Fostering a FAIR Culture – what works?

Brian Matthews

Large-Scale Analytic Facilities

Key challenges of the 21st century

- energy, global climate, health and security
- study matter at small scales in high detail
  - from single atoms (10^{-10} m) to living cells (10^{-6} m) to whole systems (10^{-3}m - 1m)

High resolution "microscopes" ⟶ Intense beams of particles ⟶ Specialist sources

These sources require large scale research infrastructures that are beyond the capability of any single university or research group

Diamond

Photons (X-Rays) "see" electric charge – high atomic number nuclei

ISIS

Neutrons "see" nucleons – including hydrogen atoms

ExPaNDS
an Open Science Cloud Photon nd Neutron Data Services

Facilities around the World



50+ institutes around the world:

- More X-ray synchrotrons
- Neutrons
  - Reactors
  - Spallation sources
- Free-Electron Lasers

- > 50,000 users worldwide

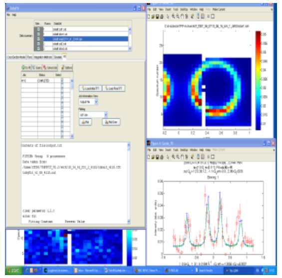



The science we do - Structure of materials

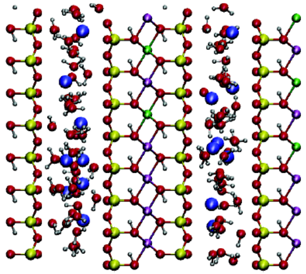
Visit facility on research campus

Place sample in beam

Diffraction pattern from sample



Fitting experimental data to model



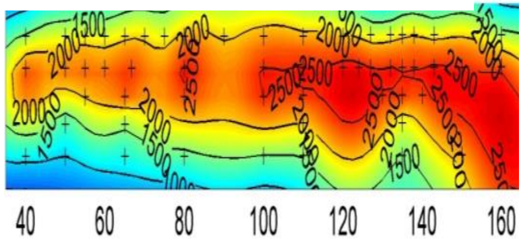
Structure of cholesterol in crude oil

- Run for external users
  - >30,000 users / year in Europe
- Support multiple techniques
  - Crystallography, tomography, SAS, spectroscopy, EM, …
- Diverse science, backgrounds and expertise
  - physics, chemistry, biology, medicine,
  - energy, environmental, materials, culture
  - pharmaceuticals, petrochemicals,

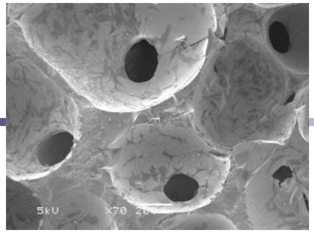
- 800+ experiments (30 parallel)
- Instruments: 6KB/sec to 6GB/sec
- Billions of € of investment
  - c. £400M for DLS
  - + running costs
- Over 5.000 high impact publications per year in Europe

