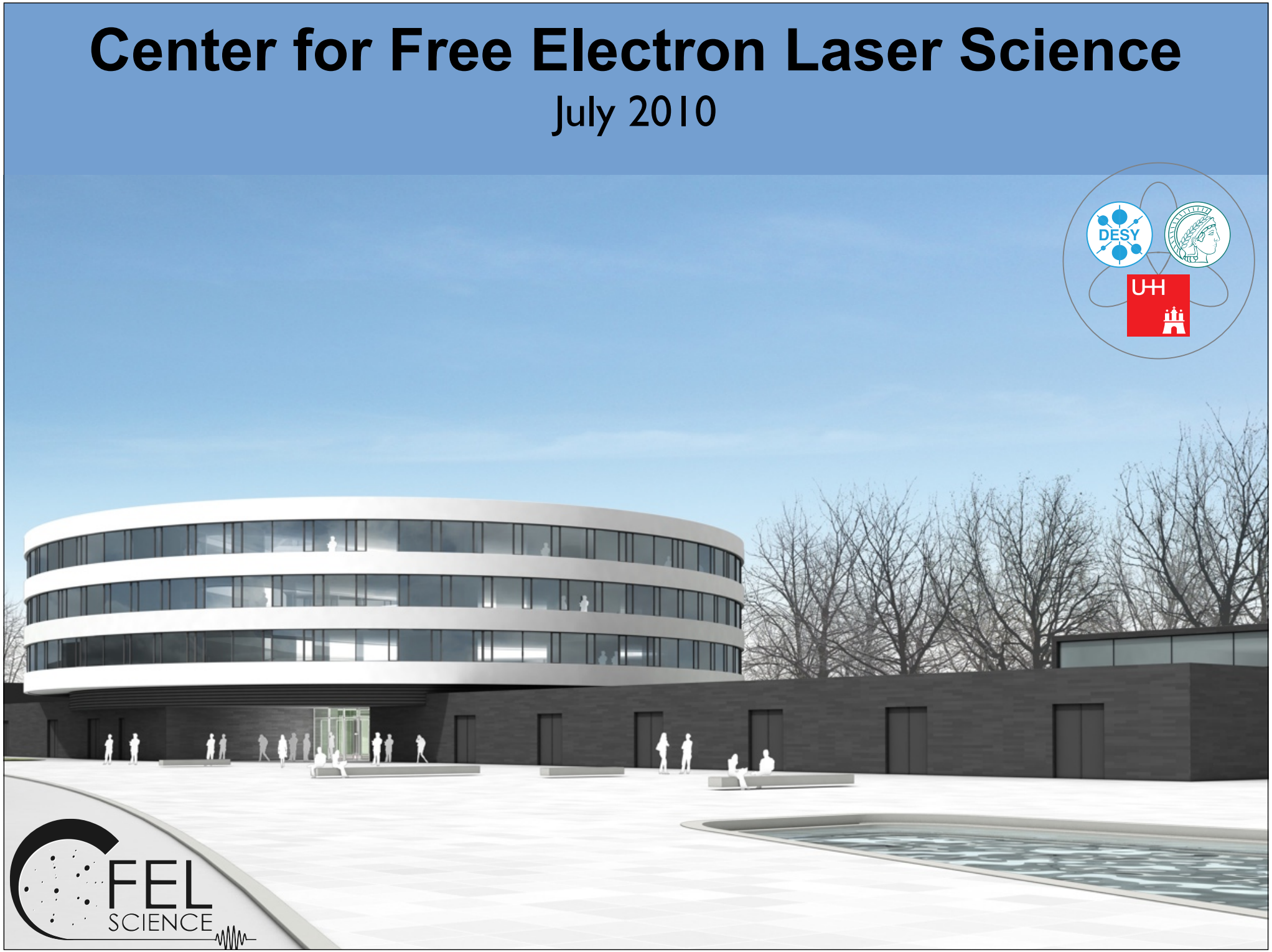
