Current Best Practice for Neuton and Photon RIs :
Data Management, Services and delivery
ESFRI Workshop on RIs and EOSC

Prof. Jonathan Taylor
Head of the Data Management and Software Centre

Where is my Data?

How do I get it

How do I look at it

What does it mean

REDUCING THE BOTTLENECK EFFECT:
"What we're trying to do here is **expedite the time to discovery.** Scientists should be able to **focus on their science** without having to become experts in data management."
—Shawn McKee
research scientist in physics

The ESS Scientific Computing Agenda

- **Scientific Discovery and Scientific Impact are the KPIs**

- FAIR is Better

- Integrating Data Curation into the whole data life cycle early.

- Make use of sensible standardisation.

- Working without a (Data) Policy is a mine field.

- Everything is better through Collaboration.

- **Open data needs sustainability.**

- Plan for scalability, both across techniques and for data rates and volume.

European Photon and Neutron Landscape

40k users per year

Multi disciplinary programmes

Individual researchers —> Large collaborations

Industrial and commercial use

Neutron and Photon RIs are **User** Driven Facilities

10000 users per year at European facilities

Many **users are not necessarily domain experts**

Many **users are not experts necessarily in e-infrastructures**

Scientific impact comes from the user programme - external users

Individual experiments —> programme access

Single investigators —> large research groups

Free at the point of access for academic use

Multi modal & cross disciplinary programmes

Access multiple sites and RIs (inc photon etc)

Rate limiting step may be another RIs data

Core Product is Data

- FAIR is a real challenge for facilities and for **users**

- 1500 + individual experiments (data sets) per year
  - Each experiment may have 200+ individual files

- Data rate has increased considerably

- Photon sources generate increasingly large volumes
  - Up to PBs per experiment on some beam lines
  - Neutron sources produce much more data than ever
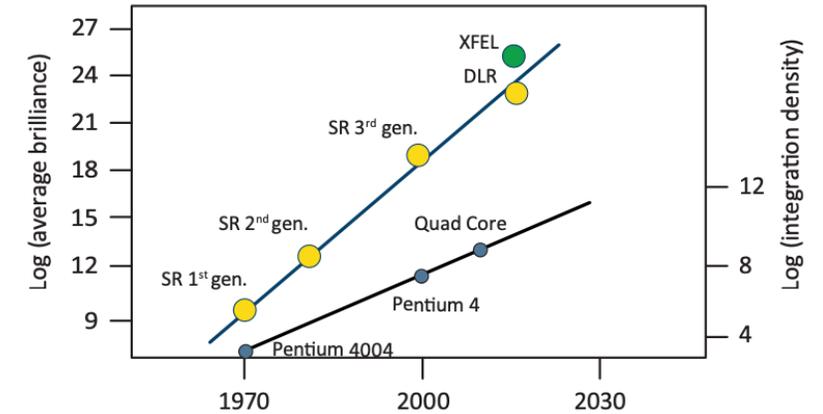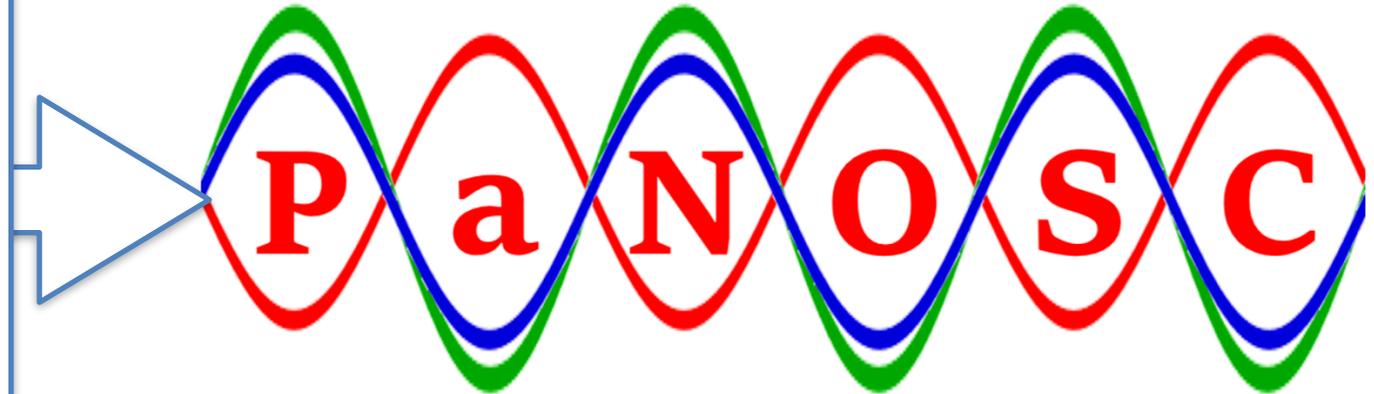 - ESS at full specification 250TB per day

Fig. 1. Comparison of the brilliance of X-rays delivered by photon facilities over the years (LEAPS' law) with the gain in transistor density in integrated circuits (Moore´s law). (DLR = Diffraction Limited Rings with ultimate performance; XFEL = X-ray free-electron laser)

Portal services

- **No commonality** between facilities
- Portal for user access, submission of proposals & safety are common
- Portal access to some open data catalogues
- Remote access to data analysis shrives – RDCs
- Some DAAS infrastructure

Photon and Neutron Open Science Cloud

- Should deliver data access and data processing
- Rich meta data capture infrastructure at current operating facilities limits use of data catalogues
- 100s of Applications for data processing and analysis
- 'Human' Portals are so far the most efficient & widely used

Services & EOSC

- **Be Domain Specific**

- Replicate current functionality

- Focused at our Community

- Federated across all facilities

- Sustainable

Current Local Services (each facility)

- Storage infrastructure
- Compute
- **Data management services**
- Data policy, Open access after embargo period

- Catalogues – ICAT, SciCAT
- DAAS services
- Data retrieval
- Remote analysis clusters

Central Infrastructure services

- Network NRENs
- National and EU level HPC, EGI, PRACE
- AAI Services
- Eduroam
- Edugain

- Umbrella
- DOI services
- Common data format