Distributed Data Sharing with Petashare for Collaborative Research
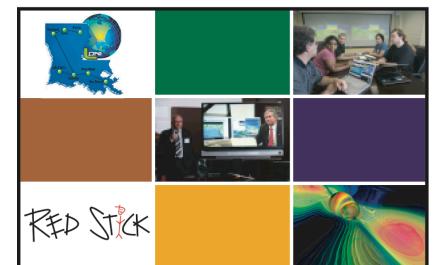
Sreekanth Pothanis
Dr. Tevfik Kosar
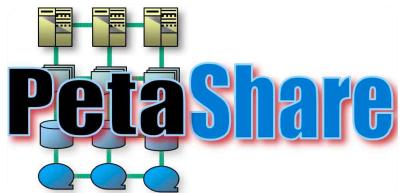
Center for Computation & Technology
Louisiana State University

March 25, 2010
iRODS User Meeting

CCT: Center for Computation & Technology @ LSU

LONI brings:
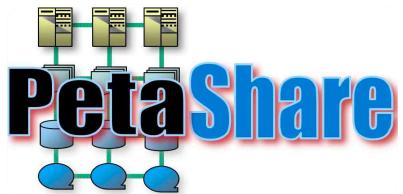
- + fat pipes (40Gb/s) and vast computing resources (100 Tflops)

- – missing a distributed data management and storage infrastructure

Our Goals:
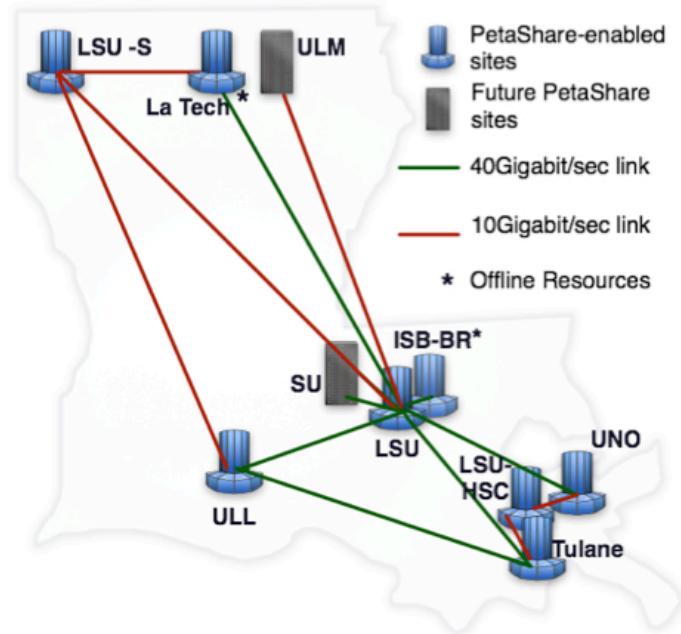
- Bring additional storage to LONI