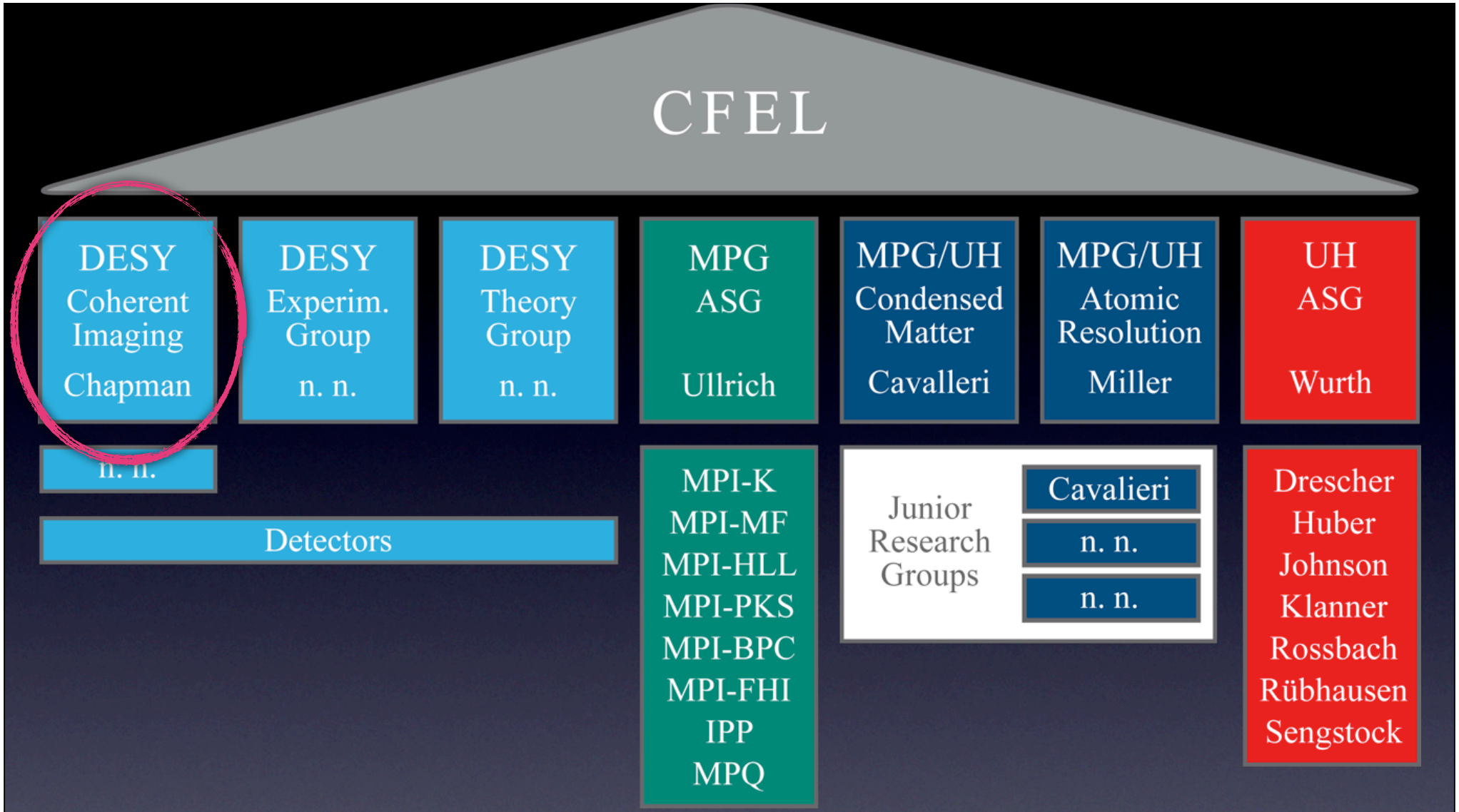
Center for Free Electron Laser Science

