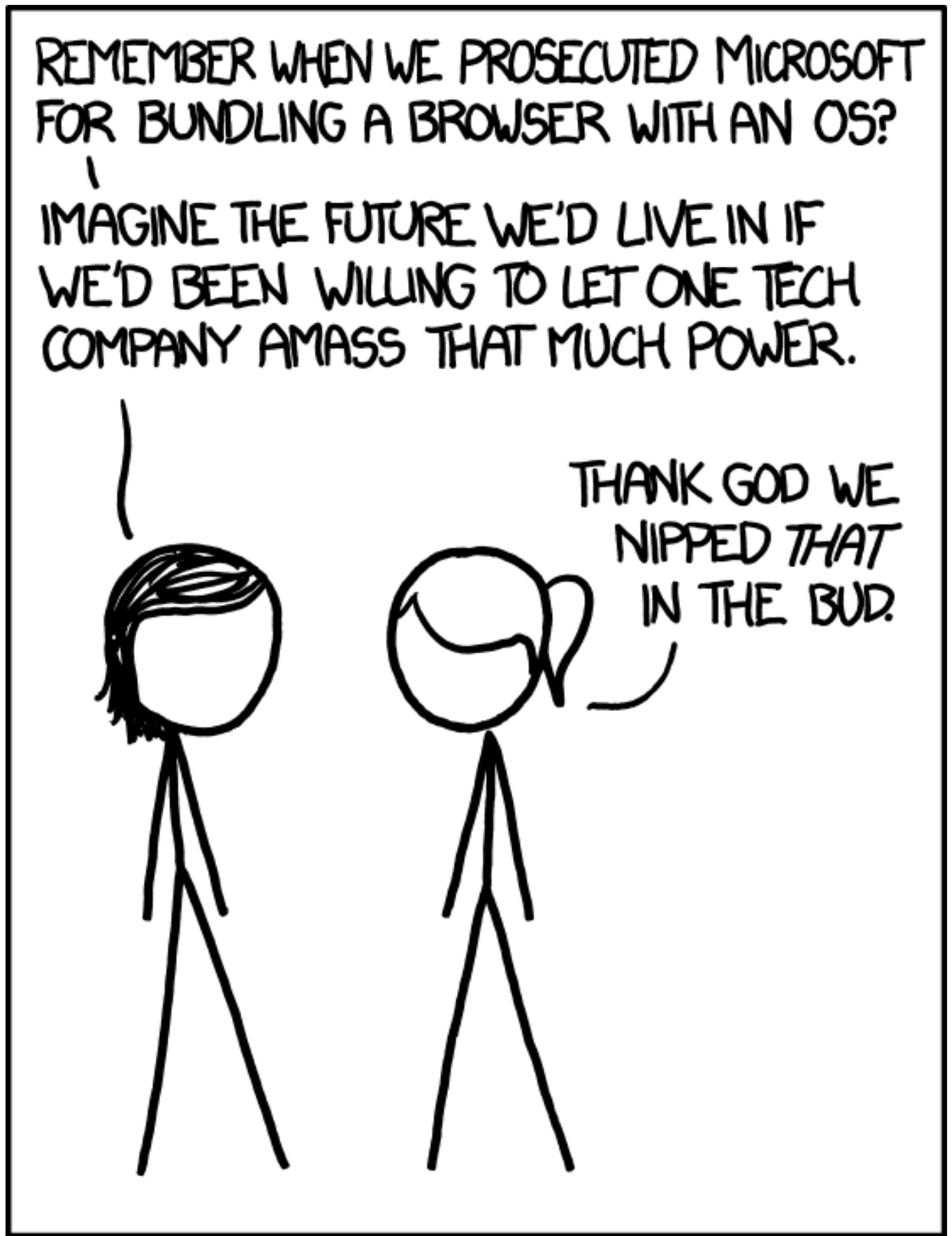
“Software”

Jon Taylor

Erice school 2018

10 June 2018

Microsoft Office



Office 365

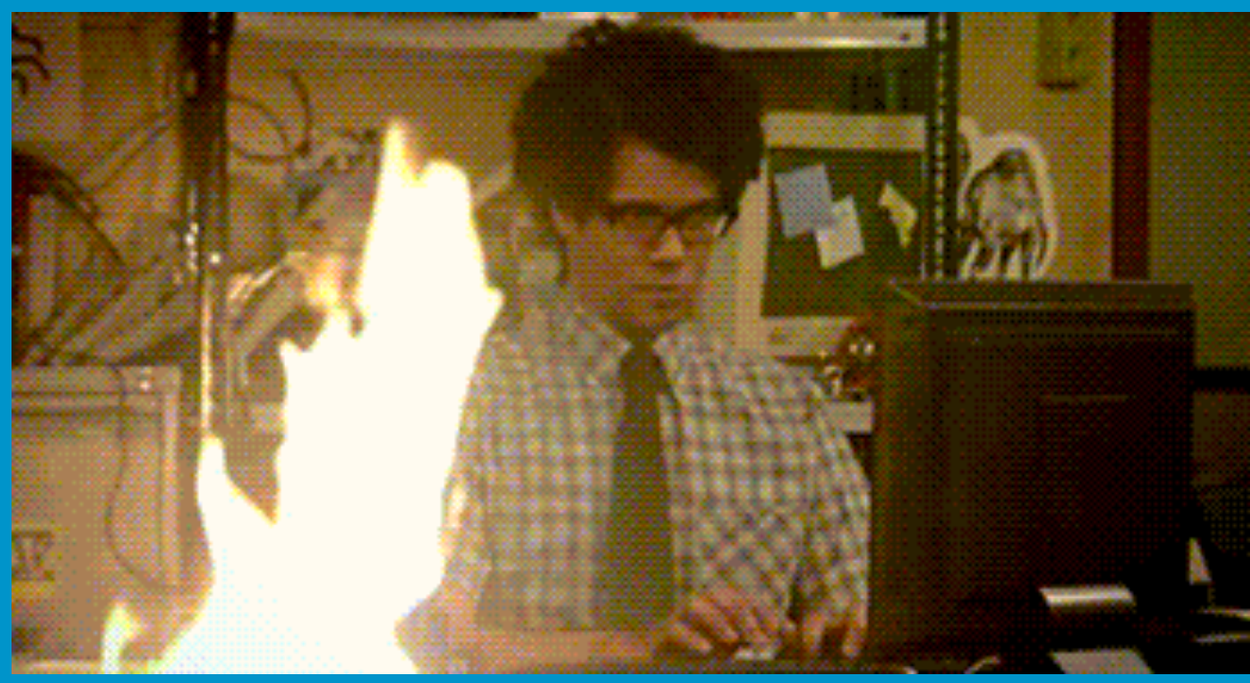
"~~Software~~"

Open Data & Open Science: How Scientific Software Development Can Make Data FAIR

Jon Taylor

Erice school 2018

10 June 2018



Jon Taylor

EUROPEAN SPALLATION SOURCE

Head of ESS Data Management and software Centre

Mari instrument scientist at ISIS facility

Project scientist for Mantid project

Build and commissioning for Merlin and LET

Neutron scatterer INS and polarised neutrons

Photon scatterer Compton scattering / XMCD

Very occasional MuSR user

Interests

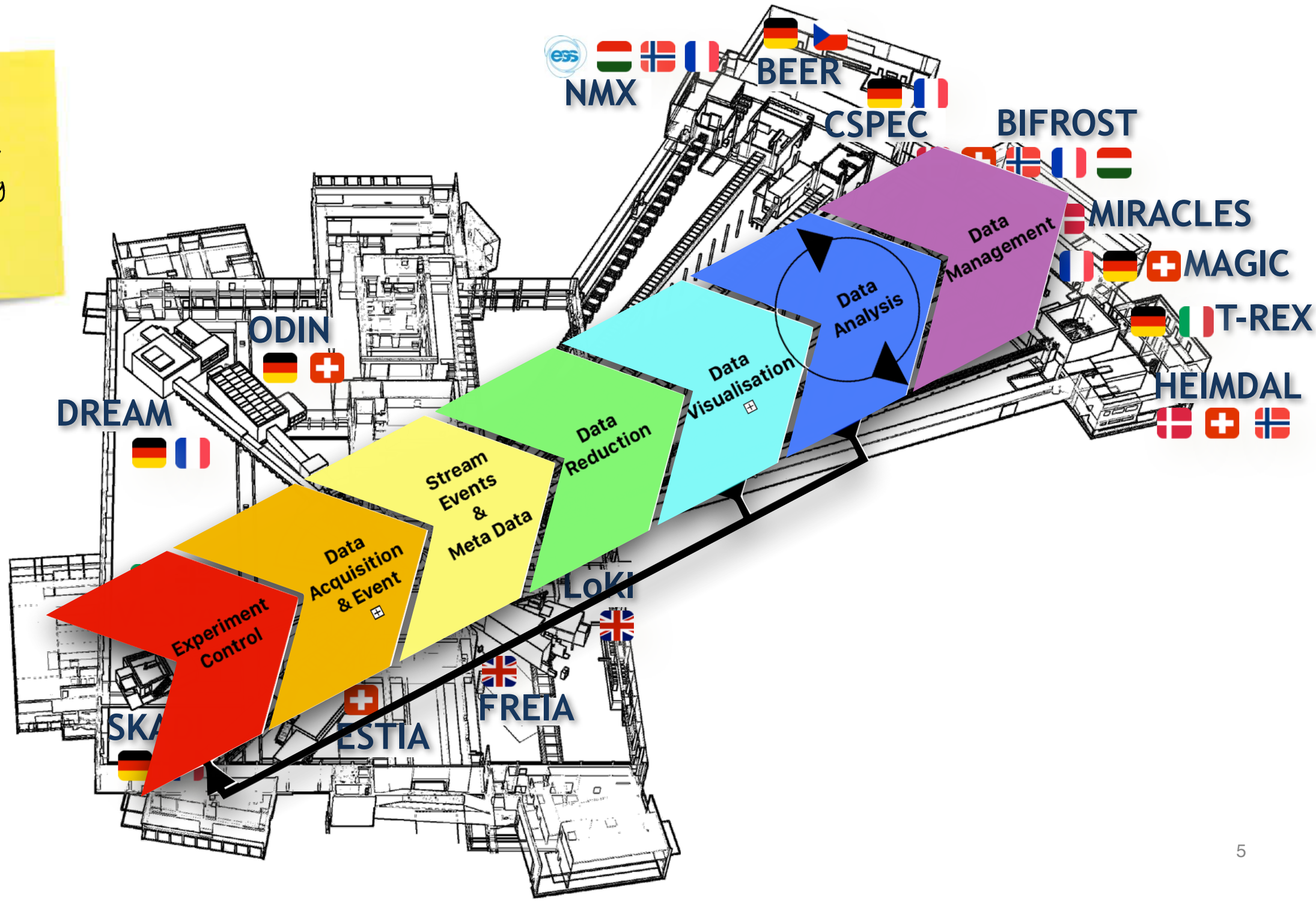
Strongly correlated electron systems

Frustrated magnetism

Scientific Computing

Data Management and Software Centre

TO DO:
Scientific computing for this



Data Management and Software Centre

EUROPEAN SPALLATION SOURCE

Provide world leading scientific software and scientific computing support for neutron scattering at ESS

- Construction budget 20M euro
- Staff 2018 27 + 8
- Staff 2028 60

Scientific Software development.

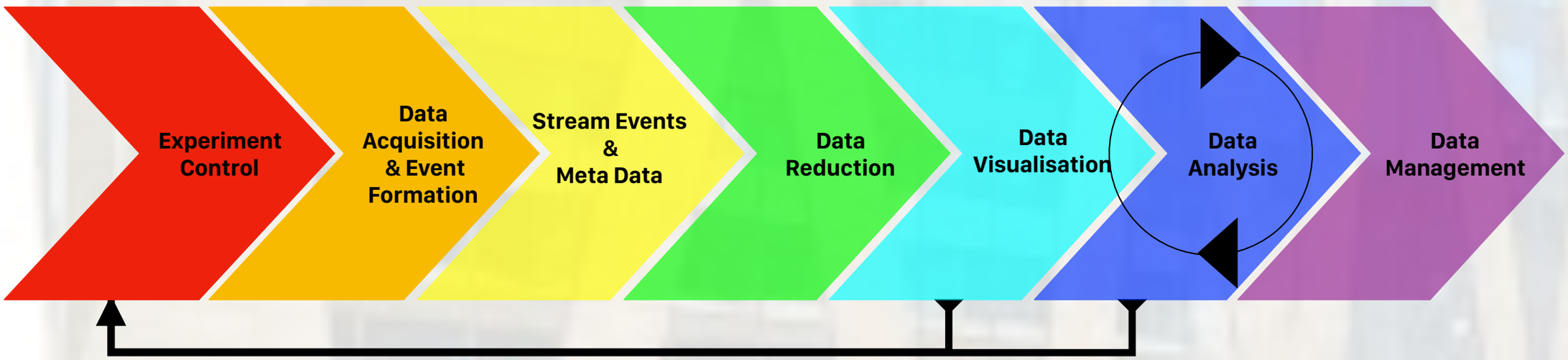
- Experiment control
- Data acquisition system
- Data reduction, analysis & modelling

Data centre operations.

- Dual location - Lund & Copenhagen
- Data management and curation

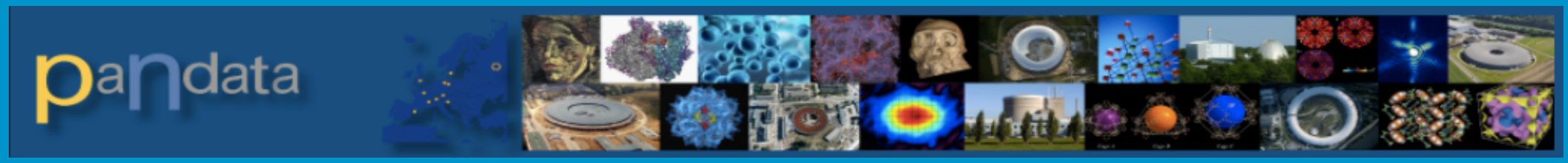
User programme support

- Instrument Data scientists
- User office software
- Remote access to data and software tools



European facilities landscape

- **10 photon sources**
- **7 neutron sources**
- **Represents ~40 000 users/year**
- **PaN is not CERN**



	alba	anka	als	ansto	aps	australian	cls	desy	dls	elettra	esrf	frm2	hzb-ber	hzb-bessy	ill	isis	lcls	lnls	maxIV	nsls	nsrrc	ornl	sinq	sls	sacla	soleil	spring8	ssls	ssrl
alba	1745	31	201	74	171	74	82	189	394	233	982	78	36	229	262	167	41	40	47	65	83	86	81	315	7	213	97	8	79
anka	31	1525	202	56	127	73	67	233	223	115	643	64	31	212	120	95	60	15	40	46	48	42	47	184	6	114	67	14	54
als	201	202	19761	682	3463	780	1675	1125	2234	968	3390	430	206	1403	937	985	1185	178	400	1042	992	1009	310	2110	107	755	1364	130	3479
ansto	74	56	682	3699	665	864	313	326	730	222	1123	479	340	339	1192	1065	134	47	111	222	398	610	386	472	40	178	561	63	381
aps	171	127	3463	665	7699	663	1392	805	1535	514	2837	371	177	612	921	930	805	113	231	779	560	1317	294	1388	93	421	792	93	2092
australian	74	73	780	864	663	6075	336	292	841	312	1147	166	83	271	412	472	174	59	119	198	398	300	123	517	22	187	430	85	531
cls	82	67	1675	313	1392	336	5593	386	776	315	1206	175	78	450	367	361	228	113	173	408	470	407	129	720	34	287	504	99	1003
desy	189	233	1125	326	805	292	386	8508	1171	780	3228	486	194	1450	749	559	834	130	472	299	292	322	276	1401	96	619	605	88	488
dls	394	223	2234	730	1535	841	776	1171	17843	979	6881	501	222	1217	1473	2337	653	223	467	574	639	649	374	2335	71	939	887	94	1163
elettra	233	115	968	222	514	312	315	780	979	7621	2519	210	93	1007	592	467	458	175	316	248	243	264	177	929	62	894	472	57	390
esrf	982	643	3390	1123	2837	1147	1206	3228	6881	2519	40207	1225	624	2939	3510	2454	977	643	1101	1146	868	1105	932	5225	176	3124	1632	125	1823
frm2	78	64	430	479	371	166	175	486	501	210	1225	3769	504	472	1558	838	111	63	96	146	209	559	615	474	27	207	257	26	208
hzb-ber	36	31	206	340	177	83	78	194	222	93	624	504	1817	364	755	532	52	45	43	60	92	305	425	231	13	68	155	10	74
hzb-bessy	229	212	1403	339	612	271	450	1450	1217	1007	2939	472	364	9272	752	585	512	141	582	323	361	369	300	1602	54	669	571	60	571
ill	262	120	937	1192	921	412	367	749	1473	592	3510	1558	755	752	10610	2630	227	183	264	333	423	1188	1049	982	54	641	763	55	441
isis	167	95	985	1065	930	472	361	559	2337	467	2454	838	532	585	2630	8442	239	120	230	358	406	1056	762	816	49	385	800	45	480
lcls	41	60	1185	134	805	174	228	834	653	458	977	111	52	512	227	239	3295	23	112	231	192	213	81	707	141	351	453	36	818
lnls	40	15	178	47	113	59	113	130	223	175	643	63	45	141	183	120	23	5802	45	40	61	59	40	162	5	165	82	9	102
max	47	40	400	111	231	119	173	472	467	316	1101	96	43	582	264	230	112	45	3220	126	123	108	72	471	25	260	160	30	201
nsls	65	46	1042	222	779	198	408	299	574	248	1146	146	60	323	333	358	231	40	126	2296	301	392	100	480	42	239	342	62	620
nsrrc	83	48	992	398	560	398	470	292	639	243	868	209	92	361	423	406	192	61	123	301	3560	478	147	523	37	184	843	101	568
ornl	86	42	1009	610	1317	300	407	322	649	264	1105	559	305	369	1188	1056	213	59	108	392	478	4766	437	551	34	186	662	63	582

Open Science and Open Data in the EU

European Commission

The Commissioner's vision

"Europe's final transition must be one from fragmented data sets to an integrated European Open Science Cloud. By 2020, we want all European researchers to be able to deposit, access and analyse European scientific data through a European Open Science Cloud.."

