# Data Management at CHESS

*Marian Szebenyi*

NIHBTR
BIOMEDICAL TECHNOLOGY RESOURCES

CHESS
Cornell High Energy Synchrotron Source

# Outline

- Background
- Big Data at CHESS
- CHESS-DAQ
- What our users say
- Conclusions

# CHESS and MacCHESS

CHESS: National synchrotron facility, 11 stations (*NSF $*)

CHESS founded 1980,
MacCHESS 1984.

MacCHESS: Structural biology support (*NIH $*)

*Historically, at CHESS:*

- Data belong to the users.
- All users use a single account (“specuser”).
- Data collected at each station goes on local disk storage, where users have full access to create, modify, and delete files.
- Users are almost always present on site and can connect their laptops to the local net and their portable hard drives to computers at the stations.

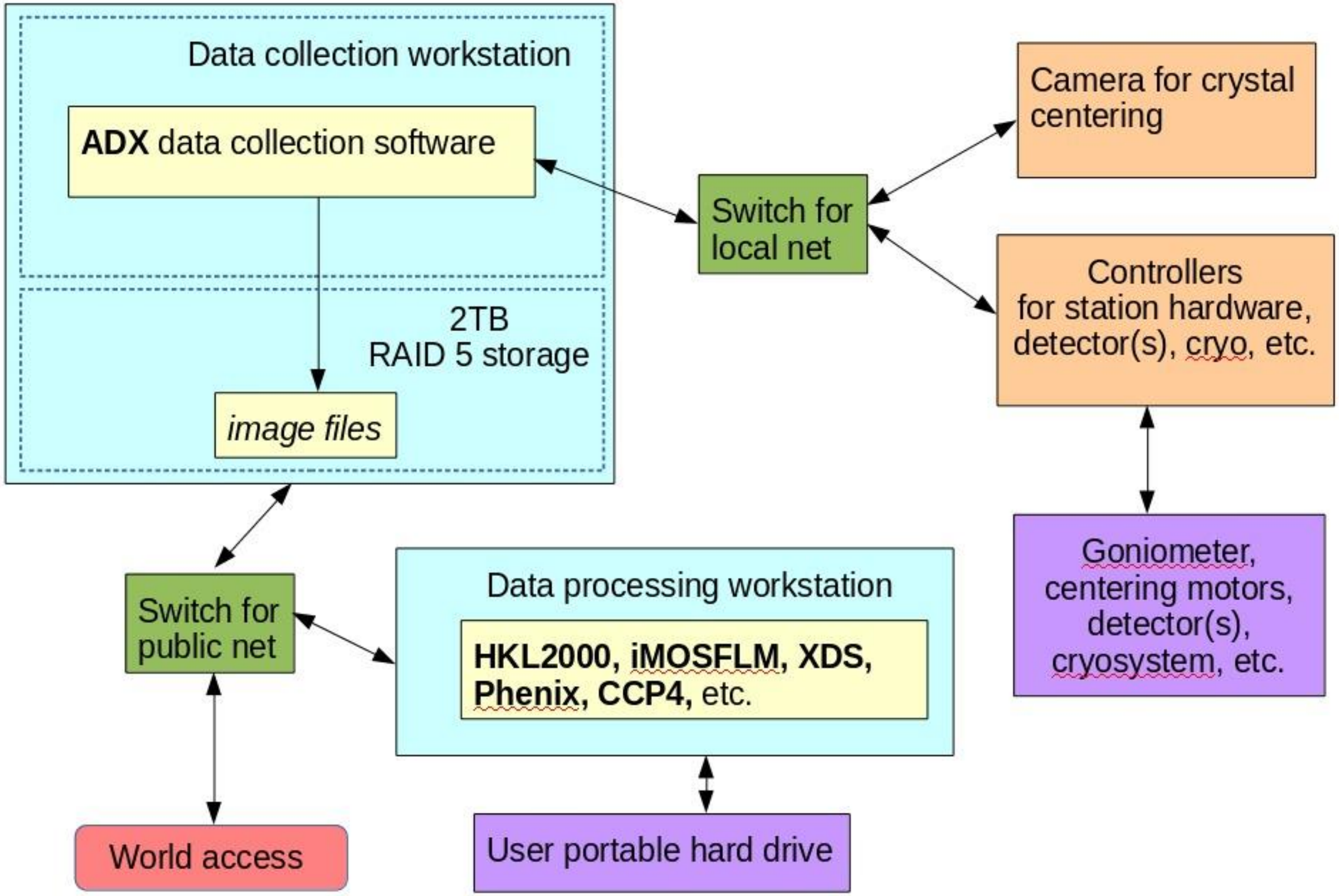
# Data backup

*Historically,*

- Users are responsible for copying data to their computer, disk, or tape, or for transferring files home by **ftp** or **rsync**.
- Data can be made available to users for a limited time after collection by copying to an external computer.
- Local backups are made periodically to offline disk, from which data can be retrieved if necessary.

