

# Data management and scientific computing at scattering facilities, what it is and why it is important

Jon Taylor  
Raciri summer school  
August 2019



# Data management and scientific computing at scattering facilities, what it is and why it is important

Jon Taylor  
Raciri summer school  
August 2019



# Introduction



- FAIR data for European Photon and Neutron sources.
- From Data Acquisition to Data analysis.
- Jupyter Notebooks for FAIR science.
- Some examples for the tutorial sessions.

<https://github.com/Jon-Taylor/raciri-2019>

- E-learning for neutrons

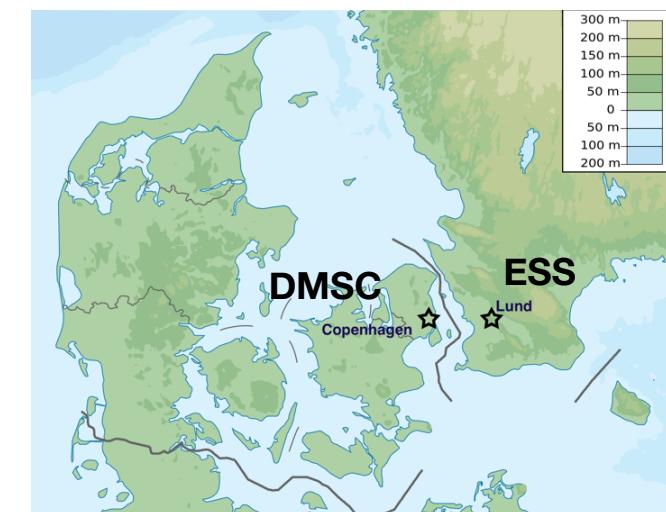
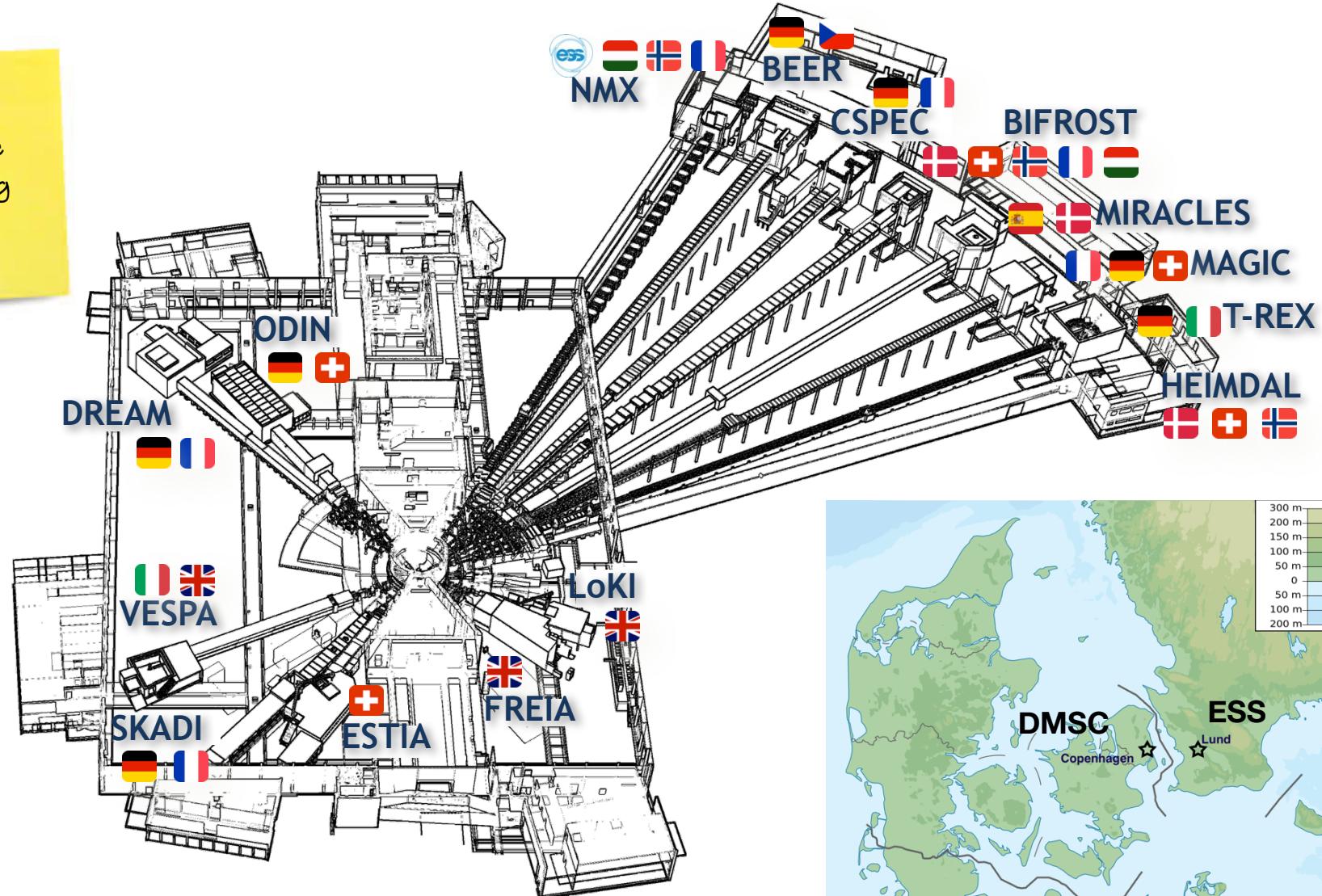
<https://www.e-neutrons.org/moodle/>



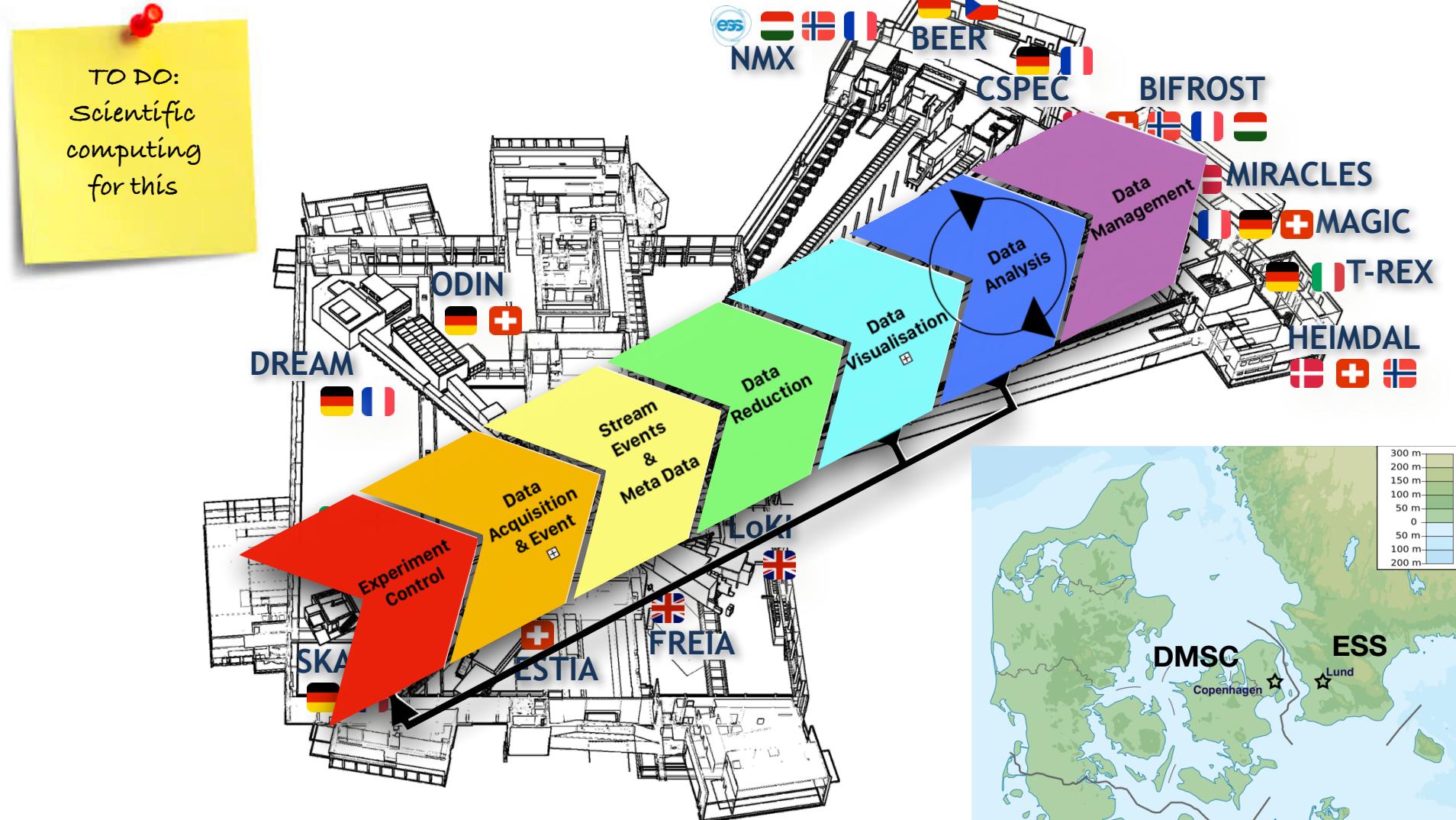
brightness



# DMSC - Data Management and Software Centre



# DMSC - Data Management and Software Centre



# The Team: ESS and InKind Partners



# Our Agenda

Collaborative Open Source Software

Developed with modern practices

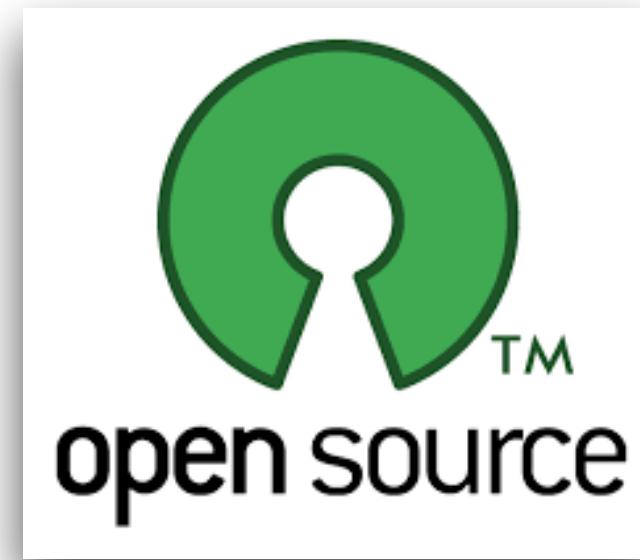
Minimise Single point failures

Community Engagement

Promote User Experience

FAIR Data

Open Science



# European Photon and Neutron Landscape



40k users per year

10's PB of data

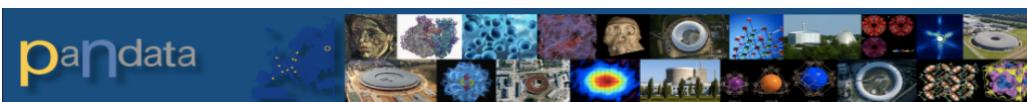
## Multi disciplinary programmes

Individual researchers → Large collaborations

## Industrial and commercial use



	alba	enike	abs	aent	ags	austri	cls	dis	gestri	erst	fern2	fls	hbessy	ill	its	kls	lns	mnstr	mls	retri	orni	sing	sbs	scsce	sobole	solis	slis	stari	
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	73	67	73	63	223	115	643	31	212	120	95	60	15	40	46	48	42	47	184	6	114	14	54		
anta	201	202	19761	682	346	769	1675	1125	2234	968	3380	430	206	1403	937	985	1185	1040	1022	992	1009	310	2110	757	1556	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	769	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	5543	386	776	315	1206	175	78	450	367	361	228	113	173	408	470	407	129	720	34	287	504	99	1003
desy	180	233	1125	326	805	292	386	850	1171	76	3228	486	194	1450	749	559	834	130	472	299	292	322	276	1401	96	619	605	88	488
ds	394	223	2234	730	1535	841	77	1174	1784	979	6881	501	222	1217	1473	2317	653	223	467	574	538	649	374	2335	71	939	887	94	1163
elektra	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	2510	40207	1225	624	2939	3510	4547	643	1101	1166	868	1030	932	5225	76	3124	1632	125	1823	
fmf	78	64	430	470	479	371	83	166	476	225	1276	3760	472	458	1520	838	111	63	96	146	209	559	615	474	47	207	257	26	208
hbz-beer	36	31	206	340	177	83	76	194	222	93	624	504	1817	364	755	532	52	45	43	60	92	305	425	231	13	68	155	10	74
hbz-bessy	229	212	1403	339	612	271	450	1450	1217	1007	2939	472	364	922	752	585	512	141	582	323	361	369	300	1602	54	669	571	60	571
ili	162	120	937	1192	921	412	367	749	1473	592	3510	1558	755	752	10610	1630	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
nsrls	65	46	1042	222	779	198	408	299	574	248	1146	146	60	323	333	358	231	40	126	2296	301	392	100	480	229	349	322	62	620
nsrs	83	48	992	596	598	306	490	274	629	243	868	209	92	361	423	406	192	61	123	300	3560	478	147	523	37	184	843	101	568
psi-sing	81	47	310	386	294	123	129	276	374	177	932	615	425	300	1049	762	81	40	72	100	147	437	2590	570	15	115	259	19	147
psi-sls	315	184	2110	472	1388	517	720	1401	2315	929	5225	474	231	1602	982	816	707	162	471	480	523	551	570	14510	106	1011	851	82	1072
sacra	7	6	107	40	93	22	34	96	71	62	176	27	13	54	54	49	141	5	25	42	37	34	35	106	454	97	436	4	62
soleil	213	114	755	178	421	187	287	619	939	894	3124	207	68	669	641	385	351	165	260	239	184	186	115	1011	97	6388	424	39	325
springb	97	1674	561	792	1130	504	605	887	472	1632	257	155	574	763	800	453	82	160	342	843	662	259	851	416	424	9056	79	732	
sts	8	14	130	63	93	85	99	84	57	125	26	10	60	55	45	36	9	30	62	101	63	19	82	4	39	79	404	20	
stsjr	79	54	3479	381	2093	531	1003	488	1163	39	1823	208	57	541	441	480	818	72	101	20	61	260	325	732	95	90	9665		



# Data & Meta Data



## data

/'deɪtə/

*noun*

noun: **data**

facts and statistics collected together for reference or analysis.

"there is very little data available"

synonyms: facts, figures, **statistics**, details, particulars, specifics, features; [More](#)

- the quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.
- PHILOSOPHY things known or assumed as facts, making the basis of reasoning or calculation.

Origin

LATIN

datum → data  
*mid 17th century*



## metadata

/'metədeɪtə/

*noun*

noun: **meta-data**

a set of data that describes and gives information about other data.



# Where is my Data?



How do I get it?

How do I look at it?

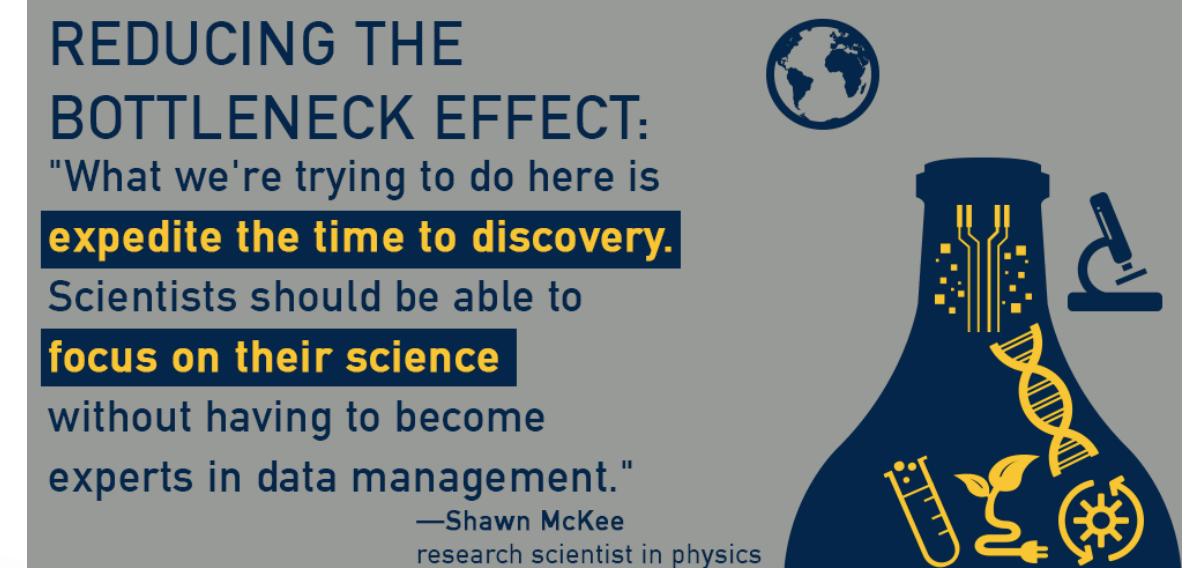
What does it mean?

Is it right?