Speech by Commissioner Carlos Moedas in Amsterdam, NL:
"Open science: share and succeed", 4 April 2016

What is FAIR

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Research data

» Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*[#]

Received: 10 December 2015

Accepted: 12 February 2016

Published: 15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Scientific Efficiency *needs* to be FAIR

EUROPEAN SPALLATION SOURCE

Findable

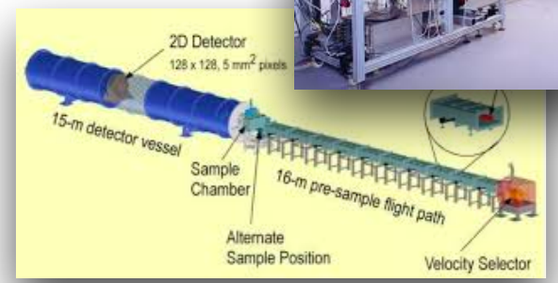
- ☑ Searchable rich meta data
- ☑ Persistant identifiers
- ☑ Experiment type
- ☑ Sample parameters
- ☑ Instrument parameters
- ☑ Applies to raw and proceed data
- ☑ What software?
- ☑ What parameters
- ☑ what is the repository
- ☑ Where is it stored

Accessible

- ☑ Data policy
- ☑ Data format
- ☑ Storage architecture
- ☑ Data catalogues
- ☑ APIs
- ☑ Authentication
- ☑ Security
- ☑ Applies to raw and processed data
- ☑ Applies to software

Interoperable

- ☑ Compatibility
- ☑ Data format
- ☑ Meta data standards
- ☑ Allow for multi modal data analysis
- ☑ Software APIs
- ☑ Catalogue interoperability
- ☑ Standards



Reusable

- ☑ Data shelf life
- ☑ Software shelf life
- ☑ Open access
- ☑ Backwards compatibility

90%* of neutron users are not computing experts

Scientific Computing adds value

* This is an estimate

At some point in everyones career …

What to think about to avoid disasters ?

Make your data and code FAIR

Collaborative Free Open Source Software

EUROPEAN SPALLATION SOURCE

Sustainability …

Will todays software be available in 2025 ?

What happens if the developers all leave or worse?

What if you're the developer?

Open Source is not ‘Free’

EUROPEAN SPALLATION SOURCE

Software developers have to eat

Licenses are Complex

EUROPEAN SPALLATION SOURCE

Software developers have to protect their IP (if they want)

Licenses of dependencies matter

Software licenses and rights granted in context of the copyright according to Mark Webbink.[1] Expanded by freeware and sublicensing.

Rights granted	Public domain	Permissive FOSS license (e.g. BSD license)	Copyleft FOSS license (e.g. GPL)	Freeware/Shareware/ Freemium	Proprietary license	Trade secret
Copyright retained	No	Yes	Yes	Yes	Yes	Very strict
Right to perform	Yes	Yes	Yes	Yes	Yes	No
Right to display	Yes	Yes	Yes	Yes	Yes	No
Right to copy	Yes	Yes	Yes	Often	No	No
Right to modify	Yes	Yes	Yes	No	No	No
Right to distribute	Yes	Yes, under same license	Yes, under same license	Often	No	No
Right to sublicense	Yes	Yes	No	No	No	No
Example software	SQLite, ImageJ	Apache web server, ToyBox	Linux kernel, GIMP, OBS	Irfanview, Winamp, *League of Legends*	Windows, *Half-Life* series, Spotify, xSplit	Server-side Games by Blizzard Entertainment, Rockstar, Activision, etc. PlayStation Network and Xbox Live TIDAL

Take a Professional Approach …

Think About Architecture

- Design Software
- Code Quality

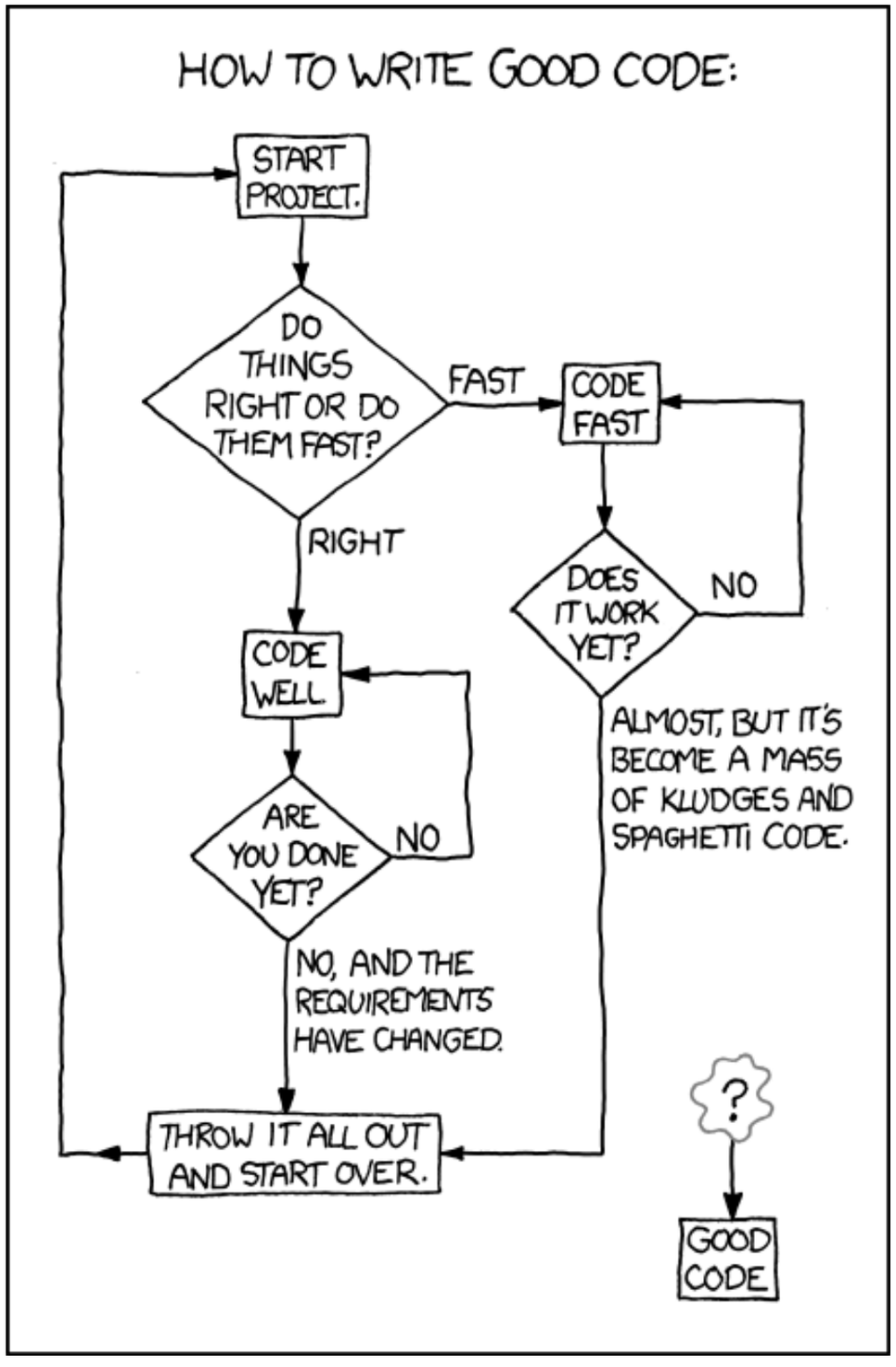
Think about infrastructure

- Version Control
- Testing
- Review

Don’t Reinvent

Keep It Simple

Be Agile ?



Architecture

There are a lot!

The point is to make code reusable / modular / testable

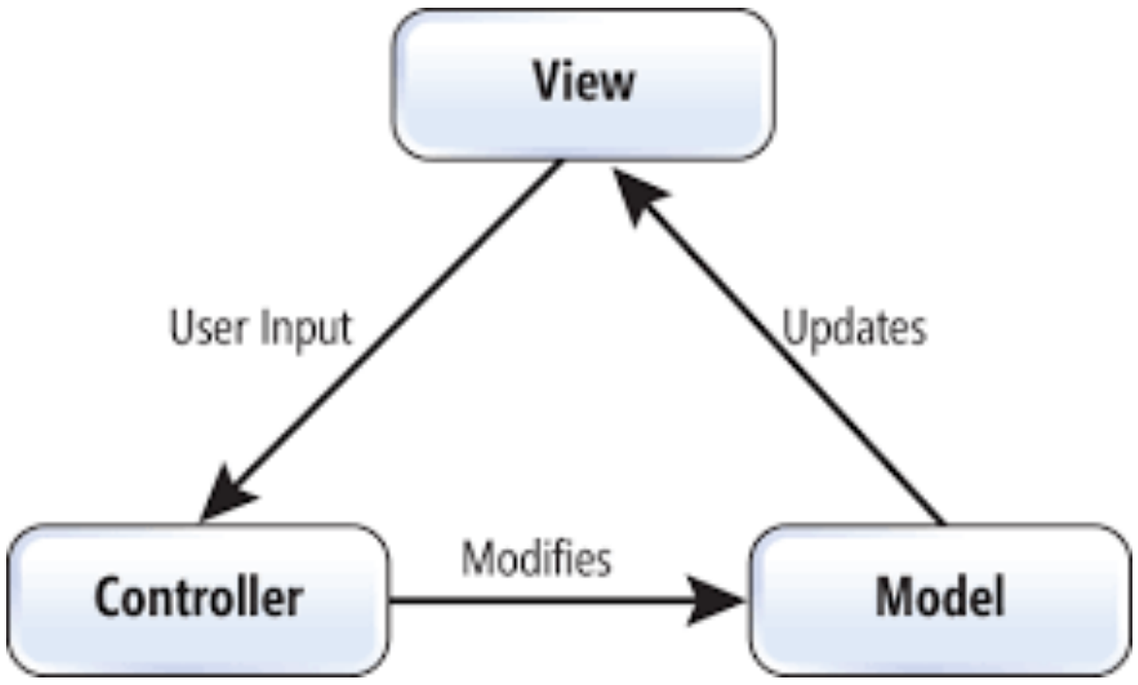
Take a look on wikipedia or at …

https://github.com/faif/python-patterns

Model View Controller design pattern

Idea : Keep the GUI code separate from the logical code

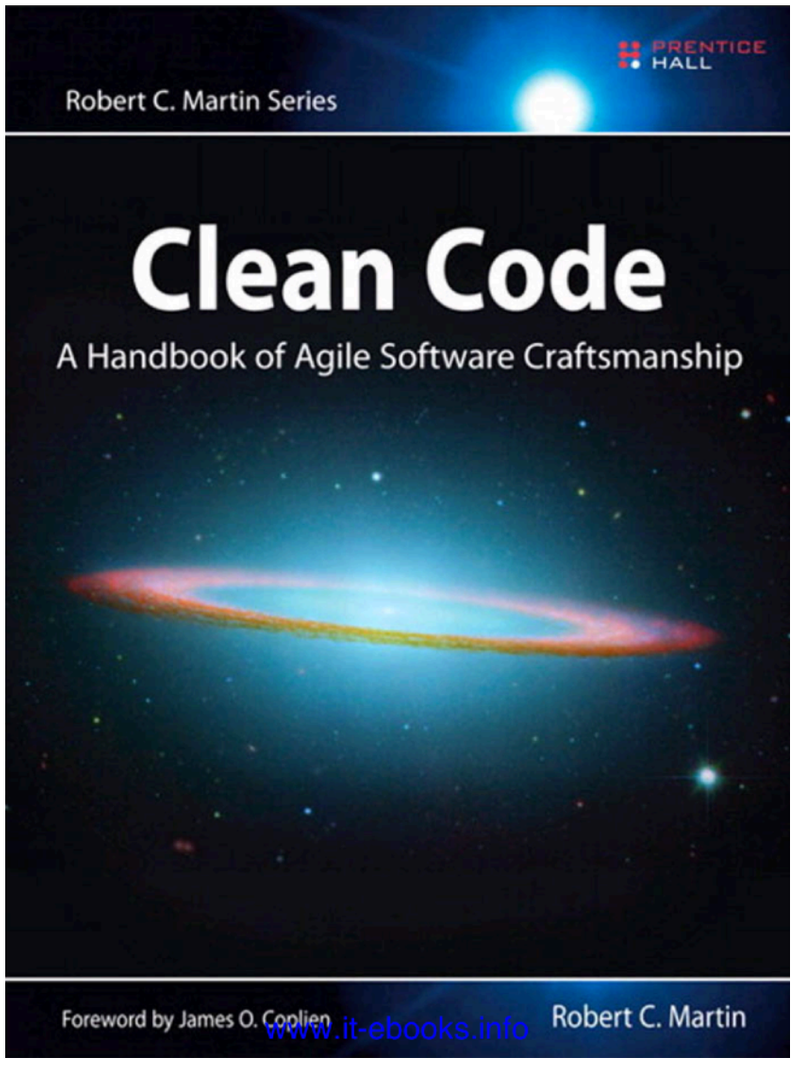
If you write a GUI – use this architecture



Code Quality

If you do one thing, write quality code

The code documents itself (or not!)



PEP 8 -- Style Guide for Python Code

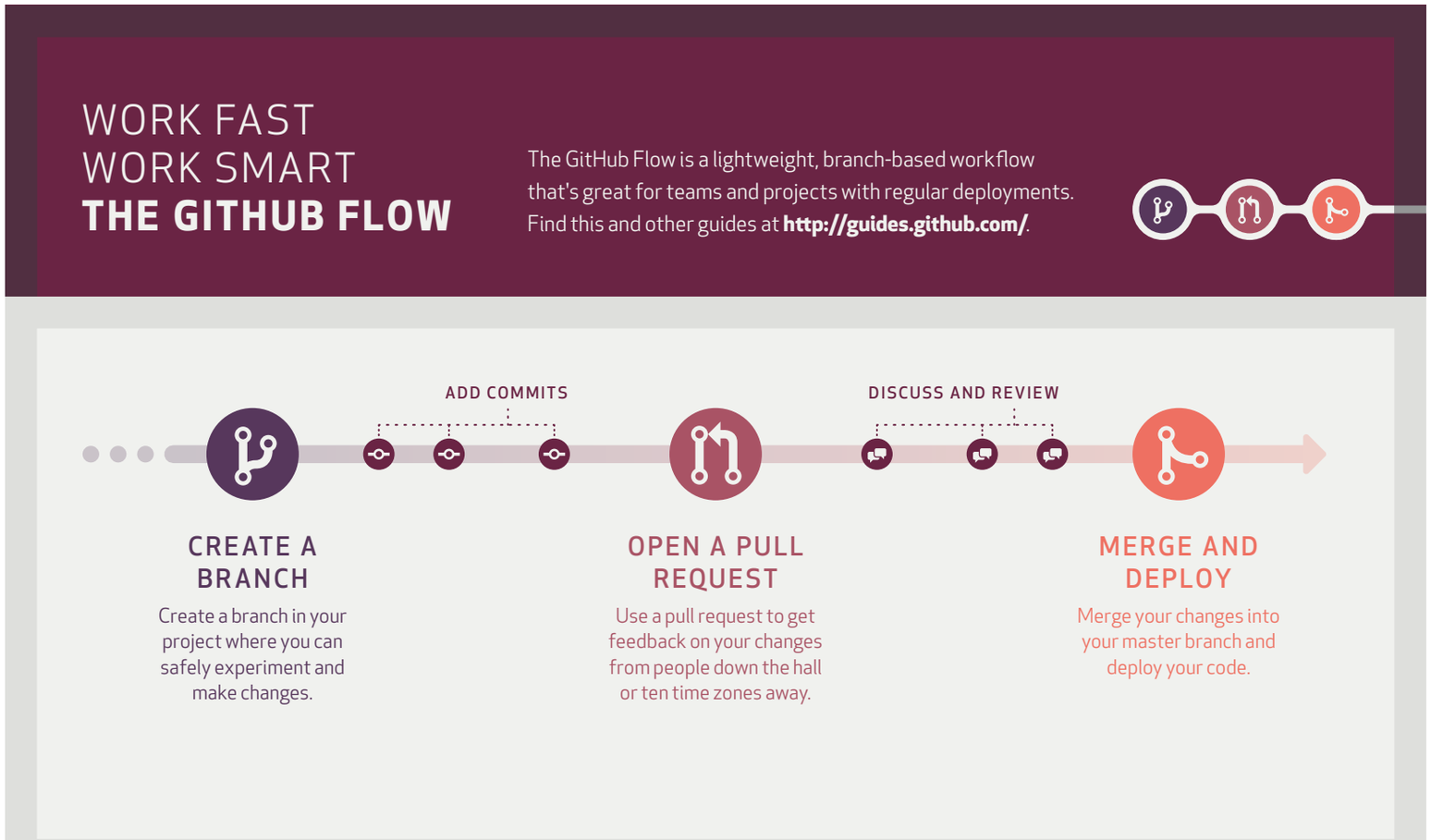
PEP:	8
Title:	Style Guide for Python Code
Author:	Guido van Rossum <guido at python.org>, Barry Warsaw <barry at python.org>, Nick Coghlan <ncoghlan at gmail.com>
Status:	Active
Type:	Process
Created:	05-Jul-2001
Post-History:	05-Jul-2001, 01-Aug-2013

