

STFC Scientific Computing Department: *e-Infrastructure to support research*

Brian Matthews

Scientific Computing Department

brian.matthews@stfc.ac.uk



Science & Technology
Facilities Council

Scientific Computing Department

- ~180 Staff between RAL and DL
- Software engineering expertise
 - Data management systems
 - Visualisation and analytics
 - Applied maths
 - CoSEC: Computational Science Centre
 - CCPs in Physics, Chemistry, Biology, Engineering
- Systems hosting and management
 - GRID-PP Tier 1
 - JASMIN
 - Facilities Computing (Data Archive, SCARF Compute cluster)



Our Science Communities

Environment



JASMIN

Life Sciences



Diamond



Astronomy

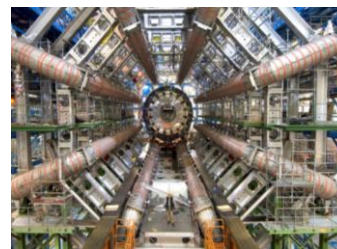


And wider research communities



ISIS

Materials
Sciences



CERN

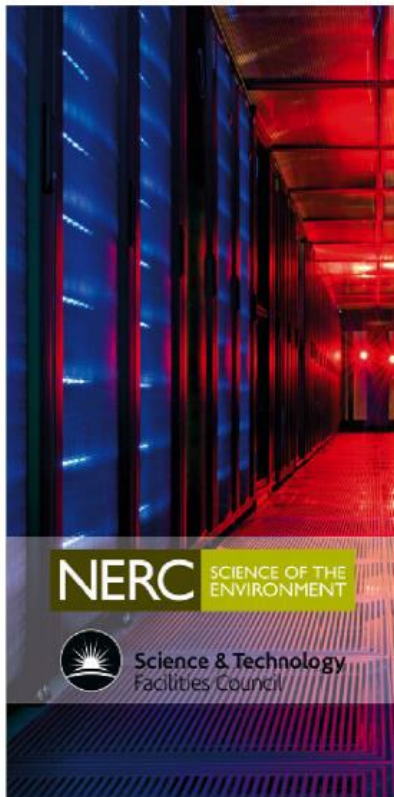
Fundamental
Physics

Urgency to provide better environmental predictions

- HPC for higher-resolution models

But...

- Massive data requirement: observational data transfer, storage, output, post-processing



- ▶ 16 PB Fast Storage
(Panasas, many Tbit/s bandwidth)
- ▶ 1 PB Bulk Storage
- ▶ Elastic Tape
- ▶ 4000 cores: half deployed as
hypervisors, half as the
“Lotus” batch cluster.
- ▶ Some high memory nodes,
a range, bottom heavy.



Safe Data, Big Data, Open Data



- **Data storage and management**
 - Petabyte data store
- **Integrated data management pipelines for data handling**
 - From data acquisition to storage
- **A Catalogue of Experimental Data**
 - Metadata as Middleware
 - Automated metadata capture
- **Providing access to the user**
 - Integrated into Analysis frameworks
- **Data archiving and Preservation**
- **Leading role in European data infrastructure**
 - Making data FAIR

