OAK RIDGE National Laboratory

# Thoughts and lessons learned on software development for the European spallation source project

Jon Taylor

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

# ESS DMSC

All scientific computing for ESS instruments, excluding low level controls

**Scope**

DAQ and experiment control

Data processing

Data analysis

Data management & Curation

Business systems (User office software, experiment scheduling)

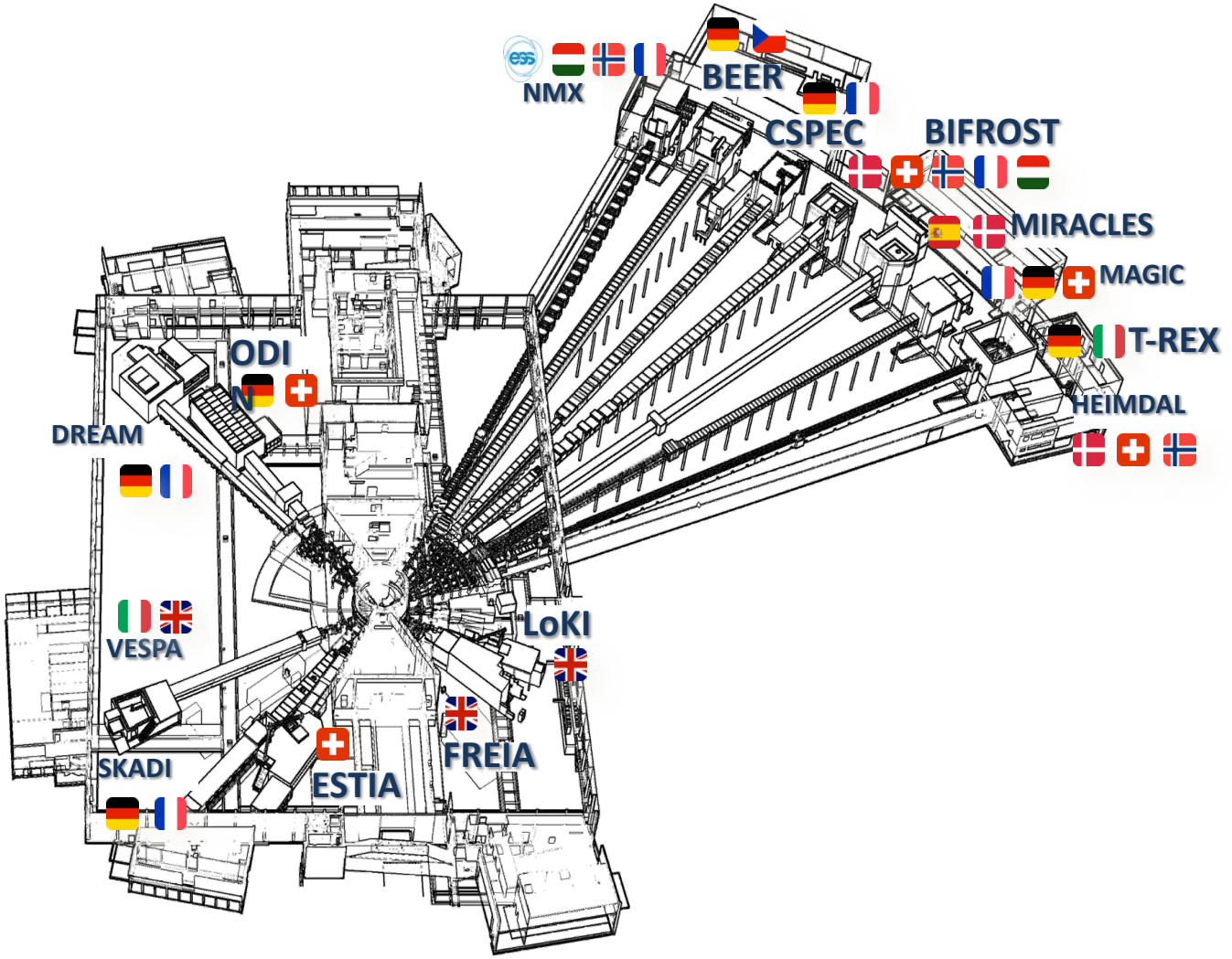
Storage & compute in 2 connected data centers