Big Facilities, Small Science

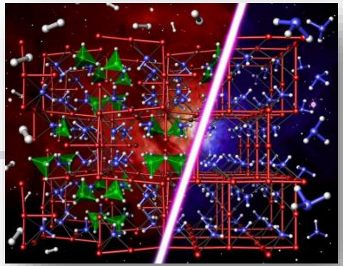
Longitudinal strain in aircraft wing



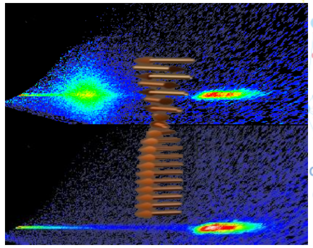
Bioactive glass for bone growth



Hydrogen storage for zero emission vehicles

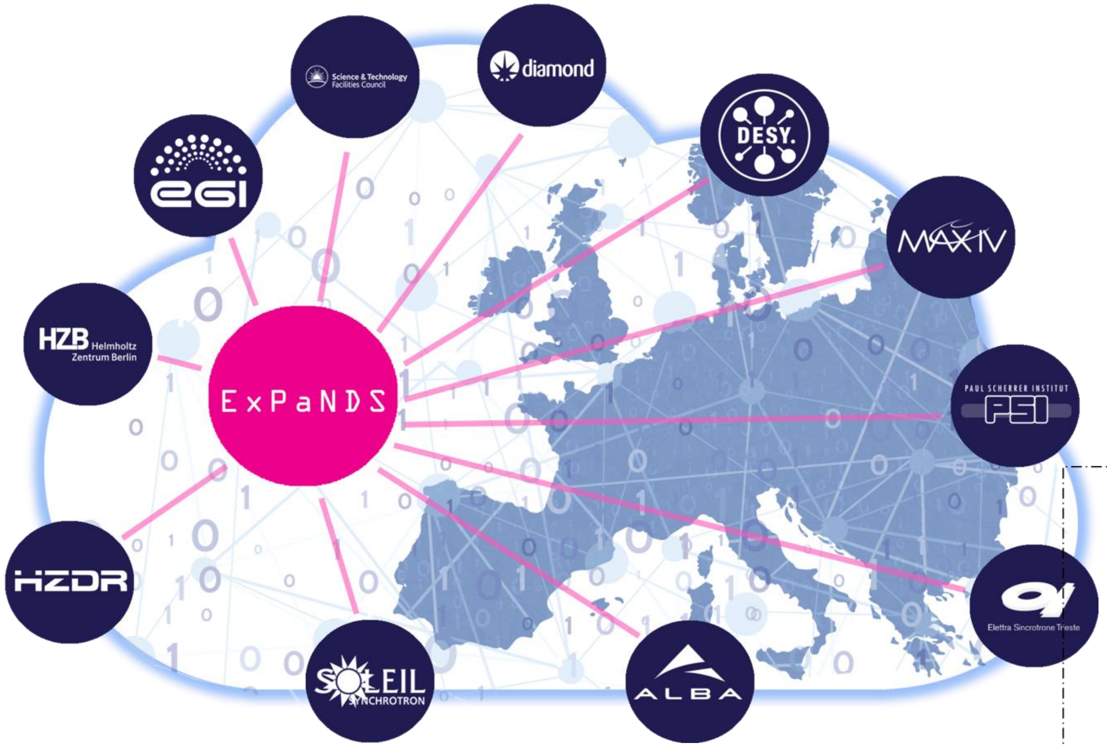


Magnetic moments in electronic storage



ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

ExPaNDS: Bringing National Facilities into the EOSC



- INFRAEOSC-5b Project
- Started Sept 2019

Enabling **FAIR** data

EOSC data **catalogue services**

EOSC data **analysis services**

Training

Outreach Dissemination Communication

Vision

ExPaNDS will contribute to the realisation of the FAIR Data Commons for Neutron and Photon science, making it a real and productive working tool for the scientists from the numerous existing and future disciplines exploiting data from Photon and Neutron sources.

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

What works for FAIR: developing common policies

- PaN-data Data Policy Framework, 2011 http://pan-data.eu/
- Came from a collaboration of data infrastructure providers in the Facilities
- Open Data Policy
  - Public data in the public domain
  - Collection and publication of metadata
  - Storage period
  - Embargoed access
  - Data Publication with PIDs
- Facilities can influence their communities
  - Need to take communities along
  - Publication of metadata can be sensitive as data
  - Persuading the community to accept open data
- Revision in the light of FAIR guidelines
  - Part of the ExPaNDS programme, in conjunction with PaNOSC

pandata
Photon and Neutron Data Infrastructure

Adoption :

ILL- PanData data policy since 2012

ISIS-PanData data policy since 2012

Elettra- PanData data policy since 2013

ESRF-PanData data policy since November 2015

HZB-PanData data policy since June 2016

MAXIV-PanData data policy since 2015

HZDR-PanData data policy since June 2016

PSI-PanData data policy since August 2016

EUXFEL-PanData data policy since August 2017

ALBA-PanData data policy since July 2017

DESY-PanData data policy since August 2017

Diamond PanData data policy since 2019

SOLEIL PanData data policy since 2018

FELIX PanData data policy since 2019

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services

What works for FAIR: Building common tools

Metadata catalogues

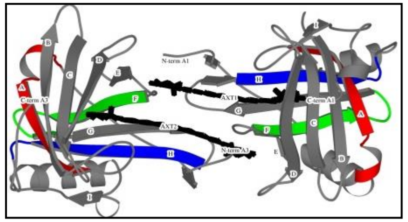
Metadata catalogues provide a metadata catalogue and related components to support Large Facility experimental data,

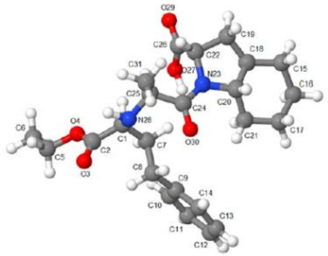
- common structural metadata
- linking all aspects of the research chain from proposal through to publication.
- Automated population
- Linked to data storage for access and retrieval

What works for FAIR: Using Community Standards

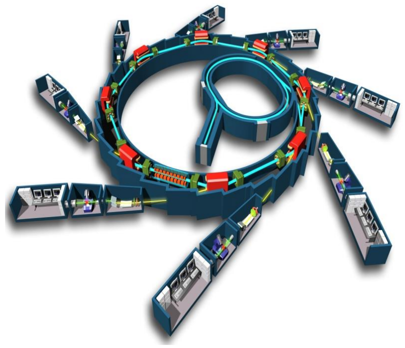
- Crystallography has had a long history of defining its own data standards
  - A Crystallographic Information Framework with a controlled vocabulary, controlled attribute values/types, and concrete formats (https://cif.iucr.org).
- Recognised data repositories for crystal structures
  - CCDC, PDB
  - Recognised by IUCr Journals for supplementary data sets
- Raw diffraction data is also recommended
  - Much of this is at synchrotrons

- ***Data validation and data quality assessment are needed***



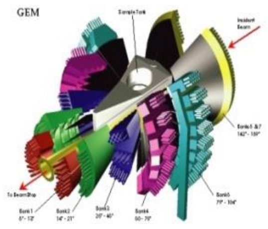


What is still needed ? Bringing FAIR to the Experiment



- Data Policy applies at the Facility level
  - This needs to be enforced for each experiment

- Data Management Planning for each experiment
  - Metadata to be collected, data storage, connections to derived data



- DMP for an experiment needs to be done in context:
  - The DMP for its instrument and technique
  - The community norms for the discipline
  - The DMP of the user's institution and funder

- Guidelines on bringing FAIR to the Experiment
  - Work of the ExPaNDS

ExPaNDS
European Open Science Cloud Photon and Neutron Data Services