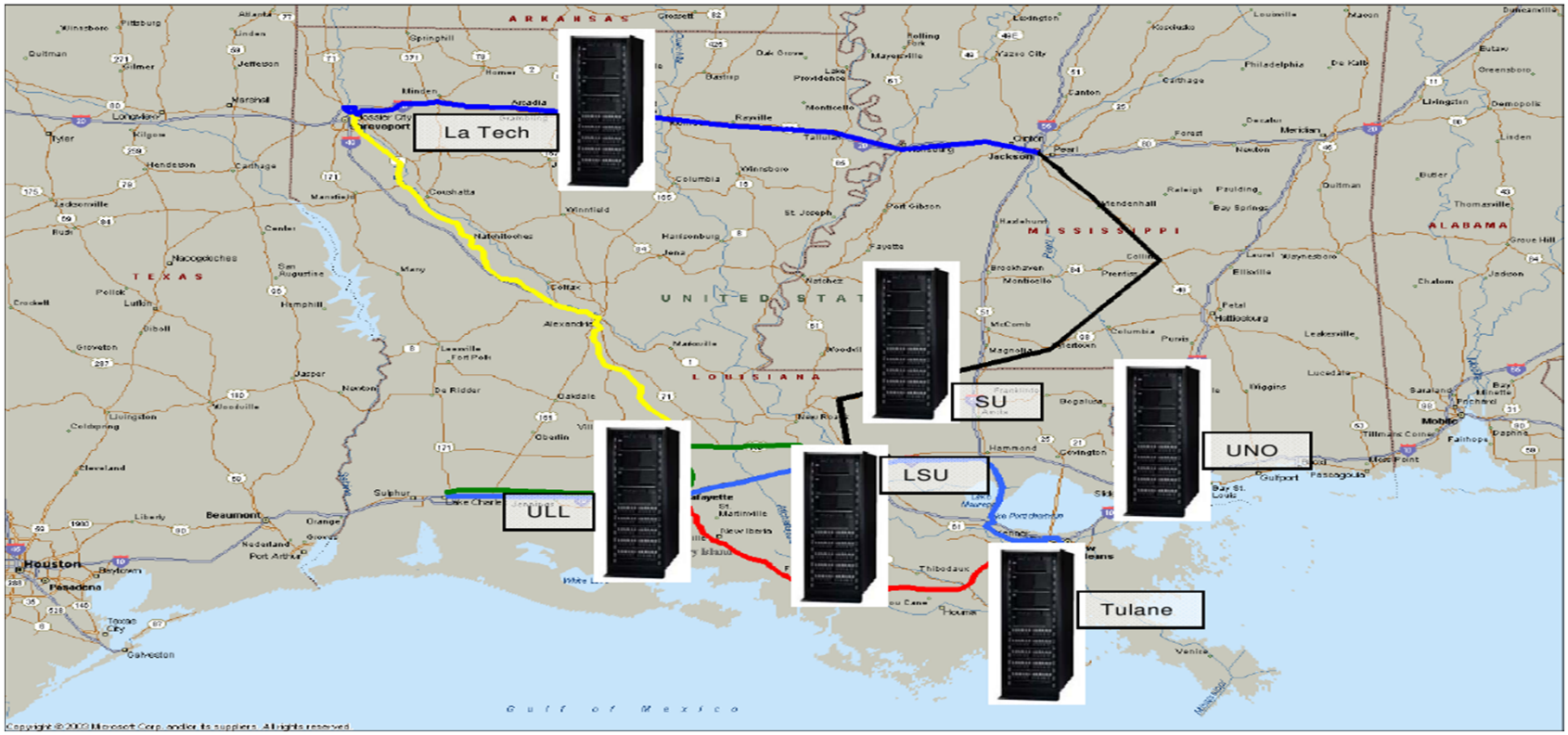
According to Operating System (OS)

Linux Clusters

- LSU's HPC
  - SuperMIC (1050 TF)

    NEW! Acquisition in progress
  - SuperMike-II (220 TF)
  - Shelob (95 TF)
  - Tezpur (15.3 TF)
  - Philip (3.5 TF)
- LONI
  - Queen Bee (50.7 TF)
  - Five (@ 4.8 TF)