MacCHESS

# MX data collection


Data collection workstation
ADX data collection software
2TB
RAID 5 storage
image files
Switch for local net
Camera for crystal centering
Controllers for station hardware, detector(s), cryo, etc.
Goniometer, centering motors, detector(s), cryosystem, etc.
Switch for public net
Data processing workstation
HKL2000, iMOSFLM, XDS, Phenix, CCP4, etc.
World access
User portable hard drive

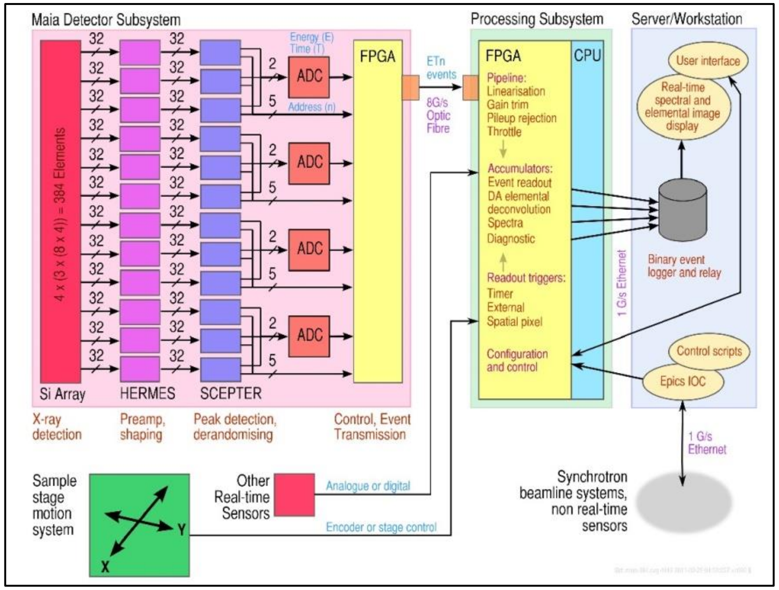
# Advent of Big Data

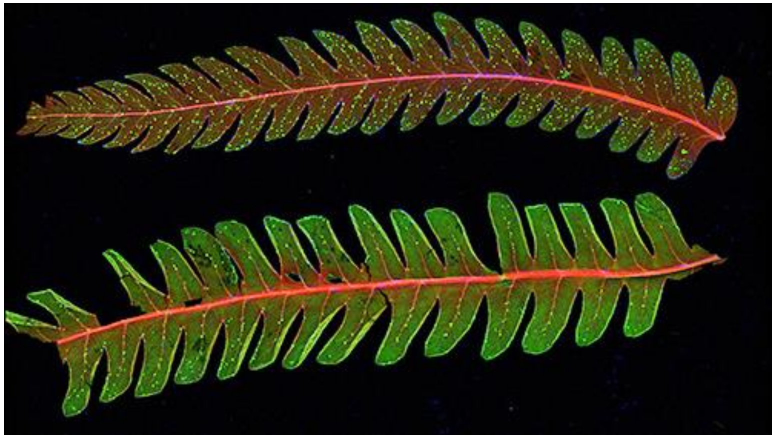
*Currently,*

- New detectors, used in new types of experiments, generate prodigious amounts of data.
- CHESS upgrade scheduled for 2018 will enhance flux, lead to shorter exposures, more time-resolved experiments, and increased data production rate.
- Improvements in computers and networking have made non-local real-time data storage practical.
- Agencies are requiring increased access to raw data.
- Administrative changes at CHESS have resulted in greatly enhanced IT support.

# Maia detector

- A key driver of developments in data management.
- 384-element energy-dispersive detector designed by physicists from CSIRO and BNL.
- Binary logger daemon (blogd) receives data from detector and writes to central disk storage.
- Data are read back for processing with **GeoPIXE**.
- 8+ TB collected in 18 months.





# CHESS-DAQ

- New data acquisition system created in response to need to handle “big data”.
- Represents a paradigm shift from distributed to central data storage.
- Includes a computing cluster as well as high-speed networking, facilitating on-site data reduction.
- Builds on experience with processing large amounts of data from CLEO detector.
- Needs to accommodate data from many different types of experiment, satisfy needs of many different users.

# DAQ Hardware

- 10 Gb storage area network (SAN)
  - Enterprise-class storage devices, servers, network switches
- 2 x Infortrend iSCSI storage devices
  - Dual controllers, redundant power
  - 24 x 4 TB drives/device configured in 2 x RAID 6
  - Total 128 TB usable