Infrastructure - Version Control

EUROPEAN SPALLATION SOURCE

Would you jump out of a plane without a parachute

Version control is like a parachute

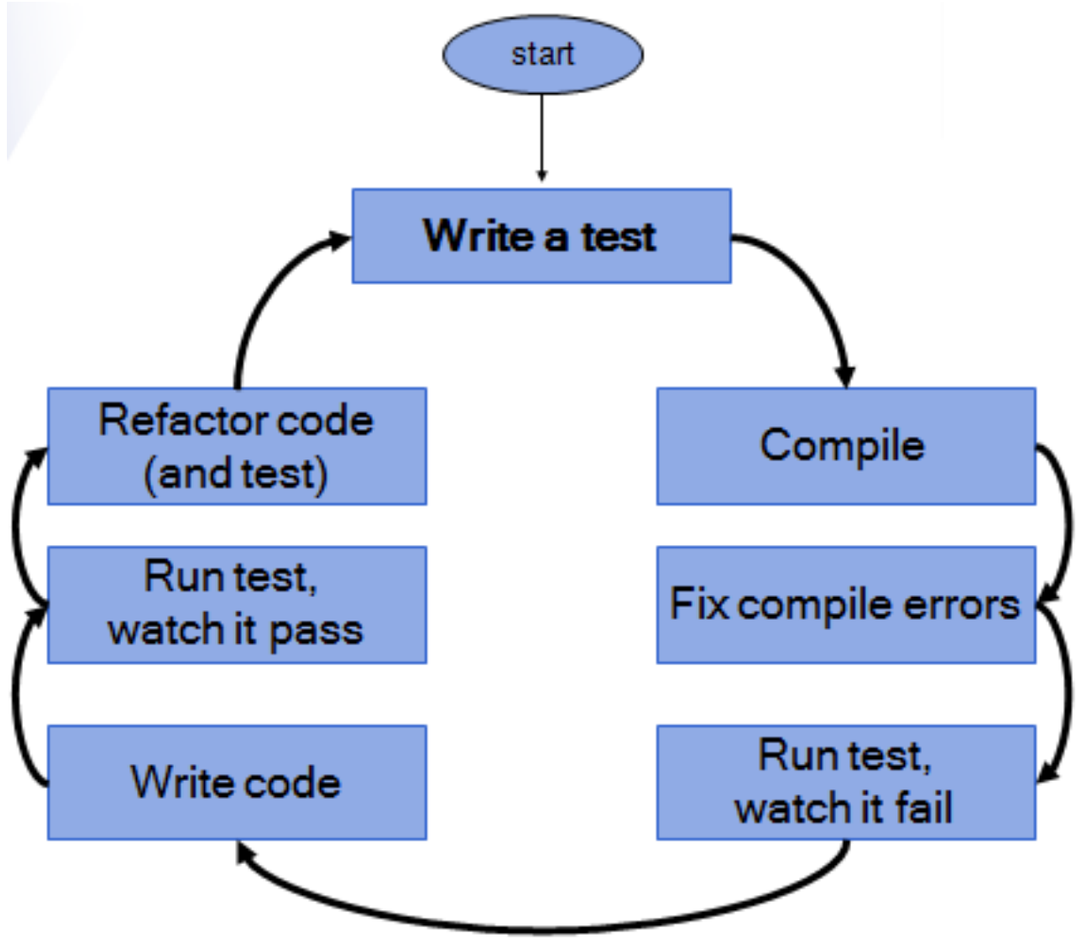


Infrastructure - Testing

How do you know its correct

Do test your code (or get Jenkins to do it, CI rocks)

* see later examples on python for why



Infrastructure - CI

EUROPEAN SPALLATION SOURCE

Everything looks fine …

Welcome to the continuous integration server for Mantid.

All | **Critical Jobs** | Imaging | Infrastructure | MSlice | Master Pipeline | ParaView | Pull Requests | Release Pipeline | Static Analysis | Valgrind

S	Name ↓	Last Success	Last Failure	Last Duration
	master_cppcheck	1 hr 13 min - #1782	4 days 21 hr - #1774	1 min 48 sec
	master_create_conda_osx_pkg	4 days 4 hr - #28	3 days 10 hr - #29	38 min
	master_doctests	3 days 11 hr - #840	2 days 12 hr - #841	29 min
	master_doxygen	2 days 12 hr - #874	N/A	5 min 29 sec
	master_flake8	2 hr 44 min - #1989	N/A	38 sec
	master_incremental	4 days 22 hr - #3817	2 hr 37 min - #3825	1 hr 27 min
	master_performancetests2	1 hr 50 min - #1719	N/A	1 hr 47 min
	python3	1 mo 25 days - #727	10 hr - #782	1 hr 3 min

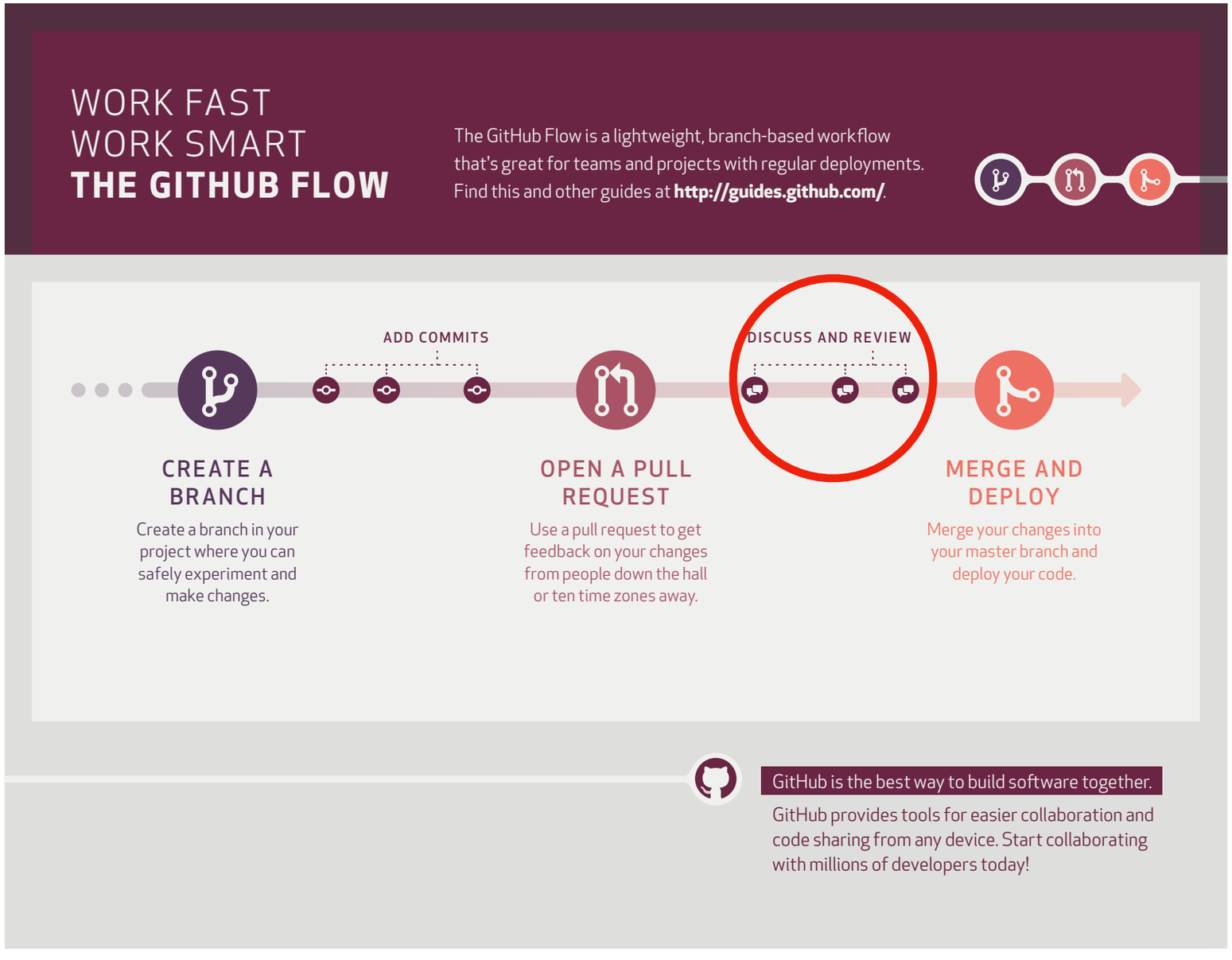
Icon: S M L

Legend RSS for all RSS for failures RSS for just latest builds

Infrastructure - Review

EUROPEAN SPALLATION SOURCE

Nothing happens anymore without looking a review website

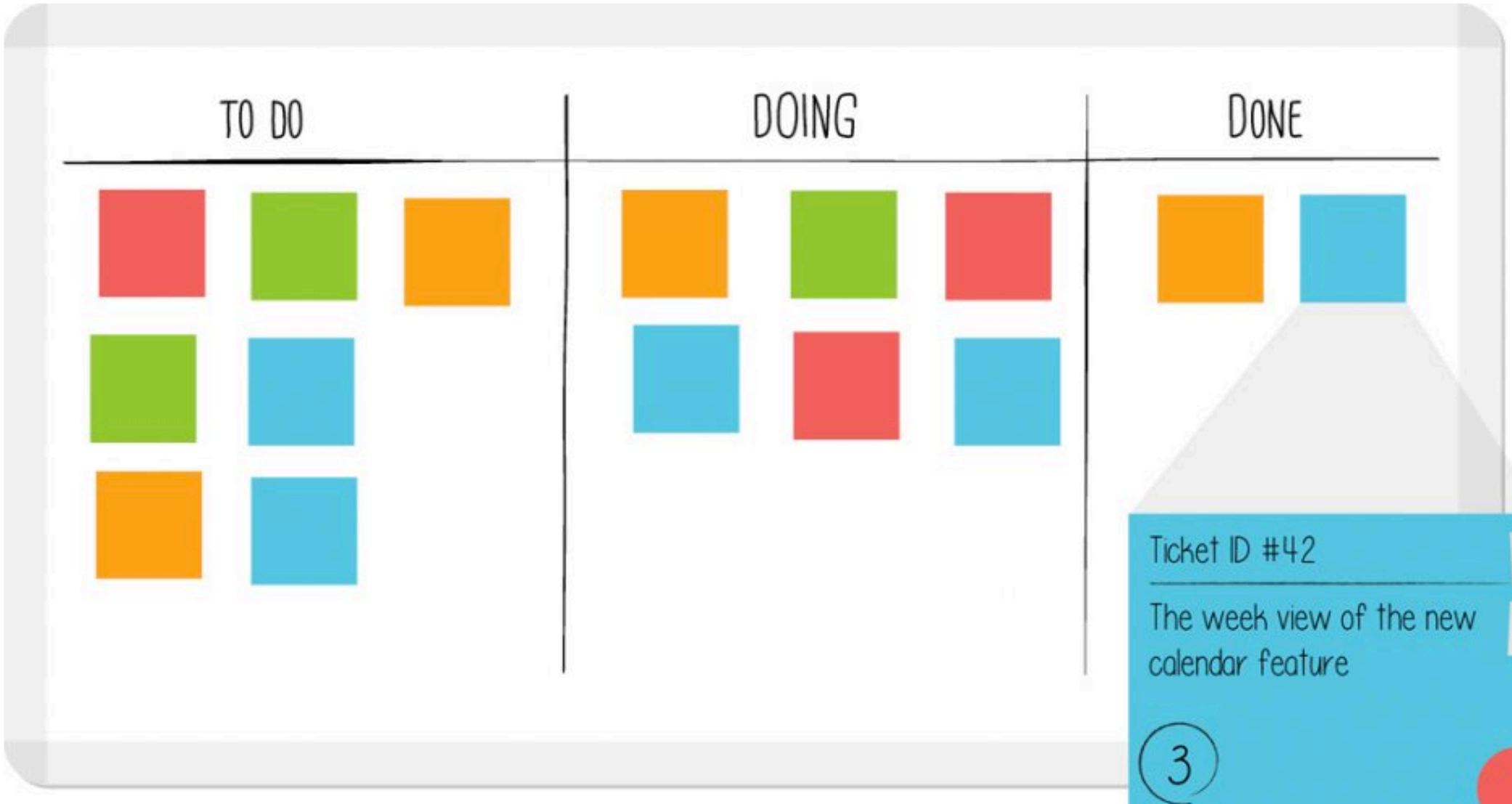


Don’t Reinvent - Be agile

EUROPEAN SPALLATION SOURCE

Agile is a ‘thing’ now.

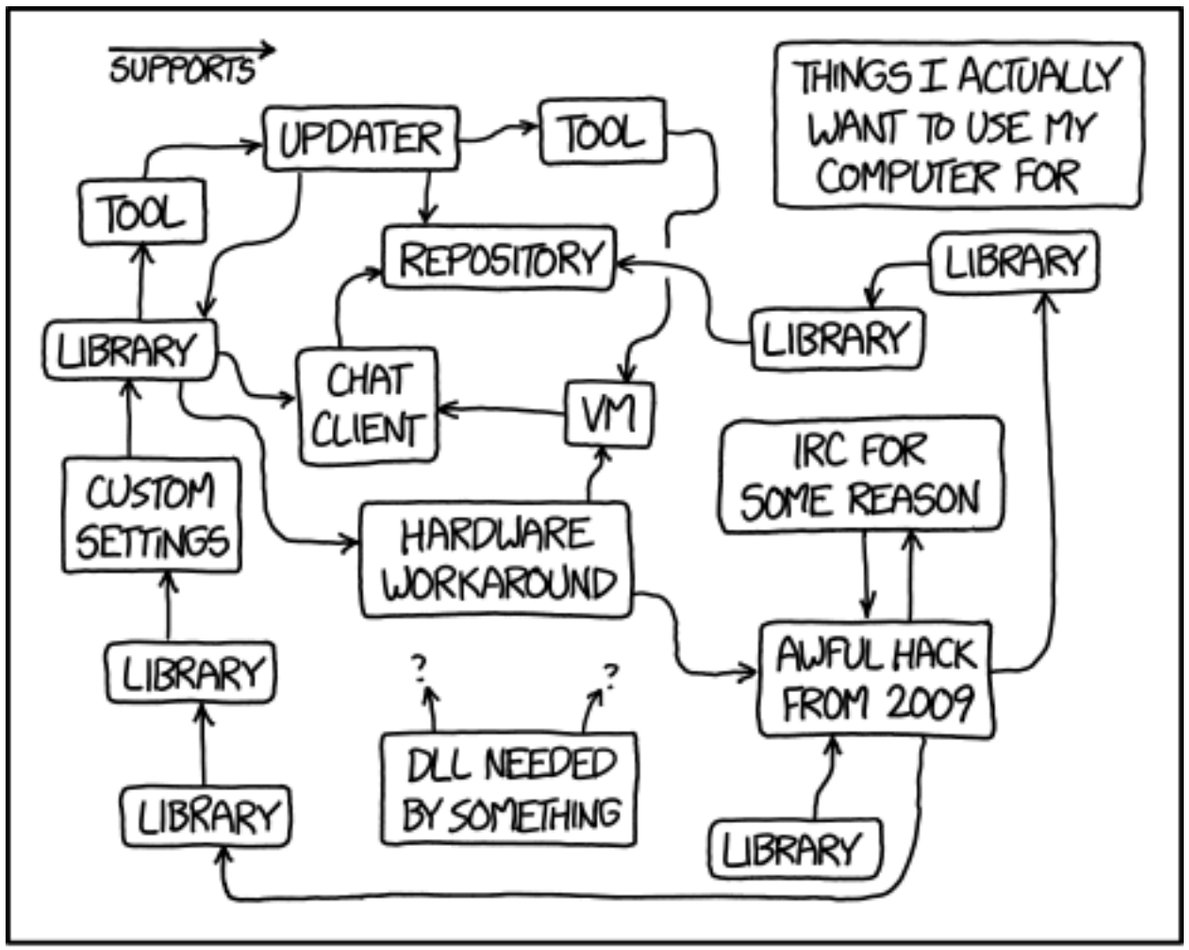
Be able to adapt to changes in requirements in a rapid way



Keep it Simple

Sounds simple right?