LONI's High-Performance Computing Grid



- Originally, 6 clusters, located at six different campuses

LONI Dell Linux Clusters

- **Queen Bee**: 50 TFlops centerpiece
  - 23rd on the June 2007 Top 500 list
  - Hosted at the state's Information Systems Building
  - 668 nodes with 8 Intel Xeons cores @ 2.33 GHz, 8 GB RAM (i.e., 5344 cores)
  - 192TB DDN storage running Lustre
- **Five 5 TFlops clusters**
  - Online: Eric(LSU), Oliver(ULL), Louie(Tulane), Poseidon(UNO), Painter (LaTech)
  - 128 nodes with 4 Intel Xeons cores@ 2.33 Ghz, 4 GB RAM
  - 9TB DDN storage running Lustre each

LONI Dell Linux Clusters

- **Queen Bee**: 50 TFlops centerpiece
  - 23rd on the June 2007 Top 500 list
  - Hosted at the state's Information Systems Building
  - 668 nodes with 8 Intel Xeons cores @ 2.33 GHz, 8 GB RAM (i.e., 5344 cores)
  - 192TB DDN storage running Lustre
- **Five 5 TFlops clusters**
  - Online: Eric(LSU), Oliver(ULL), Louie(Tulane), Poseidon(UNO), Painter (LaTech)
  - 128 nodes with 4 Intel Xeons cores@ 2.33 Ghz, 4 GB RAM
  - 9TB DDN storage running Lustre each

According to Operating System (OS)

Linux Clusters

- LSU's HPC
  - SuperMIC (1050 TF)

  NEW! Acquisition in progress
  - SuperMike-II (220 TF)
  - Shelob (95 TF)
  - Tezpur (15.3 TF)
  - Philip (3.5 TF)
- LONI
  - Queen Bee (50.7 TF)
  - Five (@ 4.8 TF)

According to Operating System (OS)

Linux Clusters

- LSU's HPC
  - SuperMIC (1050 TF)

    NEW! Acquisition in progress
  - SuperMike-II (220 TF)
  - Shelob (95 TF)
  - Tezpur (15.3 TF)
  - Philip (3.5 TF)
- LONI
  - Queen Bee (50.7 TF)
  - Five (@ 4.8 TF)

LSU's Linux HPC Systems

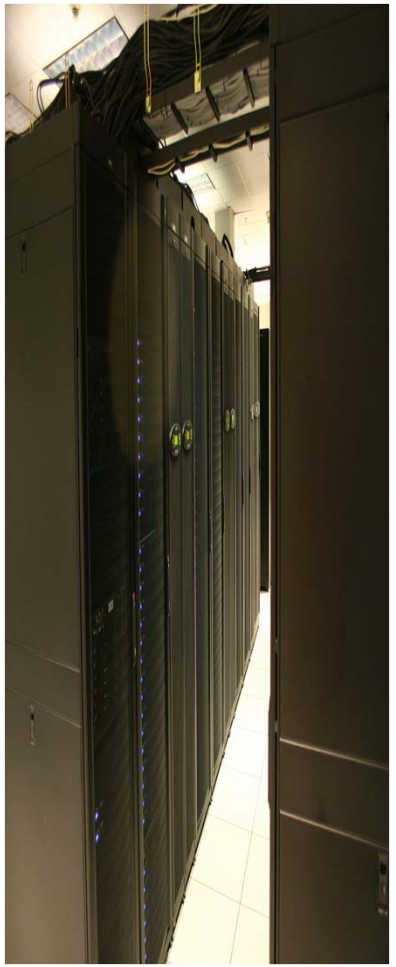
- **SuperMike-II** (mike.hpc.lsu.edu)
  - **380 compute nodes**: 16 Intel Sandy Bridge cores @ 2.6GHz, 32GB RAM, 500GB HD, 40Gb/s infiniband, 2x 1Gb/s Ethernet
  - **52 GPU compute nodes**: 16 Intel Sandy Bridge cores @ 2.6GHz, 2 NVIDIA M2090 GPUs, 64GB RAM, 500GB HD, 40Gb/s infiniband, 2x 1Gb/s Ethernet
  - **8 fat compute nodes**: 16 Intel Sandy Bridge cores @ 2.6GHz, 256 GB RAM, 500GB HD, 40Gb/s infiniband, 2x 1Gb/s Ethernet, Aggregated together by ScaleMP to one big SMP node
  - **3 head nodes**: 16 Intel Sandy Bridge cores @ 2.6GHz, 64 GB RAM, 2 x 500GB HD, 40Gb/s infiniband, 2x 10Gb/s
  - **1500TB** (scratch + long term) DDN Luster storage