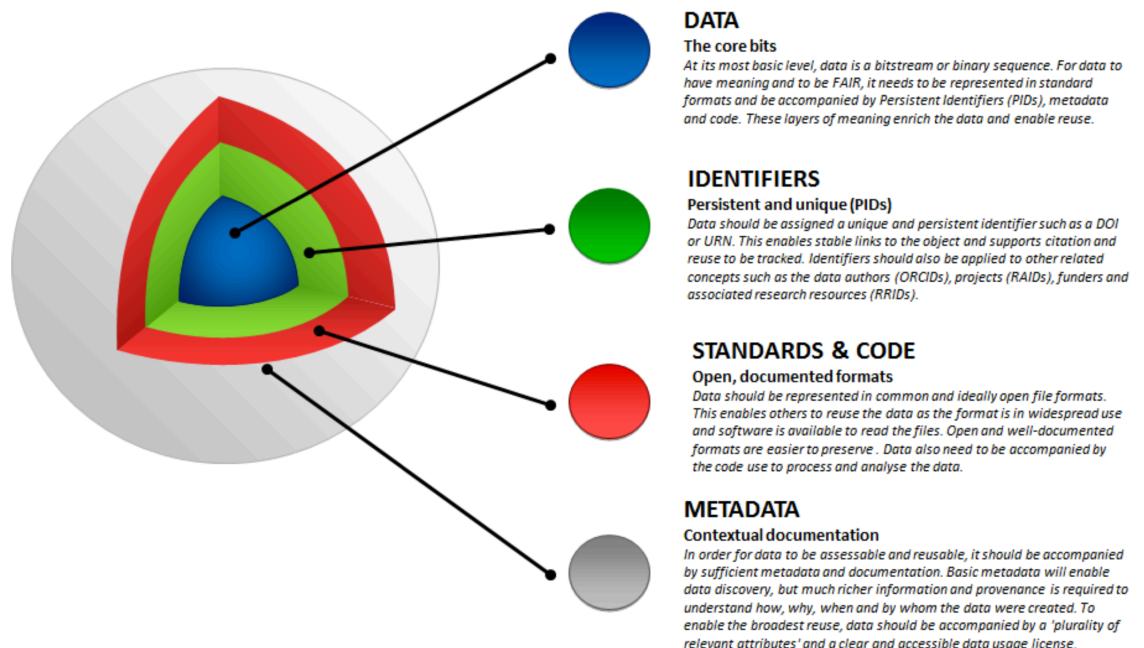


**REDUCING THE  
BOTTLENECK EFFECT:**  
"What we're trying to do here is  
**expedite the time to discovery.**  
Scientists should be able to  
**focus on their science**  
without having to become  
experts in data management."

—Shawn McKee  
research scientist in physics



# Data Is Our ‘Core Product’



- EuXFEL ~500TB - 2PB data per experiment.
- ESS ~ 250TB per day.
- Data volume is too much for user portable drives.
- Central access is needed.
- Data management is essential.

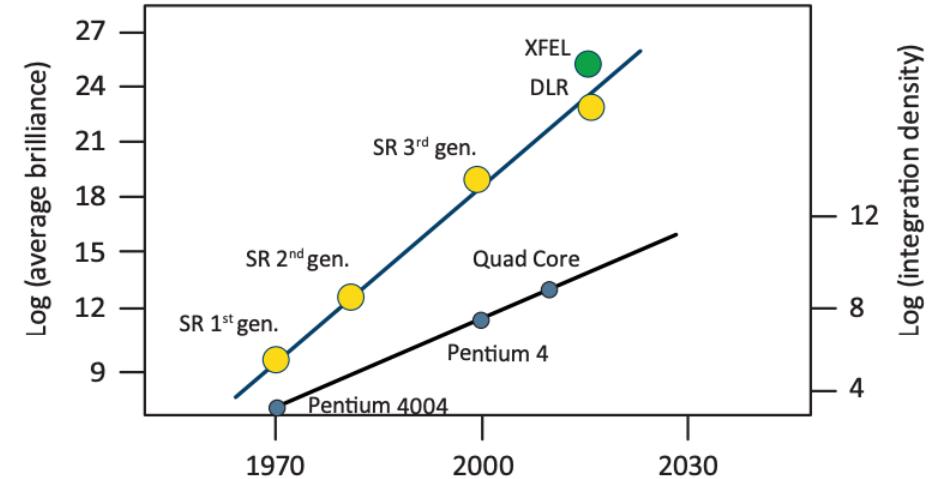


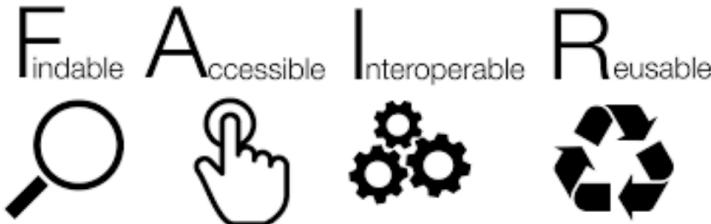
Fig. 1. Comparison of the brilliance of X-rays delivered by photon facilities over the years (LEAPS' law) with the gain in transistor density in integrated circuits (Moore's law). (DLR = Diffraction Limited Rings with ultimate performance; XFEL = X-ray free-electron laser)

# FAIR Principles.

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. doi:10.1038/sdata.2016.18

## To be Findable:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.



## To be Accessible:

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
  - A1.1 the protocol is open, free, and universally implementable.
  - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

## To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

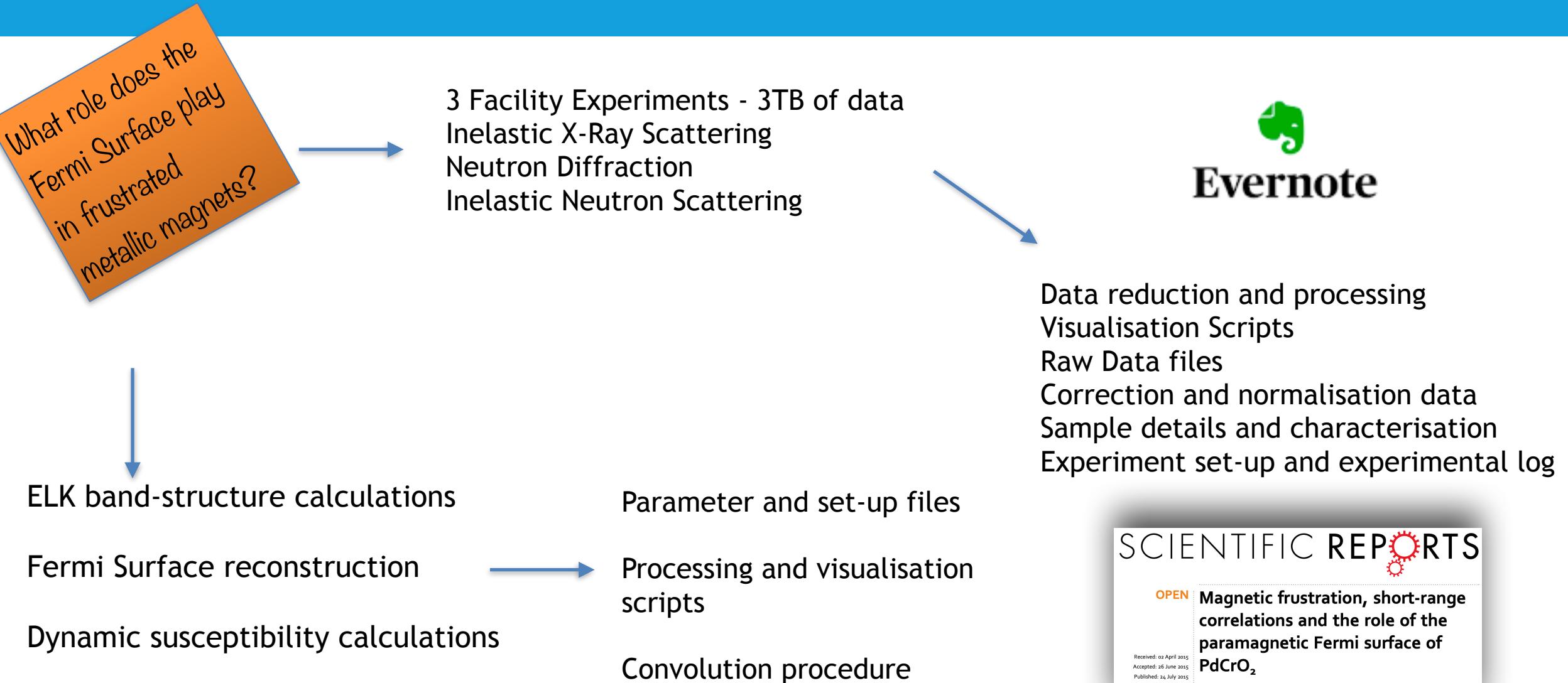
## To be Re-usable:

- R1. meta(data) have a plurality of accurate and relevant attributes.
  - R1.1. (meta)data are released with a clear and accessible data usage license.
  - R1.2. (meta)data are associated with their provenance.
  - R1.3. (meta)data meet domain-relevant community standards.

# Core Product is Data

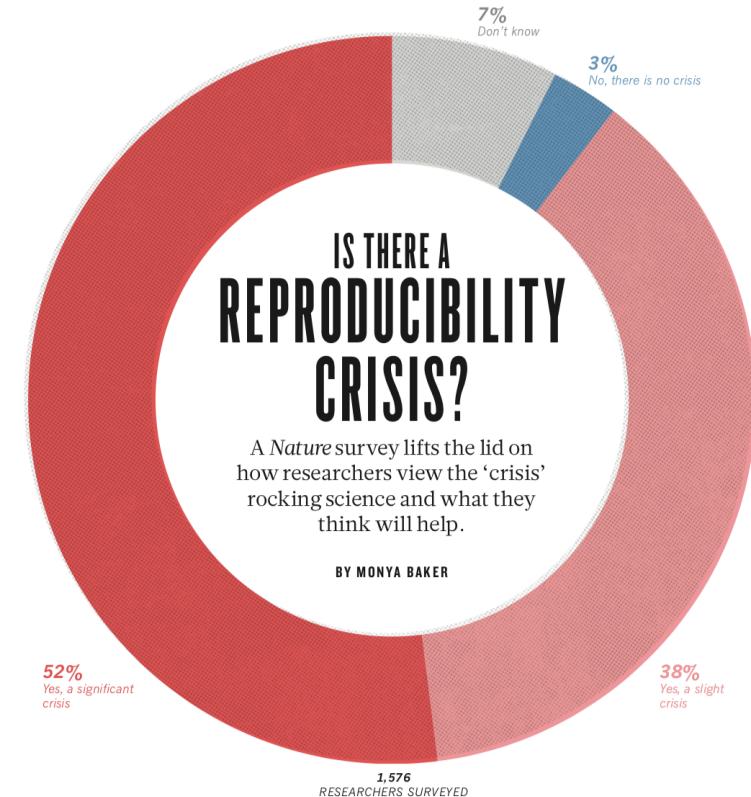
- FAIR is a real challenge for facilities and for **users**
- 1500 + individual experiments (data sets) per year
  - Each experiment may have 200+ individual files
- Data rate has increased considerably
- Photon sources generate increasingly large volumes
  - Up to PBs per experiment on some beam lines
- Neutron sources produce much more data than ever
  - ESS at full specification 250TB per day (10GB/Minute/Instrument)
    - $\sim 10^9$  Events / s each event is 32bit position + 64bit Time stamp

# The FAIR Data & Meta Data Challenge



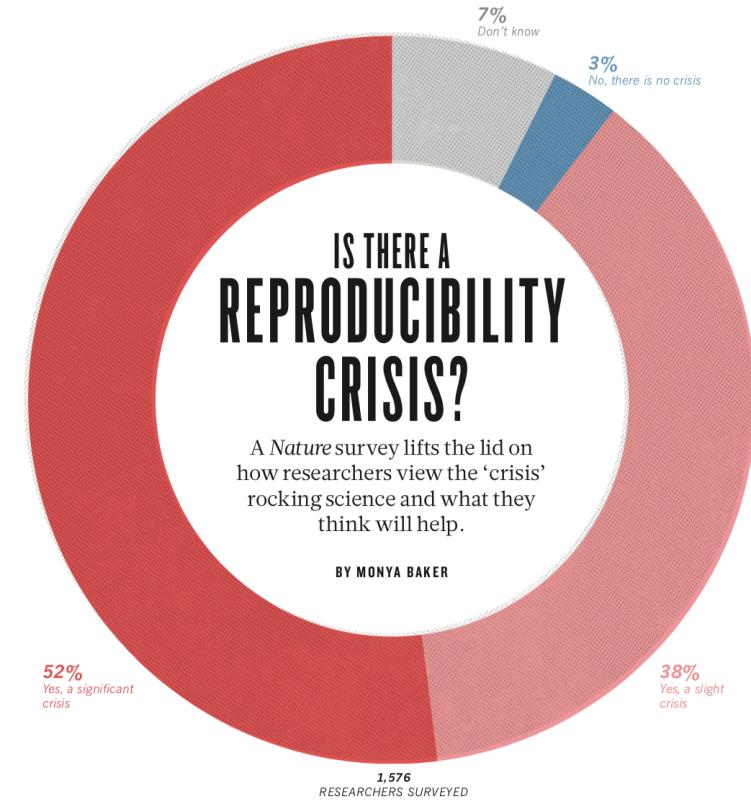
# What makes good science

- One metric is reproducibility
  - How to reproduce a measurement.
  - How to reproduce an analysis.
  - How to reproduce a calculation.
- FAIR applies to software as much as to Data



# FAIR Enables Reproducibility

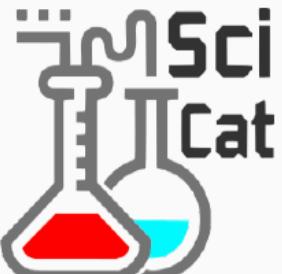
- Current tooling is not sufficient for FAIR\*
- Beneficial for original research team
- The push for Open Science drives investment that benefits all
- \* For ~75% of facility research programmes



# FAIR Data. FAIR Software.



- Common data types that are complete descriptors
- Capture Rich Meta Data
- Catalogue and Curate with a data policy
- Persistently identify Data and Software



**SciCat Project**  
SciCat is a metadata catalogue for scientists



## NeXus

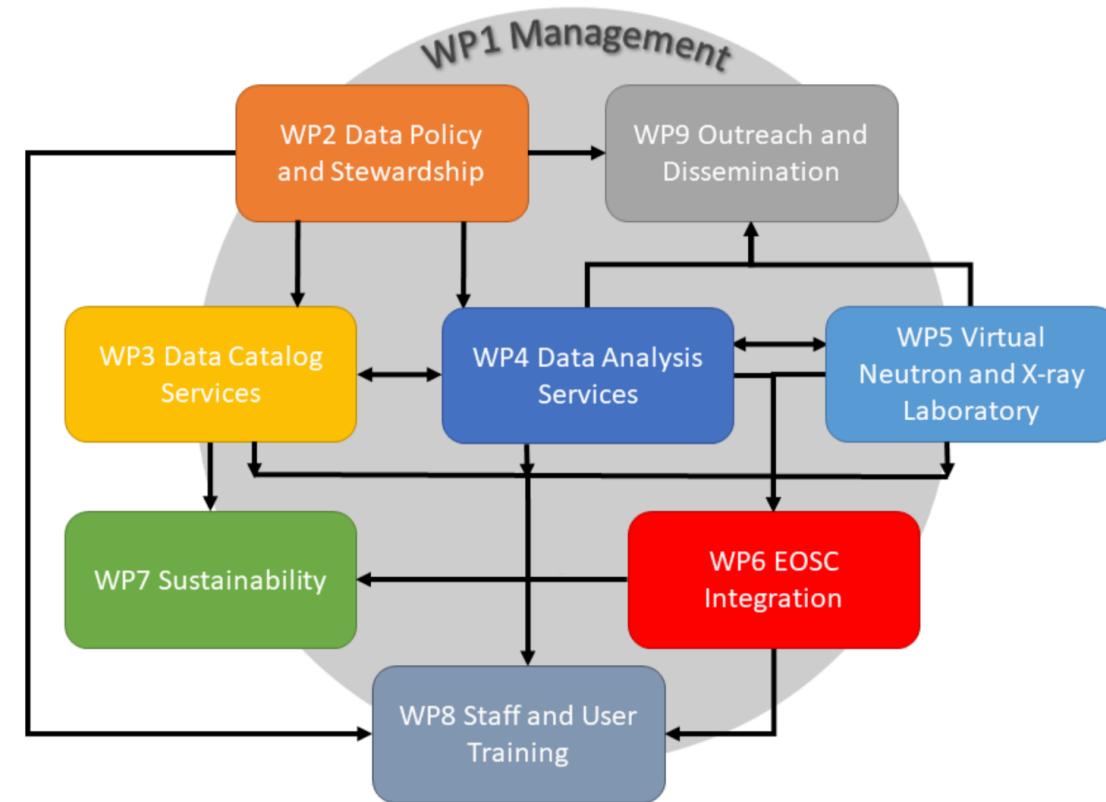
NeXus is developed as an international standard by scientists and programmers representing major scientific facilities in Europe, Asia, Australia, and North America in order to facilitate greater cooperation in the analysis and visualization of neutron, x-ray, and muon data.

## Make Fair data a reality for Photon and Neutron **ESRFI** facilities.

**FAIR - PaNOSC** will comply with the FAIR principles in the following ways:

- Findable** - all data will have a doi, rich metadata, common api for federated search
- Accessible** - api will support open protocol, metadata accessible even without data
- Inter-operable** - metadata to follow community standards (NeXus), register metadata
- Reusable** - follow community standardise metadata, clear licence (CC-BY)

**Vision- PaNOSC** will help fasttrack new Photon and Neutron facilities, improve data services for existing users, create a new class of virtual users of open data, help build and integrate the PaN RIs to the **EOSC**.

