Science & Technology Facilities Council

Data Exploitation in the CLF OCTOPUS Facility

Dan Rolfe

Central Laser Facility

STFC Rutherford Appleton Laboratory

UK

OCTOPUS

CLF central laser facility

Research Complex at Harwell

Working across the life and physical sciences

BBSRC bioscience for the future

MRC Medical Research Council

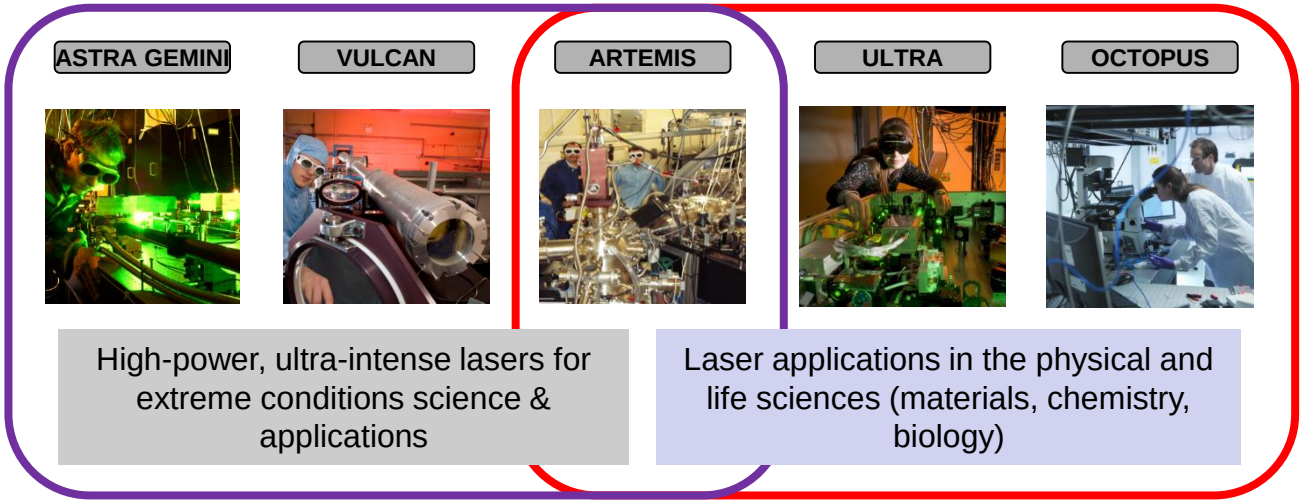
Overview

- OCTOPUS
- Examples
- Challenges
- Vision
- Approach
- UK T0 needs
- Future

Science & Technology Facilities Council

OCTOPUS

- Part of STFC Central Laser Facility at Harwell Campus



- National imaging facility with peer-reviewed, funded access
- Located in Research Complex at Harwell
- Cluster of microscopes and lasers and expert end-to-end multidisciplinary support

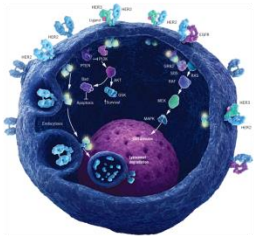
Science & Technology Facilities Council

OCTOPUS

STFC-funded

Cryo-CLEM
Correlative light & electron microscopy

Nanoscopy
>20 nm resolution



VUTARA
Fast STORM
3D system

MRC-funded

3D super-resolution SIM with PALM/STORM

20nm 3D super-resolution (under development)