Avoid Technical Debt



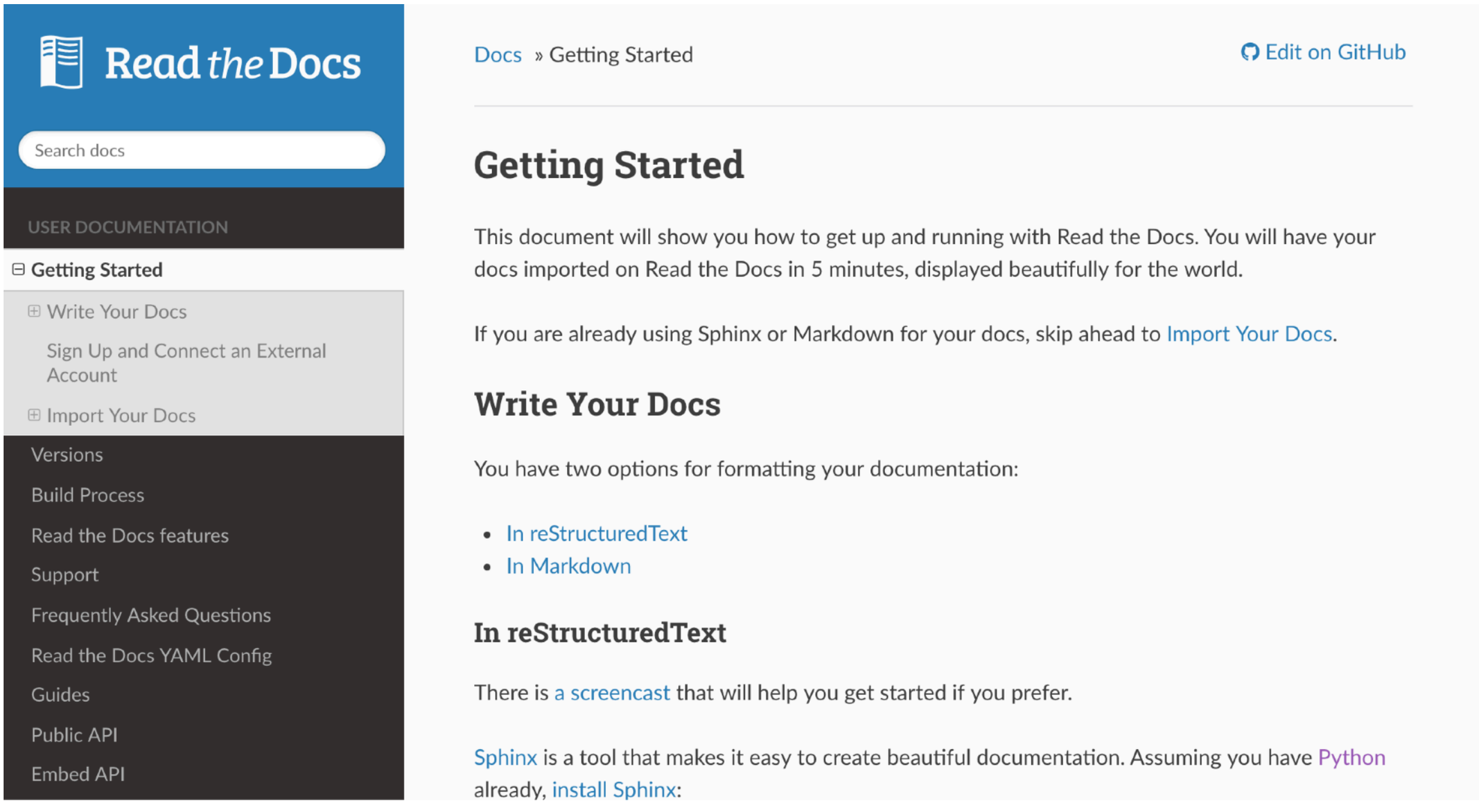
EVERY NOW AND THEN I REALIZE I'M MAINTAINING A HUGE CHAIN OF TECHNOLOGY SOLELY TO SUPPORT ITSELF.

Documentation is Difficult to Write

EUROPEAN SPALLATION SOURCE

The code documents itself

Manually Written Manuals are Mostly Moribund



Quick Tour of python™

EUROPEAN SPALLATION SOURCE

Learn Python, it's easy!

https://www.learnpython.org/

Interpreted

Not strongly typed

Object Oriented (if you want)

Fast with Numpy

Very Popular

Is OOP important in python

EUROPEAN SPALLATION SOURCE

Classes provide safety for

Data and Logic*

Most modules are OO.

*Private functions are not entirely private in Python if you know where to look

Syntax basics

Indentation delineates code

PEP8 defines 4 spaces as standard - not tabs

```
def myFunc():
“”
A doc string for myFunc
“”
    print ‘Hello World’ #Python 2.x
    print(‘Hello World’) #Python 3.x & 2.x

In[1]: myFunc()
Hello World
```

Syntax basics 2

Careful with the namespace…

```
In [16]: import numpy as np

In [17]: np.abs(-12)
Out[17]: 12
```

```
In [19]: from numpy import abs as myAbsFunction

In [20]: myAbsFunction(-12)
Out[20]: 12
```

Syntax basics 3

Careful with the types…

```
In [25]: a = 1.0

In [26]: type(a)
Out[26]: int

In [27]: b = 1.0

In [28]: type(b)
Out[28]: float

In [29]: a == b
Out[29]: True

In [30]: type(a) == type(b)
Out[30]: False

In [31]:
```

Python Essentials for Science

EUROPEAN SPALLATION SOURCE

NumPy

PyCharm

Python IDE
for Professional Developers

SciPy.org

Sponsored By ENTHOUGHT

matplotlib

LMFIT
Non-Linear Least-Squares Minimization and Curve-Fitting for Python

NumPy

EUROPEAN SPALLATION SOURCE

Fast array manipulation

Slicing and dicing n dimensional data

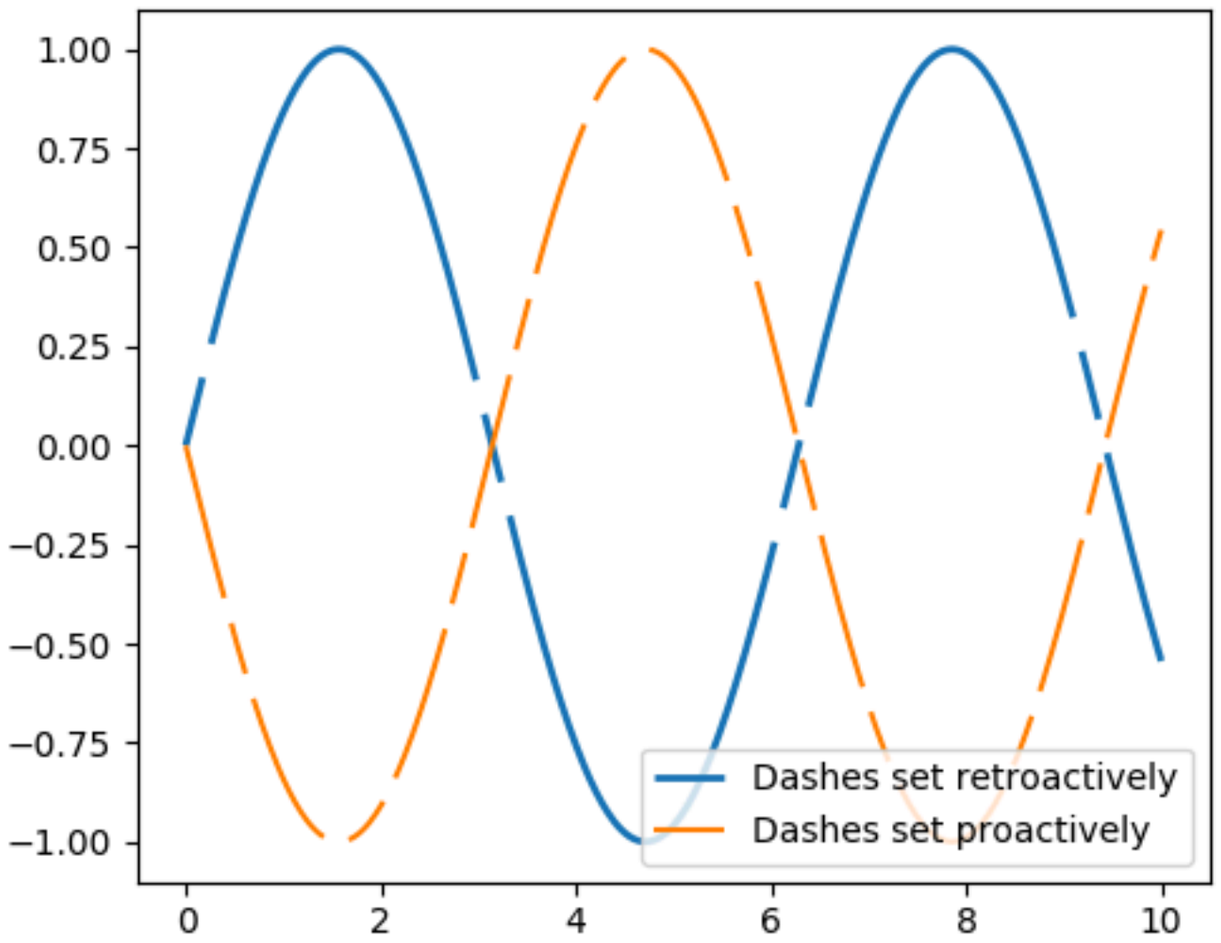
```
>>> import numpy as np
>>> a = np.array([2,3,4])
>>> a
array([2, 3, 4])
>>> a.dtype
dtype('int64')
>>> b = np.array([1.2, 3.5, 5.1])
>>> b.dtype
dtype('float64')
```

A frequent error consists in calling `array` with multiple numeric arguments, rather than providing a single list of numbers as an argument.

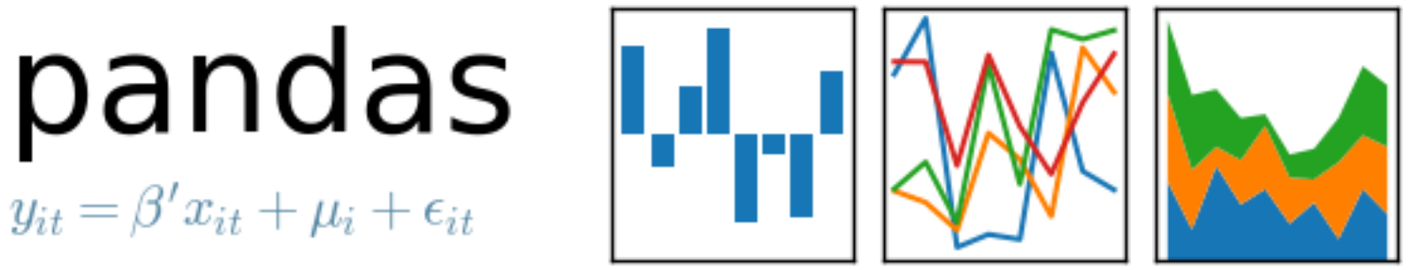
```
>>> a = np.array(1,2,3,4)    # WRONG
>>> a = np.array([1,2,3,4])  # RIGHT
```

Plotting 1D 2D & 3D data and images

```
"""
=====================================
A simple plot with a custom dashed line
=====================================

A Line object's ``set_dashes`` method allows you to specify dashes with
a series of on/off lengths (in points).
"""
import numpy as np
import matplotlib.pyplot as plt


x = np.linspace(0, 10, 500)
dashes = [10, 5, 100, 5]  # 10 points on, 5 off, 100 on, 5 off

fig, ax = plt.subplots()
line1, = ax.plot(x, np.sin(x), '--', linewidth=2,
                 label='Dashes set retroactively')
line1.set_dashes(dashes)

line2, = ax.plot(x, -1 * np.sin(x), dashes=[30, 5, 10, 5],
                 label='Dashes set proactively')

ax.legend(loc='lower right')
plt.show()
```



Python (non) Essentials for Science

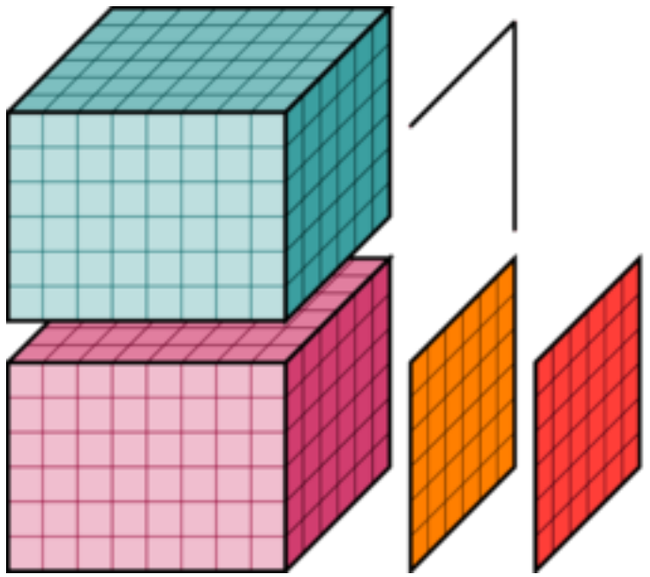


Manipulation of table like named data (spreadsheets)

Standard data type object called a data frame

```
In [1]: import pandas as pd

In [2]: import numpy as np

In [3]: import matplotlib.pyplot as plt
```

Python (non) Essentials for Science



xarray

Pandas container schema for NumPy arrays

Typing for n dimensional arrays

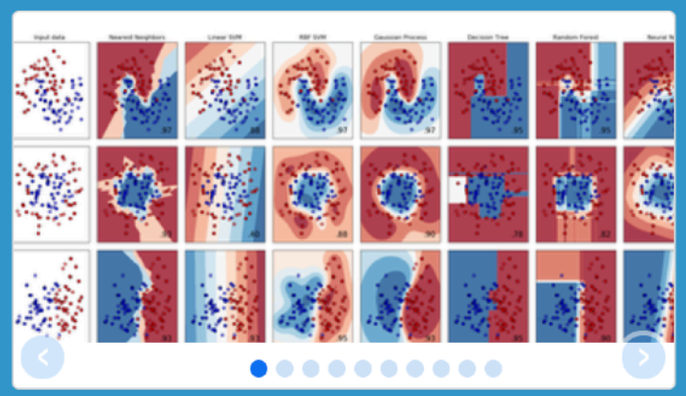
SciKits

EUROPEAN SPALLATION SOURCE

scikit learn

Home | Installation | Documentation | Examples

Google Custom Search



scikit-learn

Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

scikit-image
image processing in python

Download | Gallery | Documentation | Community Guidelines | Source

Search documentation ...

Stable (release notes)
0.14.0 - May 2018

Download

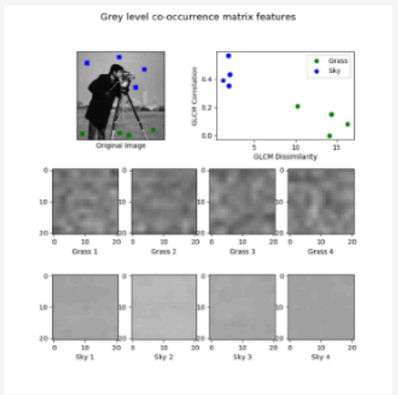
Development
pre-0.15

Download

Image processing in Python

scikit-image is a collection of algorithms for image processing. It is available free of charge and free of restriction. We pride ourselves on high-quality, peer-reviewed code, written by an active community of volunteers.

Download



Python installation

Take a distribution - enthought

Use python package manager pip
on the command line

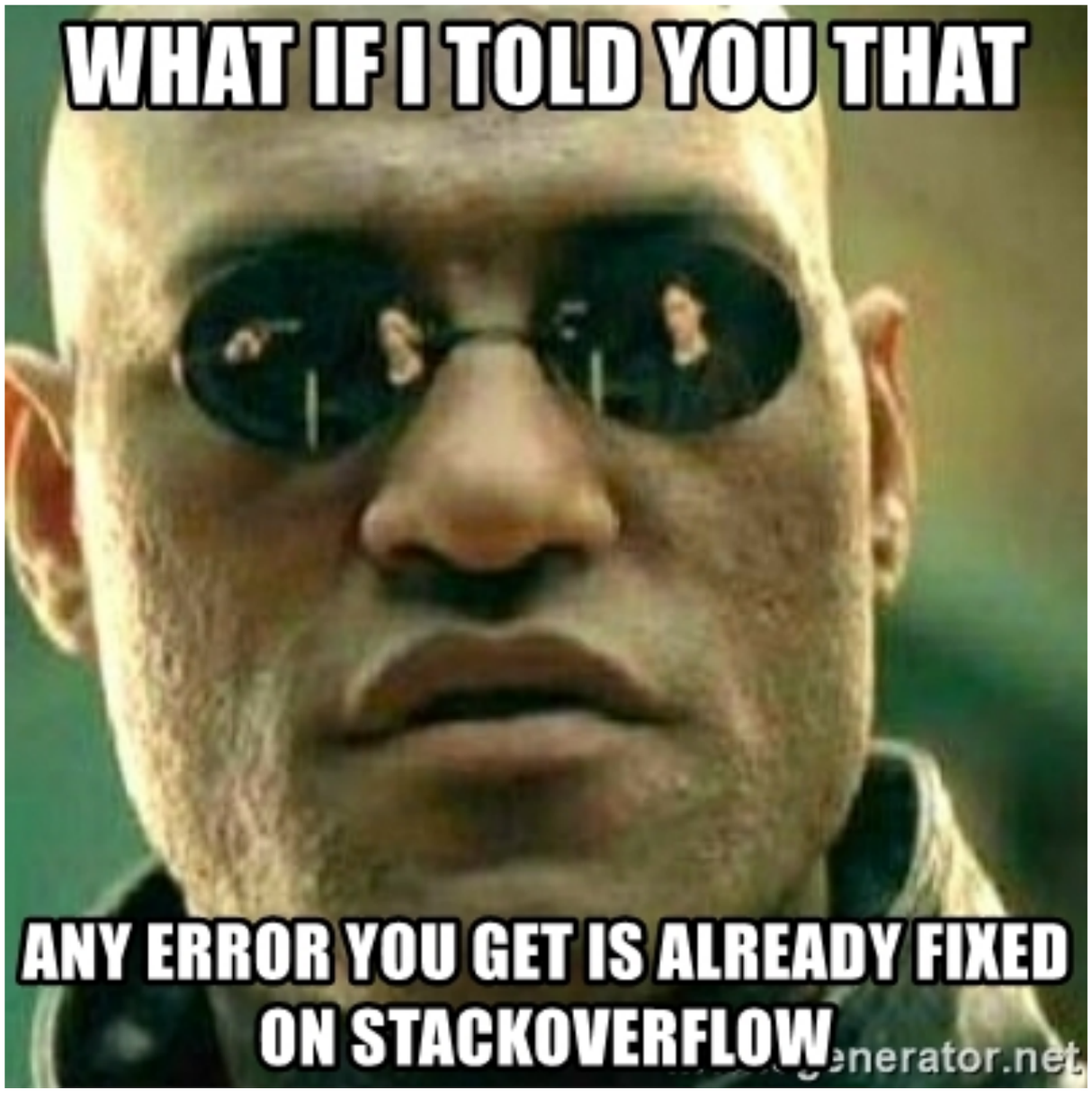
```
>> pip install numpy
```

Python in a notebook

Install Locally or Remotely

https://hub.mybinder.org/user/ipython-ipython-in-depth-vhas6tnx/notebooks/binder/Index.ipynb

where does everyone go for answers …



Examples of design decisions from ESS 1

Experiment control framework.