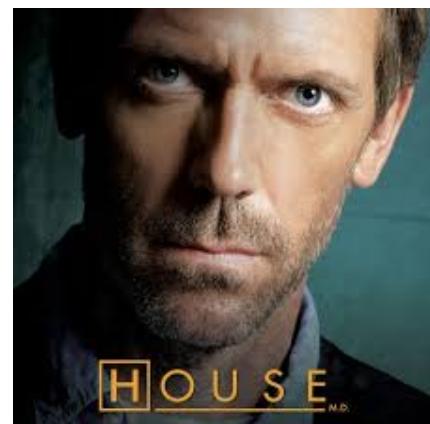
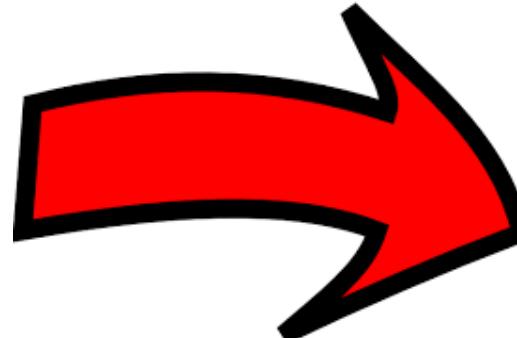
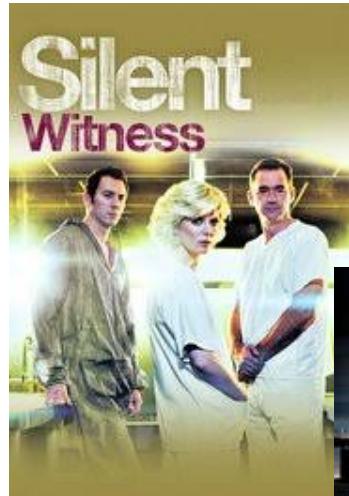


# Scientific Computing at facilities.

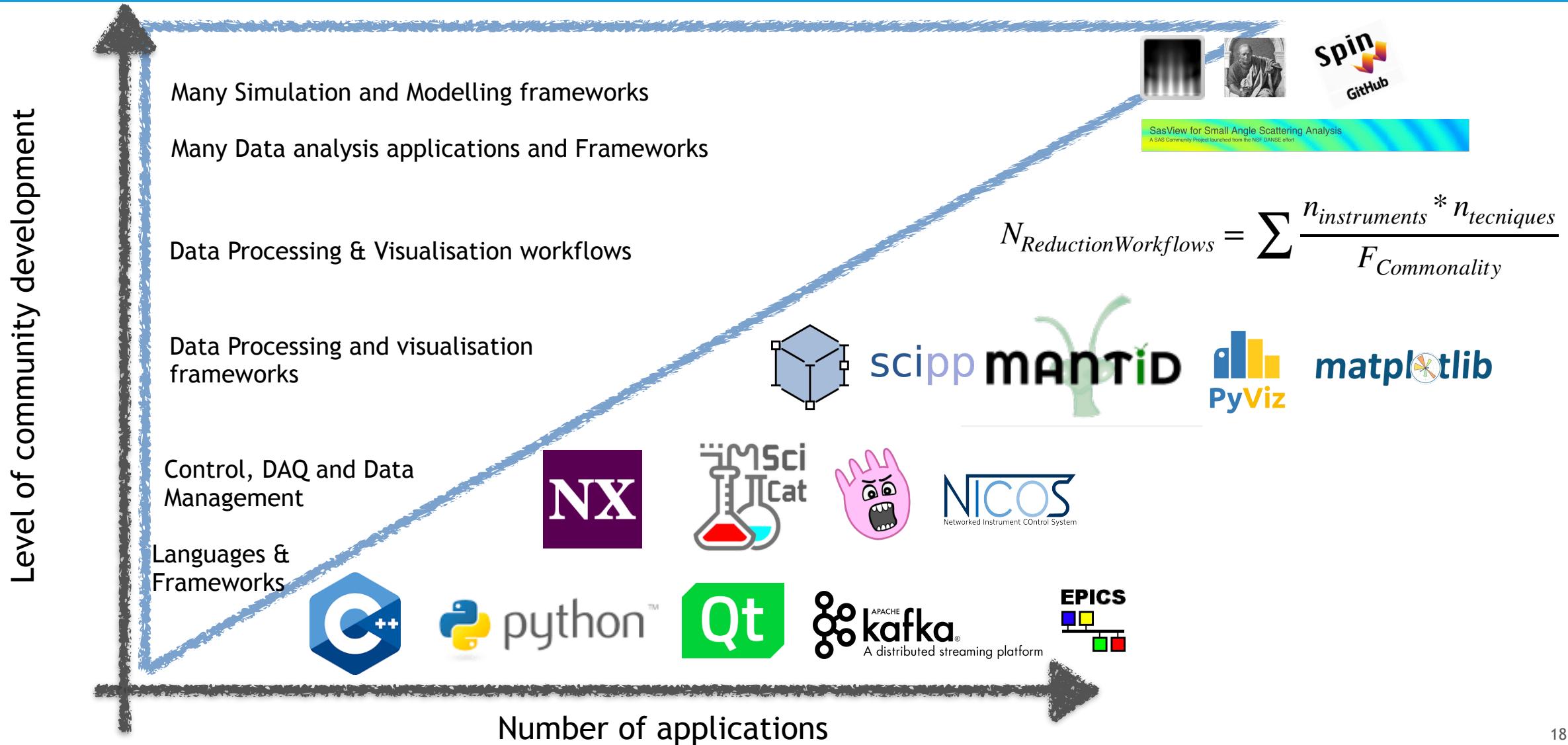
## How To Improve Scientific Efficiency (A TV series analogy)

Move from Post Mortem Analysis

To Live Analysis

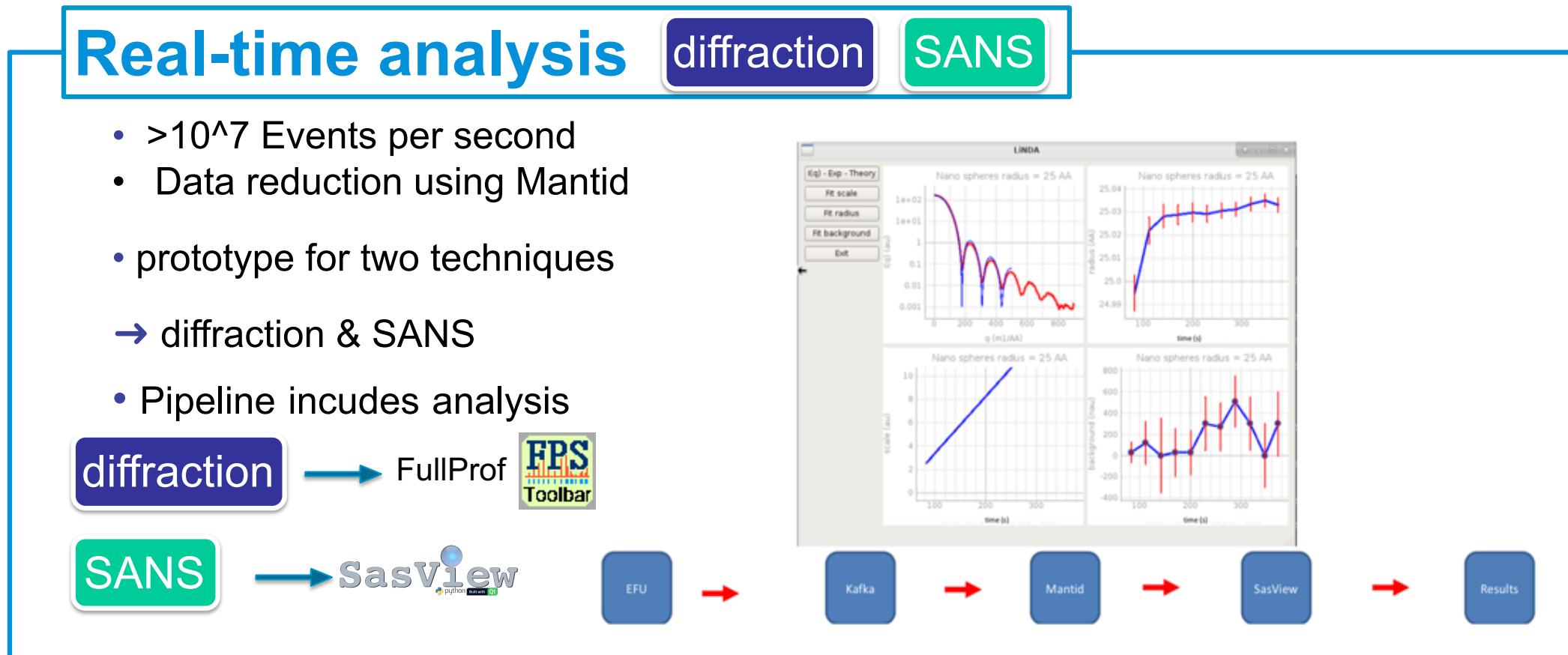


# Scientific Application Stack



# Real Time Pipelines.

- Stream data through processing to analysis.



# Whats in a pipeline

- A few million lines of code
- C++, Python
- Continuous integration automated testing, build and deploy
- 100Gb High performance networking
- Enterprise servers
- High performance storage systems
- Team of scientists, computer scientists, data scientists and developers

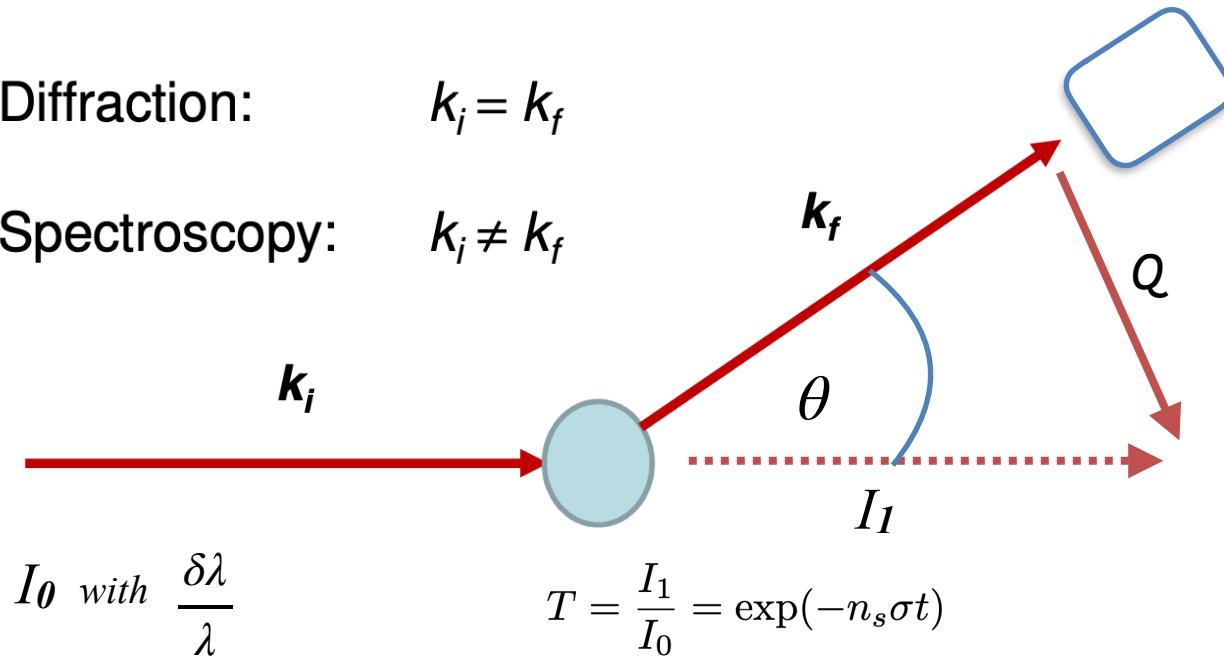


# TOF Neutron scattering instrumentation

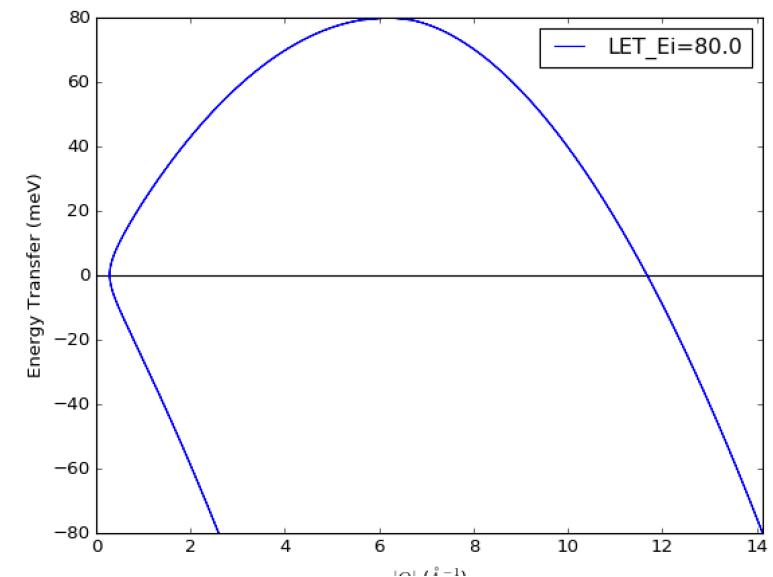
- The neutron energy is encoded in its *Time of Flight*

Diffraction:  $k_i = k_f$

Spectroscopy:  $k_i \neq k_f$



- Volume sampled in  $Q$  is limited by kinematics of scattering.
- Most experiments scatter ~10% of incident neutrons
- Detectors have a finite performance. Dead time is an issue.
- Most (not all) detectors work as single particle counters.
  - Assume Poisson statistics.



# Time of Flight Neutron Scattering

Pulsed sources rely on an accurate time system better than 1ns

Many instruments count in event mode or list mode.

Every neutron detected as good is given a timestamp.

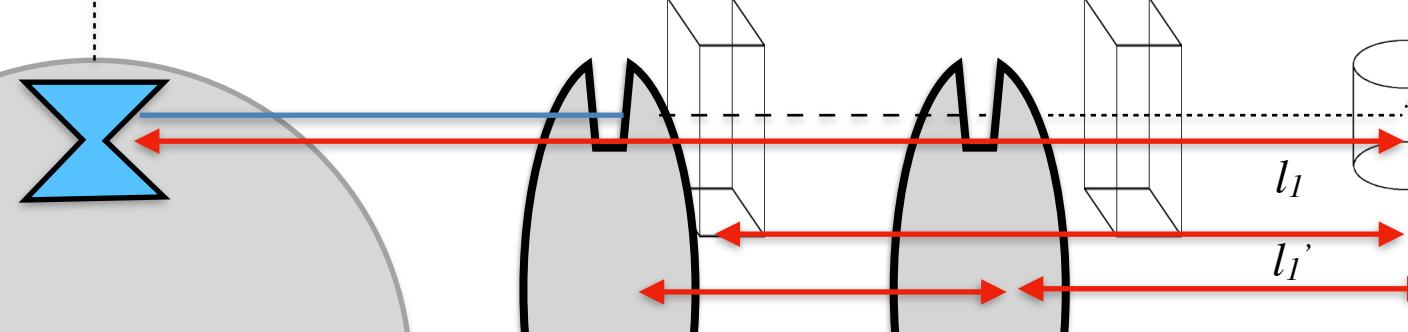
All the meta data can be timestamped

Save data lists to a hdf5 file along with the geometry of the experiment

List data can then be evaluated in software.

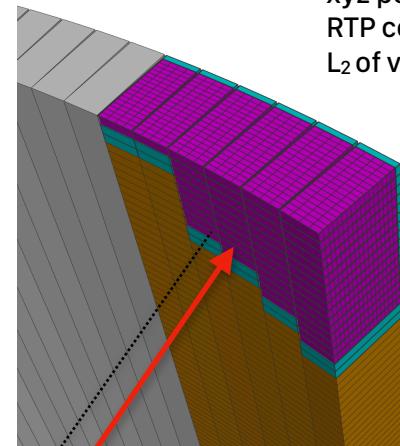
 Wall clock time proton pulse  
11:10  
Proton Pulse ID

Monitor List mode DAQ  
offsets from Wall clock  
:01  
:02  
...  
+ ID  
(Can histogram if required)



Detector List mode DAQ  
offsets from Wall clock  
:05  
:06  
...  
+Pixel ID

xyz position  
RTP coordinates  
 $L_2$  of voxel



$$\lambda = \frac{h}{p} = \frac{h \cdot \text{tof}}{m_N \cdot L_{\text{tot}}} = 2d \sin \theta$$

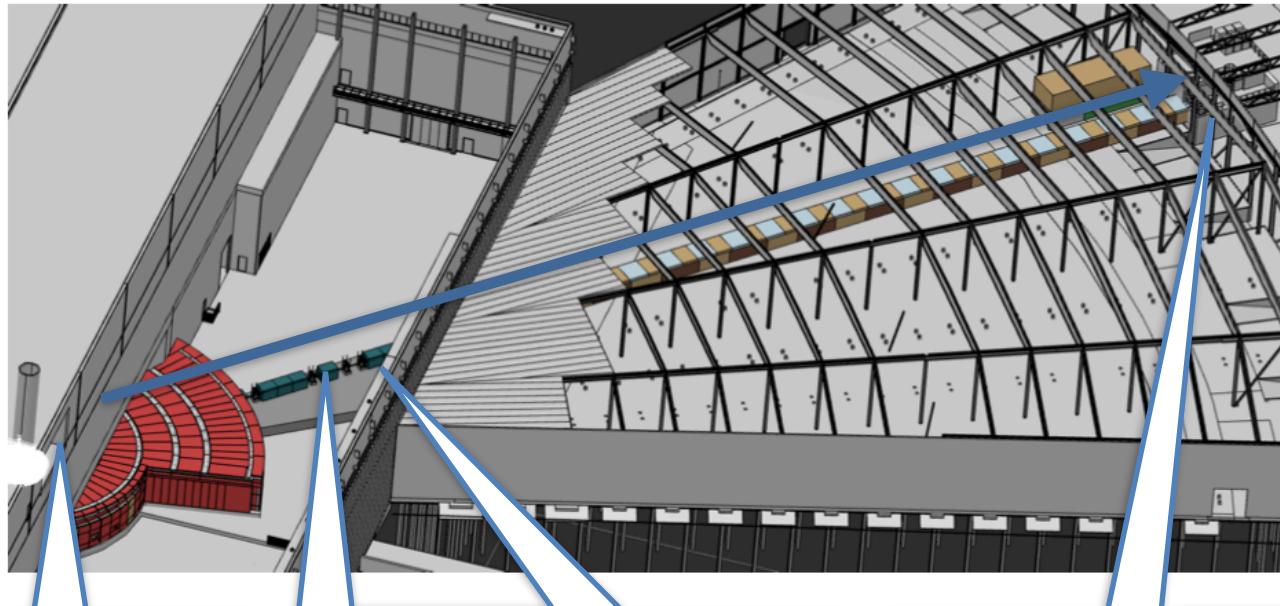
$$\text{tof} = t - t_0$$

$$L_{\text{tot}} = L_1 + L_2$$

**Calibration required for:**  
Motion axis  
chopper offsets and transmission  
Flight paths  
Detector voxel / pixel positions