**2013 – 2020** core development of frameworks and systems – Budget 20.1M

Including a project rebaseline 2018

**2020 – 2022** – Instrument specific development from operational funding ~ 5M per year

project rebaseline 2021

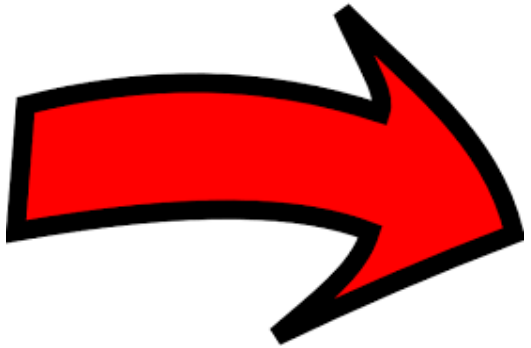


# Scientific Computing at ESS.

The objective was to enable live processing and analysis of data

How To Improve Scientific Efficiency (A TV series analogy)

Move from Post Mortem Analysis. To Live Analysis

# Strategy

## Hope is not a plan. – M. Hagan

**Prioritize.**

- DAQ & Controls -> Data reduction -> Data analysis
  - Modelling and simulation was de-scoped in 2015 and remains so.

Prioritize *Performance* & software quality. Perfection is not required, good is good enough.

Agreement, buy-in and alignment from all stakeholders

Build functional and performant frameworks

- Recognize that some things just will not work (i.e. Mantid)
- Use open source

Run to schedule and budget

- Develop a credible plan that everyone understands and agrees.

Integrate systems and staff – we developed software collaboratively across multiple sites ESS, PSI, STFC, FZJ

Build data pipelines.

Build social capital and good work environment.

# The Team: ESS staff, In-Kind Partners and Collaborative Research

Chalmers University of Technology

BROOKHAVEN NATIONAL LABORATORY

JÜLICH FORSCHUNGSZENTRUM

PAUL SCHERRER INSTITUT PSI

Science & Technology Facilities Council ISIS

# Our Agenda: Scientific Computing Enables Discovery

## Data Management & curation is prioritised

Collaborative Open Source Software

Developed with modern practices

Minimise Single point failures

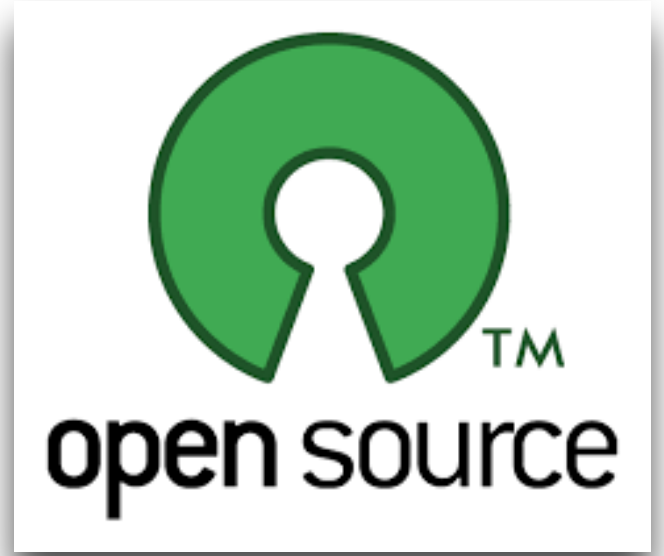
Community Engagement

Promote User Experience

Actively manage interfaces and APIs

Deliver FAIR Data

Promote Open Science


# Stakeholder Engagement, Governance, Leadership & Training

## Develop buy-in and manage expectations.

Regular advisory panel meetings 2 per year reporting to SAC and council.

Multiple technique and instruments workshops.

User training – python, Scipp, mcstas …

Major projects have specific governance and oversight.