- Provide a CI for easy and efficient storage, access, and management of data.
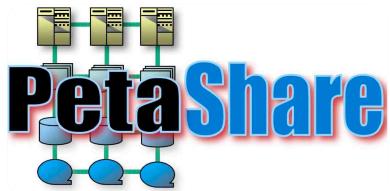
  "Let scientists focus on their science rather than dealing with low level data issues. The CI should take care of that."
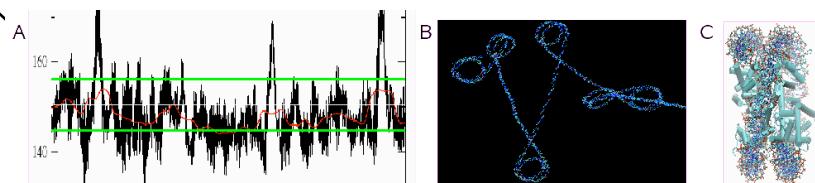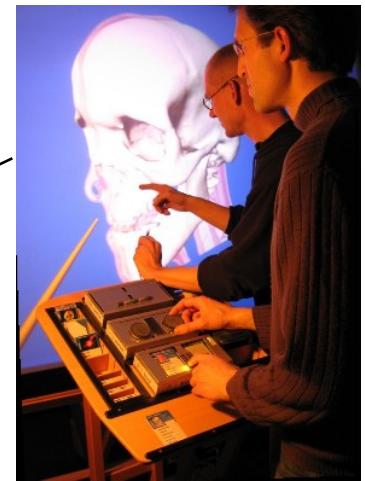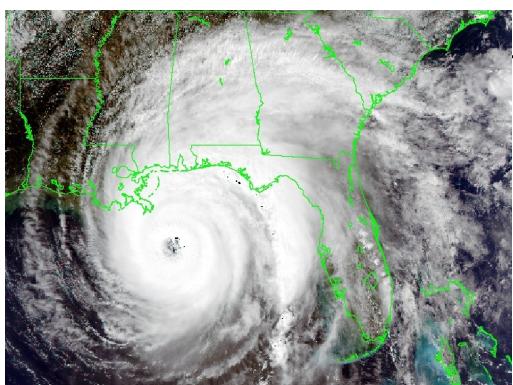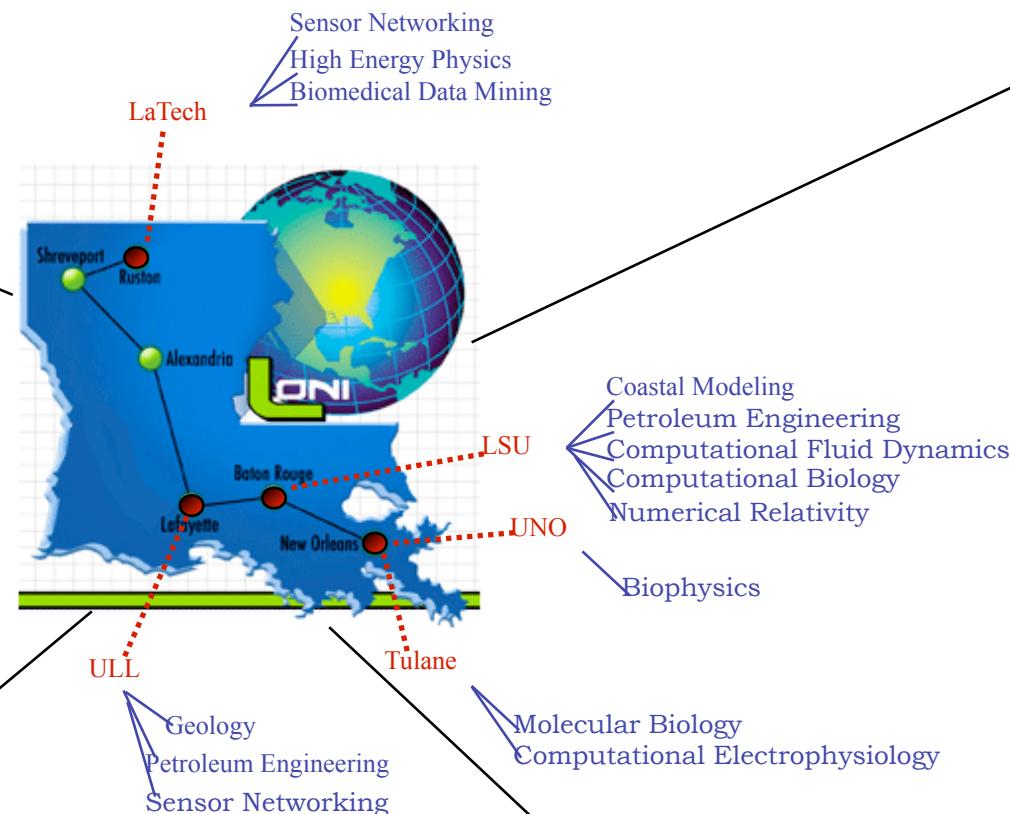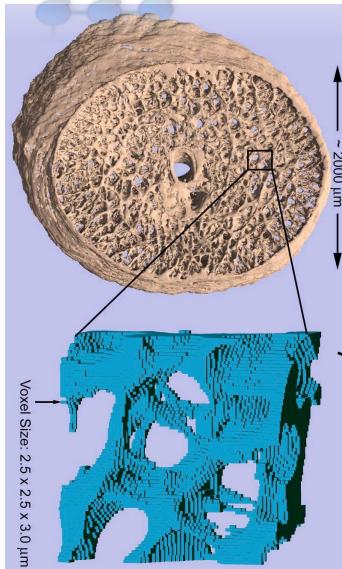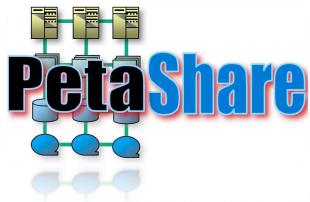
- **Goal:** Enable domain scientists to focus on their primary research problem, assured that the underlying infrastructure will manage the low-level data handling issues.

