

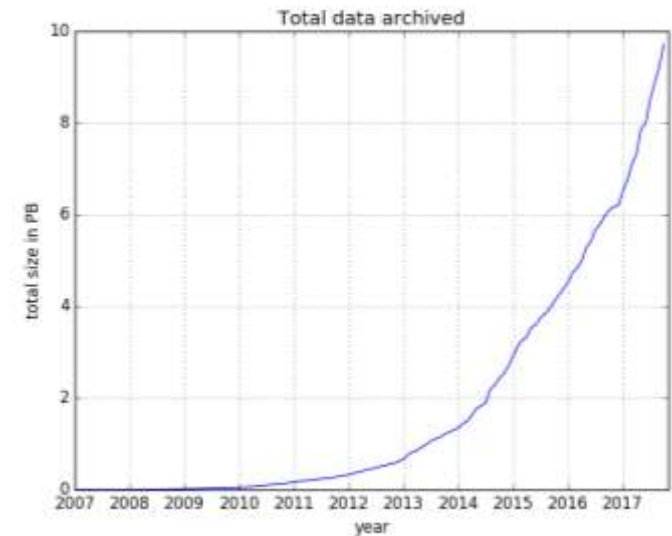
Future Direction and Plans

Increases in data volumes and
the challenges this brings

Scientific Computing



New Computer Room
(CSCR3 – Inner courtyard)



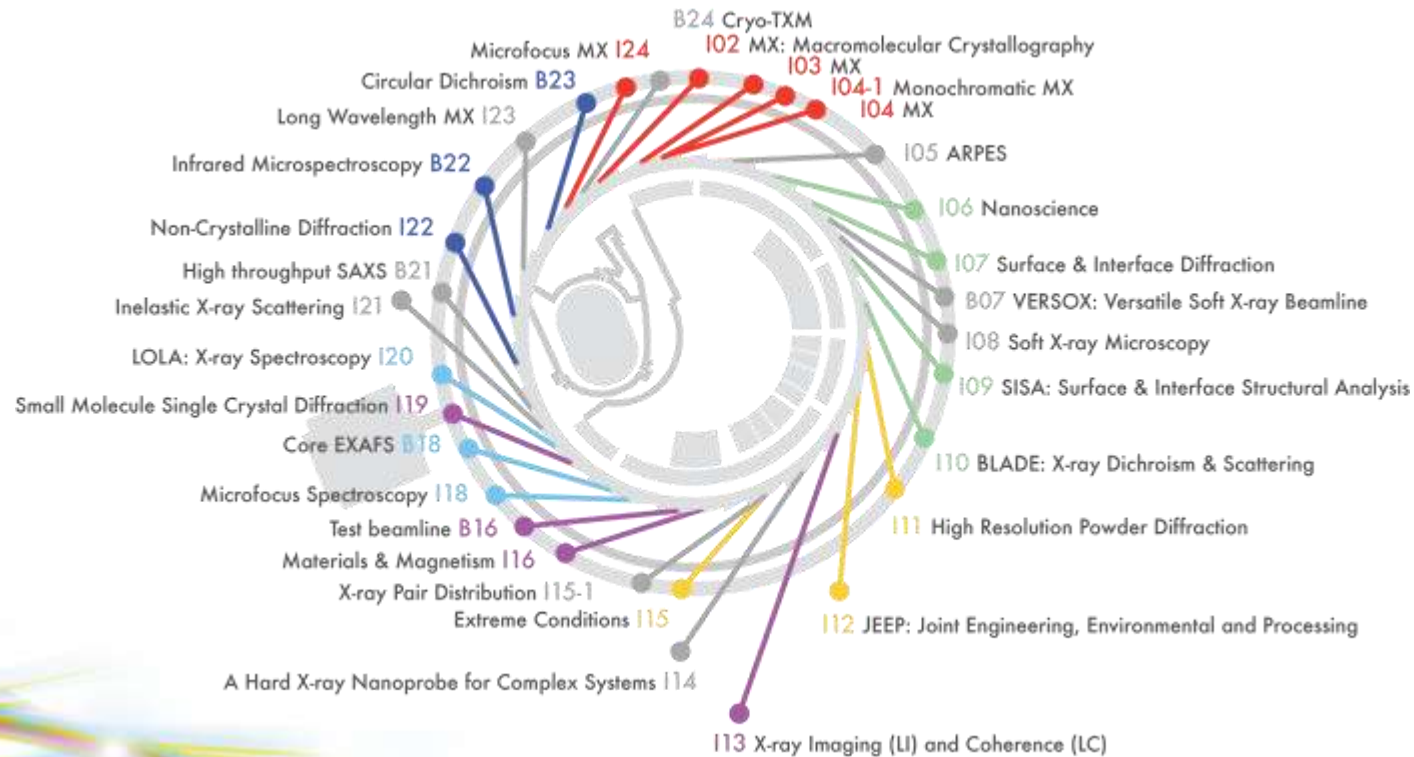
New compute and
Data Storage
On premise and
exploration of hybrid
cloud