LSU's Linux HPC Systems

- **SuperMIC** (mic.hpc.lsu.edu) Acquisition in progress!
  - The largest NSF MRI award LSU has ever received ($3,924,181 with $1,681,792 LSU match)
  - **360 compute nodes**: 20 Intel Ivy Bridge cores @ 2.8GHz, 2 Intel Xeon Phi with 61 cores @ 1.2 GHz, 64GB RAM, 500GB HD, 56Gb/s infiniband, 2x 1Gb/s Ethernet
  - **20 hybrid compute nodes**: 20 Intel Ivy Bridge cores @ 2.8GHz, 1 Intel Xeon Phi with 61 cores @ 1.2 GHz, 1 NVIDIA K20X GPU @1.3 TF, 64GB RAM, 500GB HD, 56Gb/s infiniband, 2x 1Gb/s Ethernet
  - **1 Phi head node**: 20 Intel Ivy Bridge cores @ 2.8GHz, 2 Intel Xeon Phi with 61 cores @ 1.2 GHz, 128 GB RAM, 2x 1TB HD, 56Gb/s infiniband, 2x 10Gb/s Ethernet
  - **1 GPU head node**: 20 Intel Ivy Bridge cores @ 2.8GHz, 2 NVIDIA K20X GPU @1.3 TF, 128 GB RAM, 2x 1TB HD, 56Gb/s infiniband, 2x 10Gb/s Ethernet
  - **1000TB** (scratch) DDN Luster storage

Queen Bee Replacement

- Expect a **~1.5-2 Pflops** system
- Funded by Board of Regents via LONI
- Work with vendors and researchers to configure and design the system
- We expect to acquire and install the machine this fiscal year



What System Architecture Should We Pursue?

National HPC Resources

- **NSF XSEDE**
  - Intel- or AMD-based commodity cluster
    - Tightly coupled (Kraken, Lonestar4, Gorden, Trestles, etc.)
    - Loosely coupled (Condor pool from the Open Science Grid)
  - SGI Altix SMP (Blacklight)
  - Xeon Phi acceleration (Stampede, SuperMIC)
  - NVIDIA GPU acceleration (Keeneland, Nautilus, etc.)
- **NSF Blue Waters** (Cray installation w/ NVIDIA GPU acceleration)
- **DOE INCITE**
  - Cray XT5 AMD-based with NVIDIA GPU (Titan)
  - IBM Blue Gene/Q (Mira)

Scenario #1

- **2x of SuperMIC: 720** Intel Ivy Bridge 10-core 2.8GHz or 12-core 2.4 GHz + Intel Xeon Phi 7120P @1.2 TF) nodes, 40 GPU or (GPU + Phi) nodes, 2 PFlops, $10M

Scenario #2

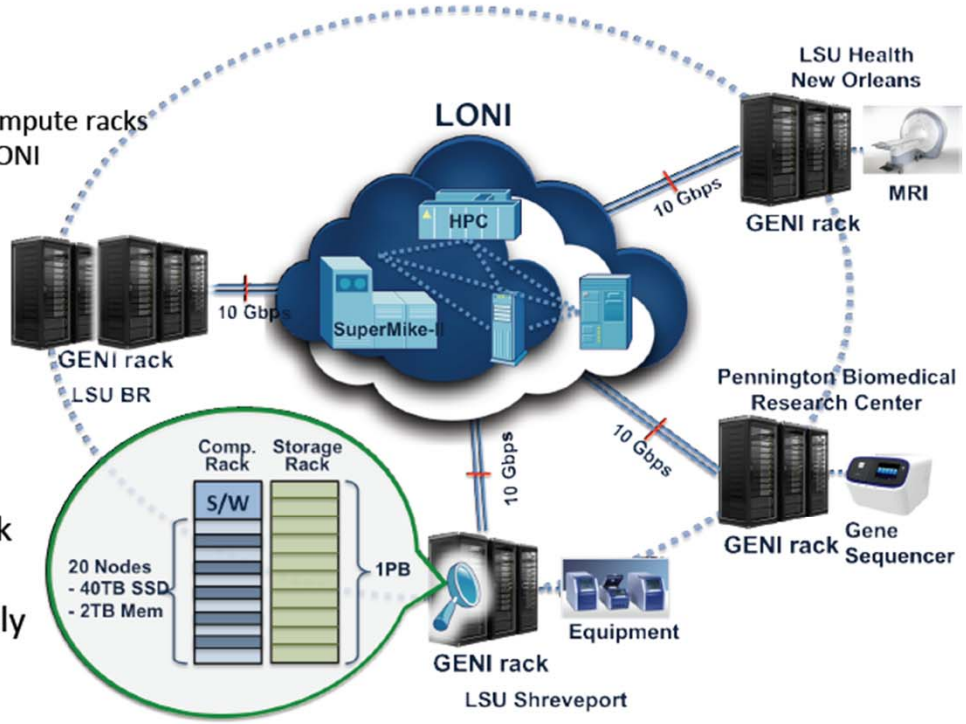
- Homogeneous CPU cluster: Intel Ivy Bridage 10-core 2.8GHz or 12-core 2.4 GHz, 1800 nodes (36000 or 43200 cores), 800 TFlops, $10M