July 2010





Dwayne Miller is in process of transferring from Toronto
Robin Santra (2nd DESY W3 position (theory) starts 1st July
Second experimental DESY W3 is close to receive the call

Our experiments

Imaging

Today: 2x 2D detectors, 60-120Hz frame rate

60Hz = 250 MB/sec data rate

External facilities

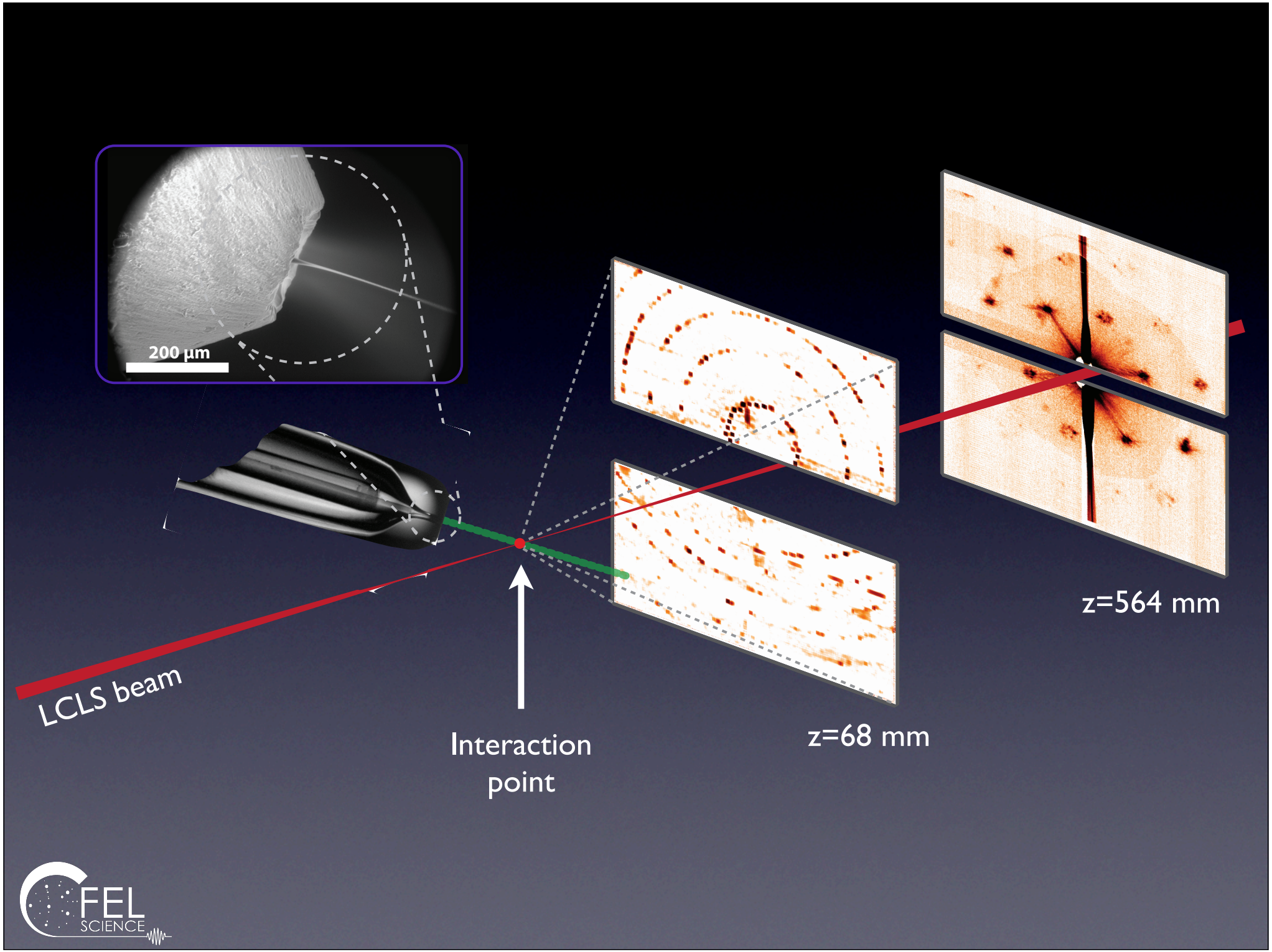
Including LCLS, SLS, XFEL

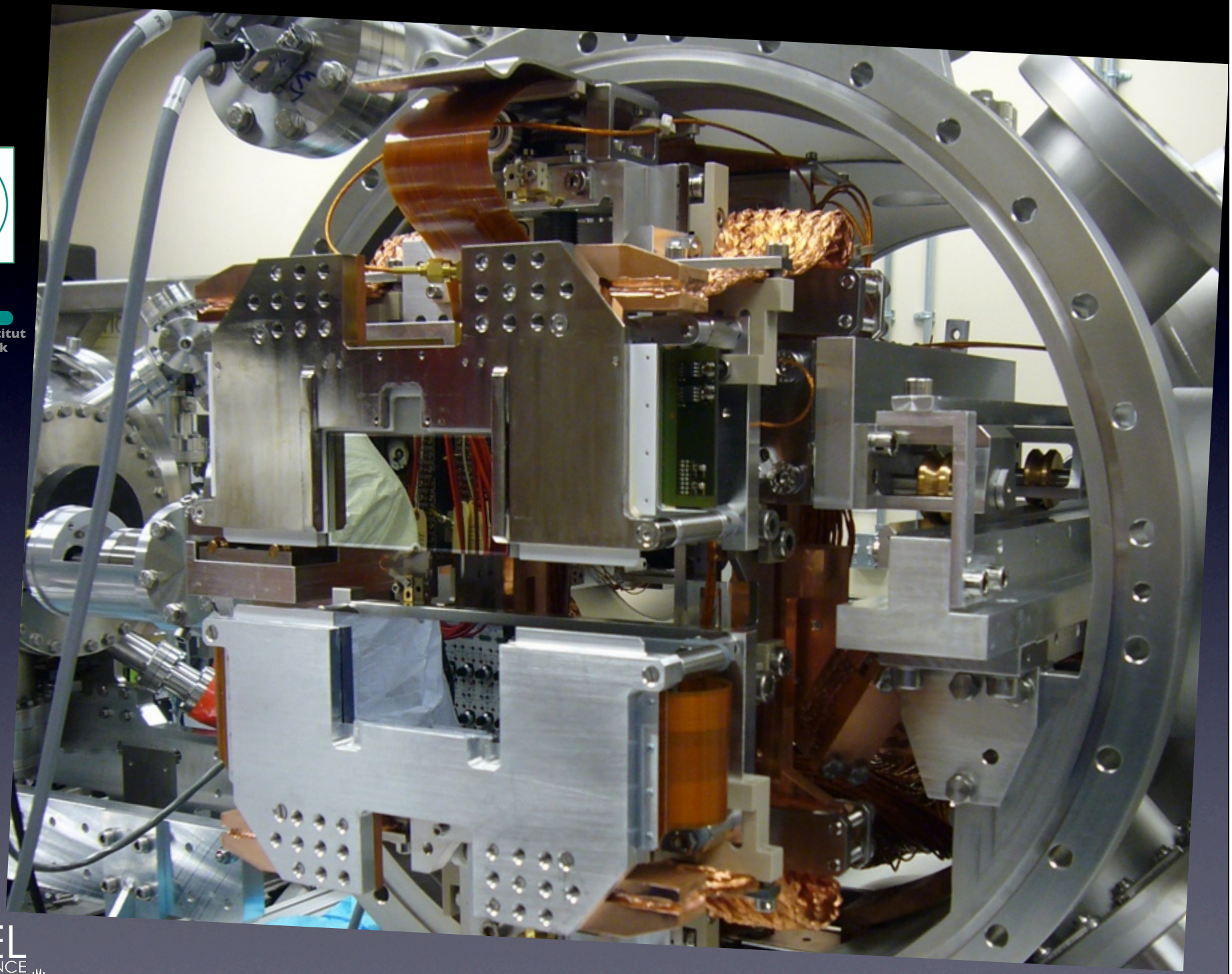
New experiments

New techniques

on-the-fly data reduction yet to be perfected

FEL SCIENCE

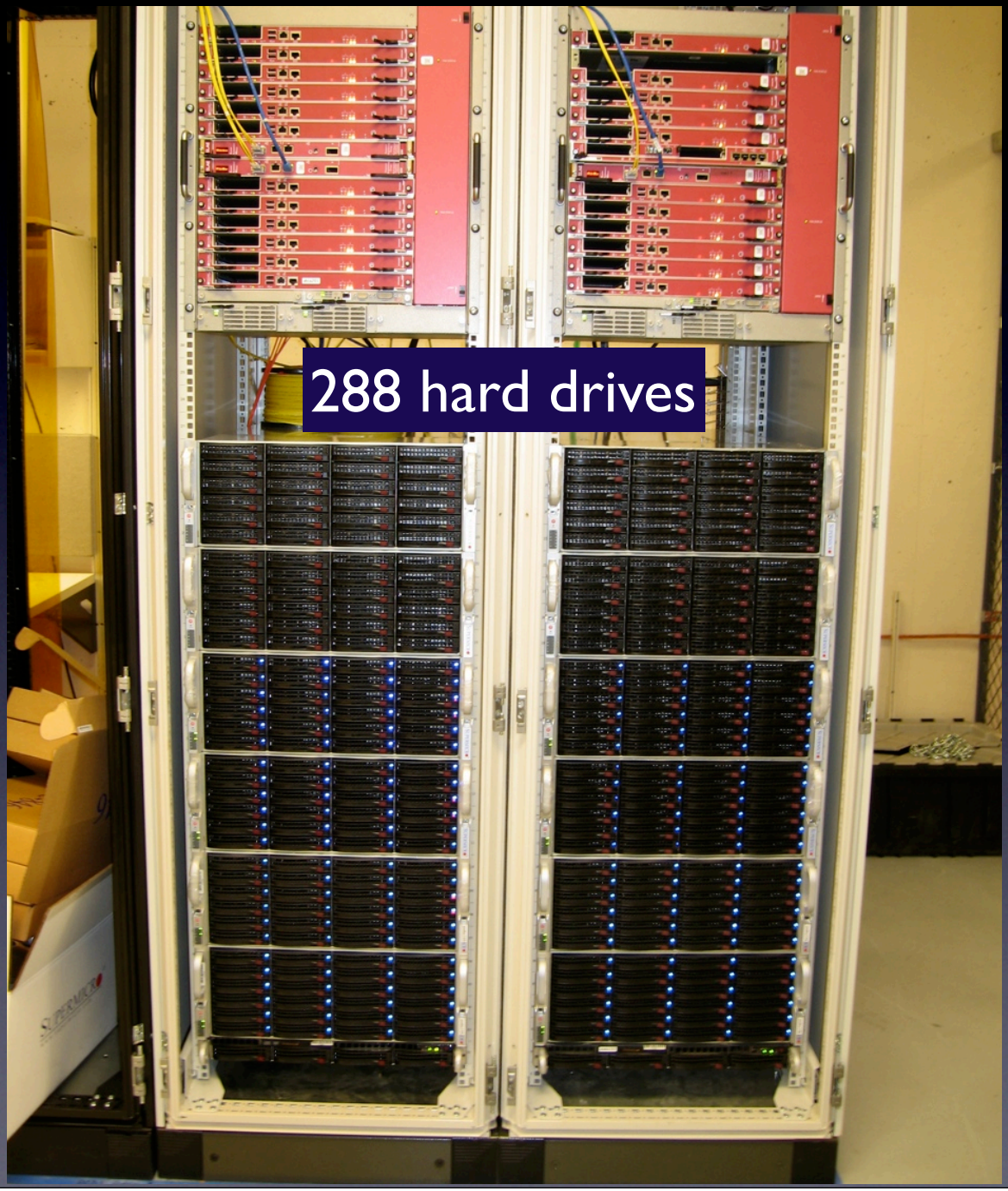

200 µm
LCLS beam
Interaction point
z=68 mm
z=564 mm
FEL SCIENCE



Max-Planck-Institut
für Kernphysik

FEL
SCIENCE

Where's my data?



30 images/sec
4 MB/image
= 430 GB/hr

Dec 2009:
1 week
>24 TB data
>6,000,000 images

May 2010:
5 weeks
60 Hz
240 TB data...

FEL SCIENCE

Our needs
I

Fast data transfer

Single experiment: 20-50 TB data
In the last 5 weeks: over 120 TB

Copy 100 TB using sftp?

There is alternate infrastructure: iRODS? GridSCP?
Help us use it.