2D super-resolution PALM/STORM

BBSRC-funded

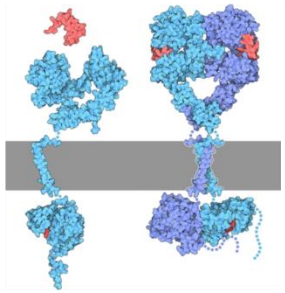
G-STED/FLIM system

OCTOPUS Imaging Village

Single Particle Imaging
>5 nm resolution

Single molecule confocal ALEX

2-colour single molecule FRET & polarisation



3-colour single molecule tracking

3-colour single molecule FRET & polarisation & ALEX

BBSRC-funded

5-colour single molecule tracking

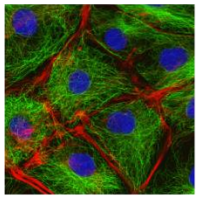
Single photon confocal FLIM/FRET

Multiphoton confocal FLIM/FRET

3-colour confocal

TIRF Tweezers

Light sheet 3D multicolour volume imaging

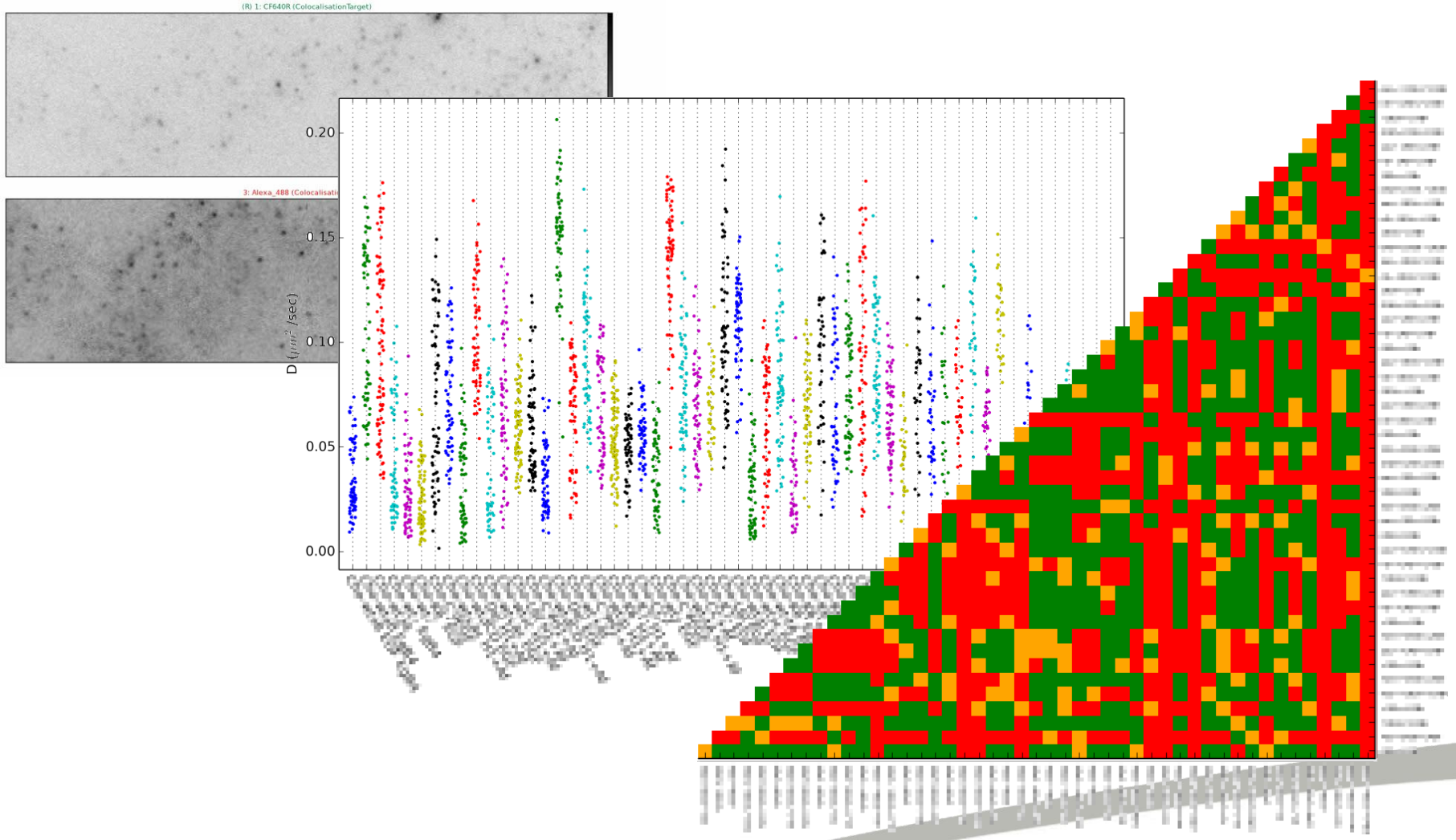


Cellular Imaging
>200 nm resolution

STFC-funded

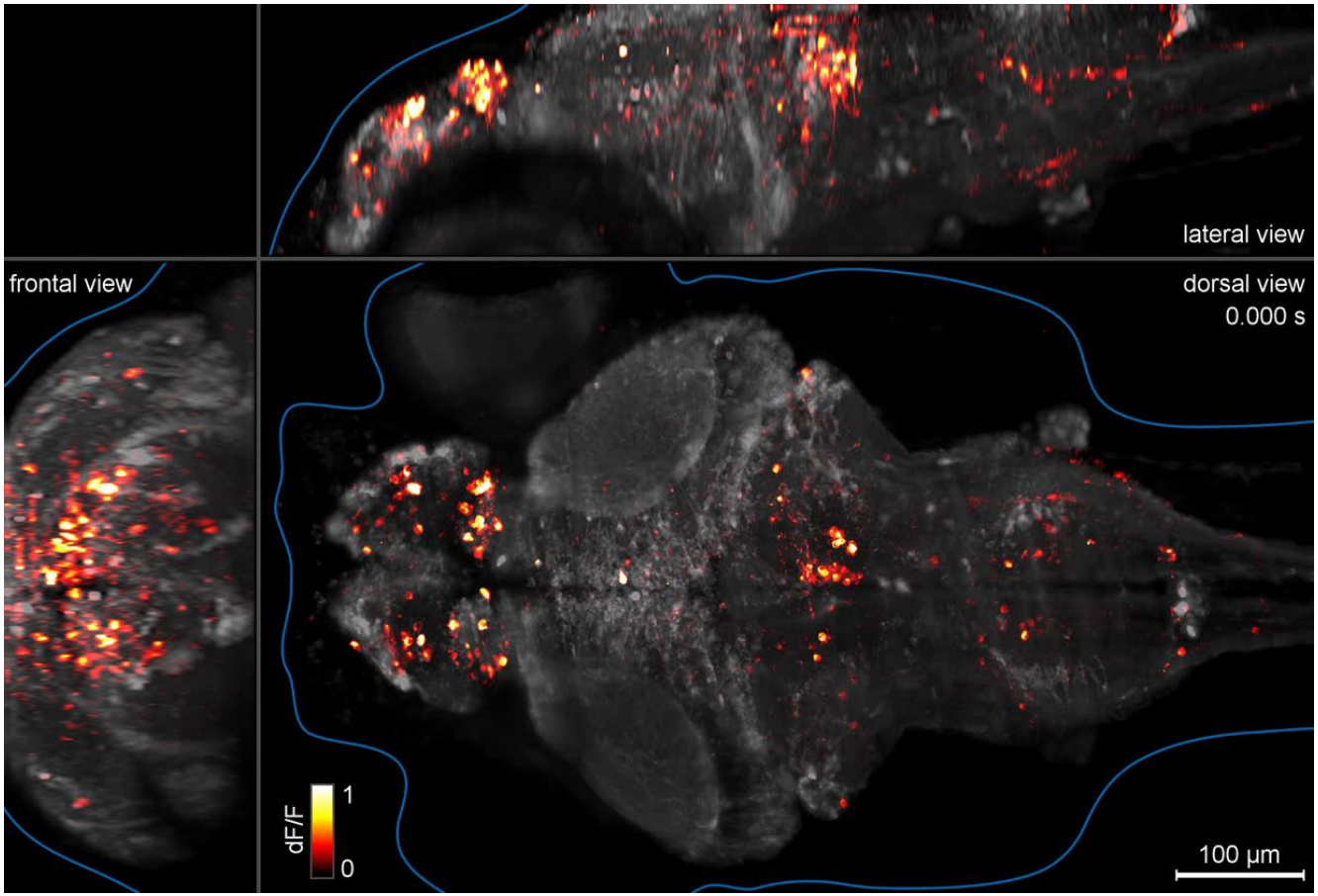
Science & Technology Facilities Council

Molecular interactions in cancer



Science & Technology Facilities Council

3D lightsheet microscopy



2048x2048px

3h30min@10Hz→1Terabyte

20min@100Hz→1Terabyte

5 days to preprocess on 100 cores

Keller and Ahrens, https://doi.org/10.1016/j.neuron.2014.12.039

Science & Technology Facilities Council

Challenge

- Effective integration and delivery across OCTOPUS
  - Changing variety and combinations of
    - studies, samples, instrumental and analysis techniques, platforms, licensing models, dataset types, computing architectures, algorithms
  - Challenging image analysis – often very manual
  - User expertise (or lack of it) in numerical/computational work
  - Metadata – often limits ability to exploit large studies

Vision

Multi-technique bioimaging solutions centre of excellence

with data exploitation to match advanced instrumentation and lead new developments

- Flexible integration of data and computation across OCTOPUS
  - From start to end of studies, across all methods in study
  - Robust, quantitative, automated approach to data analysis, visualisation and management
  - *Enable step change in capability through development and integration of varied data and algorithms*