Scenario #3

- Cray cluster ($15-25M guess):
  - XC-30 CASCADE cluster: 4 Pflops, 2,264 nodes (1 10-core 2.5GHz Ivy Bridge + 1 Xeon Phi 5120P @1.1 TF), 32 GB memory per node
  - 8 fat nodes: 4 10-core Ivy Bridge, 1.5 TB memory
  - 32 visualization nodes: 2 10-core Ivy Bridge, 2 NVIDIA K20s, 2 Xeon Phi, 256 GB memory

Scenario #4 (distributed system)

- Each campus will have a cluster consisting of
  - Compute rack (20 nodes)
    - Each node has 1TB SSD, 2 CPU (8 cores), 128GB Memory
  - Storage rack
    - 1PB, Lustre file system storage servers
  - 10Gbps Openflow switch network
    - Transferring data between storage and compute racks
    - Transferring data between clusters over LONI
- Multiple clusters will be federated with GENI technology
  - Cloud computing
  - Openflow
- Application
  - Each research group stores big data at a local storage rack and process big data at a local compute rack
  - The stored big data can be transferred to other clusters through LONI efficiently
  - If large-scale resource is required, users can make a virtual slice federating multiple clusters over LONI



How Important Is …

- High bandwidth interconnectivity of compute nodes?
- Accelerator hardware?
- Memory per core?
- Disk storage relative to DRAM?
- SSDs versus spinning disks?
- Distributed computing?
- Distributed storage?
- Other?

Discuss!

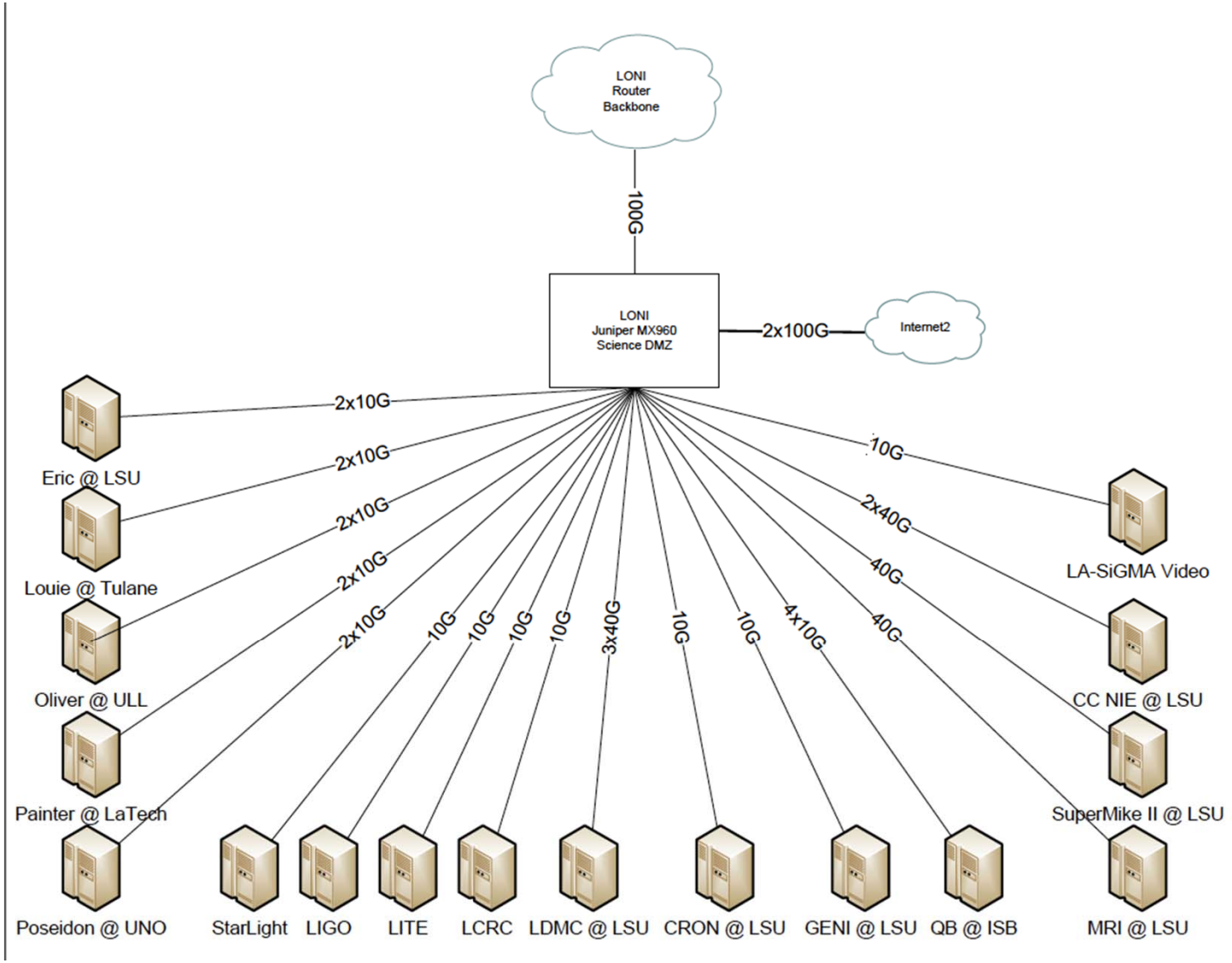
Queen Bee Replacement
(discussion 10/7/2013)