# TOF Neutron scattering instrumentation

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



Neutron source

Neutron choppers

Neutron guide

Neutron instrument

Sample position

Detector  
tank 3D  
voxel dets

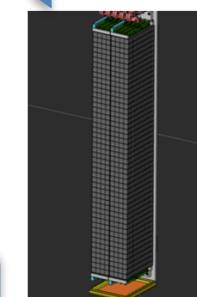
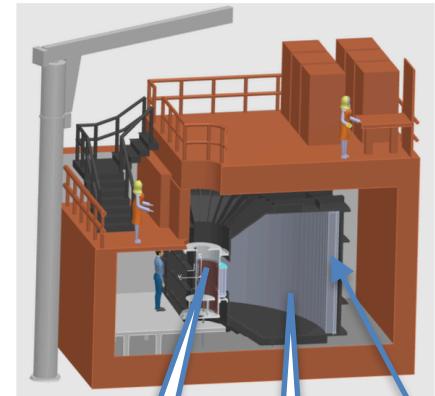


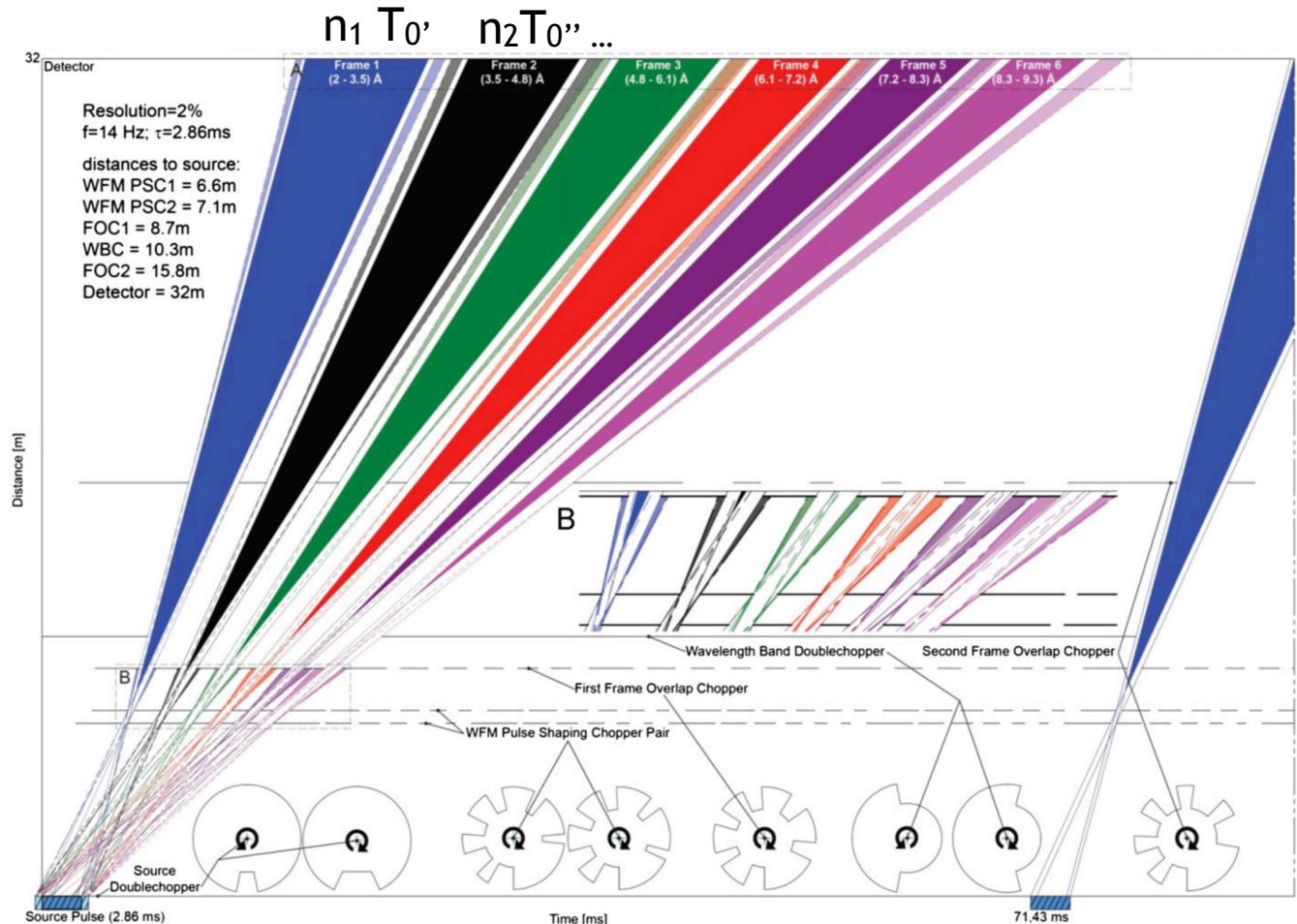
Figure 18: CAD image of a B10  
detector module.  
M.Anastopoulos

$10^6$  Detector elements  
 $>10^7$  n/s

~5TB of data per day

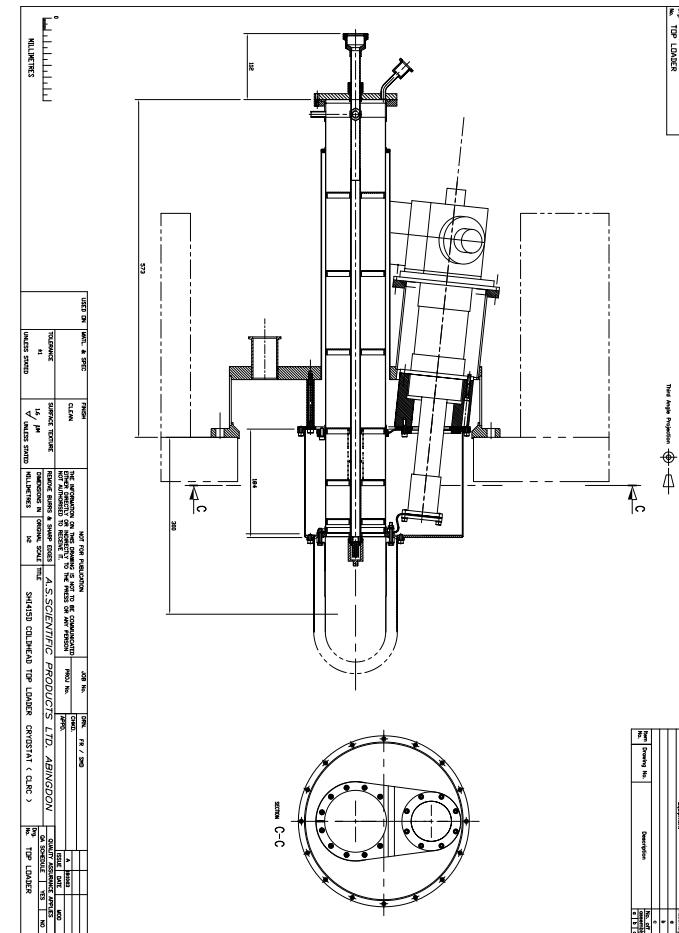
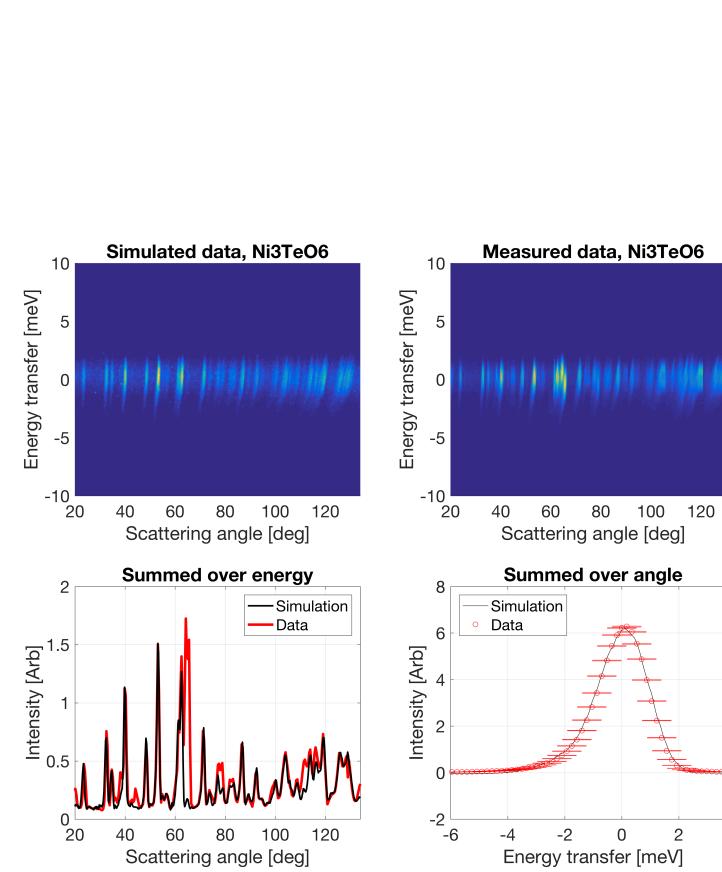
# Data reduction / processing at TOF Neutron facilities

- Loading experimental data.
- Loading calibration data.
- Normalisation.
  - Intensity normalisation,  $I / I_{\text{Monitor}}$
  - Solid Angle correction,  $I / I_{\text{Vanadium}}$
- Convert Units
  - Spectrum to Theta, Q...
  - ToF to D, meV...
- Correct for experimental effects ...



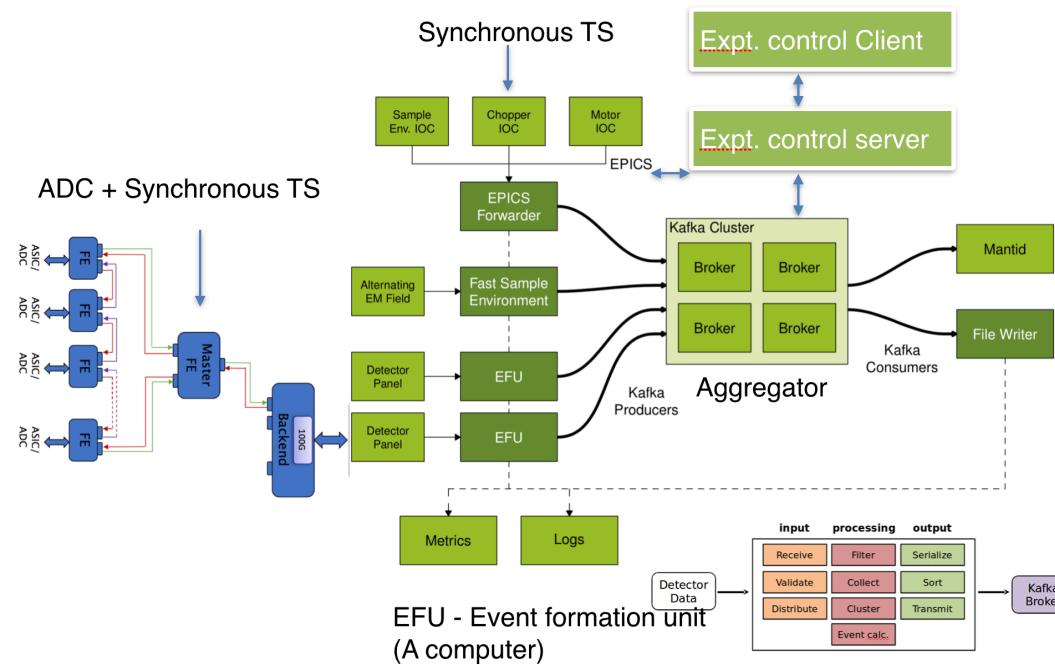
# Experimental / Instrument effects

- Detector efficiency.
  - Sample absorption.
  - Background scattering.
    - Elastic or inelastic
  - Multiple scattering.
  - Inelasticity effects.
  
  - Solve analytically or with MC
    - McStas Union component - model complex geometries
    - Source + instrument + SE +sample
    - Sample environment multiple scattering simulation
    - Get the resolution and flux for free



# Event Mode Data Acquisition

- TimeStamp everything. (Neutrons, Choppers, Motion control)
- Filter Neutron events in software.
  - The Good from the Bad and Ugly.
- Go Fast.



# Event Mode Data Acquisition examples from SNS



research papers

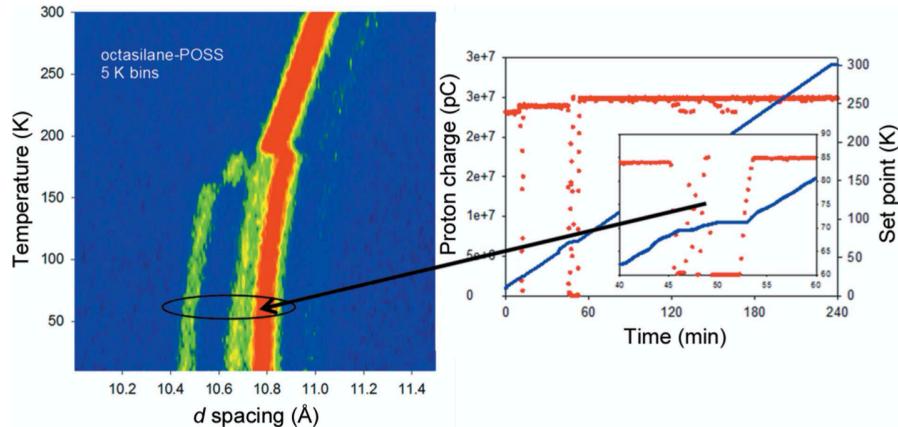
JAC  
JOURNAL OF  
APPLIED  
CRYSTALLOGRAPHY

ISSN 1600-5767

Received 13 October 2017

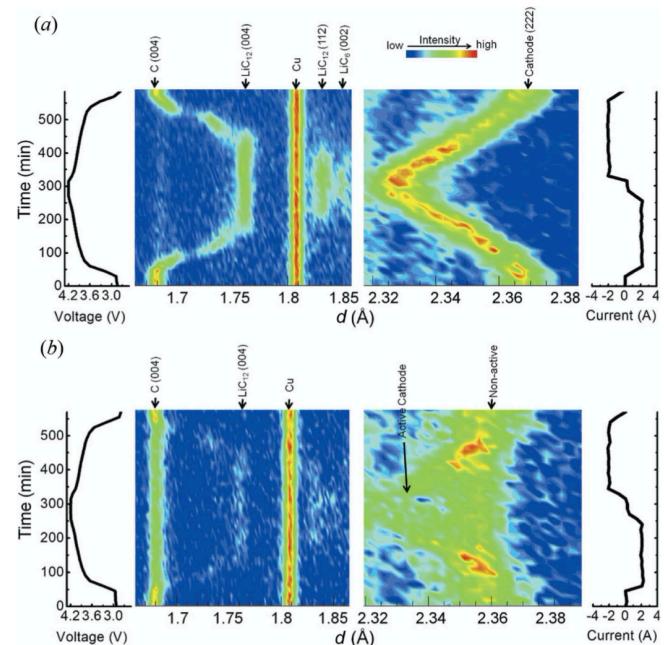
Accepted 22 March 2018

Edited by M. Arai, European Spallation Source,  
Lund, Sweden



**Figure 4**  
A contour plot of both transitions in octasilane-POSS during heating from 10 to 300 K at  $2 \text{ K min}^{-1}$  and the corresponding plot of temperature/proton charge *versus* time. The graph shows that the temperature holds when the beam has tripped. Had the temperature ramp not paused during this time there would be a gap in the data in the region indicated on the contour plot.

- Ramp control parameter - T, H/E Field.
- Collect data continuously.
- Filter events in software.



**Figure 7**  
Contour plots during charge-discharge cycles, showing signs of degradation of a large-format battery (a) before and (b) after long-term cycling, with limited  $\text{Li}^+$  insertion/removal processes in an active material (Cai, An *et al.*, 2013).