Make current tour-de-force experiments routine

Science & Technology Facilities Council

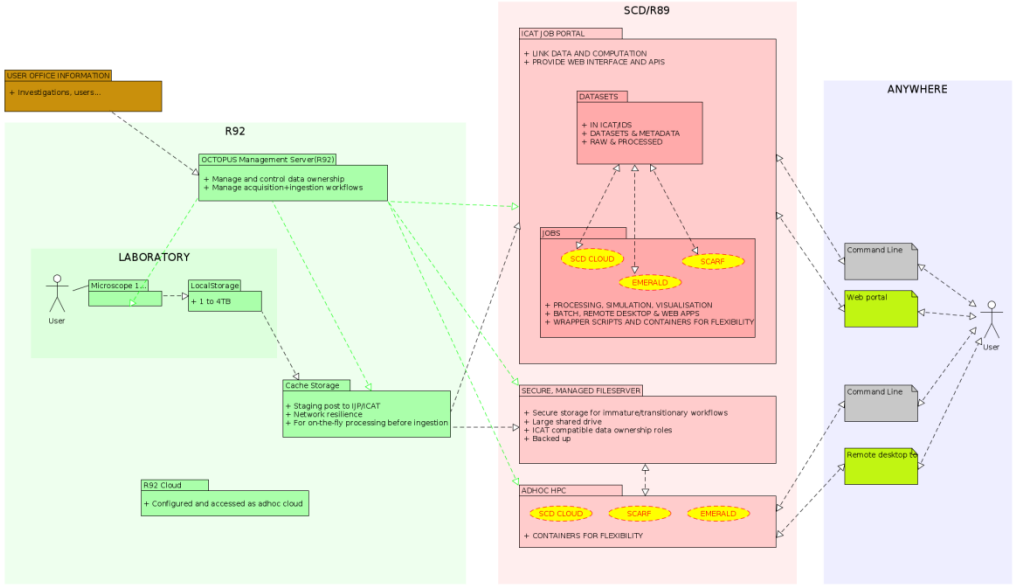
Need

- Enable facility scientists to focus on science with users
  - Access to data management, storage and computation resources
    - efficient
    - scalable
    - secure
    - reliable
    - on- and off-site access
    - convenient (for facility staff and users)
    - supported
- Predominantly a middleware challenge at present
  - ALC
  - Looking forward to hardware bottleneck…

Science & Technology Facilities Council

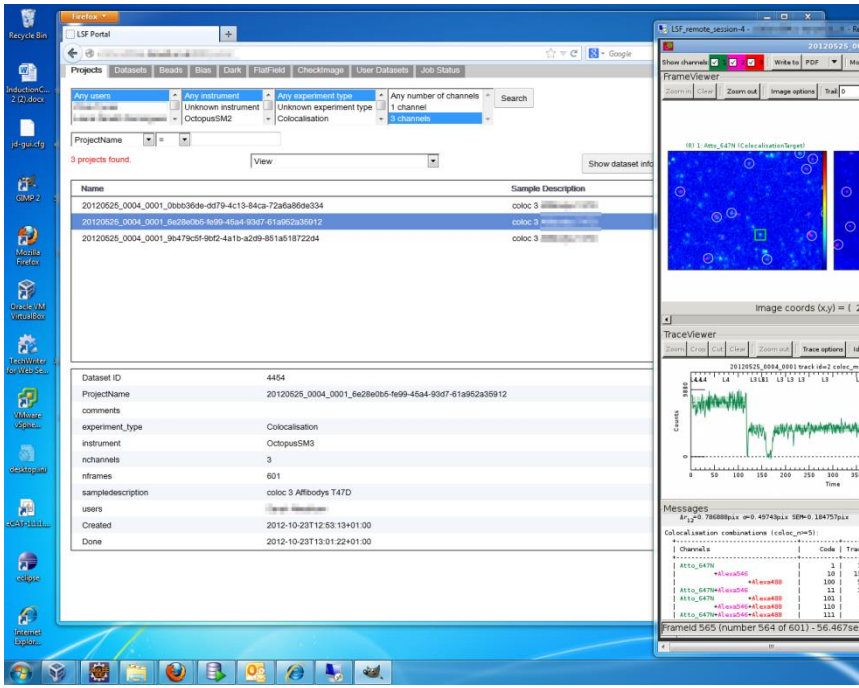
Approaches – WIP

- *Colocate data and compute infrastructure*
- *Loose coupling of components, abstraction of data types and computation needs*
- Currently ~20 in-house analysis workstations
  - Quite manual but improving
  - Remote access VPN + RDP/X2Go unreliable
  - Use the DAaaS system instead – with our workstations and cloud
- Secure fileservers following ICAT data ownership (SSSD)
- User and group management system linked to ICAT, AD, CLF User Office
- SCARF – with Singularity + SMBNETFS
- ICAT Job Portal (IJP)