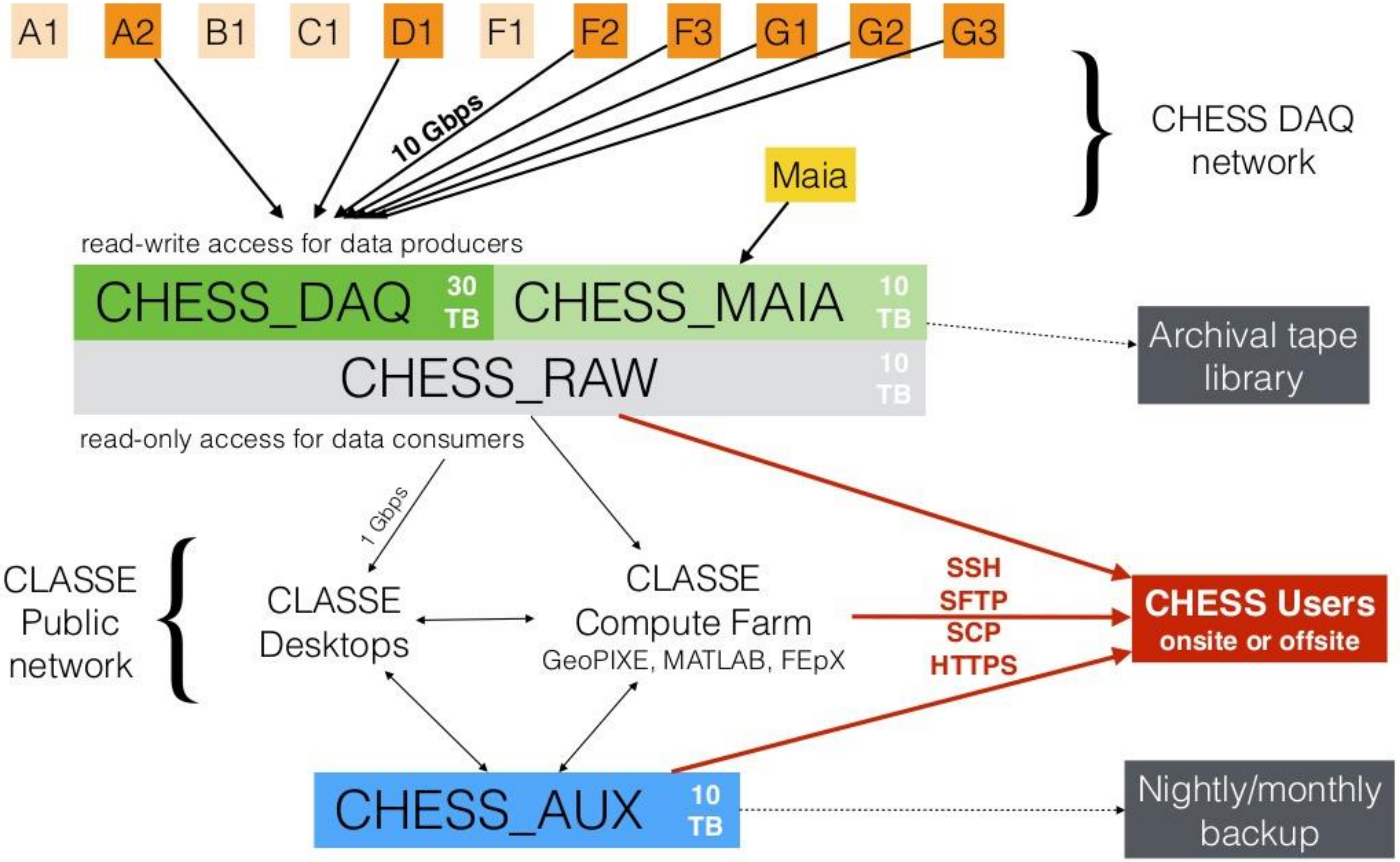
- Files served by CHESS-DAQ cluster
  - 5 x IBM x3550 M4 servers
  - Each has 2 x 6-core Intel Xeon, 128 GB RAM
- Networking
  - 10 Gb IBM Blade switches
  - New multi-mode optical fiber runs throughout CHESS

- Throughput
  - Up to 900-1000 MB/s writes (sustained 600 MB/s)
  - Average 200-300 MB/s reads
- IBM tape library, total capacity 250 TB (uncompressed)

# DAQ file systems

- Write newly-generated raw data to:
  - CHESS_DAQ (70 TB)
  - CHESS_MAIA (10 TB)
- Read raw data for processing from:
  - CHESS_RAW (Read-only interface to both current and older data)
- Save raw data from previous run in:
  - PREVIOUSDAQ (30 TB)
  - PREVIOUSMAIA (10 TB)
- Store processed data, metadata, project info in:
  - CHESS_AUX (10 TB)
- Processing software stored in:
  - CHESS_OPT (500 GB)

# DAQ data flow



# Data archiving

- Current and previous runs are always on disk.
- Archiving and backup are handled by **Symantec NetBackup**; data from older runs is on tape and can be restored on request.
- Between runs:
  - Two copies made of CHESS_DAQ and CHESS_MAIA; one is stored off-site.
  - Data from CHESS_DAQ and CHESS_MAIA moved to PREVIOUSDAQ and PREVIOUSMAIA, respectively.
  - Standard directory structure created for next run.
- Nightly incremental backups are made of CHESS_DAQ, CHESS_MAIA, CHESS_AUX and CHESS_OPT. Monthly full backups are made of CHESS_AUX and CHESS_OPT.
- Archived data are kept indefinitely.

# Access to data

- Data may be **written** to CHESS_DAQ and CHESS_MAIA only by certain special users (one for each station). Various data collection protocols are used to work with this system.
- Data may be **read** from CHESS_RAW by anyone with an account on a computer on which it is mounted. Users can use the "specuser" account. CHESS_RAW is a read-only file system that incorporates links to CHESS_DAQ and CHESS_MAIA, as well as directories in which to store any older data recovered from the archive.
- CHESS_AUX and CHESS_OPT are open to anyone in the "chess" group, i.e. staff, students, and users, for data processing.
- Data transfer kiosks are available for users to back up data; must be unlocked using a CLASSE account (staff have permanent accounts, users may obtain temporary ones).
- The preferred means of transferring large datasets to users' labs is now Globus.