# Data Streaming & Event Data



- How to process events without creating a histogram.
- In some cases it allows fast execution
- Realtime Processing of neutron data
- Modern photon detectors can count in event mode (TimePixIII)

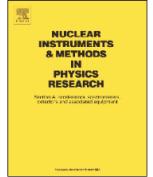
Nuclear Instruments and Methods in Physics Research A 803 (2015) 24–28



Contents lists available at ScienceDirect

Nuclear Instruments and Methods in Physics Research A

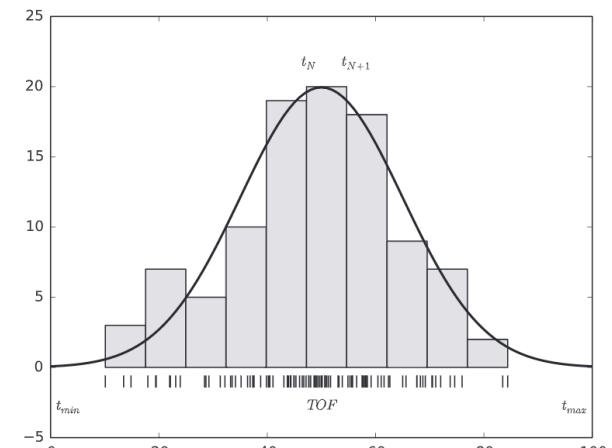
journal homepage: [www.elsevier.com/locate/nima](http://www.elsevier.com/locate/nima)



Event-based processing of neutron scattering data

Peter F. Peterson\*, Stuart I. Campbell, Michael A. Reuter, Russell J. Taylor, Janik Zikovsky

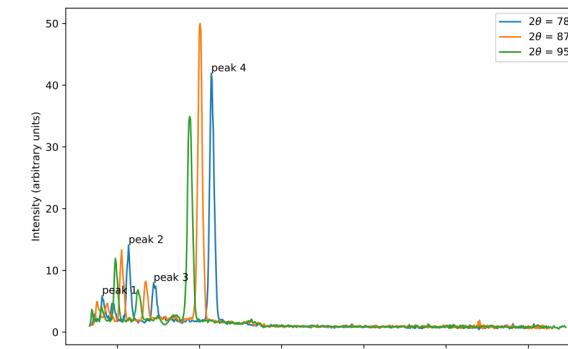
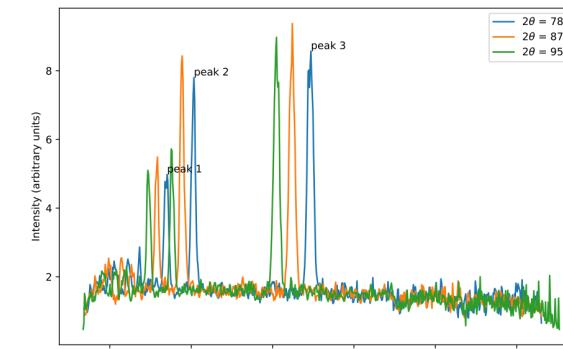
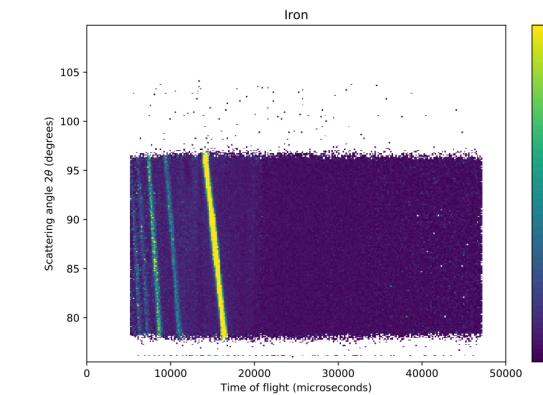
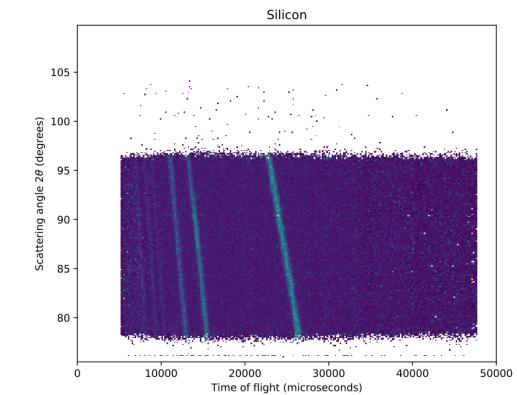
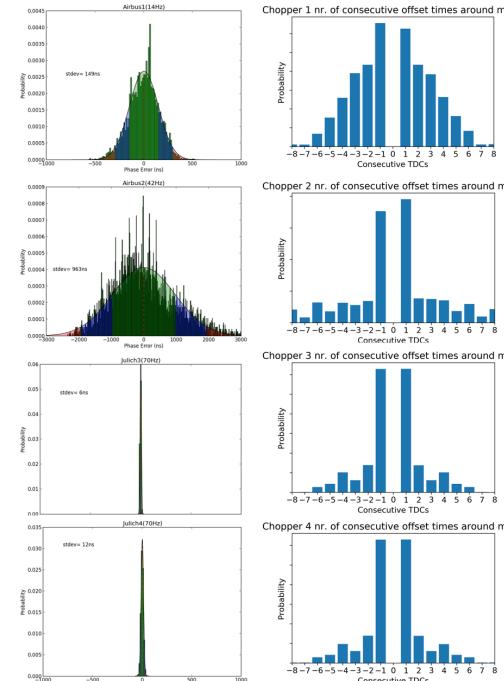
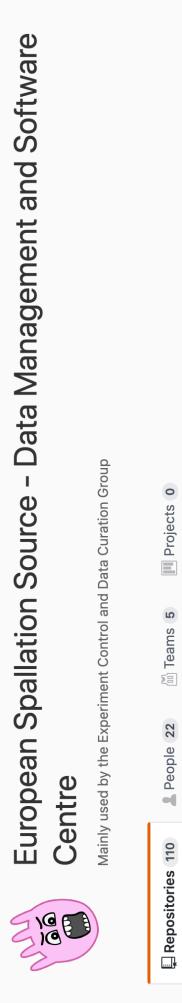
*Neutron Data Analysis and Visualization, Oak Ridge National Laboratory, Oak Ridge, TN, USA*



**Fig. 1.** Schematic diagram of histogramming. Dashes below are individual events. Boxes are the histogram representation. Curved line is the normal distribution for reference.

# Testing ESS DAQ Architecture.

- Full End to End Tests on V20 at HZB.
- CI development at DMSC and deployment into integration Lab in Lund.



Stitched in Mantid

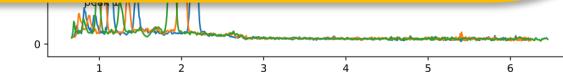
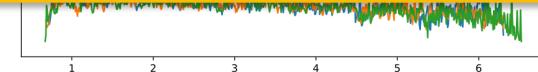


# Testing ESS DAQ Architecture.



## Event formation throughput on 1 or 2 cores

Detector	Packet Rate [ pkt/s ]	Trigger Rate [readouts/s]	Event Rate [events/s]	cores
Gd-GEM	n/a	18.6 M	500 k – 1 M	1
Multi-Grid	86.000	3.0 M	2.46 M	1
Multi-Blade	82.000	5.6 M	2.31 M	2
SoNDe	94.000	23.5 M	23.5 M	2

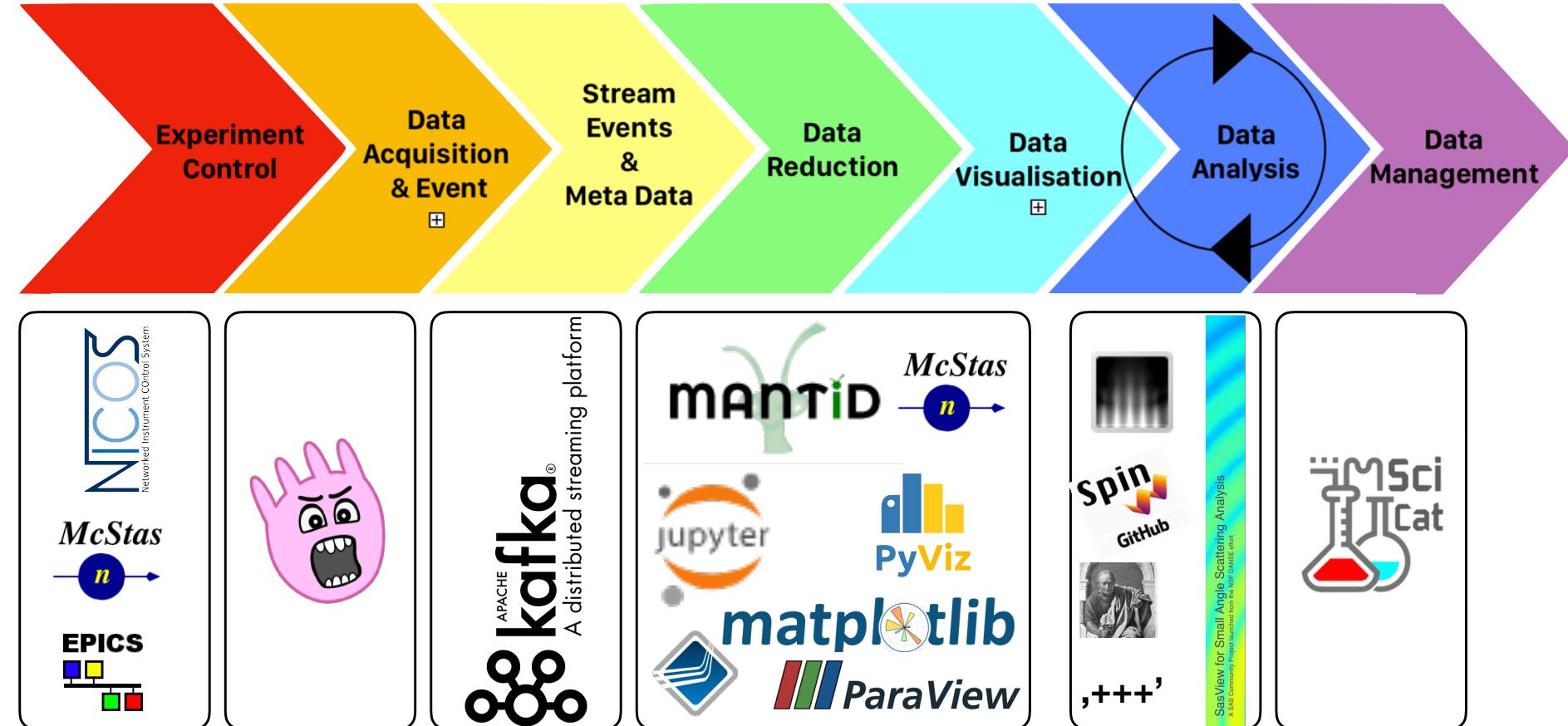


# ESS Scientific Computing Pipeline



Core acquisition, controls and reduction stack tested on V20 test beam line.

- Detectors
- Monitors
- Motion
- Choppers
- Sample environment
- Highlights successful implementation
- Areas for further development



# Data files & Data Containers for FAIR



Journal of  
Applied  
Crystallography

ISSN 1600-5767

Received 17 November 2014  
Accepted 17 December 2014

## computer programs

### The NeXus data format

Mark Könnecke,<sup>a\*</sup> Frederick A. Akeroyd,<sup>b</sup> Herbert J. Bernstein,<sup>c</sup> Aaron S. Brewster,<sup>d</sup> Stuart I. Campbell,<sup>e</sup> Björn Clausen,<sup>f</sup> Stephen Cottrell,<sup>b</sup> Jens Uwe Hoffmann,<sup>g</sup> Pete R. Jemian,<sup>h</sup> David Männicke,<sup>i</sup> Raymond Osborn,<sup>j</sup> Peter F. Peterson,<sup>e</sup> Tobias Richter,<sup>k</sup> Jiro Suzuki,<sup>l</sup> Benjamin Watts,<sup>m</sup> Eugen Wintersberger<sup>n</sup> and Joachim Wuttke<sup>o</sup>

<sup>a</sup>Laboratory for Development and Methods, Paul Scherrer Institute, 5232 Villigen-PSI, Switzerland, <sup>b</sup>ISIS Facility, STFC, Rutherford Appleton Laboratory, Didcot, Oxfordshire OX11 0QX, England, <sup>c</sup>imgCIF, Dowling College, Shirley, NY 11769, USA, <sup>d</sup>Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, <sup>e</sup>Spallation Neutron Source, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA, <sup>f</sup>Los Alamos National Laboratory, Los Alamos, NM 87545, USA, <sup>g</sup>Helmholtz-Zentrum Berlin für Materialien und Energie GmbH, 14109 Berlin, Germany, <sup>h</sup>Advanced Photon Source, Argonne National Laboratory, Argonne, IL 60439, USA, <sup>i</sup>ANSTO, New South Wales 2234, Australia, <sup>j</sup>Argonne National Laboratory, Argonne, IL 60439, USA, <sup>k</sup>Diamond Light Source, Didcot, Oxfordshire OX11 0DE, England, <sup>l</sup>KEK, Ibaraki 305-0801, Japan, <sup>m</sup>Swiss Light Source, Paul Scherrer Institute, 5232 Villigen-PSI, Switzerland, <sup>n</sup>Deutsches Elektronen-Synchrotron DESY, 22607 Hamburg, Germany, and <sup>o</sup>Forschungszentrum Jülich, JCNS at MLZ, 85747 Garching, Germany. Correspondence e-mail:  
mark.koennecke@psi.ch

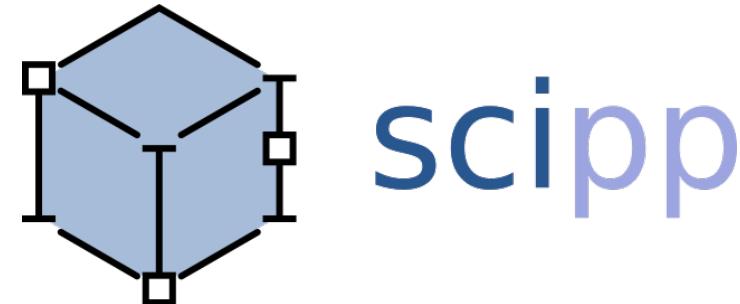
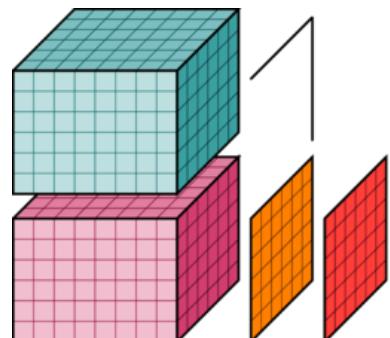
NeXus is an effort by an international group of scientists to define a common data exchange and archival format for neutron, X-ray and muon experiments. NeXus is built on top of the scientific data format HDF5 and adds domain-specific rules for organizing data within HDF5 files, in addition to a dictionary of well defined domain-specific field names. The NeXus data format has two purposes. First, it defines a format that can serve as a container for all relevant data associated with a beamline. This is a very important use case. Second, it defines standards in the form of application definitions for the exchange of data between applications. NeXus provides structures for raw experimental data as well as for processed data.

- Custom Photon and Neutron classes
- Based on HDF5 (h5Py)
- Used at ISIS, DLS, SLS, ESRF, ILL, SNS ...
- The most common data type at facilities?
- Paper is cited 41 times?
- Structured Container for data and meta data.
- Classes for Raw and processed data
- Special classes for domain types SANS → NXSAS
- Definition for Instrument Geometry and setup

# Processing Data.

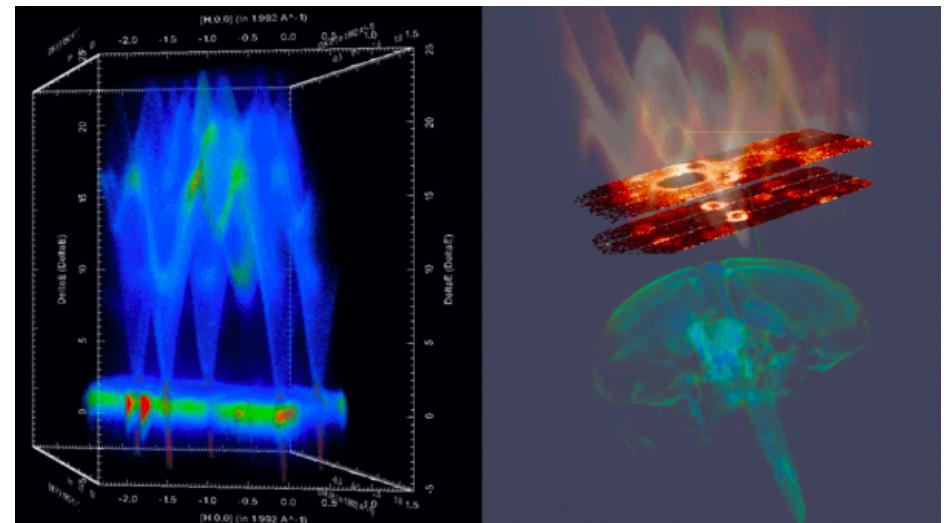
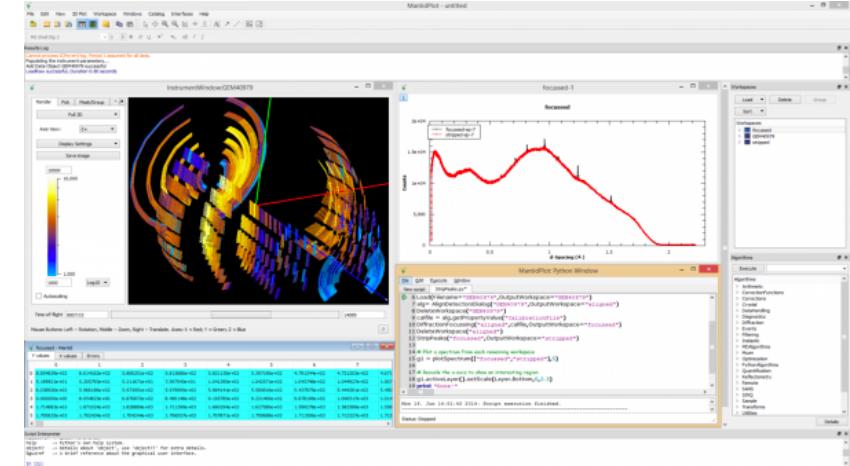
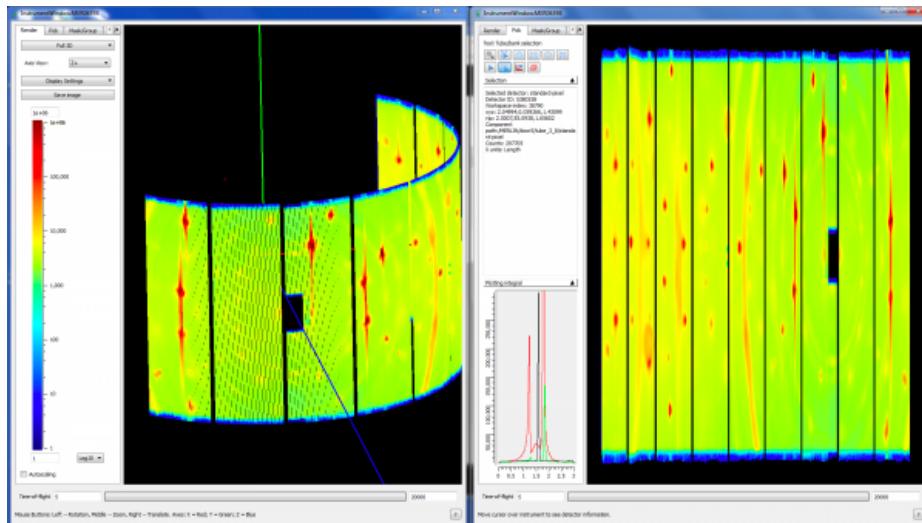


- Use a framework
- Define the processing steps
- Make Processed data documented and FAIR





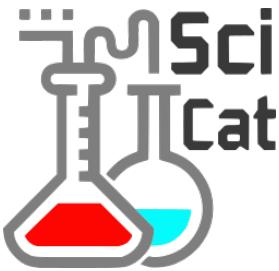
- Data Processing & Analysis
- Neutron TOF and reactor data
- Used at ISIS, ILL, SNS, ESS
- Python & C++



# SciCat - the ESS metadata catalogue



- SciCat stores metadata from experiments at ESS
- Users can search, sort and filter by types of experiment and preview images

A screenshot of a web browser displaying the SciCat ESS metadata catalogue. The page title is "SciCat ESS". On the left, there is a sidebar with search filters: "Text Search" (with "SoNDe" selected), "Location" (with "SoNDe" selected), "Group", "Type", "Keywords", and "Select a date ran...". The main area shows a table of dataset results. The columns are: Name, Source Folder, Size, Start Time, Image, Science Metadata, and Proposal ID. There are 30 items per page, and the current page is 1 of 30. Each row in the table contains a checkbox, a thumbnail image, and links to "elog\_id" and "elog\_url".