10 PB of
Archived Data



Beamlines by Village



- **Macromolecular Crystallography**
- **Soft Condensed Matter**
- **Spectroscopy**
- **Materials**
- **Engineering and Environment Surfaces and Interfaces**

A National User Facility for Biological Electron Cryo-microscopy (eBIC)

Wellcome Trust Strategic Award/MRC/BBSRC, applicants:
Helen Saibil, Kay Grünewald, David Stuart, Gerhard Materlik

- Funded initially by the Wellcome Trust, MRC and BBSRC at level of £15.6 M over 5 years, augmented to ~£25 M by additional investment by the Trust in 2016
- **The facility** currently includes:
 - 4 high-end 300kV automated cryo EMs (Titan Krios FEI)
 - 200 keV automated feeder instrument (Talos Arctica)
 - Cryo focussed ion beam instrument (SCIOS)
 - Sample prep incl. vitreous sectioning
 - Correlative fluorescence/EM
 - FEI Polara @OPIC Oxford for CAT 3 samples



New eBIC Facility

- Initially constructed with two large rooms for two Krios, remodel to house four - completed 9/16.
- Sample preparation, loading and general lab +



'Big' Data Lifecycle challenges

- How much data do you mean by BIG?
- How 'FAST' do you need to analyse the data?
- What data can be THROWN AWAY?
 - (and at what stage?)
- How LONG do you need to keep the data?
- And WHERE? Where do you want to transfer the data to/from?

Diamond Challenges

- Data rates increasing from Beamlines
- Data rates from Electron Microscopes
- Internal network needs
 - (first 40Gb to the BL connection in place)
- Compute needs for 'realtime' processing increasing.
- External network transfers
 - Support Science DMZ, Globus Online
 - Linked facilities. Links to partner institutes, UK (e.g. Oxford) and EU-XFEL (Hamburg)
- Increasing post processing
 - Need for longer term storage platforms
 - More CPU and GPU
- User expectations! Want to do more in less time, remotely, collaborate, results not data, etc.

Infrastructure challenges

- Automation, better robotics, new techniques, new detectors
- Keeping ahead of growing data storage demands:
 - Easier for 'visits' of fixed duration
 - Harder for 'post processing' of unknown duration
- End Stations:
 - Beamlines have a construction lead time.
 - EM tend to 'just arrive'. Funding dependant, faster turnaround to implementation.
 - Integrating 3rd party 'control' servers into Diamond environment. (especially Windows based systems)

Infrastructure challenges

- Infrastructure increasingly needs to be 'up' all the time:
 - Shutdowns don't apply to EM.
 - Everything does not stop in a shutdown any more – e.g. post processing.
- Increasing expectation to leave with results not just data.
- Increasing remote user usage

Archive: Things to think about

- What is an ICAT /TopCat issue and what is a local implementation, or underlying infrastructure issue.
- What are the limits of ICAT/TopCat. Do we know?
- Data volumes are increasing but data retrieval and how to present data back to the user is more challenging than just 'archiving' data.
- User issues (perceived or real), not just how to technically implement something.

Archive: Things to think about

- What is the user experience?
- What do users want to do?
- Extracting data from archive needs to be as simple as ingesting.
- Data Exploration? Is TopCat the right interface? Others (for Diamond could be ISpyB)
- Should this meeting (part of it) include users?