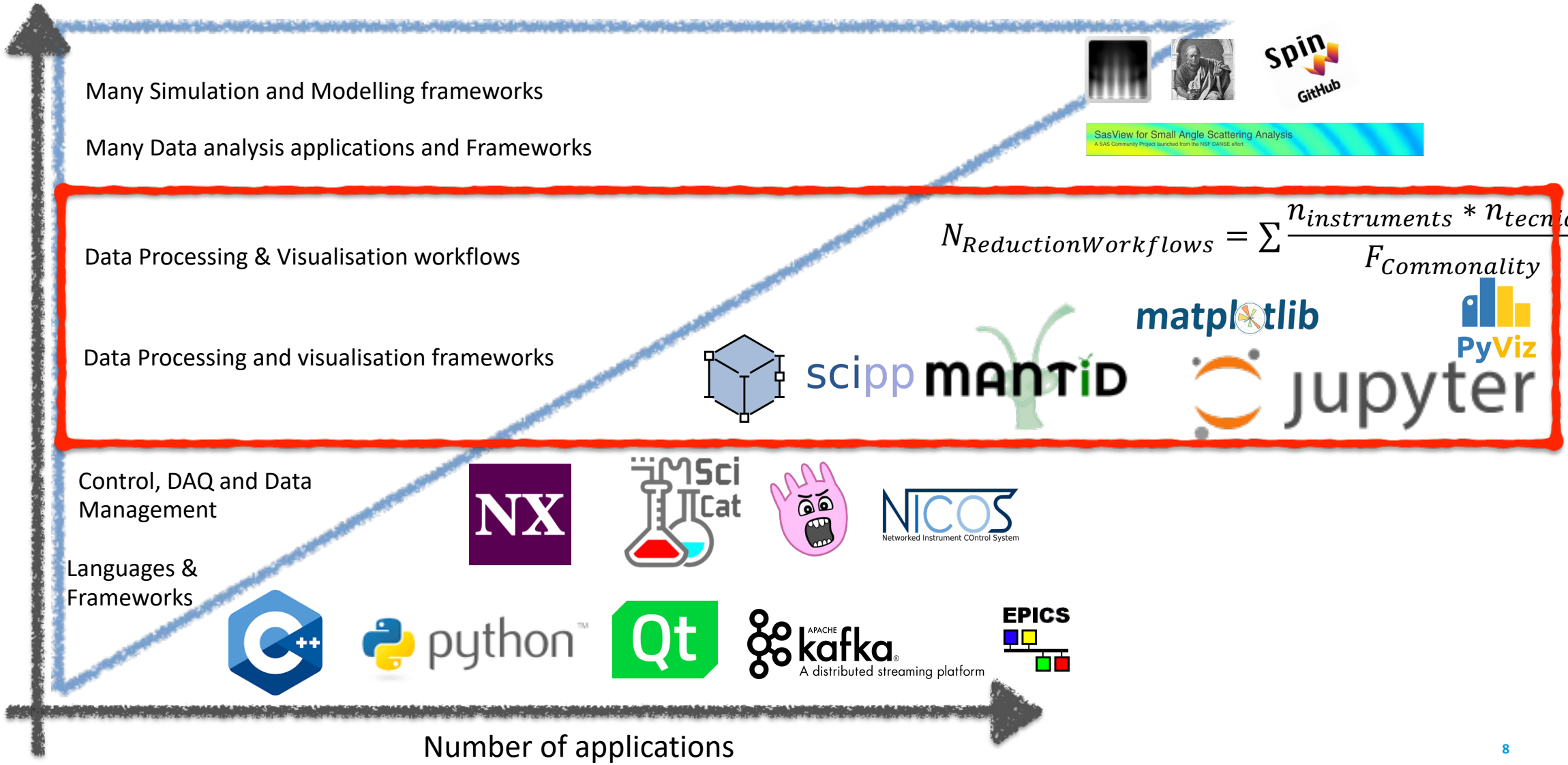
Development of facility policies play a key role for data and data services.

Project governance changes with development stage

Develop and curate a good work environment.

Provide a meaningful vision and set realistic and achievable targets.