Name	Source Folder	Size	Start Time	Image	Science Metadata	Proposal ID
ACRO mode, /w mask, thrs: 10	...8/data/S17	1 GB	2018-06-08 Fri 14:10		<a href="#">elog_id:282</a> <a href="#">elog_url:</a>	LM28IF
TOF mode, /w mask, thrs: 10	...8/data/S16	811 MB	2018-06-08 Fri 14:05		<a href="#">elog_id:281</a> <a href="#">elog_url:</a>	LM28IF
TOF mode, /w mask, thrs: 30	...8/data/S15	485 MB	2018-06-08 Fri 13:59		<a href="#">elog_id:280</a> <a href="#">elog_url:</a>	LM28IF
TOF mode, /w mask, thrs: 50	...8/data/S14	568 MB	2018-06-08 Fri 13:56		<a href="#">elog_id:273</a> <a href="#">elog_url:</a>	LM28IF
ACRO mode, no collimation, plain, thrs: 10	...8/data/S13	28 MB	2018-06-08 Fri 13:34		<a href="#">elog_id:277</a> <a href="#">elog_url:</a>	LM28IF
TOF mode, no collimation, plain, thrs: 50	...8/data/S12	27 MB	2018-06-08 Fri 11:51		<a href="#">elog_id:275</a> <a href="#">elog_url:</a>	LM28IF
TOF mode, no collimation, pixelated, thrs: 50	...8/data/S11	1 GB	2018-06-08 Fri 11:29		<a href="#">elog_id:274</a> <a href="#">elog_url:</a>	LM28IF
Pixelated scintillator scan, 2.0 Å	...8/data/S10	3 GB	2018-06-08 Fri 11:12		<a href="#">elog_id:272</a> <a href="#">elog_url:</a>	LM28IF
Beam through lead, 2.0 Å	...18/data/S9	565 MB	2018-06-08 Fri 08:55		<a href="#">elog_id:271</a> <a href="#">elog_url:</a>	LM28IF
Automated 14x14 mm2 scan (2.0 Å)	...18/data/S8	53 GB	2018-06-07 Thu 20:21		<a href="#">elog_id:270</a> <a href="#">elog_url:</a>	LM28IF

# SciCat features

- Search for your own and public data
- Publish a DOI
- Link to proposal and sample
- Coming soon ...
  - Data reduction!
  - Logbook integration!

SciCat ESS

Datasets / 20.500.12269/ed495ce2-10a8-474d-87d9-b16894c3de4a /

▽ Details    Datafiles    Attachments    Lifecycle

About the data

	Name	V20 data
	Description	V20 data
	Owner	Ioannis Apostolidis
	Keywords	[ "v20", "neutron" ]
	PID	20.500.12269/ed495ce2-10a8-474d-87d9-b16894c3de4a
	Source Folder	/nfs/groups/beamlines/v20/GH43YU

Structural information

	Type	raw
	Version	3.0.1
	Proposal	GH43YU
	Sample	-8pusdyiLK
	Orcid	orcid.org/0000-0002-7774-8995

Administrative information

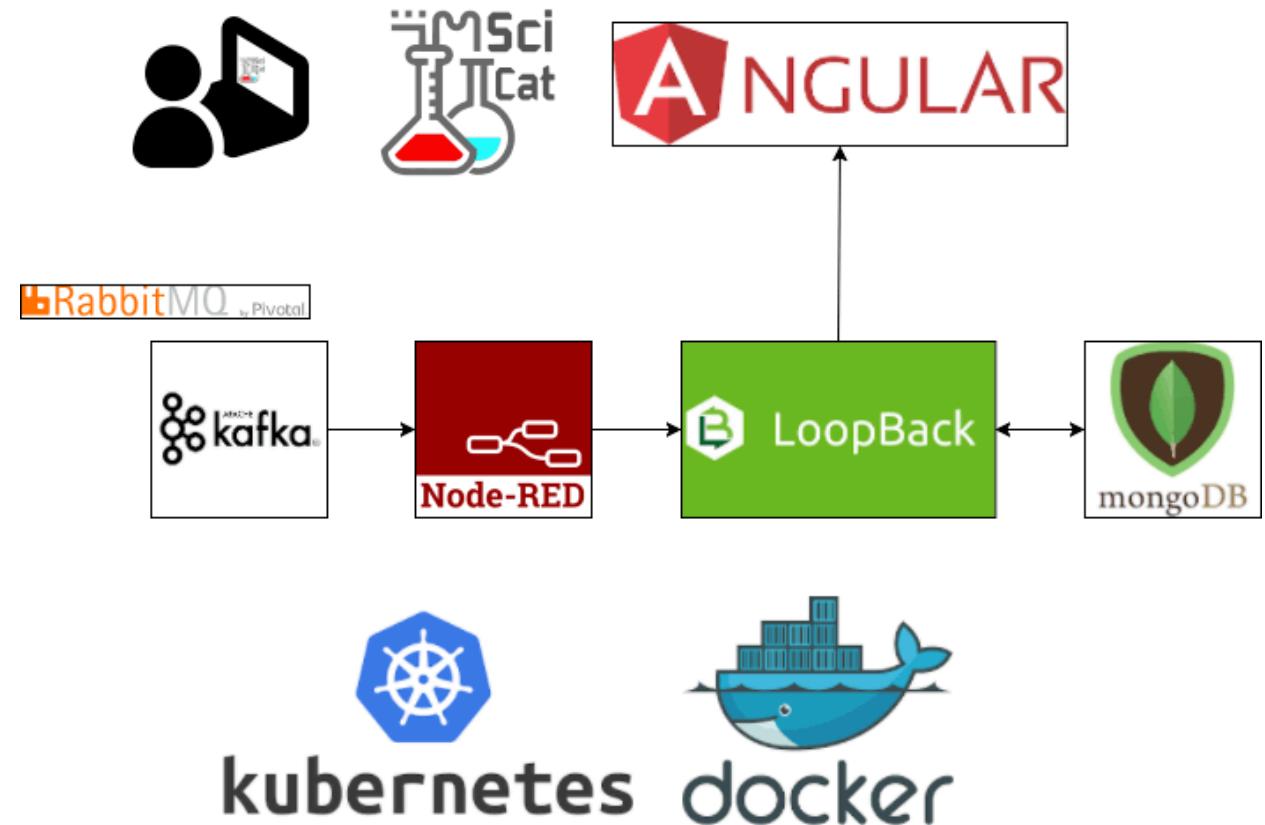
	Creation Time	2019-06-02 19:48
	Principal Investigator	Kalliopi Kanaki
	Creation Location	V20
	Owner Group	.../...

IV vs Q

CC BY-SA

# How it works

- Data is fed in from Kafka and stored in a MongoDB database by the scicat metadata server
- Users access a web application written in Google's Angular framework
- The SciCat server runs on Docker image in a Kubernetes cluster





# python™ Is A Preferred Language.



EUROPEAN  
SPALLATION  
SOURCE

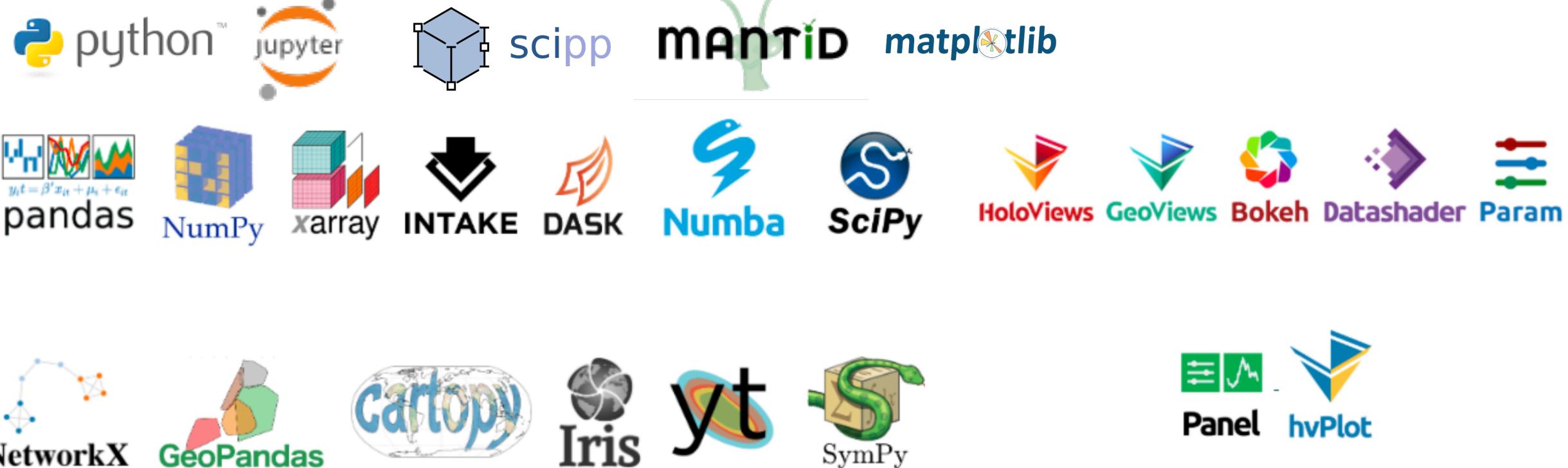
- Easy to learn.
- Fast enough for most problems.
- Object oriented.
- Scientific packages.
- Learn Python
- Write code
- Share your work on Github

```
In [8]:  
  
from mantid.simpleapi import *  
def workspace2dToDataSet(workspace):  
    signal = mtd[workspace].extractY()  
    error = mtd[workspace].extractE()  
    TOF = mtd[workspace].extractX()[0]  
    wkspDim=signal.shape  
  
    TOFAxis=(TOF[0:wkspDim[1]])  
    d = sc.Dataset(  
        {'data': sc.Variable(dims=[Dim.X, Dim.Y], values=signal),  
         },  
        coords={  
            Dim.X: sc.Variable([Dim.X], values=np.arange(signal.shape[0]), unit=sc.units.m),  
            Dim.Y: sc.Variable([Dim.Y], values=TOFAxis, unit=sc.units.us)},  
        )  
    print(d)  
    return d  
  
In [ ]:  
  
In [9]:  
d = workspace2dToDataSet('data')  
  
<scipp.Dataset>  
Dimensions: {{Dim.Y, 1900}, {Dim.X, 922}}  
Coordinates:  
    Dim.X           int64      [m]          (Dim.X)  
    Dim.Y           double     [μs]          (Dim.Y)  
Labels:  
Data:  
    data           double   [dimensionless]  (Dim.X, Dim.Y)  
Attributes:
```

# Managing Environments

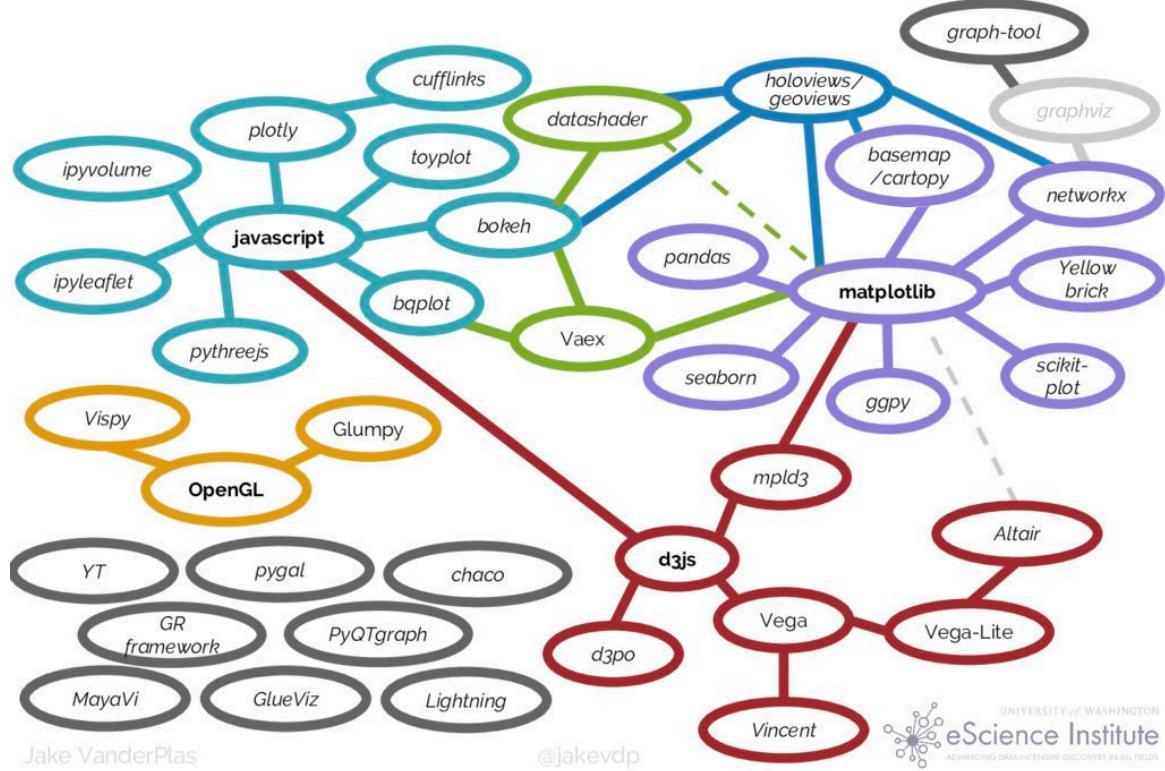
# CONDA

- Conda manages environments and installation
- Conda install ...

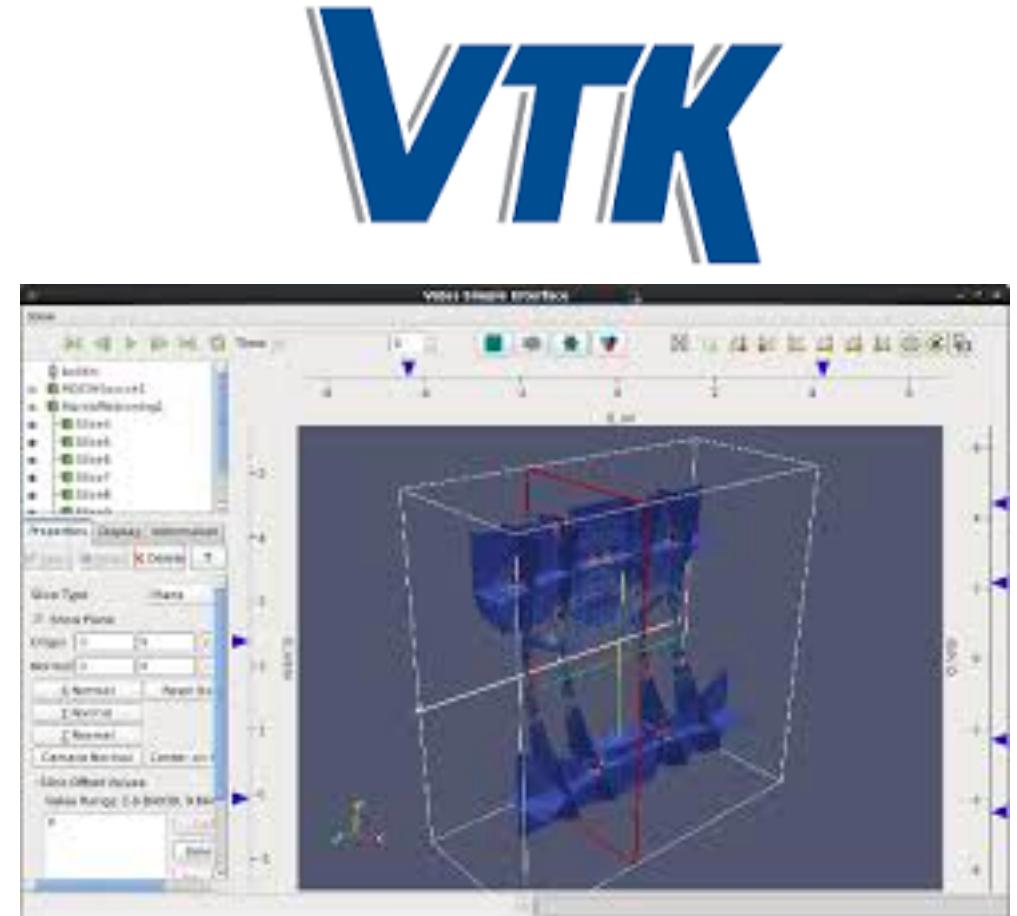


# Visualising Data in Python

## 1D and 2D



## 3D



# The jupyter EcoSystem



A screenshot of the 'nature' magazine website. At the top, there is a red header bar with the word 'nature' and 'International journal of science'. Below the header are buttons for 'Subscribe', 'Search', and 'Login'. The main content area has a white background with the title 'Why Jupyter is data scientists' computational notebook of choice' and a subtitle 'An improved architecture and enthusiastic user base are driving uptake of the open-source web tool.'.