- **Novel approach:** Treat data storage resources and the tasks related to data access as first class entities just like computational resources and compute tasks.

- **Key technologies** being developed: Data-aware storage systems, data-aware schedulers (i.e. Stork), and cross-domain meta-data scheme.

- **Provides** an additional 300TB disk, and 400TB tape storage (and access to national storage facilities)

CCT: Center for Computation & Technology @ LSU
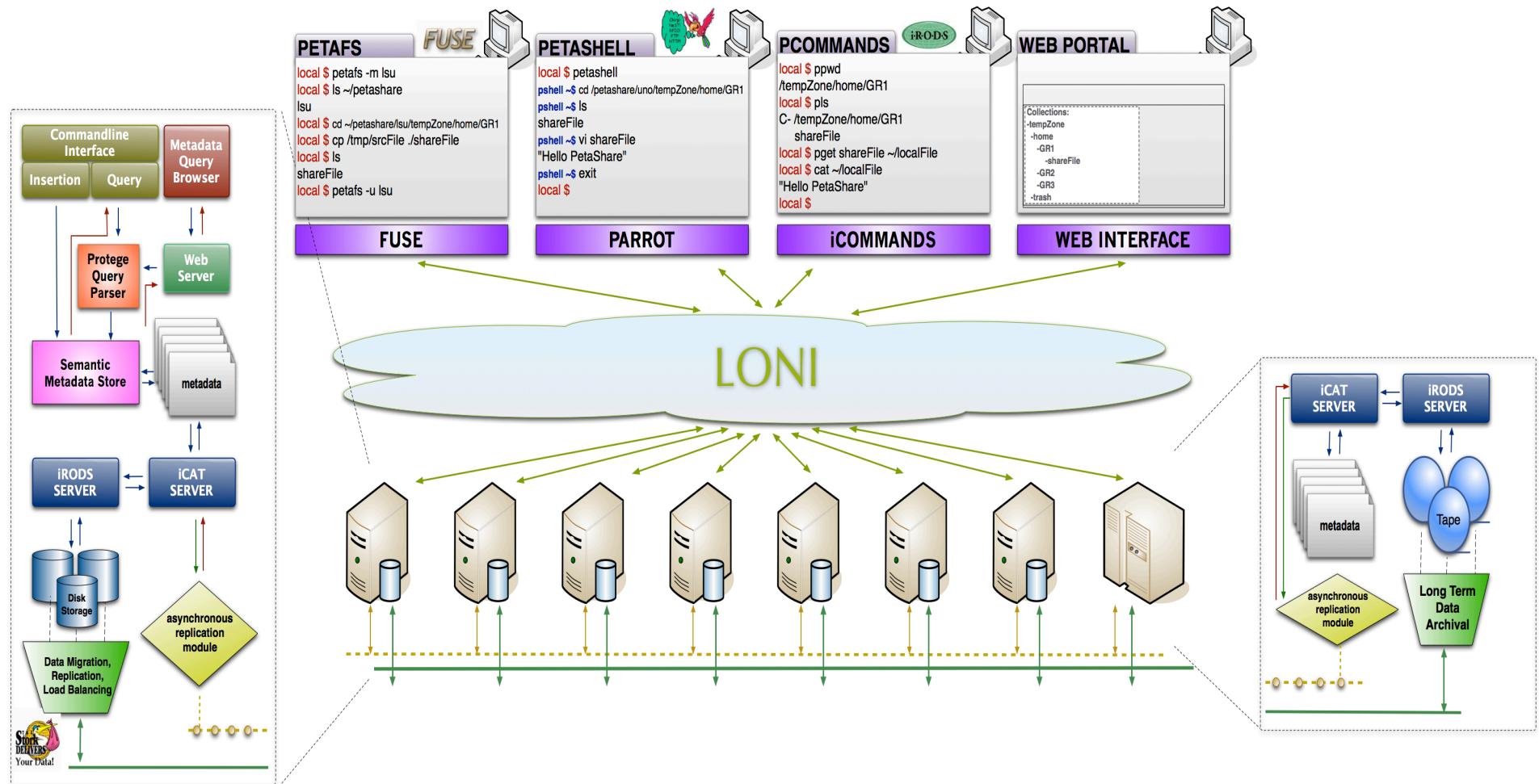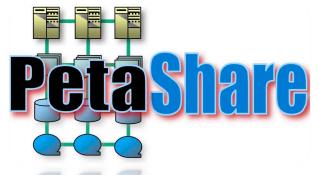
Participating institutions in the PetaShare project, connected through LONI. Sample research of the participating researchers pictured (i.e. biomechanics by Kodiyalam & Wischusen, tangible interaction by Ullmer, coastal studies by Walker, and molecular biology by Bishop).