# Scientific Application Stack ESS (& Neutron scattering)



# SCIPP Data Processing (~$4.5M special in-kind rate ☺)

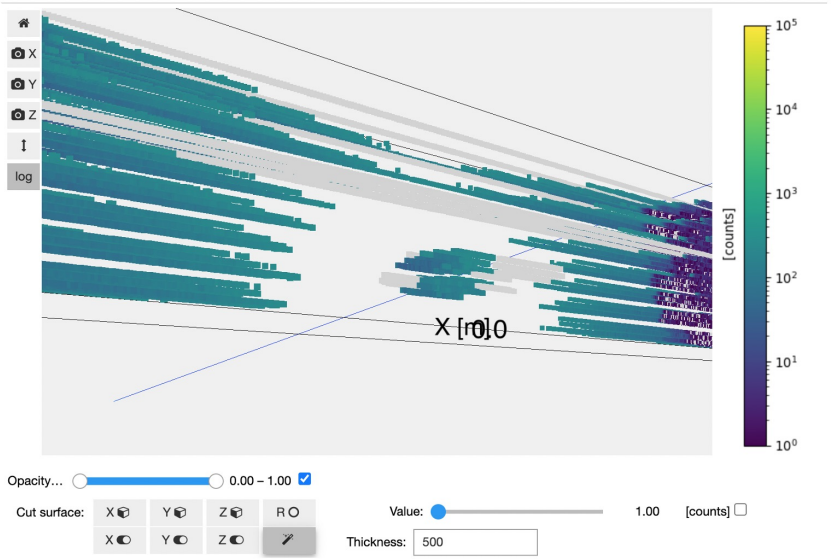
## ESS has developed a fit for purpose processing Framework

Mantid (>$50M) was not performant for ESS and too costly to develop.

Scipp is C++ / Python

Engineered using the lessons learned from Mantid

- Modularity and separation of functionality
- Collaboration with NSLSII
- Interest from other Neutron centers.
- Developed with modern Python interfaces in mind.
- Interoperable with the python Ecosystem
  - Including Mantid