Findable
Accessible
Interoperable
Reusable

Brian Matthews



EOSC pilot
The European Open Science
Cloud for Research Pilot Project

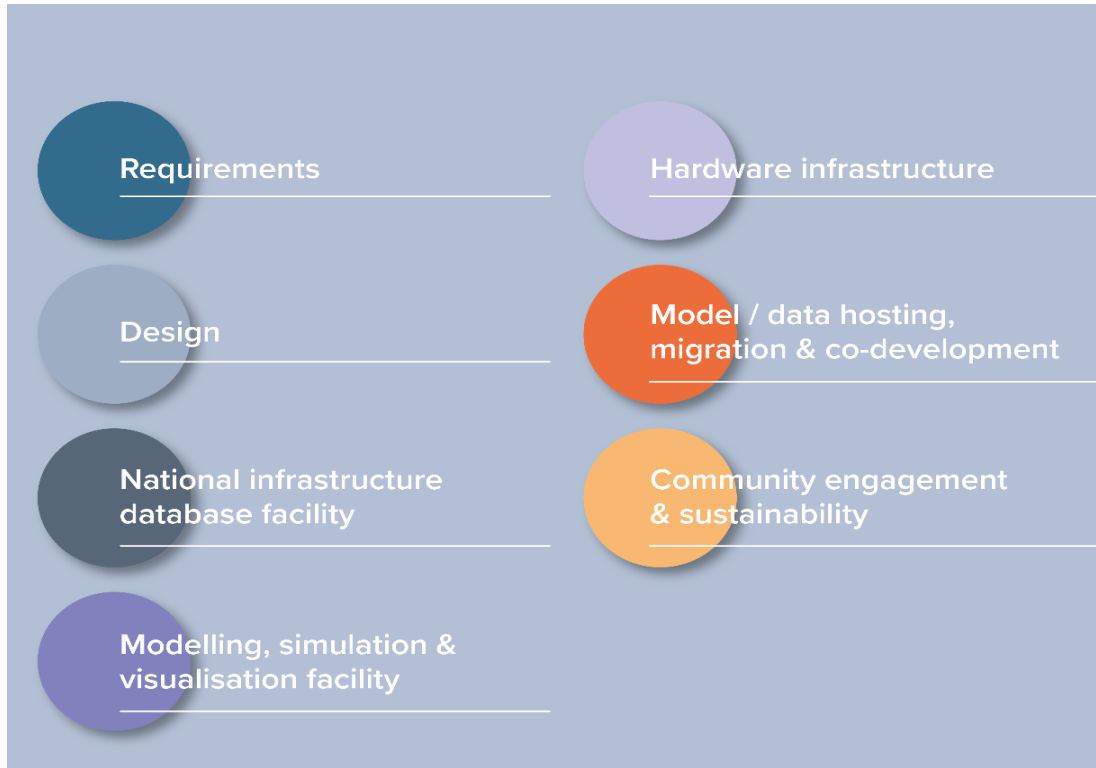


Science & Technology
Facilities Council



DAFNI

Data & Analytics Facility
for National Infrastructure



DAFNI Construction and Delivery Programme (2017-2021)

Total Project Cost: £8M
Led by Oxford University,
Constructed by STFC

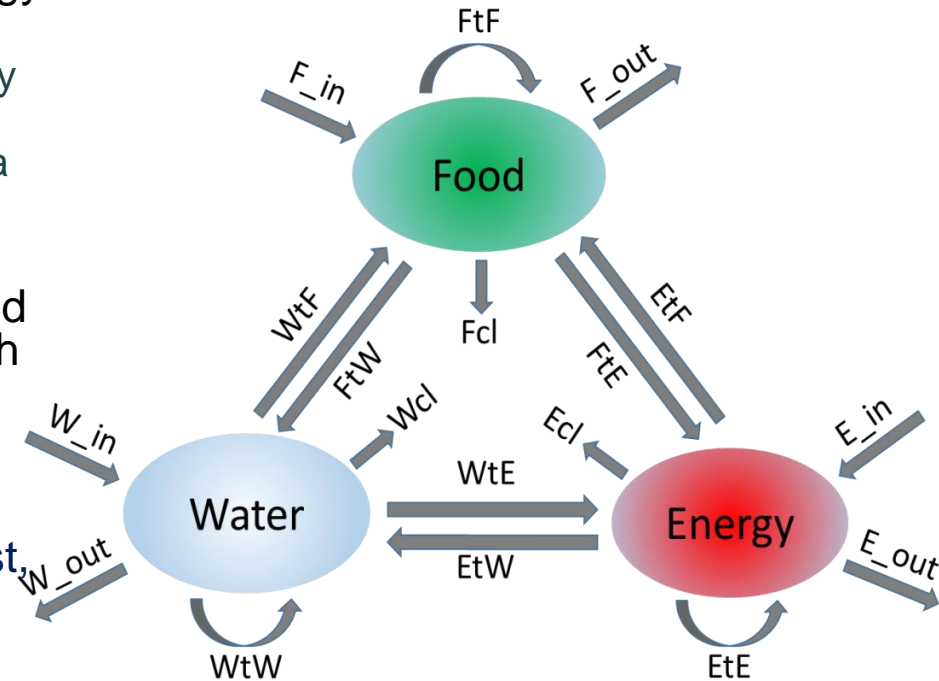
Erica Yang



WefWeb



- Modelling sustainable supply of water, energy and food
 - Case studies based in Oxford, Tamar Estuary (Devon) and London
 - Stakeholders analysis and engagement, data collection, regulation landscape
- Funded by EPSRC and STFC
- Glasgow, Cambridge, Exeter, Newcastle and Oxford, UCL, ICSTM, Rothamsted Research
- Agent: an autonomous system with certain intensities of Water, Energy and Food ingest, production, consumption, loss and supply.
 - Strong favour towards balance-based modelling, compared to impact-based ones
 - Resulted model should suit computer simulation or real data analysis, or a mixture