- $6.75 M computing hardware + $1.25 M infrastructure (including storage)
- Architecture very similar to SuperMIC, but 50% larger ➔ 1.5 Pflops (i.e., approx. 570 nodes)
  - 555 nodes, 2 Xeon-phi (8 Gbytes/card)
  - 15 FAT nodes (256 Gbytes/node), 2 NVIDIA (6 Gbytes/card)
  - 64 Gbytes/20 core
  - SSDs???

Queen Bee Replacement
(discussion 10/7/2013)

- $7.88 M computing hardware + infrastructure via other sources
- Architecture very similar to SuperMIC, but 75% larger ➔ 1.75 Pflops (i.e., approx. 665 nodes)
  - 645 nodes, 2 Xeon-phi
  - 20 FAT (256 Gbytes/node), 2 NVIDIA
  - 64 Gbytes/20 core
  - SSDs???

LONI “Science DMZ”



Now, let’s talk about data storage and management …

CyberInfrastructure of a Vibrant Research University

How does a university's strategy for handling the storage and curation of digital data relate to NSF's "Data Management" mandate?

Data Storage & Curation

CyberInfrastructure of a Vibrant Research University

How does a university's strategy for handling the storage and curation of digital data relate to NSF's "Data Management" mandate?

NSF Data Management Plans may include (taken from the *NSF Grant Proposal Guide*):

- Types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
- Standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
- Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- Policies and provisions for re-use, re-distribution, and the production of derivatives;
- Plans for archiving data, samples, and other research products, and for preservation of access to them.