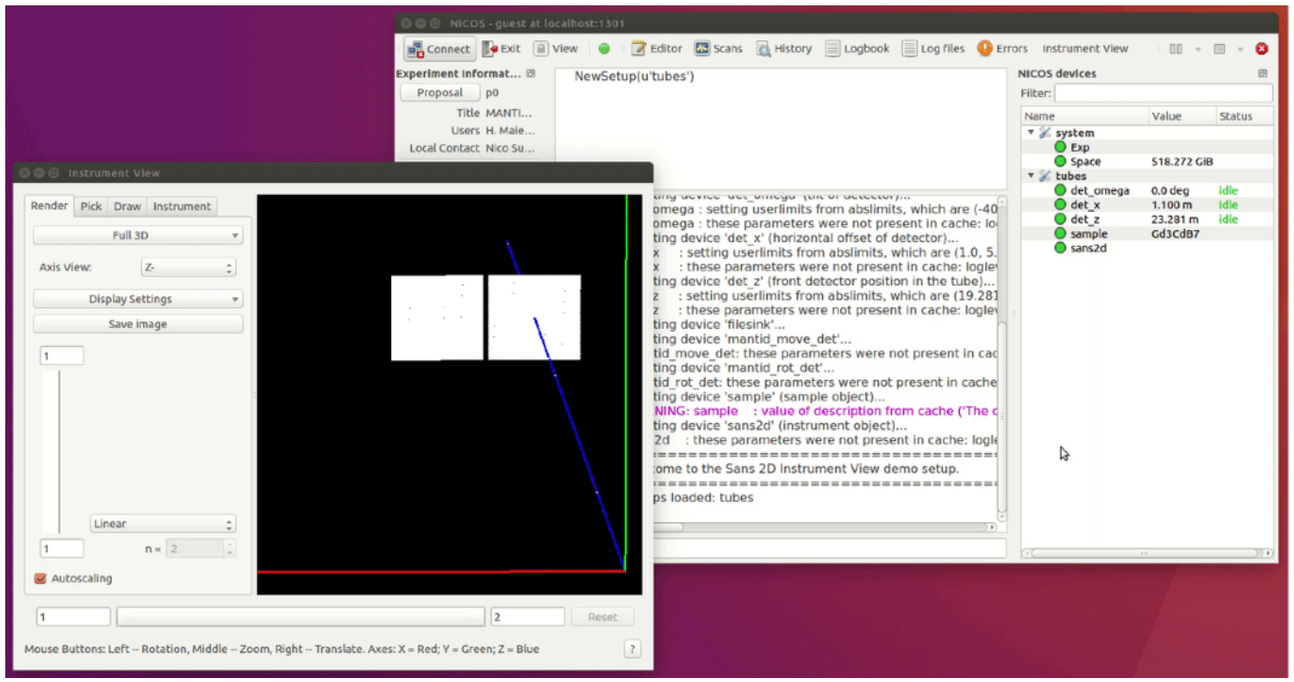
- All facilities have their own developed from scratch.
- ESS did not want to generate another.

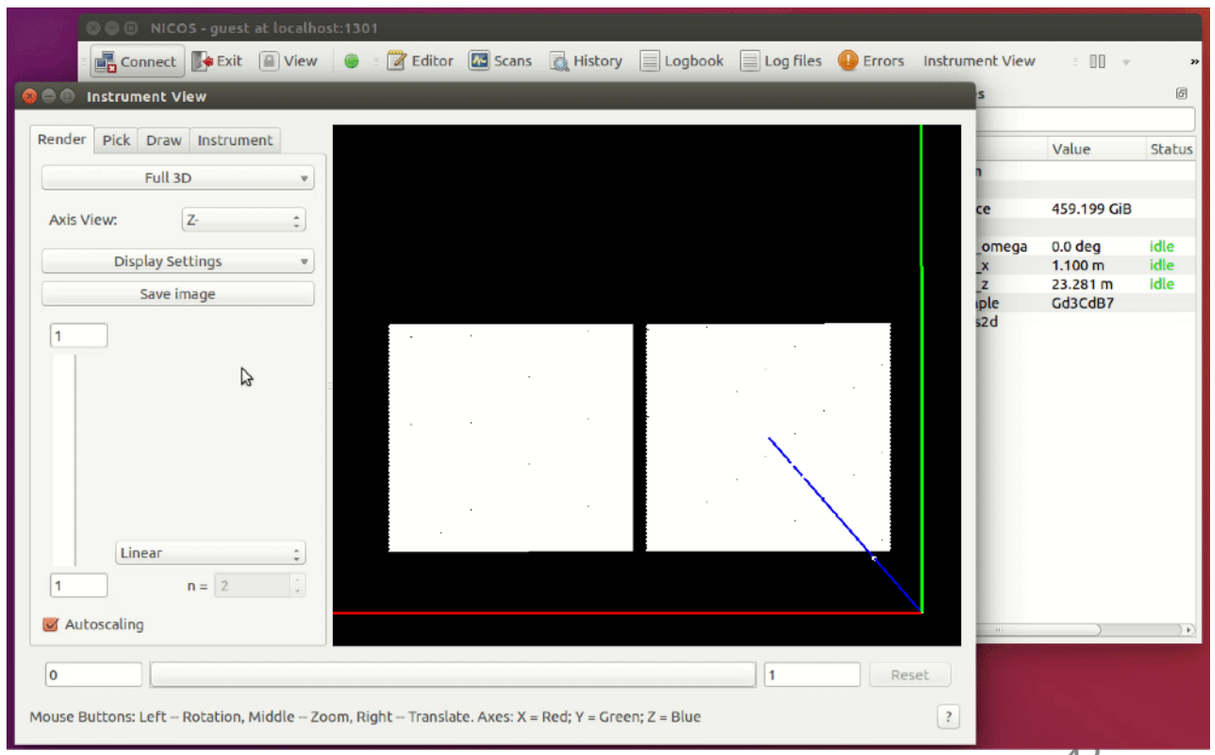
High level requirements (based on software best practice and functionality generate)

Existing open source developments were reviewed against requirements.

Experiment Control

- Feature complete solution from FRMII
- Python & Qt
- Acts as a high level interface to low level controls
- High quality python code base





Examples of design decisions from ESS 2

Data acquisition data streaming system.

- All facilities have their own developed from scratch.
- ESS did not want to generate another.

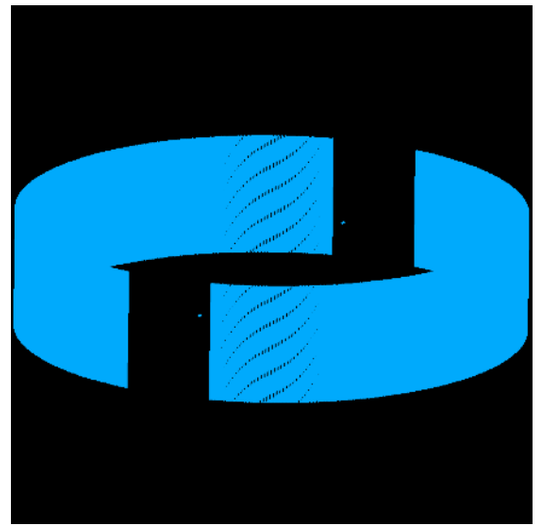
High level requirements

- Based on software best practice and required functionality

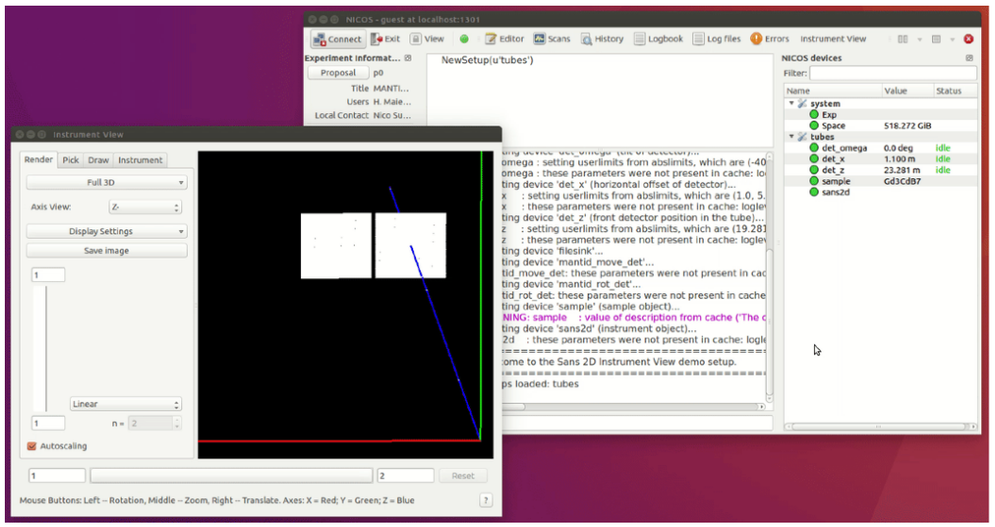
Existing open source developments reviewed against requirements.

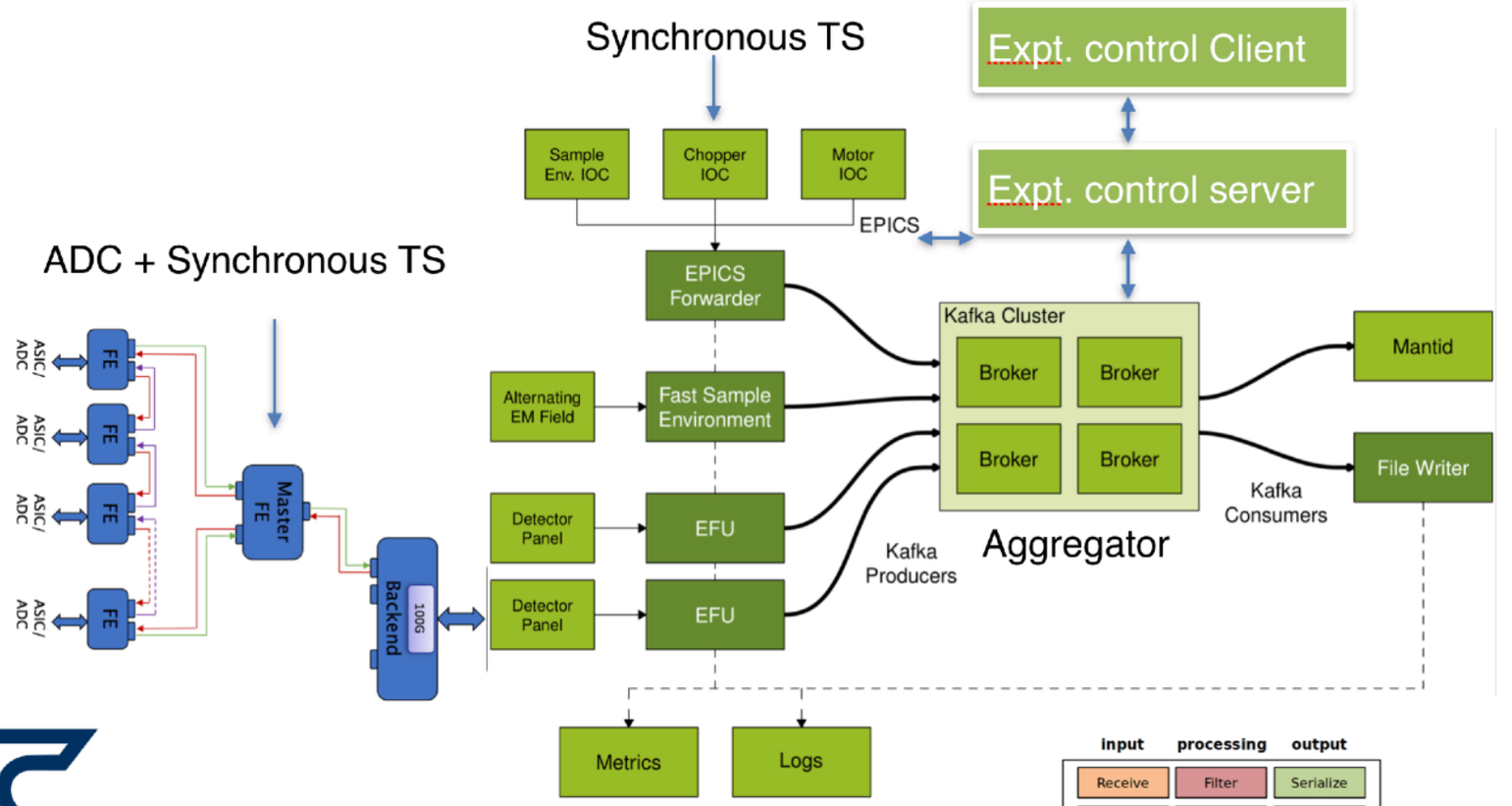
Next generation data acquisition

- Event mode data collection using Kafka streaming
- Fast capture of experimental metadata
- Big data technology
- FAIR data from the start
- High performance infrastructure, software data storage & data management



- Remote access to infrastructure





NICOS
Networked Instrument COntrol System

Why Apache Kafka

EUROPEAN SPALLATION SOURCE

- It’s open source
- It’s actively developed
- It’s the technology used by Netflix and Linkdin

Where is my data?

How do I get it

How do I look at it

What does it mean

Is there software to read it

FAIR Findable & accessible

EUROPEAN SPALLATION SOURCE

Findable

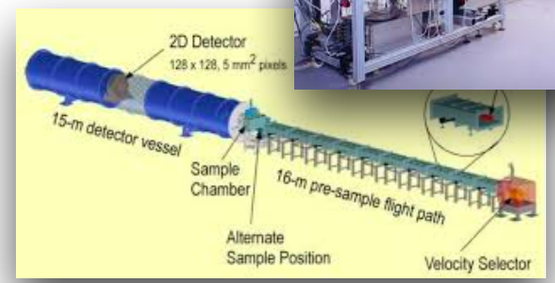
- ☑ Searchable rich meta data
- ☑ Persistant identifiers
- ☑ Experiment type
- ☑ Sample parameters
- ☑ Instrument parameters
- ☑ Applies to raw and proceed data
- ☑ What software?
- ☑ What parameters
- ☑ what is the repository
- ☑ Where is it stored

Accessible

- ☑ Data policy
- ☑ Data format
- ☑ Storage architecture
- ☑ Data catalogues
- ☑ APIs
- ☑ Authentication
- ☑ Security
- ☑ Applies to raw and processed data
- ☑ Applies to software

Interoperable

- ☑ Compatibility
- ☑ Data format
- ☑ Meta data standards
- ☑ Allow for multi modal data analysis
- ☑ Software APIs
- ☑ Catalogue interoperability
- ☑ Standards



Reusable

- ☑ Data shelf life
- ☑ Software shelf life
- ☑ Open access
- ☑ Backwards compatibility

Digital Object Identifiers

https://www.doi.org/

Essential for FAIR

DOIs are minted again digital artefacts

Store meta data and persistent location

DOIs cost money (borne by RIs & University services)