*scipp visualisation of LoKi detector prototype.*

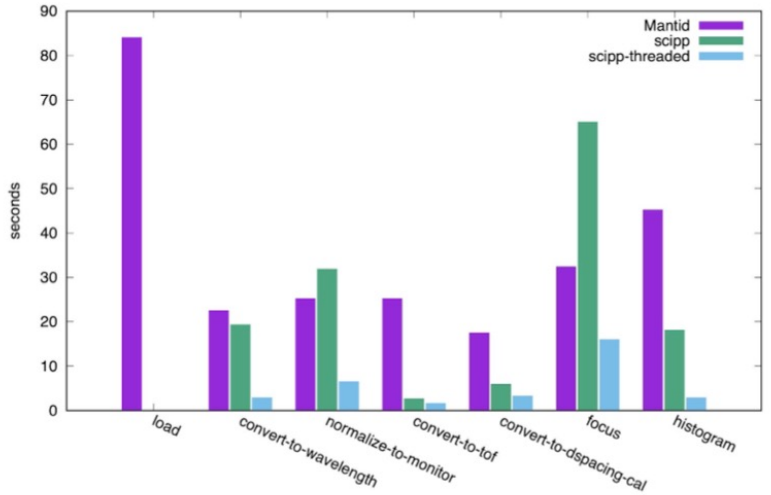


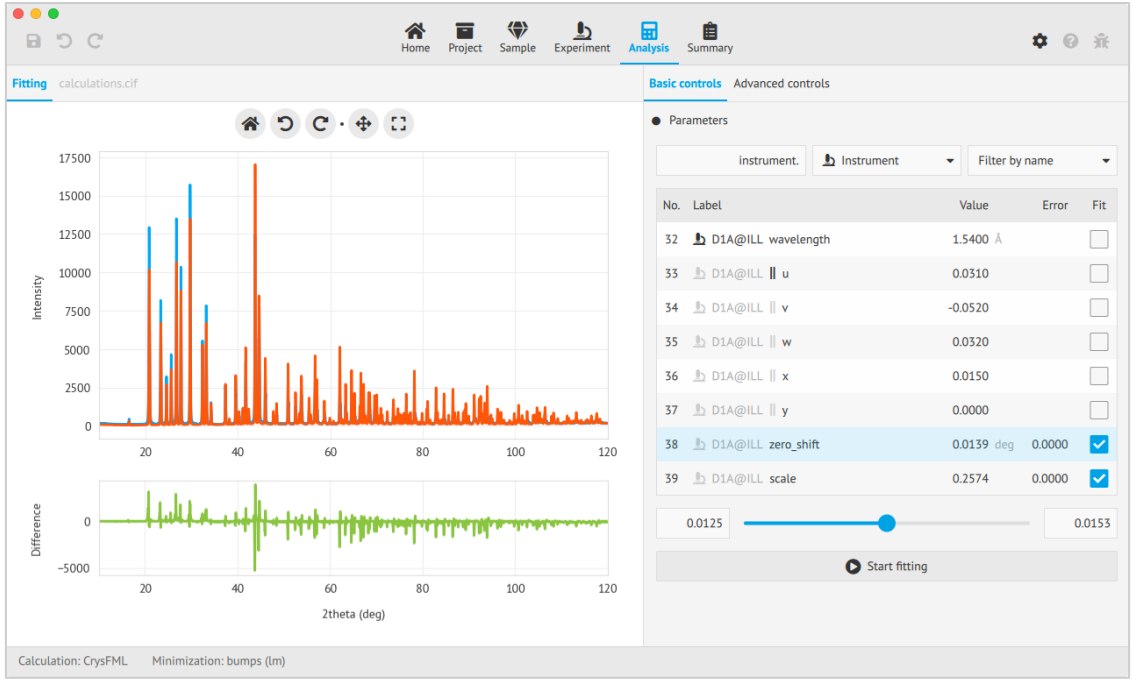

***Figure 2.*** *Timing of major steps in data reduction workflow for powder diffraction with $3\times10^9$ events.*

# Easy Science Framework (~ $1.4M)

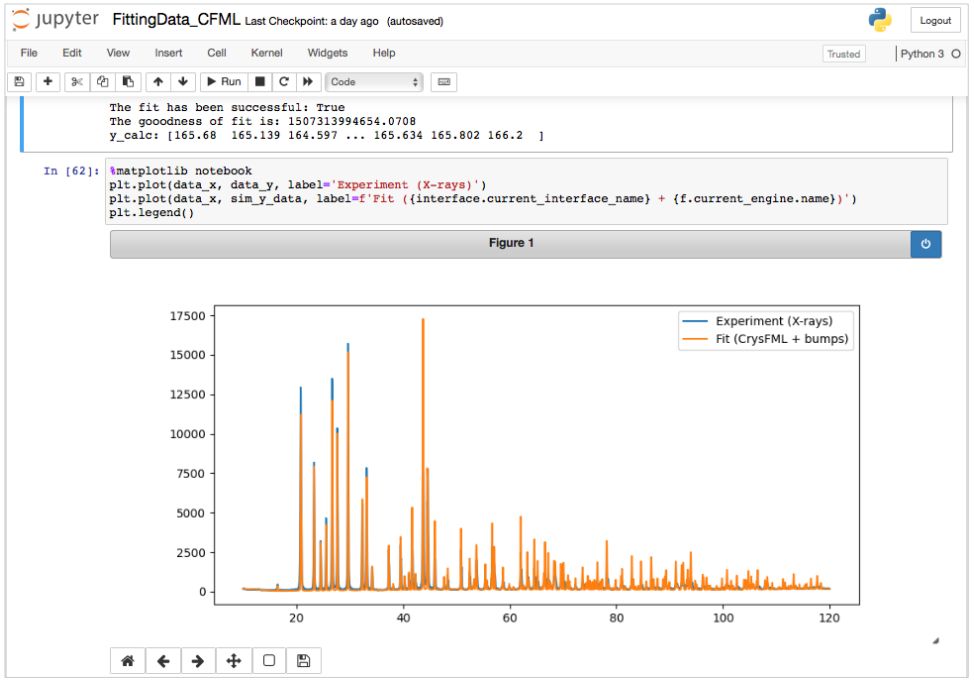
## A framework for data analysis: Provides a sustainable path forwards

- Modular structure: easyCore, easyDiffractionLib, easyDiffractionApp easyReflectometry, etc.
- Multiple minimizers (easyCore): Lmfit, Bumps, DFO-LS
- Multiple calculation engines (easyDiffraction): CrysPy, CrysFML, GSAS-II
- and much more: github.com/easyScience