ICAT Job Portal

- Robust, secure and flexible data and job management system
- Loosely coupled
  - Configurable dataset and job types (options, metadata)
  - Wrapper scripts for existing software
- Data and computation located on-site
- Batch and interactive jobs
- Underlying data storage ICAT
- Stores raw and processed data
- On- and off-site access via web-portal and integrated remote desktop



Fisher, Phipps & Rolfe, IWSG, 2013

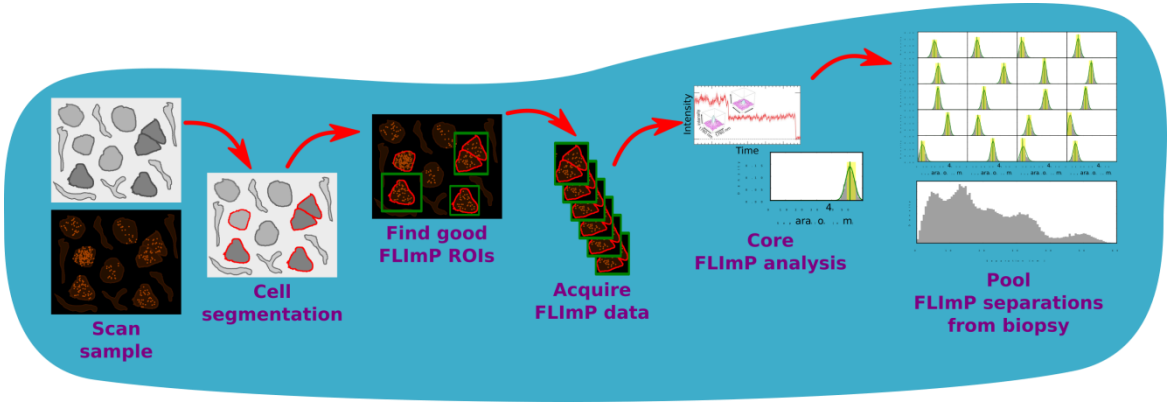
Science & Technology Facilities Council

Needs for UK T0

- Network - speed, reliability and testing
  - On campus, between labs and datacentres
  - To end users
  - For reliable, usable remote desktop and data transfer
- Flexible storage, backup, archiving solutions
  - Capacity and performance
  - Currently ~30TB fileserver backup takes ~5 weeks…
- Cloud
  - High-end nodes with and without GPU
- SCARF
  - Maybe less important than cloud in the future (not much MPI)

Future

- Automation of experiments
  - machine learning, instrumentation and data acquisition



- High throughput experiments, on-the-fly analysis
- Link to/develop richer metadata – generic and subject specific
- Make larger scale studies routine
- Greater integration of different methods, more correlative approaches
- Scalable software will move bottlenecks to hardware…

Science & Technology Facilities Council

People

- CLF
  - **Dan Rolfe**
  - Marisa Martin-Fernandez
  - Michael Hirsch
  - Jianguo Rao
  - Teodor Boyadzhiev
  - Chris Duncan
  - Michalis Vrettas
  - Sarah Needham
  - Laura Zanetti-Domingues
  - Selene Roberts
  - Chris Tynan
  - Stephen Webb
  - Benji Coles
  - Alessia Candeo
- Cambridge
  - Mike Hobson
  - Sumeetpal Singh
  - Rich Wareham
  - Charles McLachlan
- SCD
  - Steve Fisher
  - Kevin Phipps
  - Brian Ritchie
  - Rebecca Fair
  - Brian Matthews
  - Martyn Winn
  - Cheney Ketley
  - Jens Jensen
  - Derek Ross
  - Alastair Duncan
  - Frazer Barnsley
  - Gordon Brown
  - Tony Hey
  - Catherine Jones
  - Alison Packer
  - Tom Byrne
- ISIS
  - Tom Griffin
  - Martyn Gigg
- King's College London
  - Peter Parker, Simon Ameer-Begg
- RCaH IT
- OCTOPUS users
- Funding
  - BBSRC
  - MRC
  - STFC

Science & Technology Facilities Council