Data Management Plans

Required for external funding

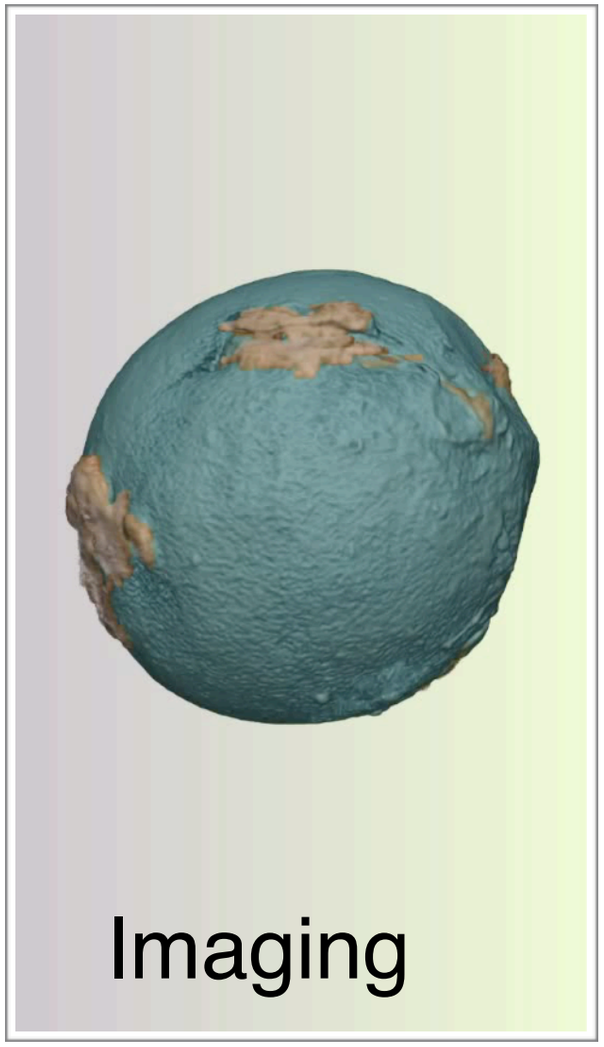
Depth of detail depends on agency

There are online services to generate DMPs

https://libraries.mit.edu/data-management/plan/write/

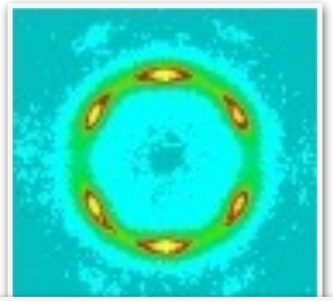
Often provide as a service by Libraries

Complexity defines the ESS science case

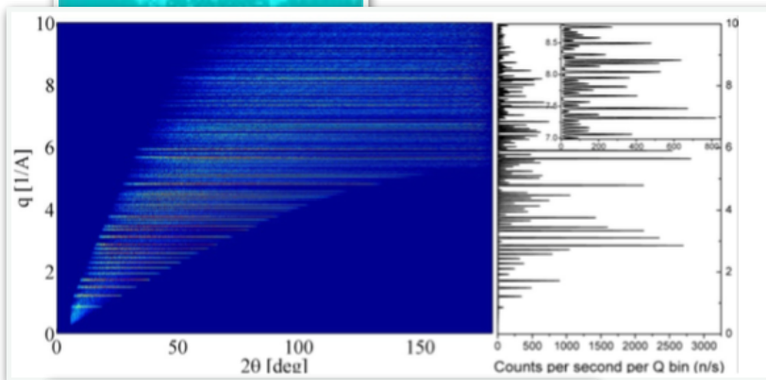


Courtesy of PSI NIAG

SANS

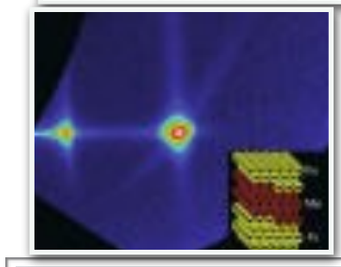


Powder diffraction

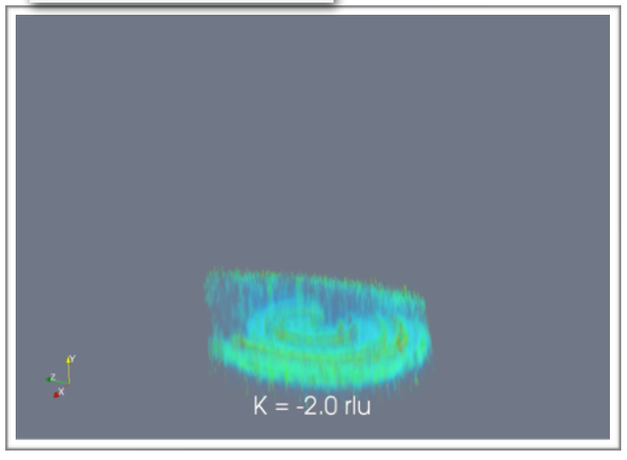


Single crystal diffraction

Reflectometry



Spectroscopy



Data policies at research infrastructures

- Most European facilities are moving to an open data policy
- Data is open access after an initial embargo period (3 years)
- Requires infrastructure and a catalogue to make data FAir
  - Infrastructure is still not federated for Neutrons and Photons in Europe - that is the task of EOSC
- Findable and Accessible are not not the same.

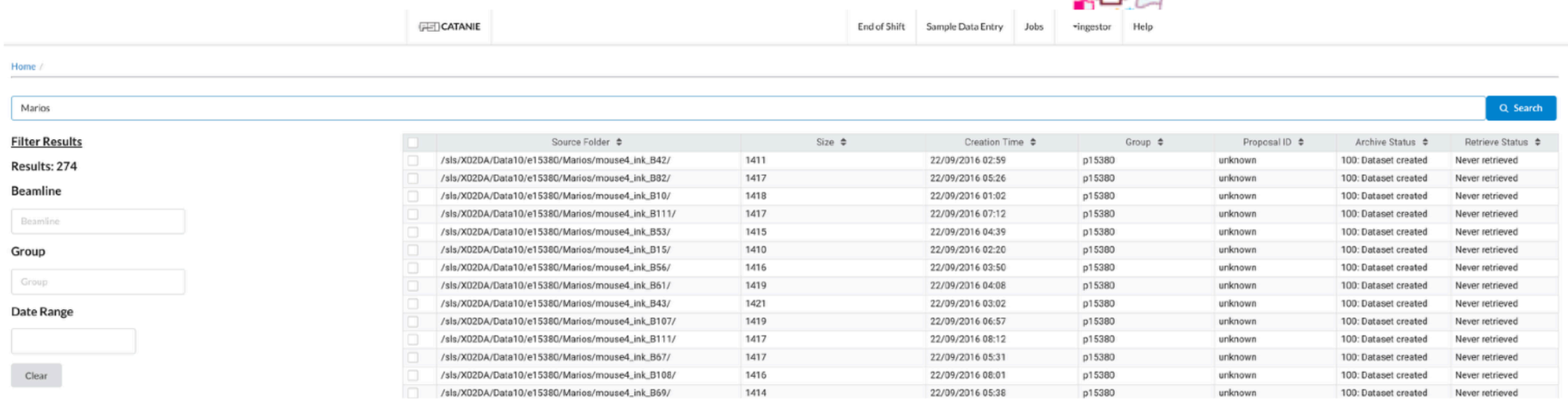
Open data & software

Fair applies to software as well as data

Everyone Benefits from Open Data

A new open source meta data catalogue

- Raw Data
- Meta Data
- Analysed Data
- Optimised for ESS
- Deployed at MAXIV & PSI



European Open Science cloud

EUROPEAN SPALLATION SOURCE

Pan European FAIR Data

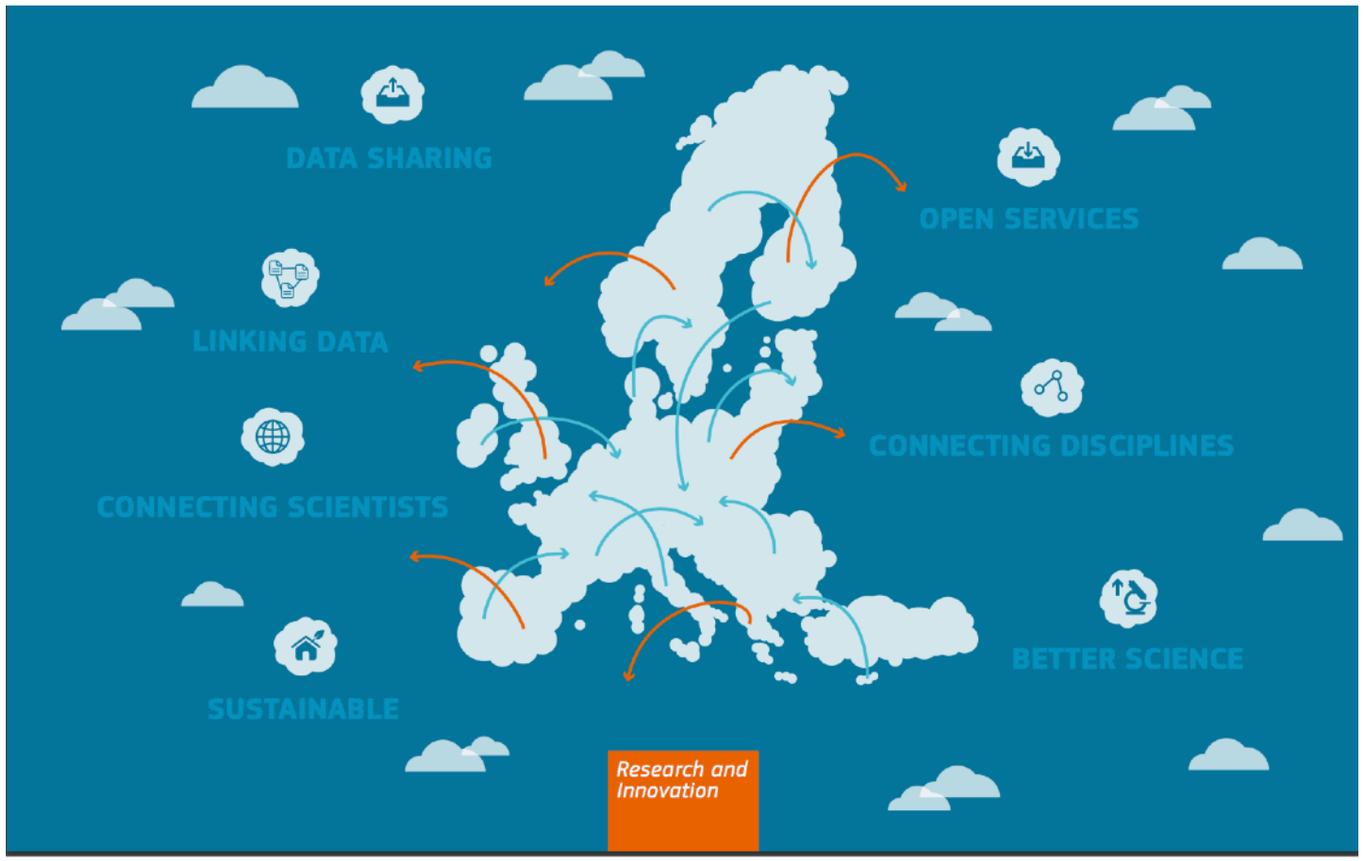
- My data are your data…
- Photon and Neutron Open Science Cloud
  - Data Federation
  - Open Services for data treatment and analysis
  - Access to compute services

NEUTRONS FOR SOCIETY

European XFEL

ESRF The European Synchrotron

eli delivery consortium



Interoperability and Reuse

- Interoperability starts with data formats

- Photon and neutron sources converge towards a common format

- Based on HDF5

About the NeXus Data Format

NeXus is a common data format for neutron, x-ray, and muon science. It is being developed as an international standard by scientists and programmers representing major scientific facilities in order to facilitate greater cooperation in the analysis and visualization of neutron, x-ray, and muon data.

Documentation:

- Most recent publication to cite:
  J. Appl. Cryst. (2015). **48**, 301-305 doi:10.1107/S1600576714027575
- User Manual:
  - Introduction to the concepts behind the NeXus data format

Nexus and HDF

- Hierarchical Data Format

- Tree based Data model
- Stores data and meta data

- various APIs including python

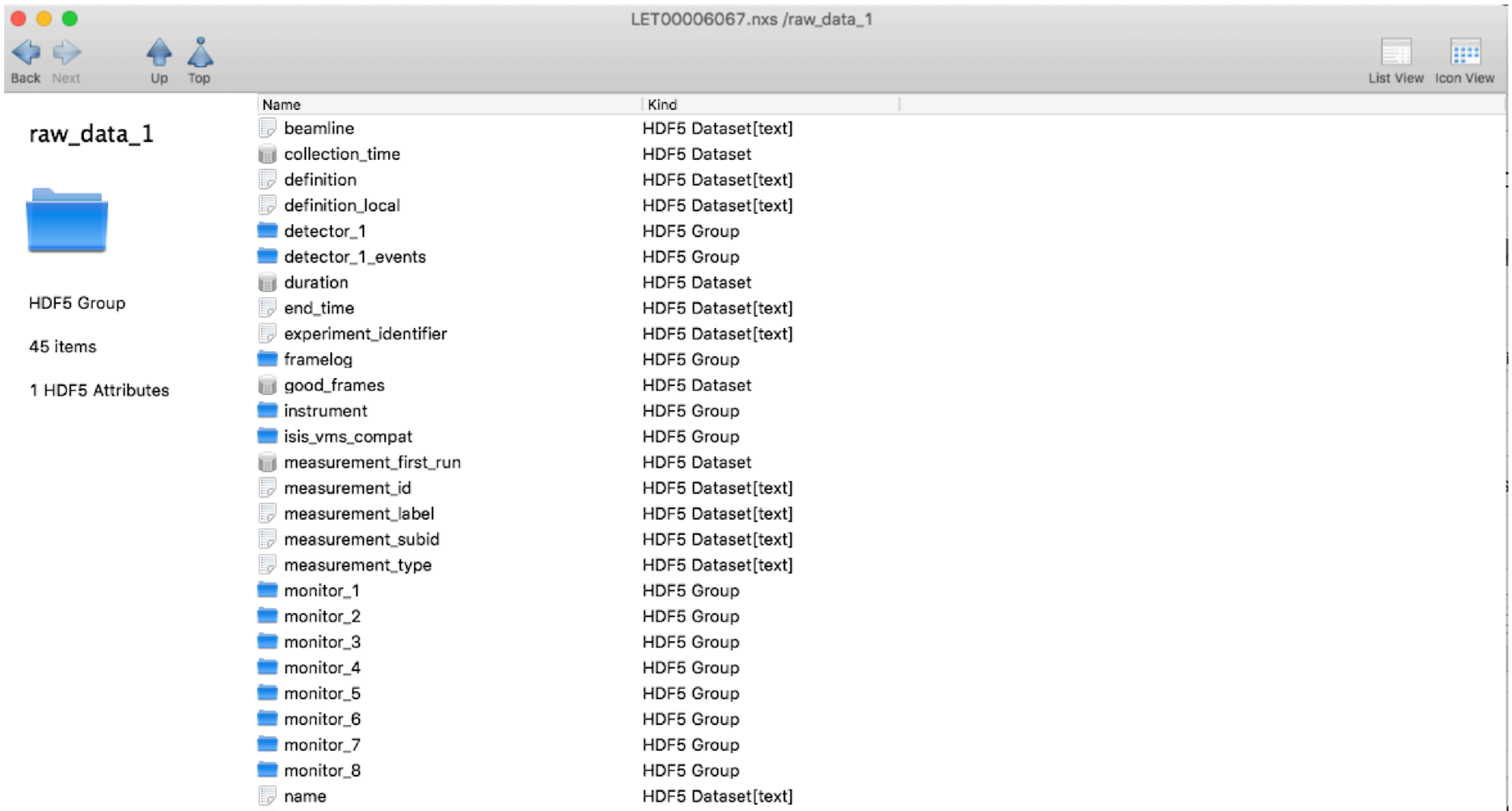
- H5Py (http://docs.h5py.org/en/latest/quick.html)

```
>>> import h5py
>>> f = h5py.File('mytestfile.hdf5', 'r')
```

- NEXUS is a set of scattering specific classes to standardise neutron photon and muon data
- including geometry meta data and experiment meta data

HDF Compass

- HDF group application to view HDF files
- Works with nexus files

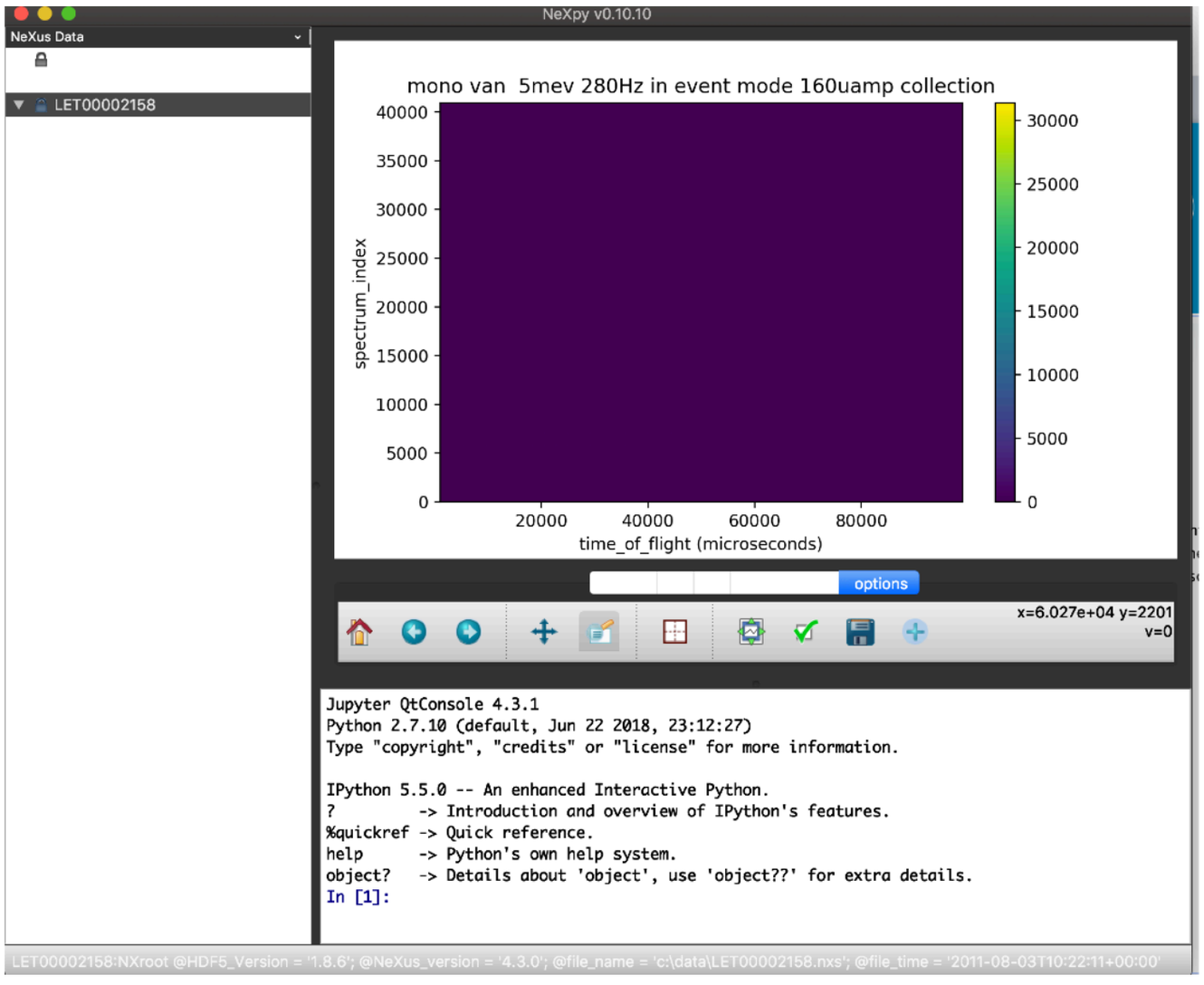


NexPy - Ray Osborne

EUROPEAN SPALLATION SOURCE

NeXpy: A Python GUI to analyze NeXus data

NeXpy provides a high-level python interface to HDF5 files, particularly those stored as NeXus data, within a simple GUI. It is designed to provide an intuitive interactive toolbox allowing users both to access existing NeXus files and to create new NeXus-conforming data structures without expert knowledge of the file format. The underlying Python API for reading and writing NeXus files is provided by the nexusformat package, which utilizes h5py. The Python API is also described here.