Re-use and re-cycle existing HEP tools.
Beware of compatibility issues at both ends.

FEL SCIENCE

Our needs
2

Data archiving and storage

Need to retain data for 10 years
(the facilities do not do this for us)

Need fast access to data until papers are published
(disk, not tape, direct attached)

Need data home within 1-2 weeks of experiment

FEL SCIENCE

Our needs
3

Data processing

Fast access to data
not tape or GigE
direct attached storage best solution

Diverse data processing steps
Some steps have many single-threaded jobs
Others have large multi-threaded processes

Both single-thread and multi-thread performance (50-100 threads or more)
70-100 cores, extensible path to 200+ cores, 4GB/core, 280GB total or more
Large shared memory machine most flexible for us

FEL SCIENCE

Our needs
4

Collaboration

Share data with collaborators

Process and reduce data where it is stored
moving TB around the world is slow

Remote access/authentication; minimal firewall intrusion
Expiring tokens are a nuisance, even if secure
'Standard' accounts, low pain threshold for collaborators (students...)

Collaborative tools
Wikis, code repositories, online spreadsheets...

FEL SCIENCE

Our needs
5

As facility users

Rapid access to collected data (minutes delay)
Data access control (unix groups work)

Cluster / queue for data pre-processing
Ability to use the cluster (ie: install stuff)
Easily accessible from home institute

Fast transfer of data home
data archiving option (disk/ tape)

Transparent power, widely compatible
low pain threshold for use

FEL SCIENCE

Our immediate needs

Data storage

Currently need 500TB disk-accessible, extensible path to PB
Backed by stable archive (dCache)

Data processing

Fast access to data (ie: direct attached storage)
Both single-thread and multi-thread performance
70-100 cores, extensible
4GB/core, 280GB total or more

Data transfer and data sharing

Better (faster) infrastructure
Stable tools, widely compatible, low pain threshold for use

The user (customer) comes first

The technical details should easy to use and transparent

FEL SCIENCE