TOOLBOX · 30 OCTOBER 2018

## Why Jupyter is data scientists' computational notebook of choice

An improved architecture and enthusiastic user base are driving uptake of the open-source web tool.

- Notebooks capture process, data, and visualisation.
- Great for documenting complex data processing.
- Can be Central or Local
- Easy to use ...



### Language of choice

Jupyter supports over 40 programming languages, including Python, R, Julia, and Scala.



Project Jupyter exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages.



### Share notebooks

Notebooks can be shared with others using email, Dropbox, GitHub and the [Jupyter Notebook Viewer](#).



### Interactive output

Your code can produce rich, interactive output: HTML, images, videos, LaTeX, and custom MIME types.

# Jupyter examples for Neutron data.

- Some example jupyter & Python 3 notebooks at
  - <https://github.com/Jon-Taylor/raciri-2019.git>
- Data Processing with mantid, data visualisation, ipywidgets for UI Holoviews and data shader.
- Atomic simulation environment for calculation of PDoS

In [26]:

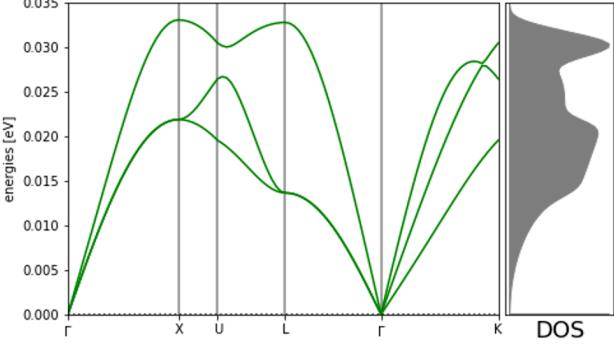
```
# Plot the band structure and DOS:
%matplotlib inline
import matplotlib.pyplot as plt

fig = plt.figure(1, figsize=(7, 4))
ax = fig.add_axes([.12, .07, .67, .85])

emax = 0.035
bs.plot(ax=ax, show=False, emin=0.0, emax=emax)
aa=bs.todict()
aa.keys
dosax = fig.add_axes([.8, .07, .17, .85])
dosax.fill_between(dos.weights[0], dos.energy, y2=0, color='grey',edgecolor='k', lw=1)

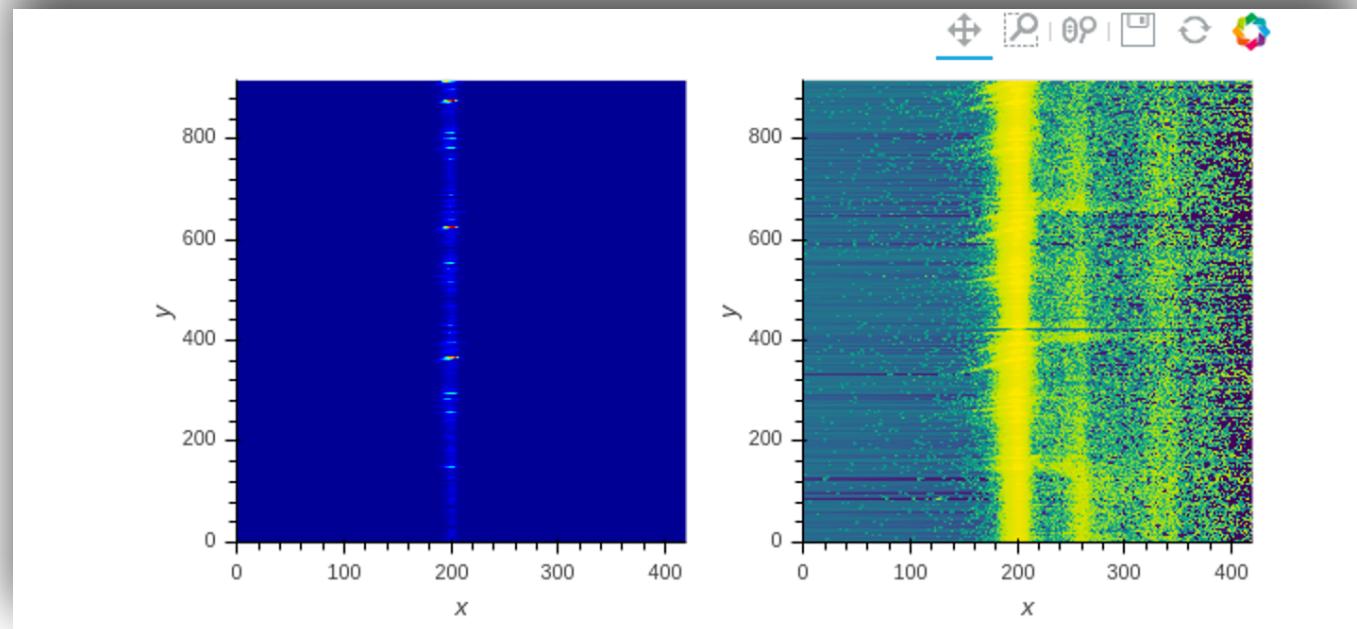
dosax.set_ylim(0, emax)
dosax.set_yticks([])
dosax.set_xticks([])
dosax.set_xlabel("DOS", fontsize=18)
```

Out[26]: Text(0.5, 0, 'DOS')



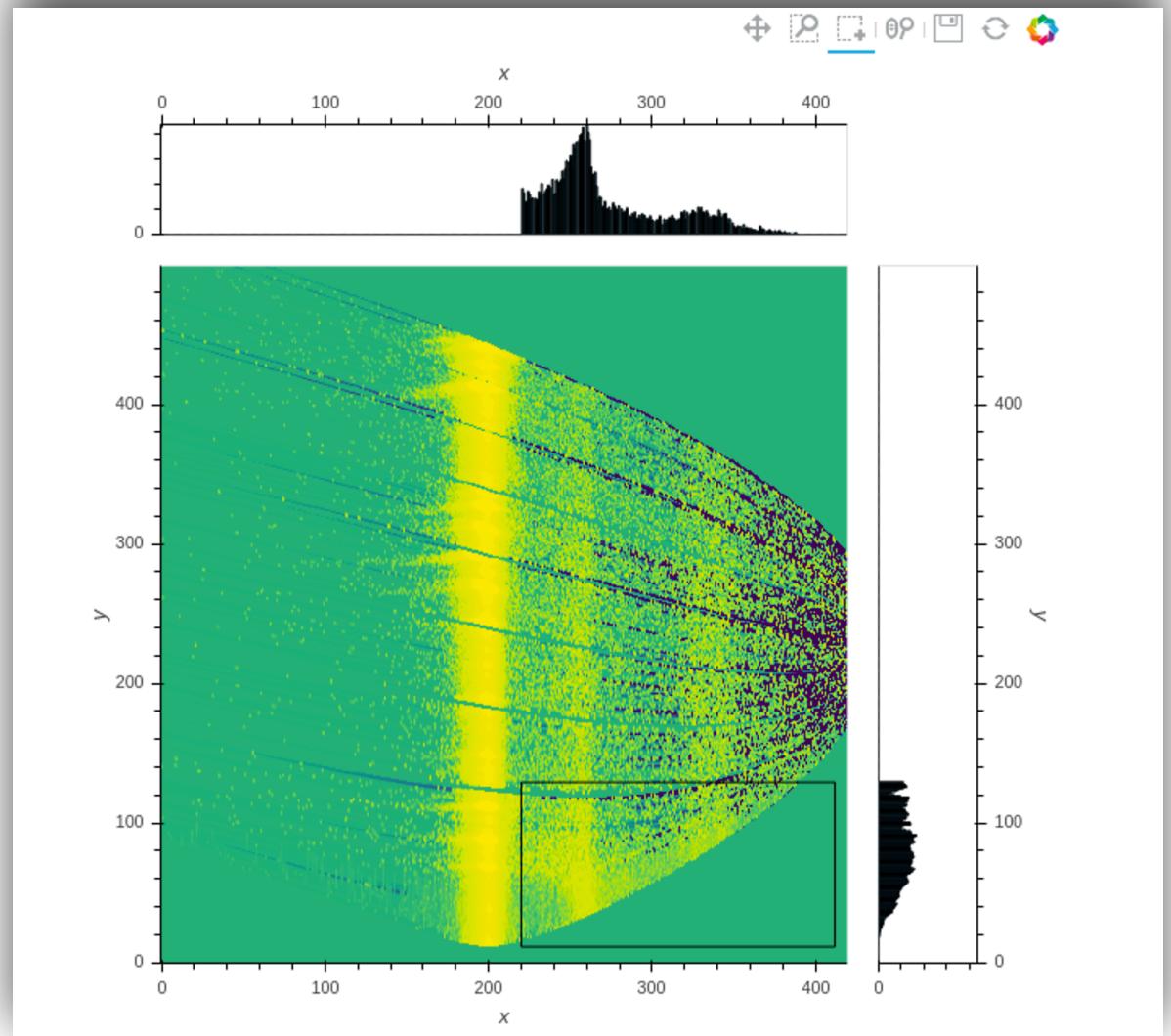
# Jupyter examples for Neutron data.

- Some example jupyter & Python 3 notebooks at
  - <https://github.com/Jon-Taylor/raciri-2019.git>
- Data Processing with mantid, data visualisation, ipywidgets for UI Holoviews and data shader.
- Atomic simulation environment for calculation of PDoS



# Jupyter examples for Neutron data.

- Some example jupyter & Python 3 notebooks at
  - <https://github.com/Jon-Taylor/raciri-2019.git>
- Data Processing with mantid, data visualisation, ipywidgets for UI Holoviews and data shader.
- Atomic simulation environment for calculation of PDoS



# Jupyter & SciCat

- SciCat can be accessed directly from your jupyter notebook!
- You can search for metadata about different experiments at ESS

The screenshot shows a Jupyter Notebook interface with the title "Search SciCat". The notebook contains the following code:

```
In [91]: import requests
import json
import urllib

def search_scicat(text, max_number_results):
    fields = {'text': text}
    limit = {'limit': max_number_results, 'order': "creationTime:desc"}
    fields_encode = urllib.parse.quote(json.dumps(fields))
    limit_encode = urllib.parse.quote(json.dumps(limit))
    dataset_url = \
        "https://scicatapi.esrf.fr/api/v3/Datasets/anonymousquery?fields=" + \
        fields_encode+"&limits="+limit_encode
    r=requests.get(dataset_url).json()
    print(len(r), "result found!")
    return r
```

Search for e.g. hexagonal boron nitride (hbn)

```
In [92]: r=search_scicat("hbn",1)
```

1 result found!

Show the sample description

```
In [93]: print(r[0]["scientificMetadata"].get("sample_description"))
```

hBN target with 1.0 mm diameter hole

Print the location of the data

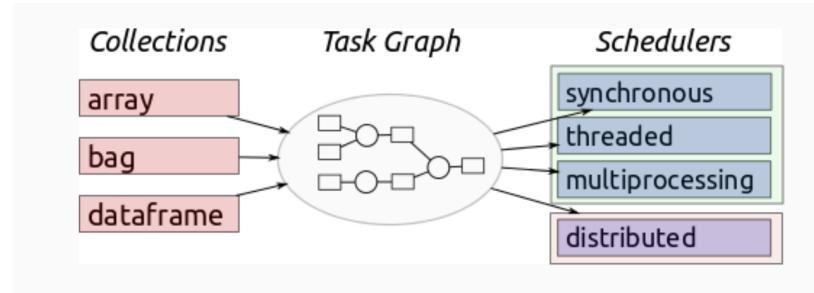
```
In [90]: print(r[0]["sourceFolder"])
```

/nfs/groups/beamlines/v20/67JH32

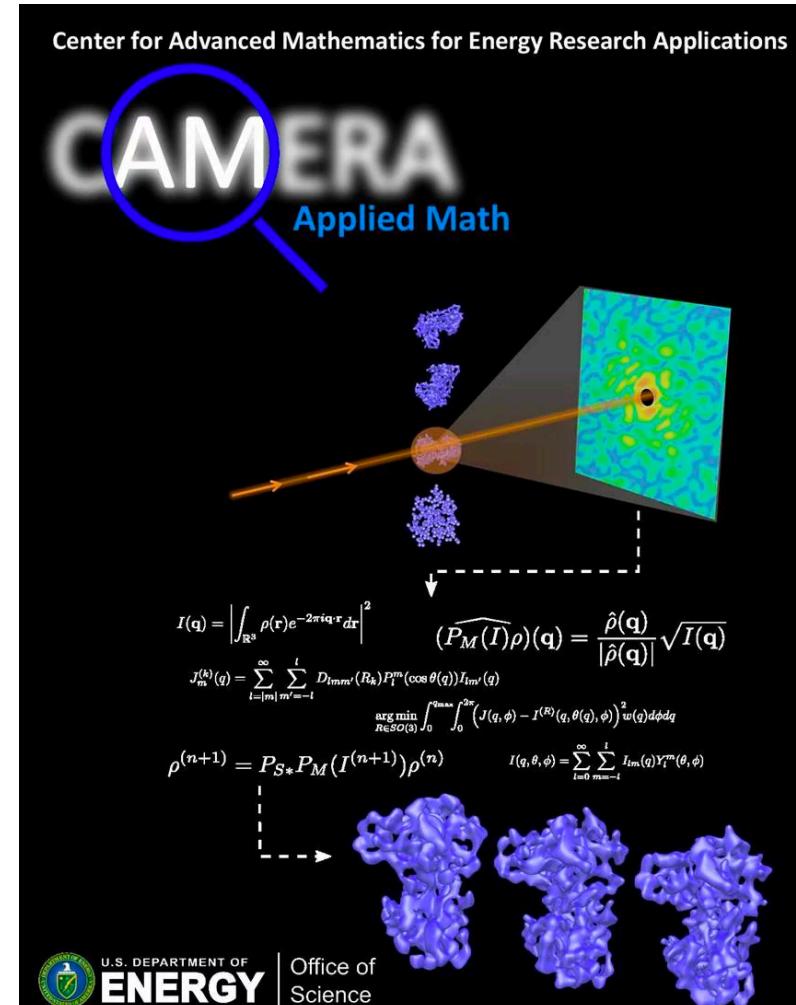
# Scaling Python and Machine Learning



- Some Problems are too large to fit on a single machine.

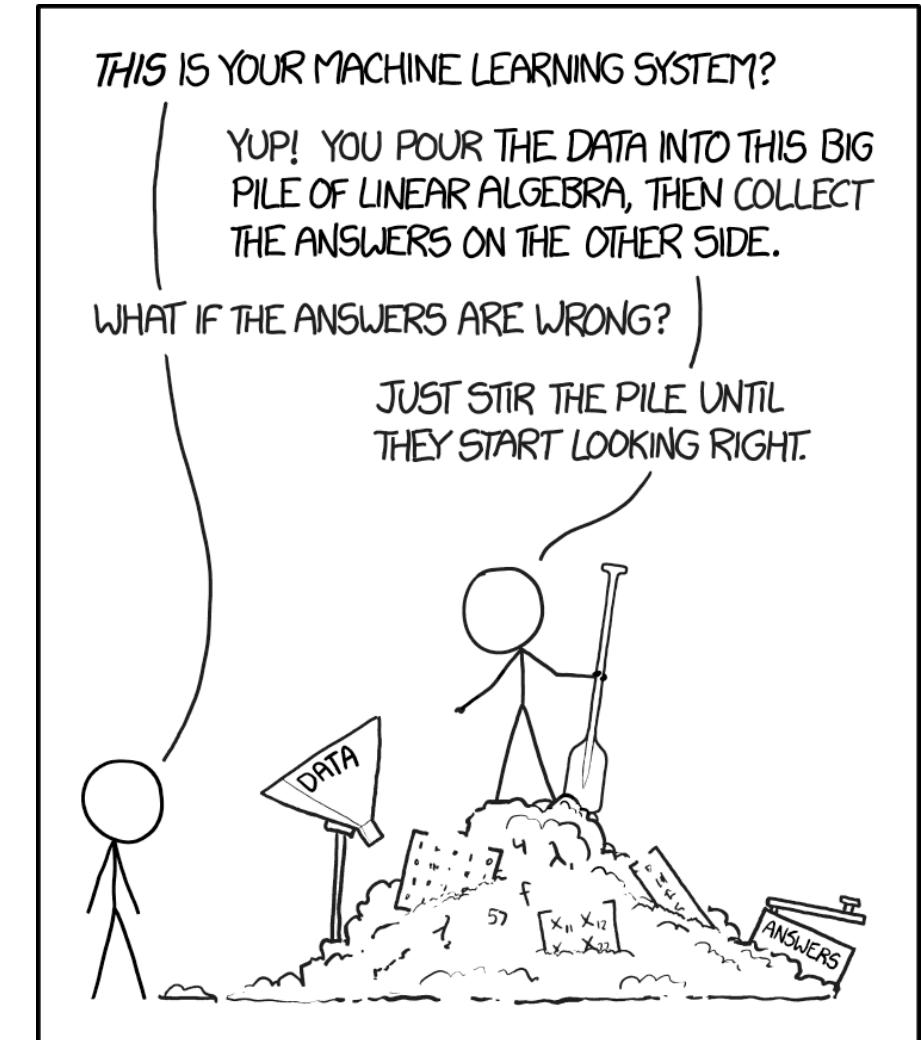


- Some Problems can be automated with Machine learning.
- <https://www.camera.lbl.gov/>



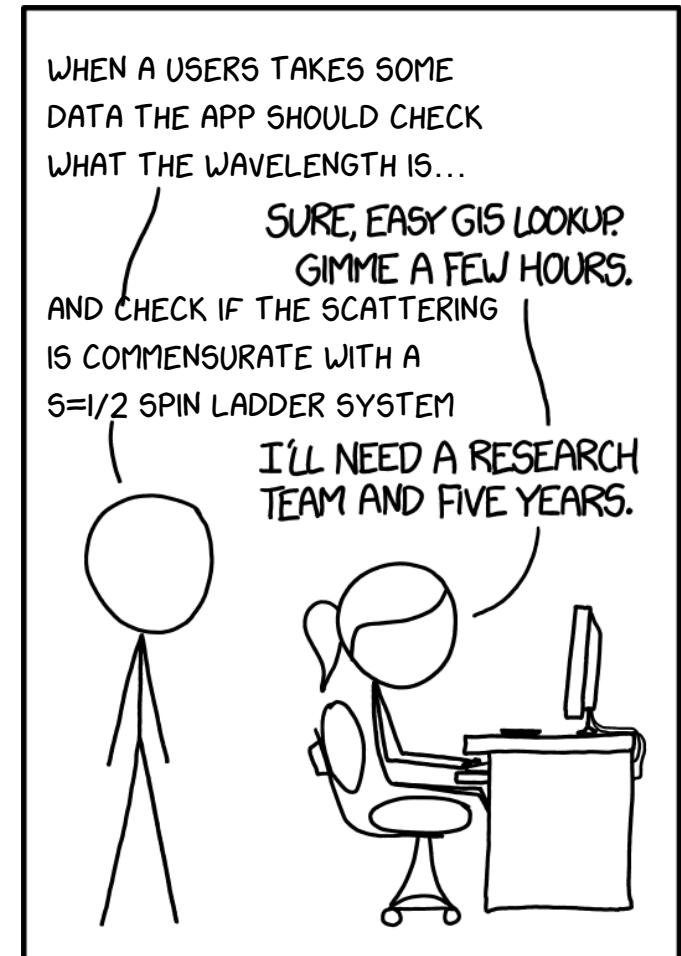
# How could Deep Learning and AI be used for scattering experiments.