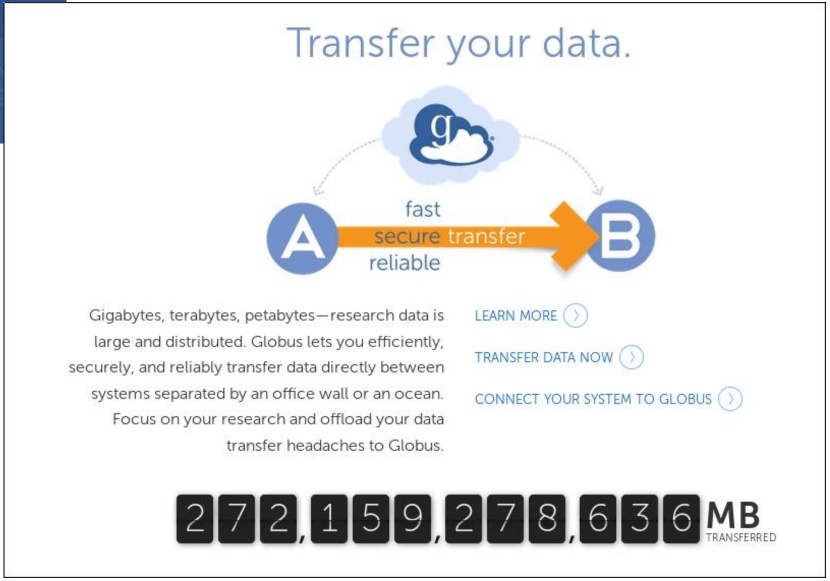
# Globus

globus

Products ▾ Pricing Developers Support ▾ Log In

Research data management simplified.

Globus (www.globus.org) is a system designed for accelerated transfer of large datasets around the world via a user-friendly interface. It is widely used in the grid computing and high-energy physics communities.



Transfer your data.

Gigabytes, terabytes, petabytes—research data is large and distributed. Globus lets you efficiently, securely, and reliably transfer data directly between systems separated by an office wall or an ocean. Focus on your research and offload your data transfer headaches to Globus.

LEARN MORE

TRANSFER DATA NOW

CONNECT YOUR SYSTEM TO GLOBUS

272,159,278,636 MB TRANSFERRED

Requires:

- Installation of (free) Globus Connect Personal software.
- CLASSE account; users may obtain a temporary account from CHESS User Office.