### easyDiffractionApp GUI




### easyDiffractionLib in Jupyter



# Research Data management (~ $1M)

## FAIR, Data catalogue, Data Policies, Data Management Plans

ESS collaboratively developed SciCAT

https://scicat.ess.eu/

DMSC developed Sci Chat

Integrations within the user office system

Understanding the facility workflow is essential for effective Research data management

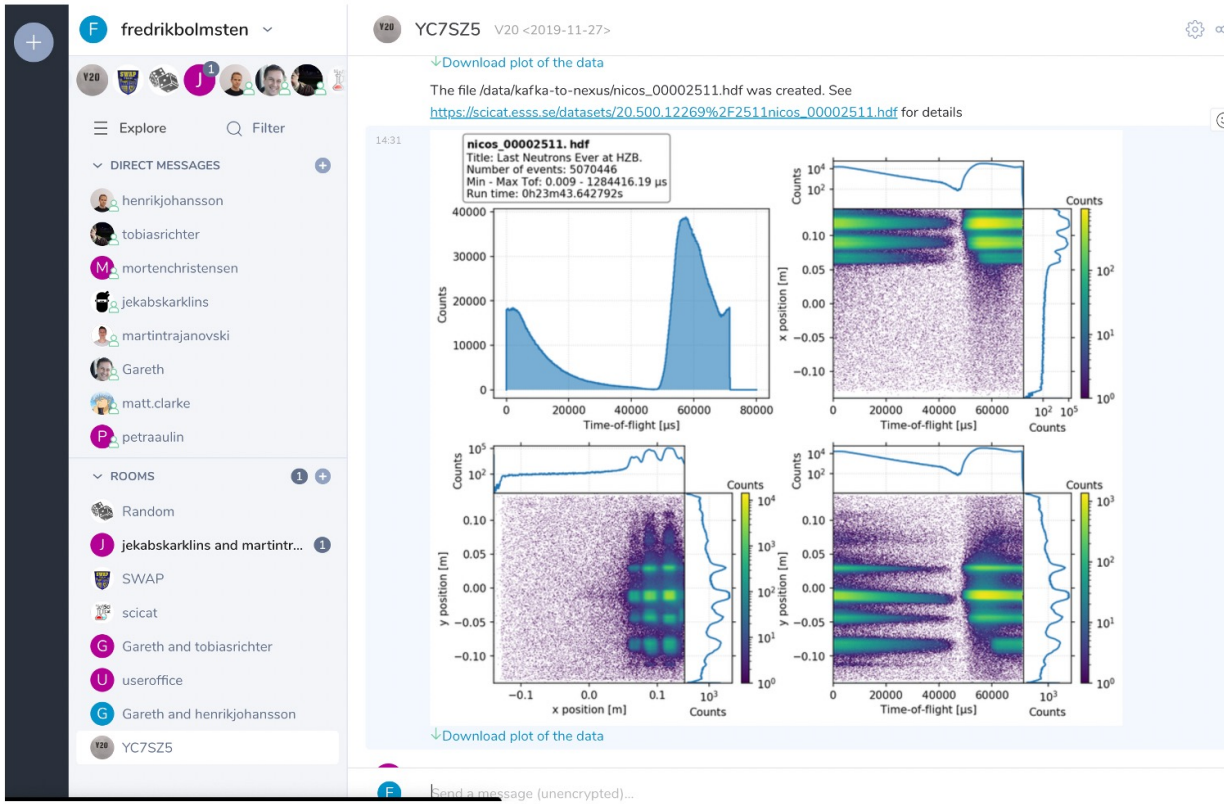



**Figure 4. Automated graphs posted to SciChat during an experiment**

# Data Management and Business Software ($0.7M)

## Tying together Data management plans and data Services

ESS proposal system development.