Sensor Networking
High Energy Physics
Biomedical Data Mining

LaTech

Coastal Modeling
Petroleum Engineering
Computational Fluid Dynamics
Computational Biology
Numerical Relativity

LSU

UNO

Biophysics

ULL

Geology
Petroleum Engineering
Sensor Networking

Tulane

Molecular Biology
Computational Electrophysiology

CCT: Center for Computation & Technology @ LSU

Infrastructure Overview

PETAFS — FUSE
```
local $ petafs -m lsu
local $ ls ~/petashare
lsu
local $ cd ~/petashare/lsu/tempZone/home/GR1
local $ cp /tmp/srcFile ./shareFile
local $ ls
shareFile
local $ petafs -u lsu
```

PETASHELL — PARROT
```
local $ petashell
pshell ~$ cd /petashare/uno/tempZone/home/GR1
pshell ~$ ls
shareFile
pshell ~$ vi shareFile
"Hello PetaShare"
pshell ~$ exit
local $
```

PCOMMANDS — iCOMMANDS
```
local $ ppwd
/tempZone/home/GR1
local $ pls
C- /tempZone/home/GR1
    shareFile
local $ pget shareFile ~/localFile
local $ cat ~/localFile
"Hello PetaShare"
local $
```

WEB PORTAL — WEB INTERFACE
```
Collections:
-tempZone
 -home
  -GR1
    -shareFile
  -GR2
  -GR3
 -trash
```

LONI

Replica & Meta
Data Management

Caching/
Prefetching

Data
Migration

HSM

Scheduled Data
Movement

Teragrid

PetaShare

The Stork DELIVERS Your Data!

A system driven by the local needs (in LA), but has potential to be a generic solution for the broader community!

For more information on **PetaShare**:
http://www.petashare.org

CCT: Center for Computation & Technology @ LSU