Physical Information File

A standard for materials information

http://citrineinformatics.github.io/pif-documentation/index.html

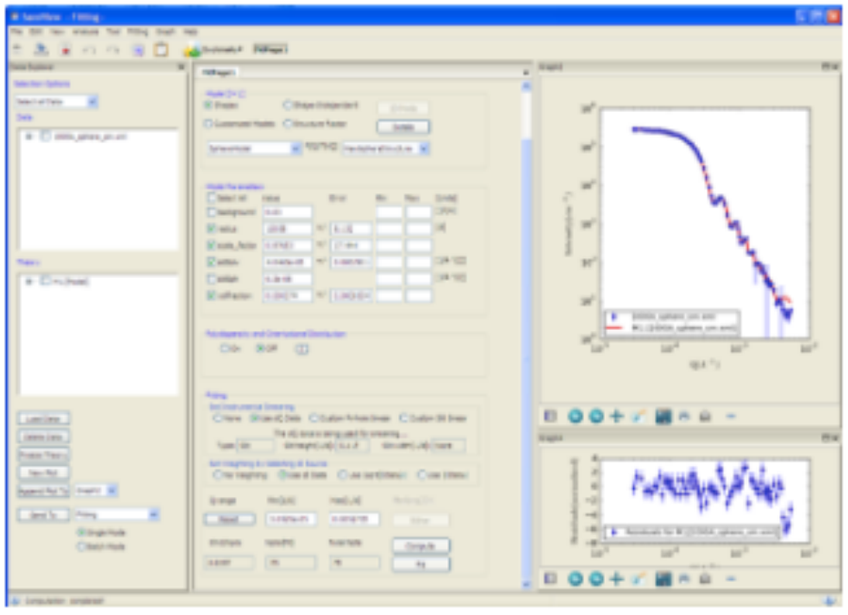
A schema to store materials structure and meta data from calculations and measurements

Versatile

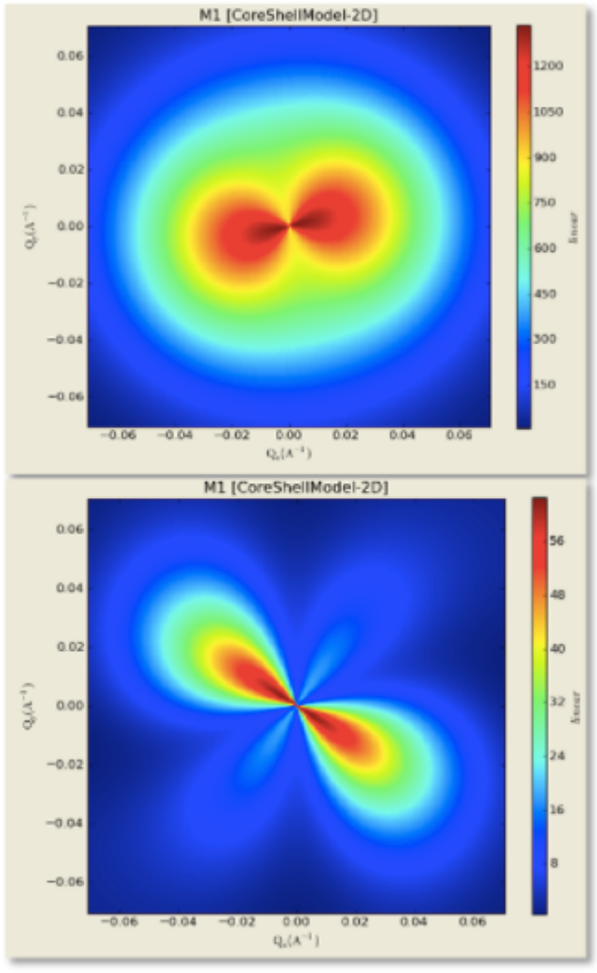
Python interface
https://github.com/CitrineInformatics/pypif

Interoperable software example - SASView

- Small angle scattering analysis
- Fitting and visualisation
- Photons and Neutrons



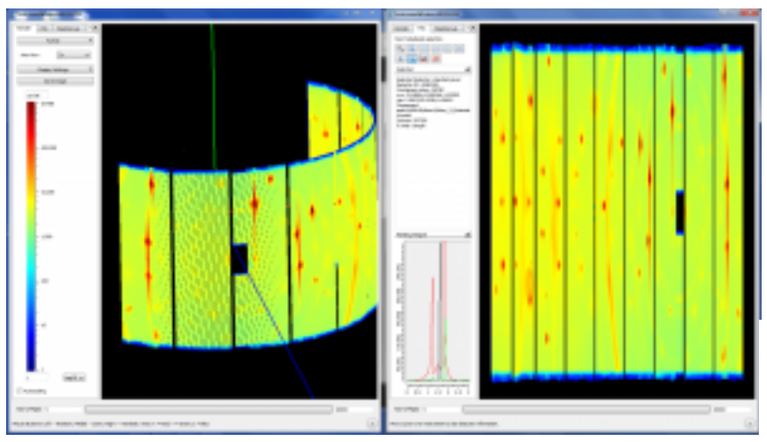
1D fitting screenshot.

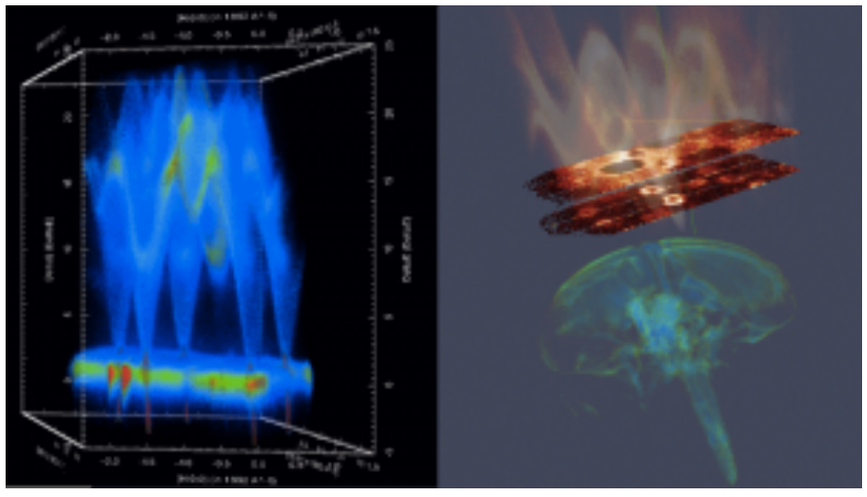


CoreShell simulation with Mcore parallel to the x-axis and Mshell canted away. Top figure up,up bottom figure up,down

faiR software example

- Meta data saved to file for data processing
  - History of data processing preserved
- DOI against each build for version compliance





Data management is key for success

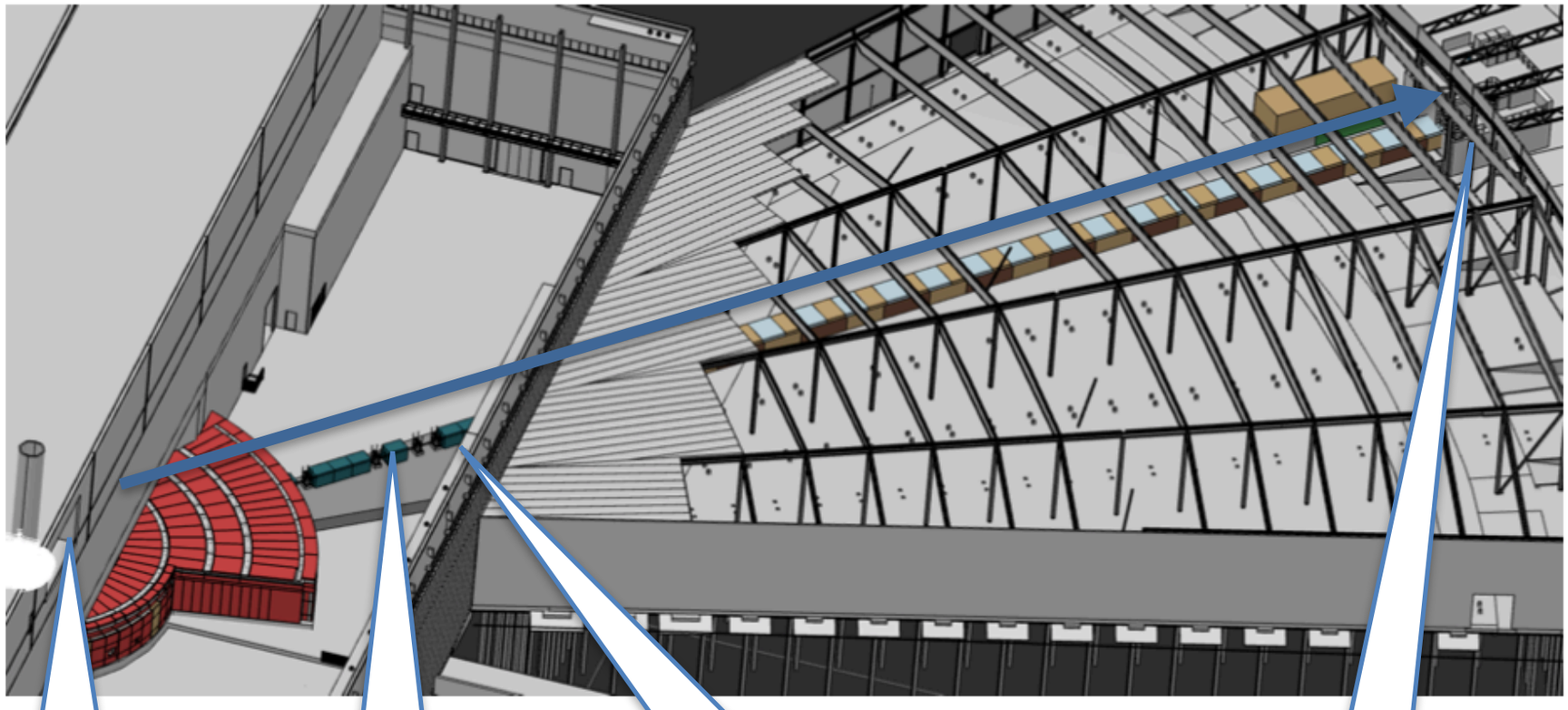
EUROPEAN SPALLATION SOURCE

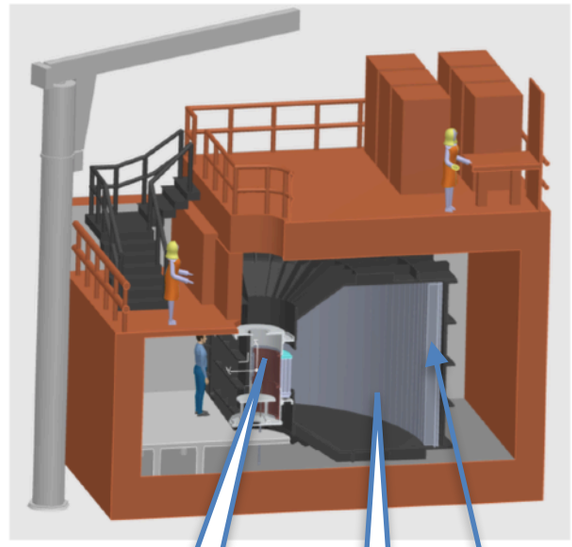
Ensure your data, software and results are FAIR

TOF Neutron scattering data treatment

EUROPEAN SPALLATION SOURCE

The neutron energy is encoded in its *Time of Flight*





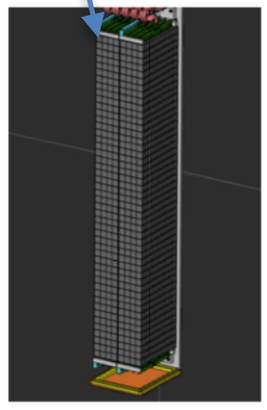


Figure 18: CAD image of a B10 detector module. M.Anastasopoulos

Neutron source

Neutron choppers

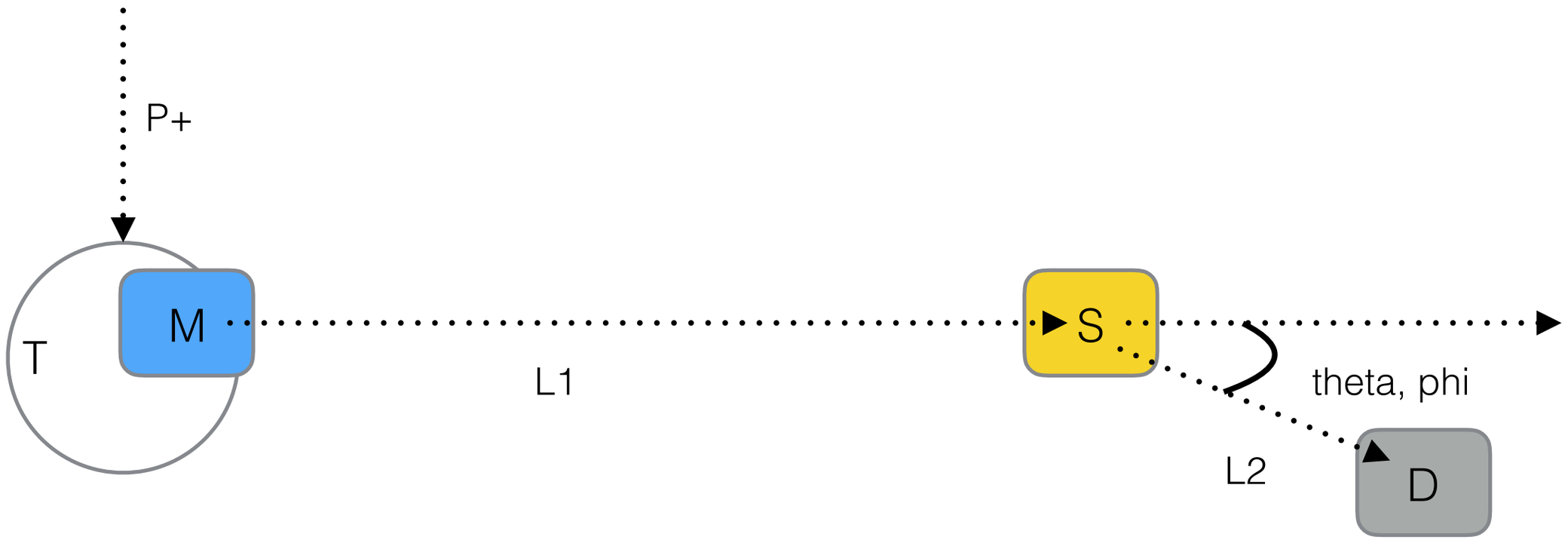
Neutron guide

Neutron instrument

Sample position

Detector tank 3D position voxel dets

Data processing for TOF experiments



- Convert T.O.F to energy, wavelength, momentum transfer, d-space.
- Precise knowledge of flight paths
- Precise knowledge of scattering angle
- Geometry information is essential

Data Acquisition

Two types

- Data are collected as histograms

  DAQ system has to configure histogram storage for each pixel ID

- Data are collected in event mode (list mode)

  Each detected neutron is assigned a Pixel ID and time stamp.