- Modern architecture & Open Source

- Includes data management plan functionality

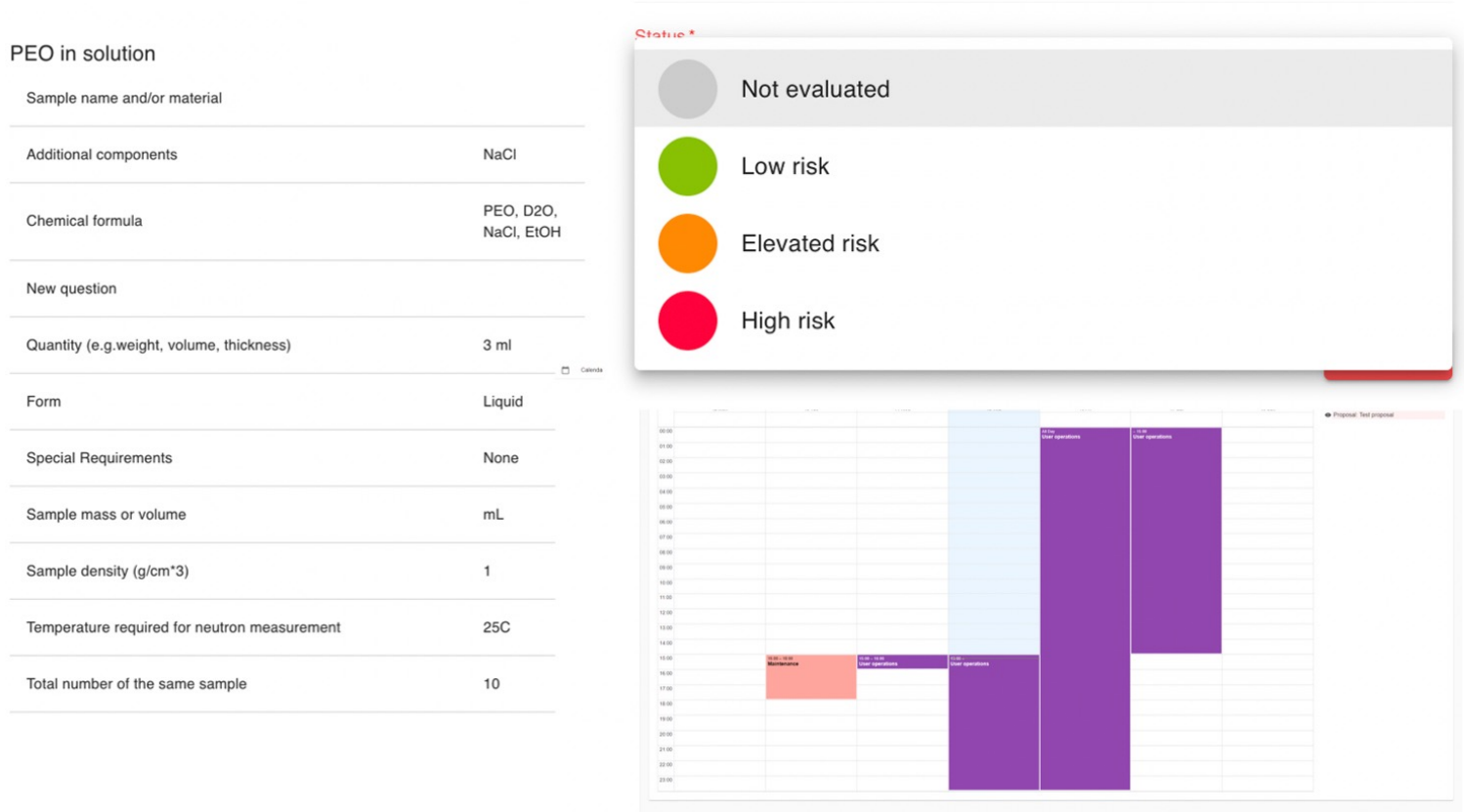



**Figure 3. Showing the interface for sample review and first prototype of scheduling software(bottom right)**

# Lessons Learned

## Know the schedule, understand the tipping point.

DMSC completed its construction scope ~ 2021

- i.e. systems were available to operate instruments and support early operations.

ESS beam on target (not early operations) will be in 2025 -2026

Too early for certain areas?

- Some systems will need to be refactored before ESS operates
- Data catalogue, some areas of the user office software.
- Considerable advantage in co development of controls and timing. The instrument controls drove a lot of the ICS early development

Slowing down is hard once momentum is gained.

Manage stakeholders get agreement and alignment on scope and functionality.

Integrate with the other key areas of the organization.

Develop systems that can be maintained and operated within the ops budget

Learn from other domain projects

Try not to reinvent the wheel