Agent-based
modelling approach

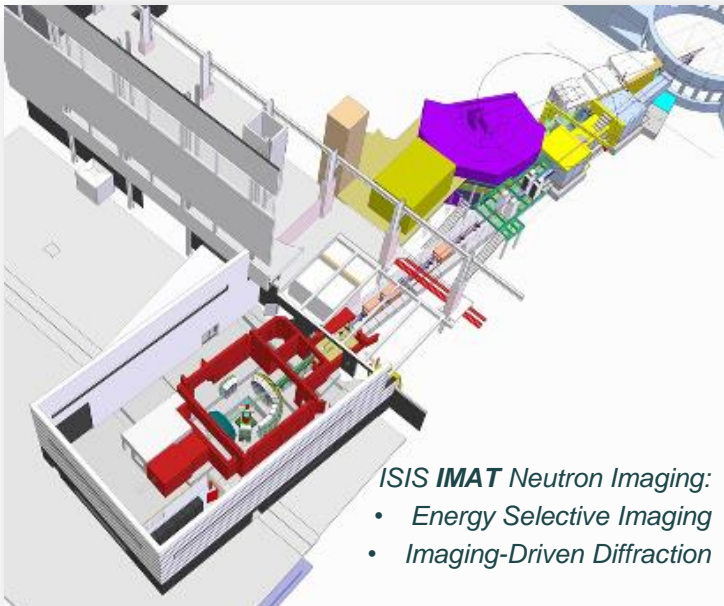
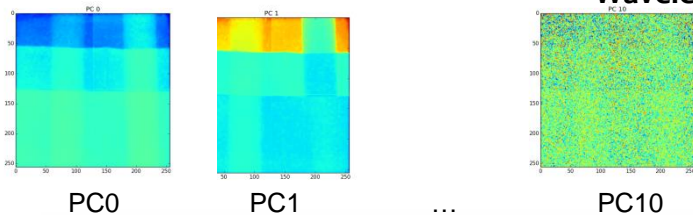
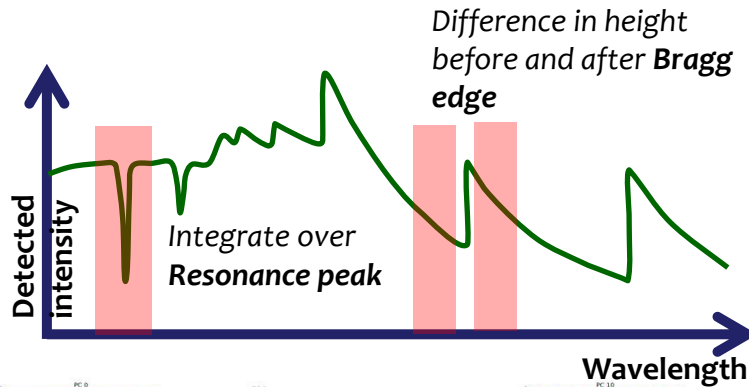
Simon Lambert, Vasily Bunakov

<https://www.gla.ac.uk/research/az/wefwebs/>



Science & Technology
Facilities Council

Visual Analytics and Machine Learning



ISIS *IMAT* Neutron Imaging:
• Energy Selective Imaging
• Imaging-Driven Diffraction