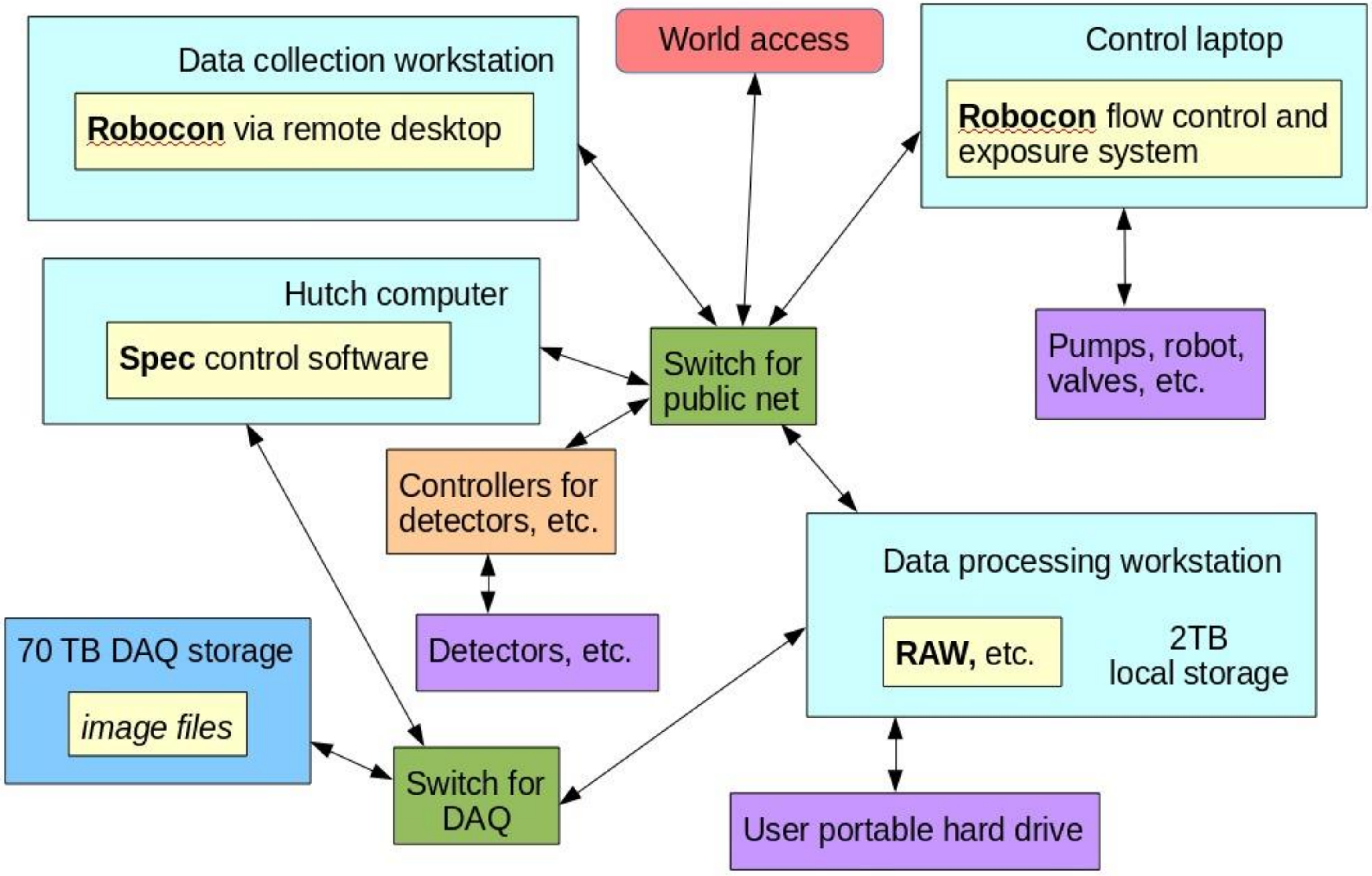
# Remote connections

- CHESS is protected from general world access by a firewall.
- While at CHESS, users may connect their devices to the CHESS public net using **eduroam** or **Red Rover** (wireless) or by a wired connection. Users are cautioned to use the net responsibly.
- Data can be transferred out of the facility by **ftp** or **rsync,** if allowed by the receiving computer.
- Users (and staff) can log in from outside to a computer designated for that purpose if they have a CLASSE account. CHESS_RAW is available from this computer.
- Remote MX data collection is possible using a temporary CLASSE account, **OpenVPN** software, and the NoMachine **NX** remote desktop.
- In special cases, a temporary hole may be made in the firewall between specific external and internal computers.

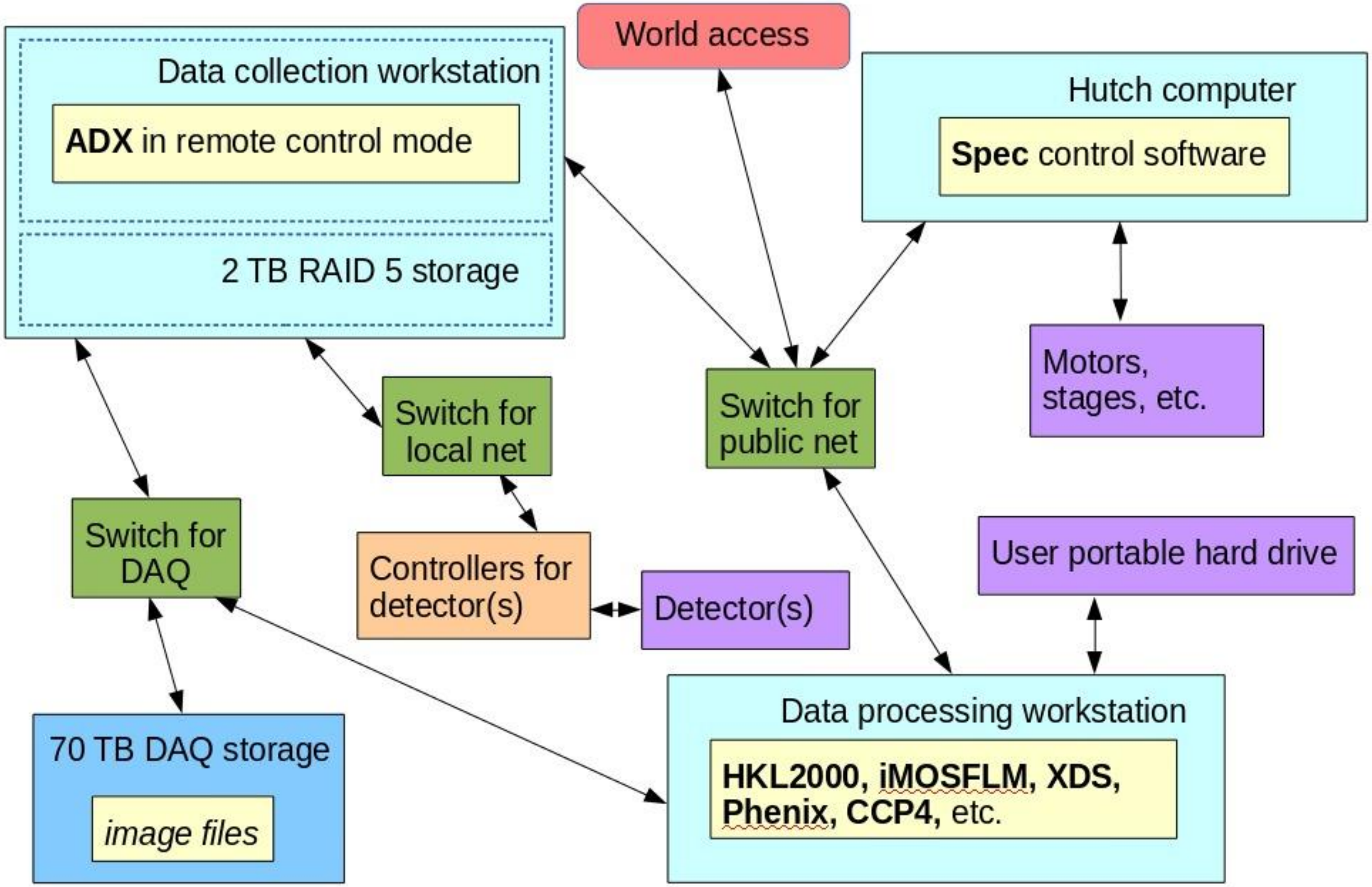
# Compute Farm

- Central resource available to users on request.
- More than 60 Linux nodes (about 400 cores), with queuing system to distribute jobs across nodes.
- Supports interactive, batch, parallel and GPU jobs; has access to DAQ file systems.
- Commonly used software packages are installed, and more can be added on request.
- Used regularly for fluorescence computed tomography, preprocessing of diffraction images, analysis of SAXS data, and finite element polycrystal modeling.
- An integral part of the Maia data analysis workflow.