  Meta data is also collected in event mode

  The event list can be filtered to generate histo data

Instrument Geometry

- Dependent on installation and technique
- Ask how is the instrument calibrated
- Mantid stores Geometry in xml format as x,y,z
  - Instrument definition file.
  - Timestamped files to account for variance over time
  - All instrument components can be described.
  - Mantid framework handles conversation to r,t,p

```
<instrument xmlns="http://www.mantidproject.org/IDF/1.0"
            xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
            xsi:schemaLocation="http://www.mantidproject.org/IDF/1.0 http://schema.mantidproject.org/IDF/1.0/IDFSchema.xsd"
            name="ARCS"
            valid-from="1900-01-31 23:59:59"
            valid-to="2100-01-31 23:59:59">
```

```
<type name="main-detector-bank">
  <component type="main-detector-pixel" >
    <location x="-0.31" y="0.1" z="0.0" />
    <location x="-0.32" y="0.1" z="0.0" />
    <location x="-0.33" y="0.1" z="0.0" />
  </component>
</type>
```

Data reduction Workflows

- Loading data
- Filtering events
- Correct for counting efficiency
- Background
- Detector efficiency
- Normalisation to monitor / time / proton charge
- Units conversion
- Detector grouping
- Saving output data in analysis application format
- Visualisation
- Workflows wrapped as a script or a GUI
- Technique & facility dependent

Data Corrections

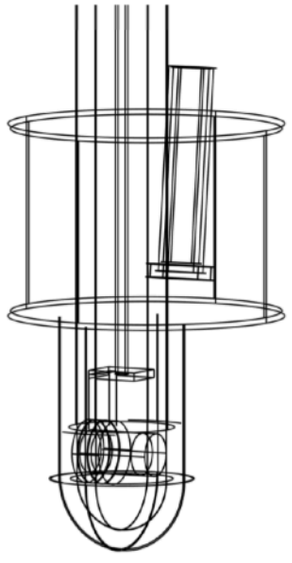
- Multiple scattering

  Sample

  Instrument and Sample environment

- Absorption correction

- MonteCarlo ray tracing proves quite useful



Mads Bertelsen
Nano-Science Center
Niels Bohr Institute
University of Copenhagen

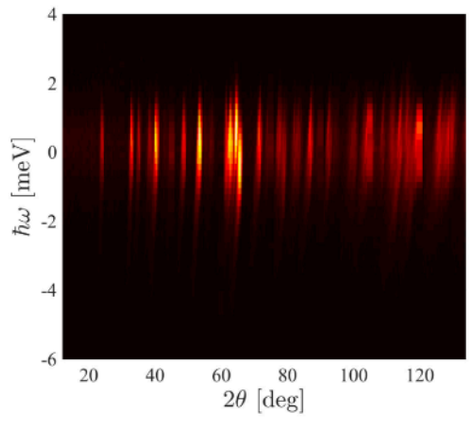
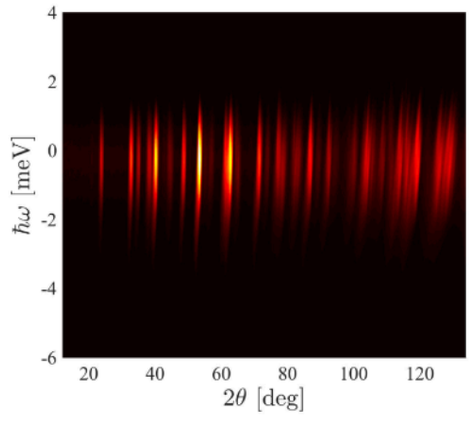


Figure 4.39: Measured scattering from cryostat and Ni_3TeO_6 sample on MARI with a selected energy of 35.19 meV from the Gd chopper running at 250 Hz.



Import required python modules

```
from qtiGenie import *
from PyStar2 import *
```

Define run numbers, Ei guess, rebin, mapfile and other keyword arguments for reduction

```
inst=iliad_setup(inst)
ext='.raw'
mapfile='mari_res'
#det_cal_file must be specified if the reduction sends out put to a workspace
cal_file='MAR16637.raw'
whitebeamfile="16637"
ei=100
rebin_params='-10,.2,95'
runs=[16654]
```

Execute python line

```
w2=iliad("whitebeamfile","runs",ei,rebin_params,mapfile,det_cal_file=cal_file,norm_method='current')
```

Create instance of reduction class for selected

Load Instrument parameter file

load White beam run

```
LoadRaw(Filename=r'/Users/jon/mprogs/mari_data/RAW/MAR16637.raw',OutputWorkspace='wb_wksp',LoadLogFiles='0')
```

Load Detector

Normalise

Monitor / Current
Monitor Integral

DIAG WB
Zero counts in frame (OFF)
Median Rate in bank --> Reject outside of pre-determined range

Apply Hard Mask

Load sample run

```
LoadRaw(Filename=r'/Users/jon/mprogs/mari_data/RAW/
MAR16652.raw',OutputWorkspace='run_wksp',LoadLogFiles='0')
```

Load Detector Info

Normalise

```
NormaliseByCurrent(InputWorkspace='run_wksp',OutputWorkspace='run_wksp')
```

Find bad spectra for sample run
Zero counts in frame (OFF)
Median Rate in bank --> Reject outside of pre-determined range
Run bleed test

```
FindDetectorsOutsideLimits(InputWorkspace='_wksp.spe-
white',OutputWorkspace='white_masks',HighThreshold='10000000000',LowThresh
old='1e-10')
MaskDetectors(Workspace='_wksp.spe-white',MaskedWorkspace='white_masks')
MedianDetectorTest(InputWorkspace='_wksp.spe-
white',OutputWorkspace='white_masks',SignificanceTest='0')
MaskDetectors(Workspace='_wksp.spe-white',MaskedWorkspace='white_masks')
MaskDetectors(Workspace='_wksp.spe-white',MaskedWorkspace='white_masks')
MaskDetectors(Workspace='background_int',MaskedWorkspace='_wksp.spe-
white')
MedianDetectorTest(InputWorkspace='background_int',OutputWorkspace='mask_
bkgd',SignificanceTest='3',LowThreshold='0',HighThreshold='2',LowOutlier='0',High
Outlier='1e+100',ExcludeZeroesFromMedian='1')
MaskDetectors(Workspace='_wksp.spe-white',MaskedWorkspace='mask_bkgd')
ExtractMask(InputWorkspace='_wksp.spe-white',OutputWorkspace='diag_mask')
RenameWorkspace(InputWorkspace='diag_mask',OutputWorkspace='masking')
```

Collect Masked Spectra into a workspace called masking

Units Change

```
ConvertUnits(InputWorkspace='_wksp.spe-white',OutputWorkspace='_wksp.spe-white',Target='Energy')
```

ReBin

```
Rebin(InputWorkspace='_wksp.spe-white',OutputWorkspace='_wksp.spe-white',Params='20,160,100')
```

Apply mask to

```
MaskDetectors(Workspace='_wksp.spe-white',MaskedWorkspace='masking')
```

ReMap

```
GroupDetectors(InputWorkspace='_wksp.spe-white',OutputWorkspace='_wksp.spe-white',
MapFile=r'/Users/jon/Work-computing/mantid_test_data/mari/mari_res.map',Behaviour='Average')
```

Generate Masked WB Integral workspace

Ei Guess

Calculate Ei

M2&M3 spectrum #s & flightpaths

```
GetEi(InputWorkspace='run_wksp',Monitor1Spec='2',Monitor2Spec='3',EnergyEstimate='100')
```

Move instrument components to define T0 as M2 position

```
ChangeBinOffset(InputWorkspace='run_wksp',OutputWorkspace='_wksp.spe',Offset='-2379.33846378')
MoveInstrumentComponent(Workspace='_wksp.spe',ComponentName='Moderator',Z='-1.4419999999999999',RelativePosition
```

Move detectors to position specified in calibration file

```
LoadDetectorInfo(Workspace='_wksp.spe',DataFilename=r'/Users/jon/mprogs/mari_data/RAW/MAR16637.raw',RelocateDets='
```

Units E trans meV

```
ConvertUnits(InputWorkspace='_wksp.spe',OutputWorkspace='_wksp.spe',Target='DeltaE',EMode='Direct')
```

ReBin

```
Rebin(InputWorkspace='_wksp.spe',OutputWorkspace='_wksp.spe',Params='-10,0.2,95',PreserveEvents='0')
```

Tube

Detector efficiency correction

```
DetectorEfficiencyCor(InputWorkspace='_wksp.spe',OutputWorkspace='_wksp.spe')
```

Ei

```
CorrectKiKf(InputWorkspace='_wksp.spe',OutputWorkspace='_wksp.spe')
```

Apply mask to

```
MaskDetectors(Workspace='_wksp.spe',MaskedWorkspace='masking')
```

ReMap

```
GroupDetectors(InputWorkspace='_wksp.spe',OutputWorkspace='_wksp.spe',
MapFile=r'/Users/jon/Work-computing/mantid_test_data/mari/mari_res.map',Behaviour='Average')
```

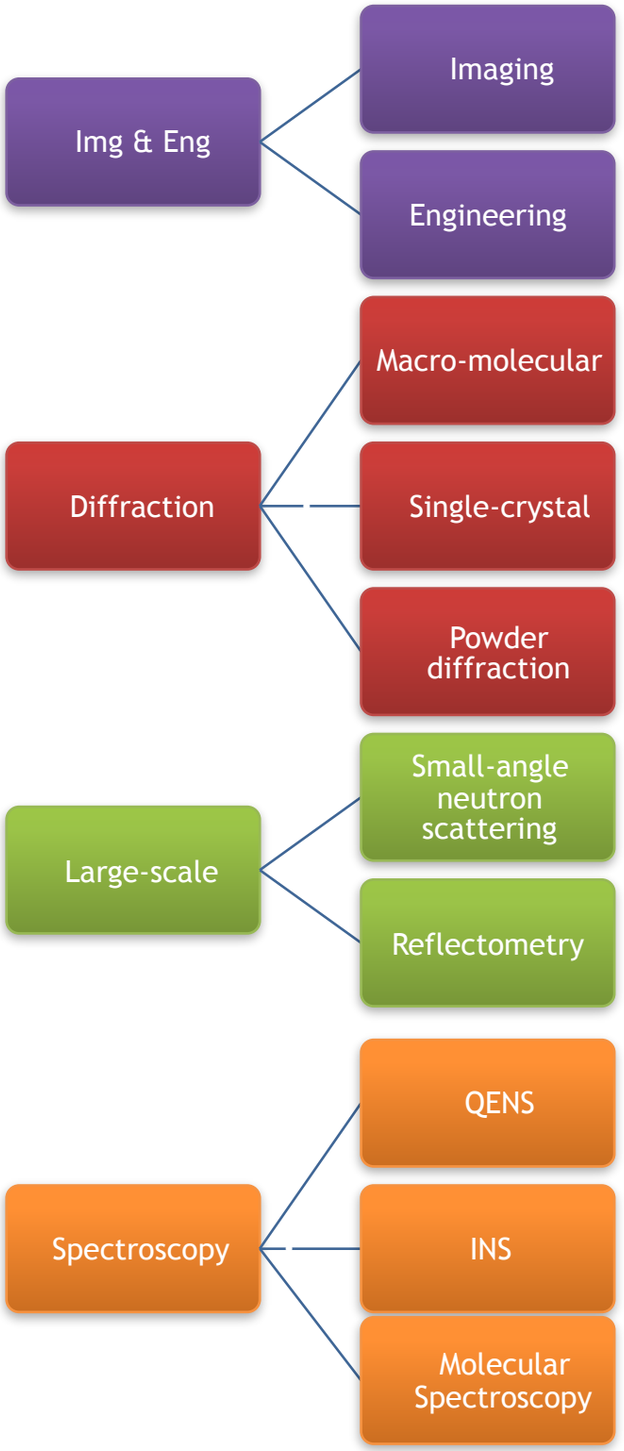
WB Integral workspace

Solid Angle Correction

```
Divide(LHSWorkspace='_wksp.spe',RHSWorkspace='_wksp.spe-white',OutputWorkspace='_wksp.spe')
```

Data analysis overview for ESS

EUROPEAN SPALLATION SOURCE



MuRec, KipTool, Savu, Octopus, VGStuido

MAUD, GSASII, Fullprof, SScanSS

DIALs MXCUBE

CCSL, NSXTool, Mantid

FullProf,GSASII MAUD, Topaz

SASView

GenX, MotoFit, Refl1D, BornAgain

Mantid functionality, multiple MD codes

SpinW, Horace, PACE

A/O Climax, Mantid, Multiple modelling codes

Important to Capture Analysis meta data

Data analysis codes

Many key areas need support to ensure sustainability

Fullprof, GSASII
Single crystal diffraction (& polarised neutron diffraction)

Collaborative development across facilities is very common

Evaluation and inclusion of instrument resolution will benefit from MC development

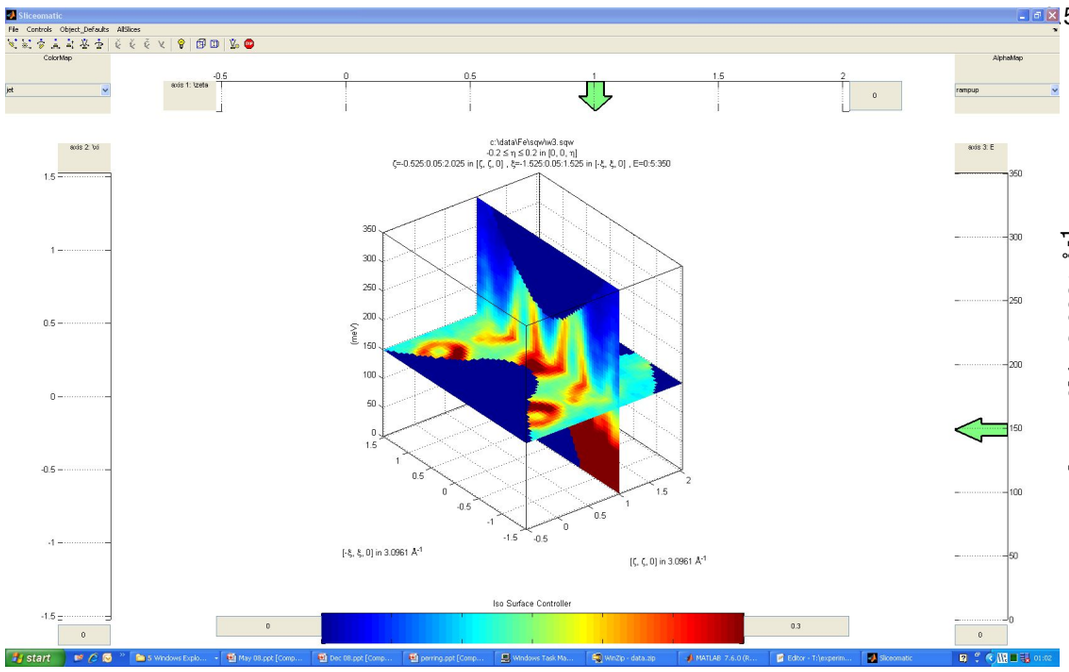
Significant expertise is within the user community

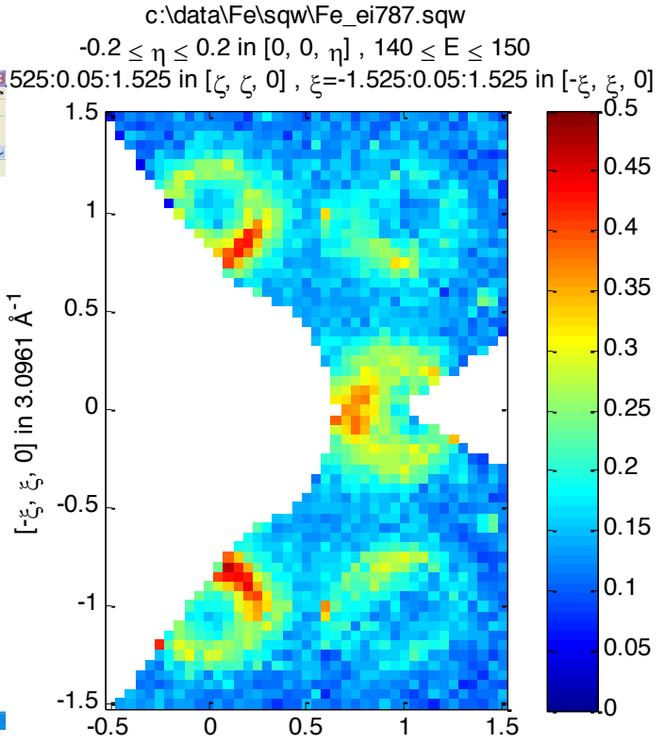
Analysis of DG INS data

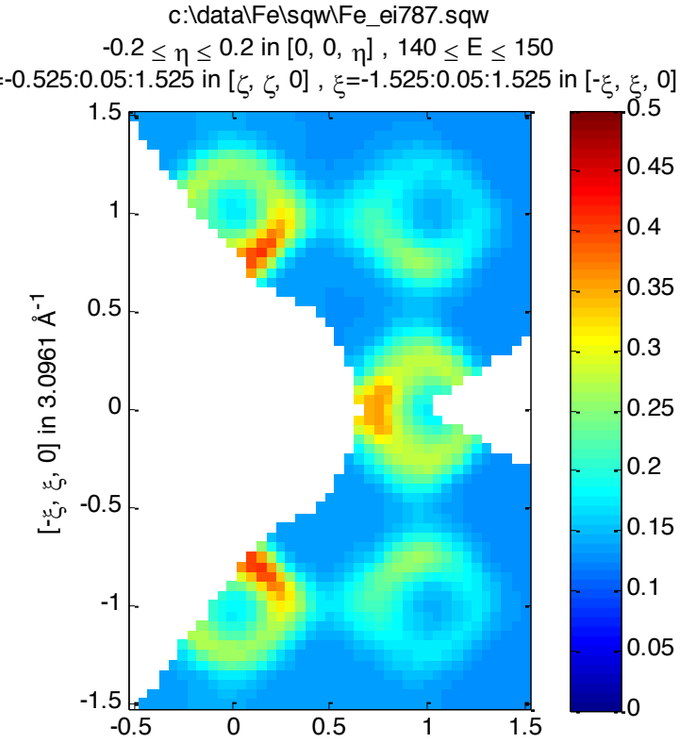
ESS - PSI - ISIS Collaboration

"Rietveld for inelastic" required for over 60% of experiments

High performance generation of 4D datasets
Model fitting & resolution convolution on distributed architecture







Summary

- Write sustainable open source software
- Think about the design and IP issues ahead of time
- Learn python it is becoming very common for all code in the data chain
- Develop a data management plan
  Especially for meta data from analysis

Caveat Emptor

There are no guarantees

Sweat the code and the data