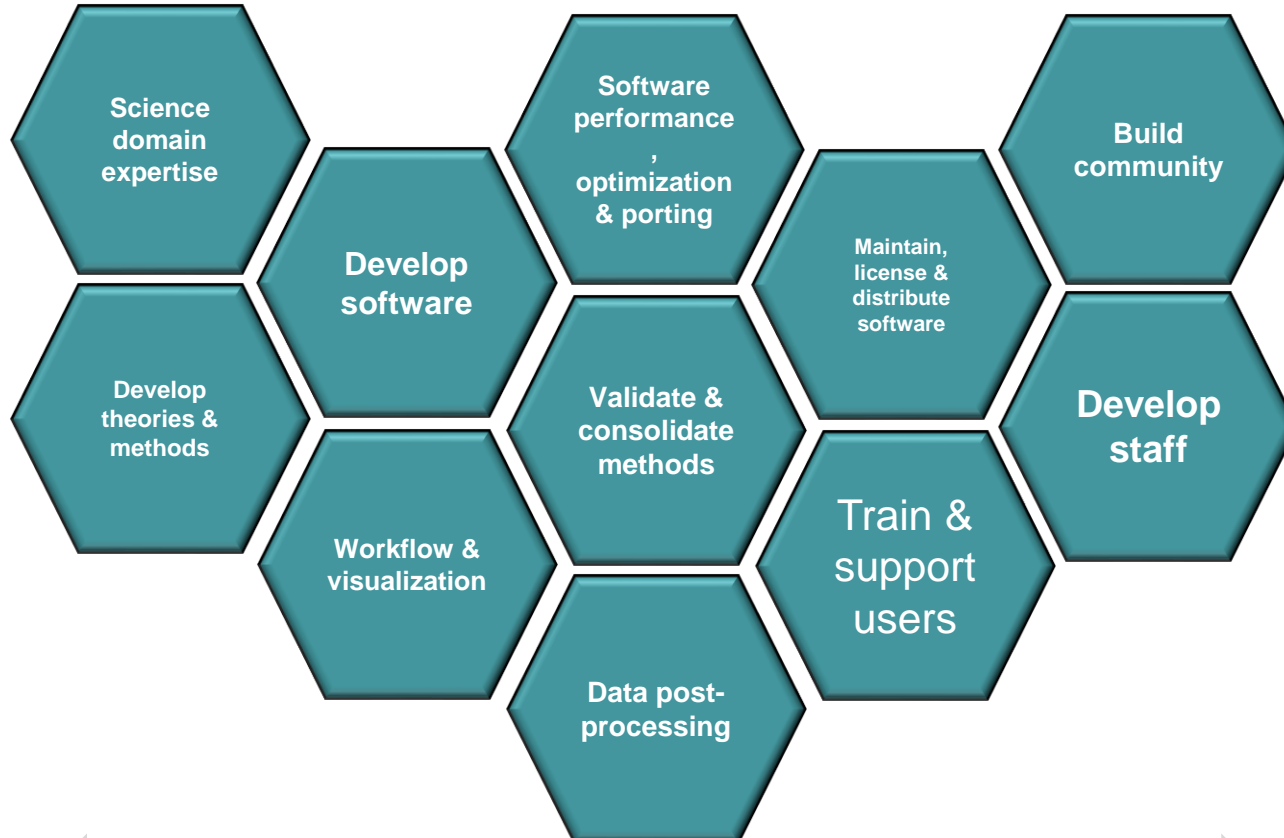
- **Data Intensive Science: Hyperspectral Imaging (IMAT/ISIS neutron imaging)**
 - 3,000 energy channels (2017) -> 375 billion voxels/3D dataset
 - 10,000 energy channels (2018 onwards - planned) -> 1.25 trillion voxels/3D dataset
 - Hundreds of concurrent analysis needed for experiment steering
- **Feature extraction** from full-spectrum neutron images using Principal Component Analysis – PCA (Joe Kelleher, Genoveva Burca, NeuWave'16, UK)
- **Developed:** fast feature extraction and exploration methods based on accelerated machine learning algorithms for hyperspectral image analysis
- **Results:** optimised processing performance from hours down to minutes, critical for in-situ visual analysis and exploration

Erica Yang



Science & Technology
Facilities Council

A Centre of Expertise in Computational Science



CoSeC supports the advancement of research by:

- developing software in multiple disciplines
- providing a hub for exchanging knowledge through training and outreach
- nurturing strong collaborations among researchers

Biology Engineering Physical Sciences Biochemistry Medical Science

So what can we bring ?

E-Infrastructure = Data + Compute + Expertise

- **Technology**
 - Tools, expertise and infrastructure for bringing data together
 - Metadata: data sharing and integration
 - Access to compute
 - Integrated solutions
 - Bring data and compute to bear, cloud systems
- **Software**
 - Image analysis and machine learning
 - Computational science : Modelling and simulation
 - Physics, Chemistry, Biology, Engineering
- **Expertise**
 - Software Development
 - Systems integration
 - Data science



Frameworks for data management and processing

- What metadata associated with a dataset for reasoning about its applicability?
- What types of access and transferability of datasets are possible, and how can they be harnessed for data reuse?
- How can data be integrated to apply to problems ?
- How data provenance models and systems can be applied ?
- How do you allow access to computing resources, remotely ?
To end users ?
- What are the best algorithms and models to apply to solve problems ?