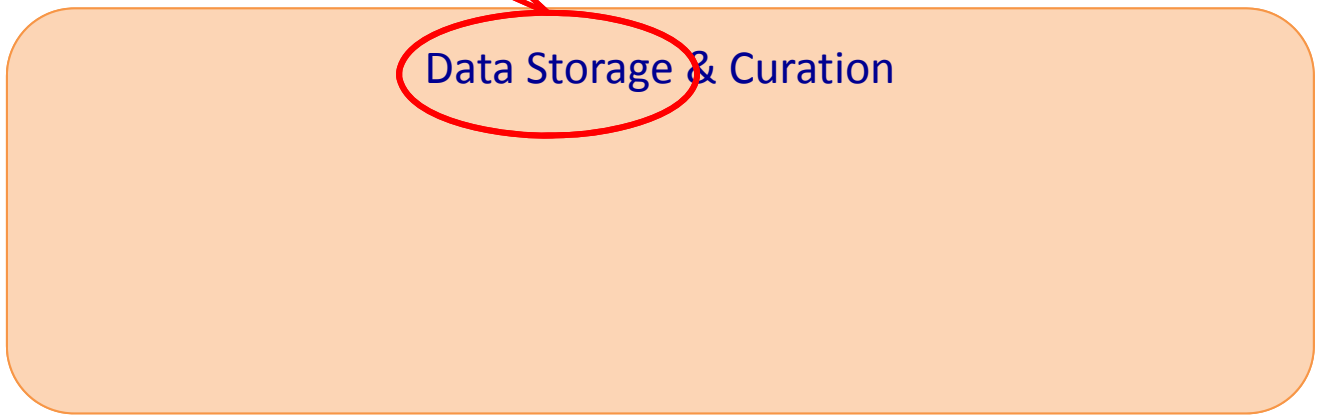
Data Storage & Curation

CyberInfrastructure of a Vibrant Research University

How does a university's strategy for handling the storage and curation of digital data relate to NSF's "Data Management" mandate?

NSF Data Management Plans may include (taken from the *NSF Grant Proposal Guide*):

- Types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
- Standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
- Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- Policies and provisions for re-use, re-distribution, and the production of derivatives;
- Plans for archiving data, samples, and other research products, and for preservation of access to them.