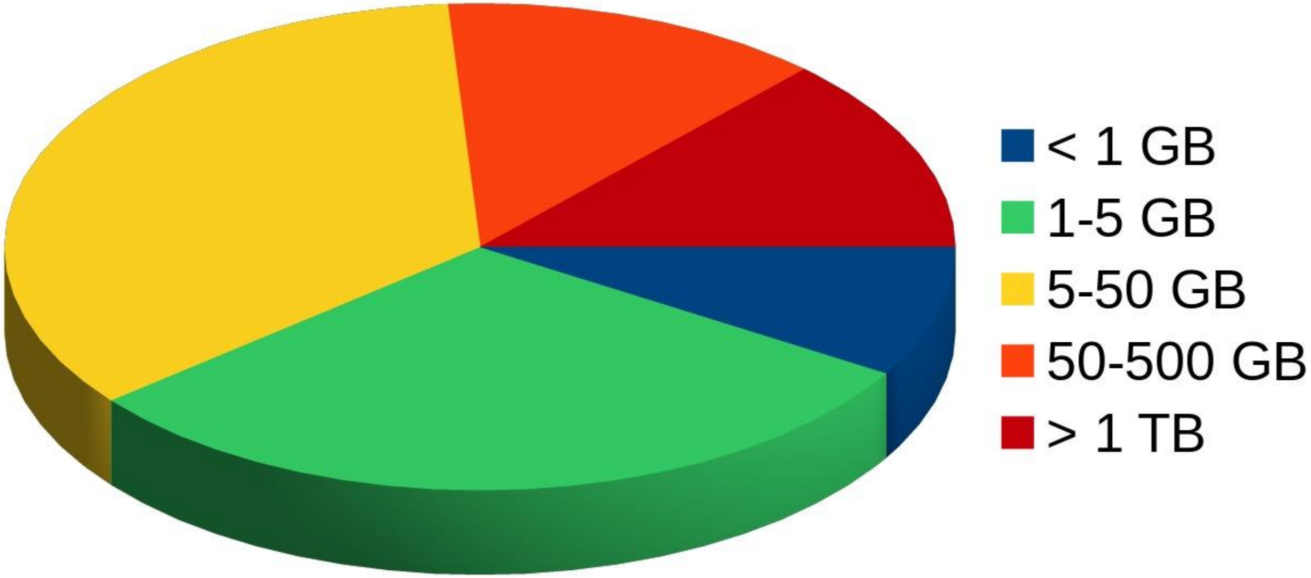
# BioSAXS scheme


Data collection workstation
**Robocon** via remote desktop
World access
Control laptop
**Robocon** flow control and exposure system
Hutch computer
**Spec** control software
Switch for public net
Pumps, robot, valves, etc.
Controllers for detectors, etc.
Data processing workstation
**RAW,** etc.
2TB local storage
70 TB DAQ storage
*image files*
Detectors, etc.
Switch for DAQ
User portable hard drive

# Possible MX scheme

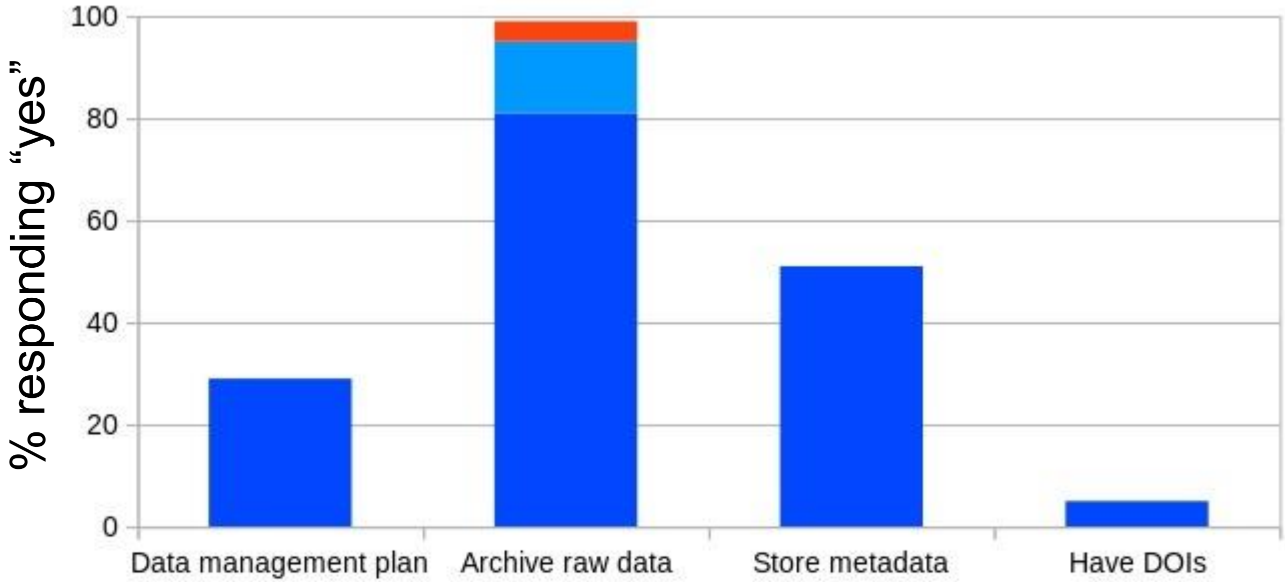


# User data per visit



Amount of user-generated data has greatly increased in recent years, and users want to take all of it home with them.