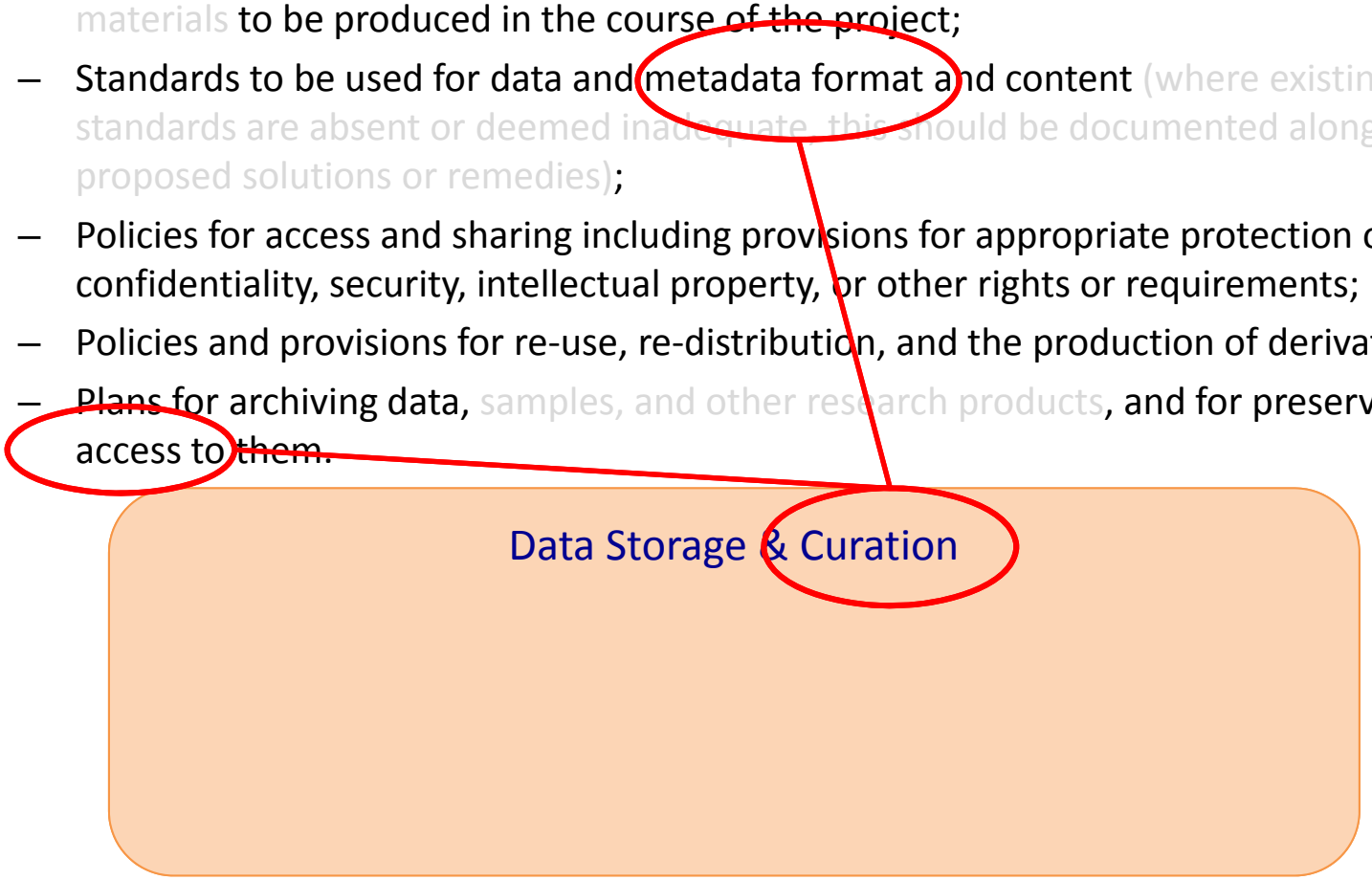


CyberInfrastructure of a Vibrant Research University

How does a university's strategy for handling the storage and curation of digital data relate to NSF's "Data Management" mandate?

NSF Data Management Plans may include (taken from the *NSF Grant Proposal Guide*):

- Types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
- Standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
- Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- Policies and provisions for re-use, re-distribution, and the production of derivatives;
- Plans for archiving data, samples, and other research products, and for preservation of access to them.



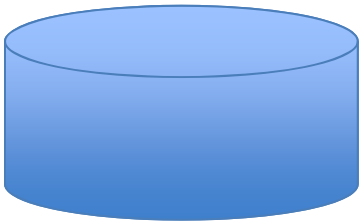
LSU's CyberInfrastructure (status in early 2013)

- We've been discussing a 3-tiered approach to storage where a limited amount of space is associated with each storage tier.
  - Tier 1: Disk storage with fast read/write access to support high-performance computing (HPC) activities; this is not intended to serve as long-term, archival storage.
  - Tier 2: Easy to use/access disk storage available to all faculty research projects; university ensures long-term viability (permanency?) of this data; a rich software interface supports data management and curation.
  - Tier 3: Very inexpensive (and, likely, very slow access) long-term storage; may be viewed as extension of tier-1 and tier-2, but without rich curation features.

Data Storage & Curation

Tier 1
[HPC users]

Default:
20 Tbytes/group
(at no direct cost)

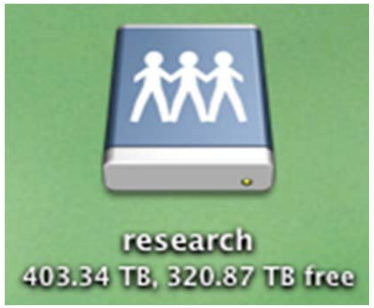


Expandable @ a cost of
$125/TeraByte/year

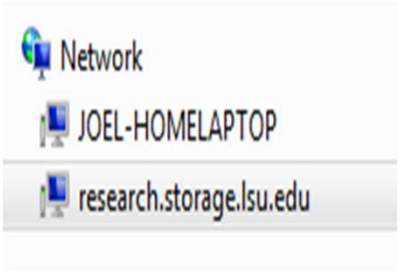
Tier 2
[1000 Faculty]

Default:
100 Gbytes/faculty
(at no direct cost)

Mac



PC



Expandable @ a cost of
$0.07/Gbyte/month
= $840/TeraByte/year

Tier 3
[1000 Faculty]

Default:
250 Gbytes/faculty
(at no direct cost)

Expandable @ a cost of
$0.06/Gbyte/month
= $720/TeraByte/year

Curation Layer

- Only provided with Tier 2a (or only Tier 2b?) or w/ Tier 3 also.
- Initial curation layer will be BE Press (?), managed by LSU Library faculty/staff
  - Server managed by library staff
  - Each entry will point back to file stored on LSU's Tier 2a system; this will avoid the need to purchase/maintain double the storage
- Long term curation solution?

Storage Options (2011)

- 200TB DataDirect Network (DDN) high performance storage running Lustre with IB connection at $300-320K
  - **Roughly $1.5-1.6K/TB**
- 200TB Panasas high performance storage with IB connection at $320-340K
  - **Roughly $1.6-1.7K/TB**
- 200TB Fusion Manager File system with IB connection at $500-600K
  - **Roughly $2.5-3K/TB**
- 192TB IBM GPFS storage with IB connection at $150-200K
  - **Roughly $0.78-1.04K/TB**