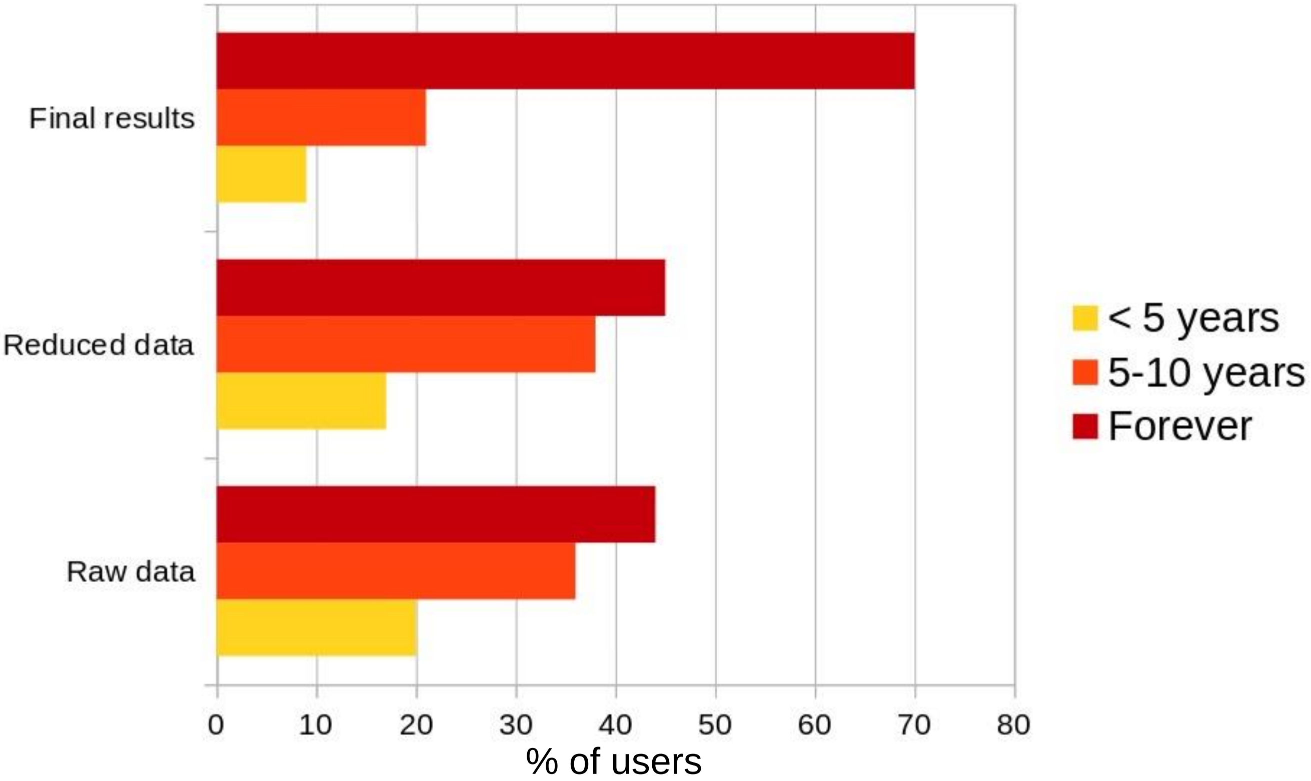
# User data handling



We asked users:

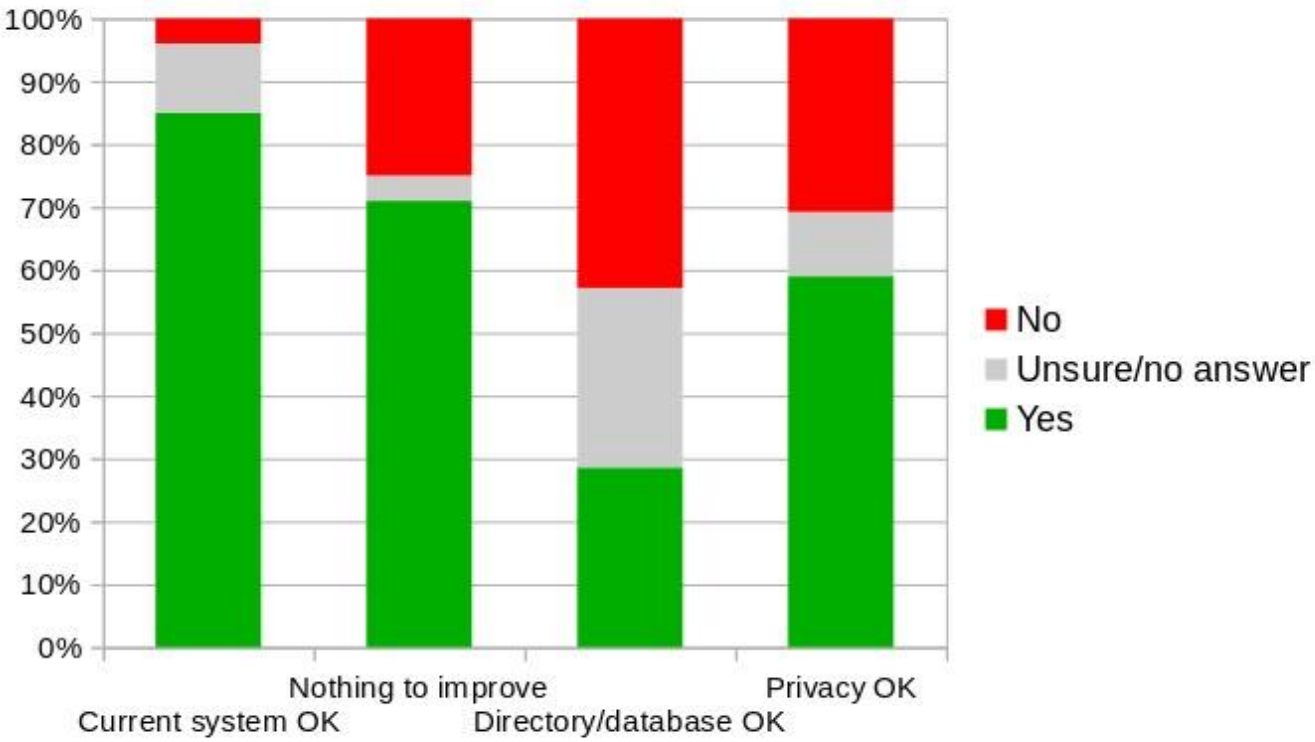
- Do you have a data management plan?
- Do you archive raw data (dark blue, in lab; light blue, institutional storage; orange, in on-line repository)?
- Do you store metadata with the data?
- Do your datasets have DOIs?

# How long to keep data



Most users think all data should be stored for at least 5 years, and nearly half would like raw data kept “forever”.

# User satisfaction



We asked users:

- Are you happy with the current CHESS data management plan?
- Is there anything we can do to improve your experience?
- Would a more organized directory structure/database be useful?
- Are you concerned about maintaining the privacy of your data?

# User comments

- Most complaints (and there were not that many) about the current system involved data transfers:
  - Transfer is too slow.
  - Not enough kiosks for data transfer.
  - Globus is awkward to use, some would specifically prefer **ftp**.
- One user found the DAQ naming conventions confusing.
- A few users had software issues (probably not MacCHESS users): file format not accepted by program, needed more processing advice.
- One user requested advice on data management strategies, and to share the burden of data storage/backup.
- "External access, DOI, metadata" from one user.
- On the other hand, "CHESS cannot afford to become a long term public data depository, and, in my opinion, should not be goaded into this position."

# Enhanced database

- No consensus on this topic.
- User comments included:
  - “Automatically generated summary file created during experiment could be helpful (run/scan, time/date, sample, temperature, pressure, condition, etc).”
  - “Correlation of metadata with specific detector image sets and/or **spec** scans would be immensely helpful. At the moment the onus is on the user to manage this separately while collecting data, adding to the workload.”
  - “I think it would have to be the same at all synchrotrons to be truly useful.”

# Security

- A substantial number of users were at least somewhat concerned about others being able to access their data.
  - The current system allows anyone with an account on CLASSE computers to view any raw or processed data on the system.
  - Users can delete processed data easily, but raw data are intended to be saved. They can be deleted by staff if requested.
- Some of those who were not worried commented that the data would not be useful to anyone else without metadata that were not stored along with the raw data.
- Restriction of access to a particular user or group is under consideration, but would be difficult to implement in practice.

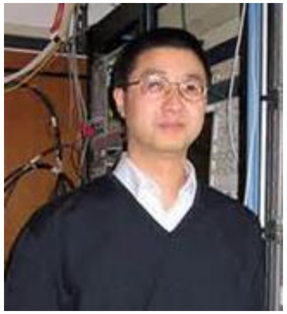
# Conclusions

- CHESS has implemented a data storage system which provides:
  - The ability to write directly to a large central storage facility under a systematic directory structure; sufficient capacity to store all users' data for a 6-8 week run.
  - Automatic nightly incremental backups, and long-term archiving of data; protection against accidental deletion of data through read-only access from most computers.
  - Freedom for users to store metadata, and to process their data, but no requirement for them to do so.
- In general, users are happy with the current arrangement. They take most or all of their data home and archive it there.
- We propose to maintain the CHESS system, with incremental improvements. We can also provide users with information to assist them in depositing data, with metadata, in appropriate repositories.

MacCHESS

# Acknowledgements

**CLASSE IT Group:** Werner Sun, Devin Bougie, and others



NIH and NSF for funding

**User Office:** Kathy Dedrick, Nahla Minges

**Maia Development:** BNL and CSIRO teams, Arthur Woll

J. Thieme (BNL), C. Ryan (CSIRO), R. Kirkham (CSIRO), F. Werner (Munich), P. Siddons (BNL). Seated: M. LeVaillant (CSIRO), G. Iancono (CNRS)