- A trained neural network can define a likelihood that a requested parameter is true or exists, in an automated workflow.
- There could be real benefits for automation and high throughput processing of data.
  - Is the sample well aligned
  - Is the instrument configured for best resolution at ...
  - What is the space group point symmetry
  - Is the sample poly disperse
  - What is the value of the exchange interaction
  - and so forth



# Machine / Deep learning and AI for Neutron Science

- Deep learning is a fast moving research field.
- There are both technical and ethical challenges for science.
  - Zero Bias is not guaranteed by default



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

# Any questions?



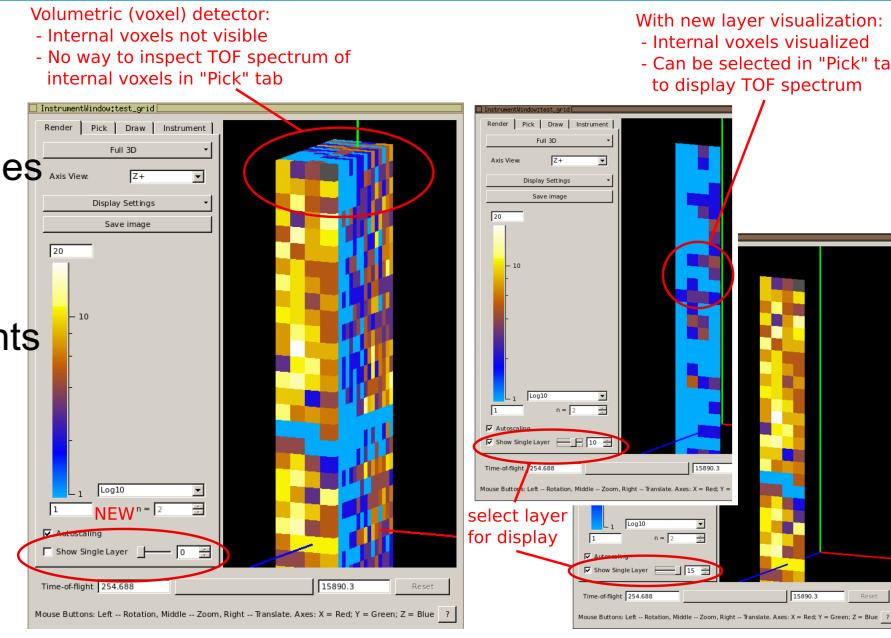
- AI & Deep learning workshops for neutron & photon scattering
  - Copenhagen September 24 <https://indico.nbi.ku.dk/event/1256>
  - Grenoble November 12 <https://workshops.ill.fr/e/ai2019>

# Data Reduction, Data Analysis and Modelling.

## Mantid @ ESS

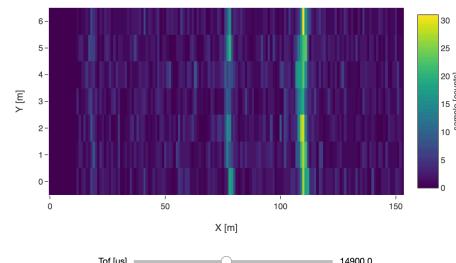
- volumetric slicing (see figure)
- support for novel detector geometries (Band-Gem).
  - support for NeXus Geometry
- significant speed-up for instruments with many pixels
- speed-ups from Instrument-2.0

### Facilities involved



<http://www.mantidproject.org>

- X Array for Neutron Scattering data
- Jupiter & Python (PaNOSC)
- Multi Dimensional Data structure.
- Units
- Errors
- Geometry

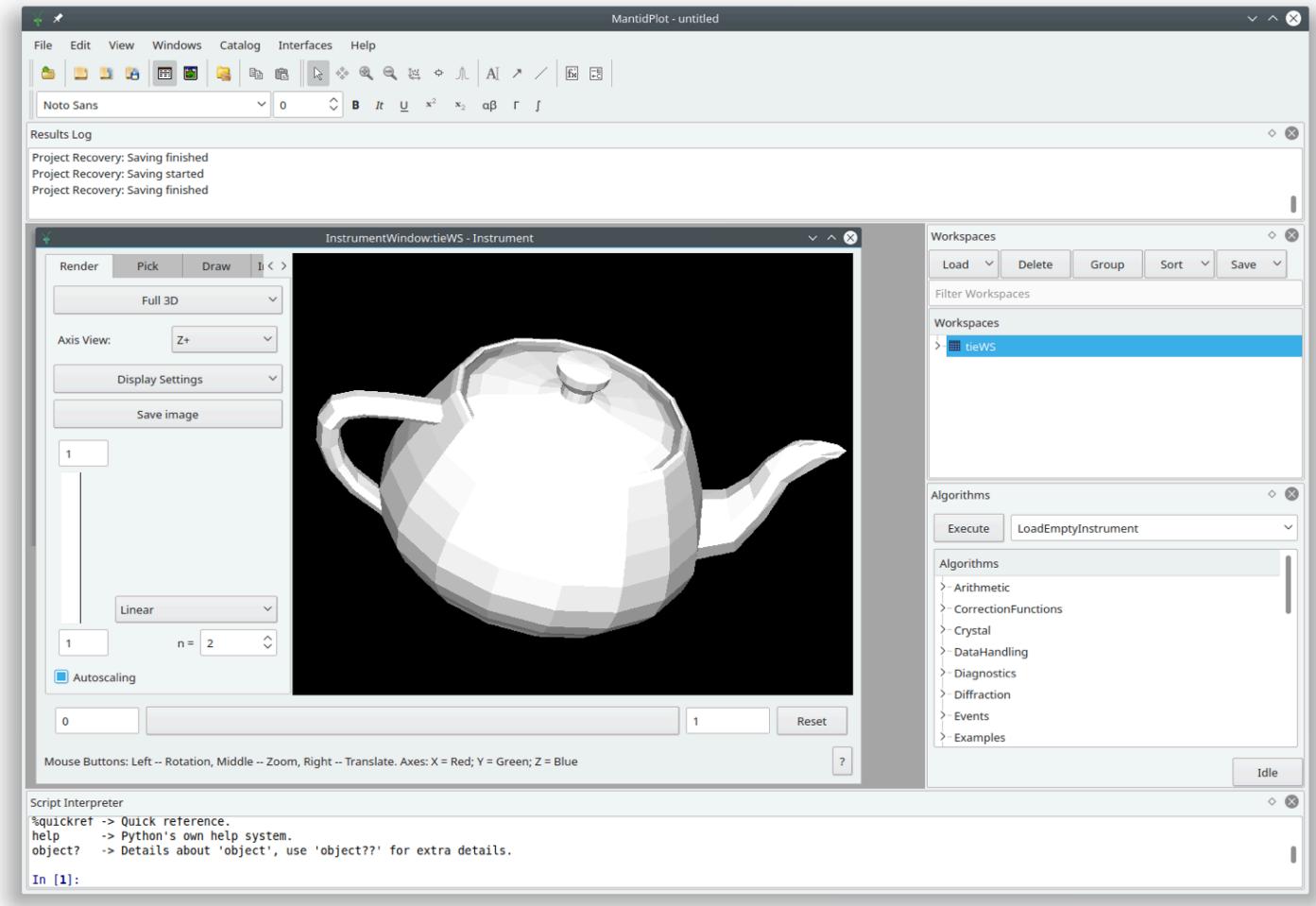


# Flexible Instrument Geometry



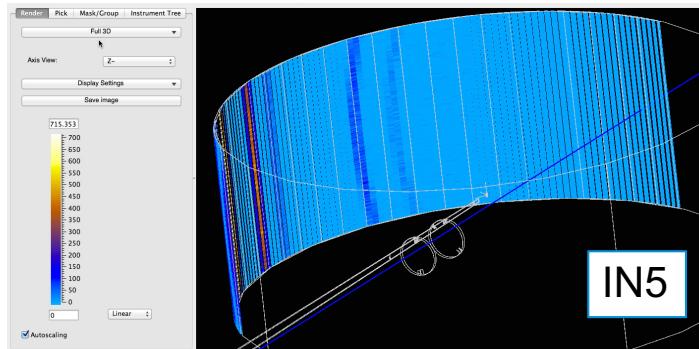
## NeXus

NeXus is developed as an international standard by scientists and programmers representing major scientific facilities in Europe, Asia, Australia, and North America in order to facilitate greater cooperation in the analysis and visualization of neutron, x-ray, and muon data.

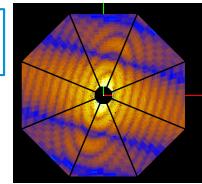


# Data Reduction, Data Analysis and Modelling.

## McStas

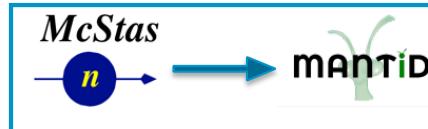
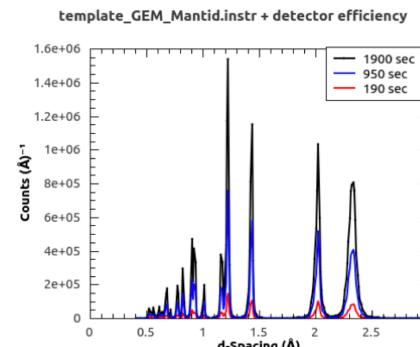


## LoKI



EUROPEAN  
SPALLATION  
SOURCE

NEUTRONS  
FOR SOCIETY

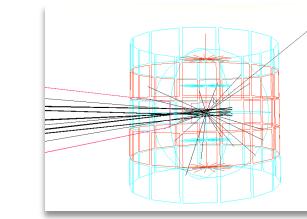


← SasView 2D scattering kernel

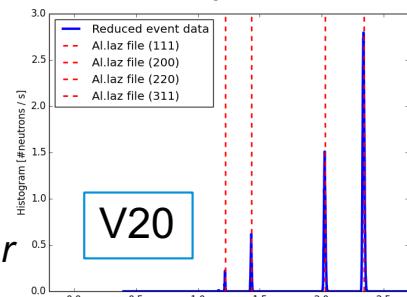
← MANTID instrument view

- MANTID reduced McStas event data

- virtual experiment of ESS instruments moderator → detector
- models of samples
- instrument model repository
- preparation for reduction & analysis



Single crystal diffractometer  
with cylindrical detector  
geometry



<http://www.mcstas.org/>

# Data Reduction, Data Analysis and Modelling.

## SpinW

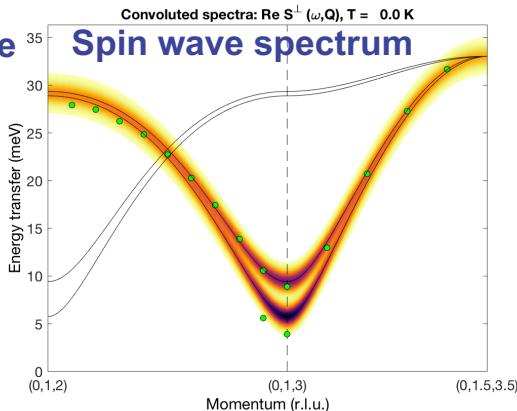
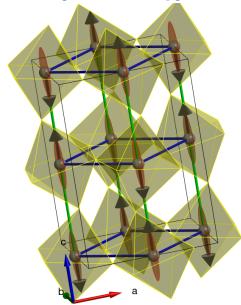
modularizing with C++ core and Python interface.



PAUL SCHERRER INSTITUT  
**PSI**



### Magnetic and crystalline structures ( $\text{LuVO}_3$ )



<https://www.psi.ch/spinw>

## Fitting QENS data

fitting potential to QENS data → dynamical version of EPSR method



Science & Technology Facilities Council  
**ISIS**



EUROPEAN SPALLATION SOURCE



CHALMERS



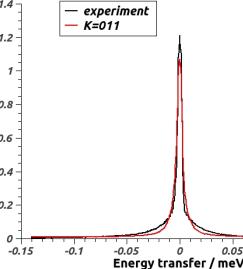
UNIVERSITY OF COPENHAGEN



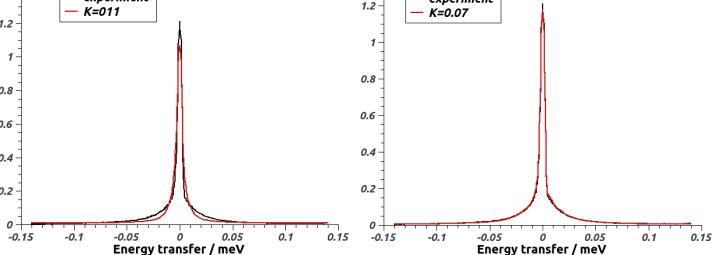
Vetenskapsrådet

### Original force-field

$T=200\text{K}, Q=1.9\text{\AA}^{-1}$



### Optimized force-field



methyl rotations in methyl-Polyhedral oligomeric silsesquioxanes

*M. Hagen, J. Borreguero (SNS/NDAV), M. Crawford (Dupont), N. Jalarvo (Jülich)*

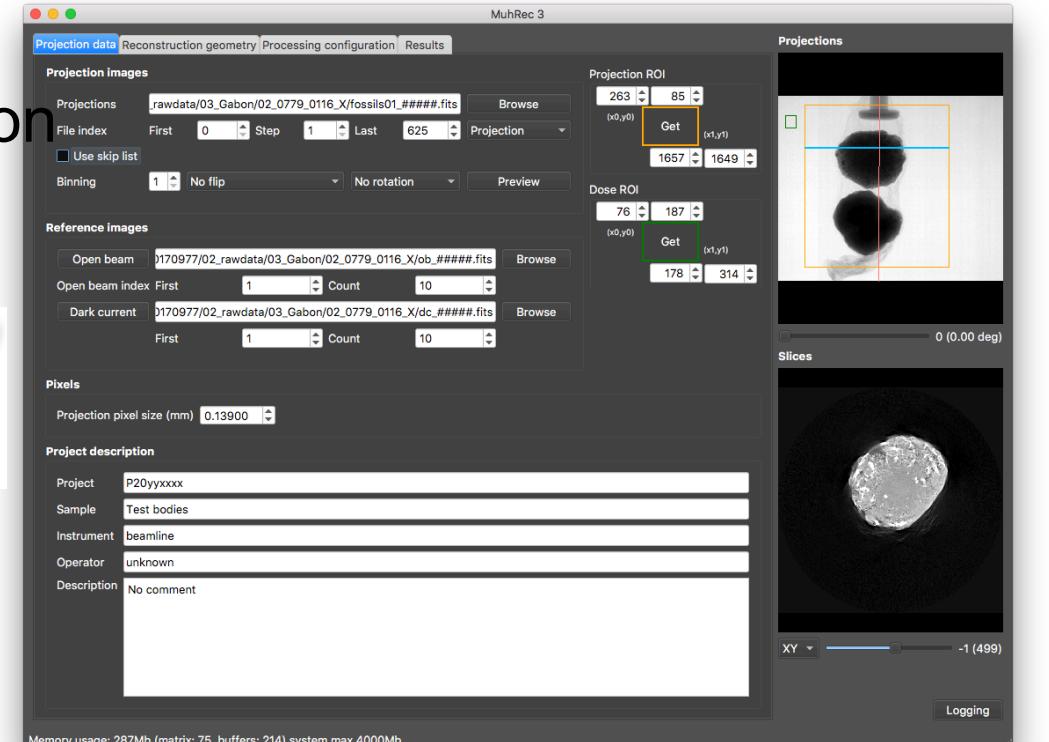
# Data Reduction, Data Analysis and Modelling.

## MuhRec imaging

- support for cone beam reconstruction
  - algorithm to correct biases due to background scattering
  - more pre-processing options
  - revised user interface
  - stability improvements
  - support for NeXus file formats



SINE  
2020

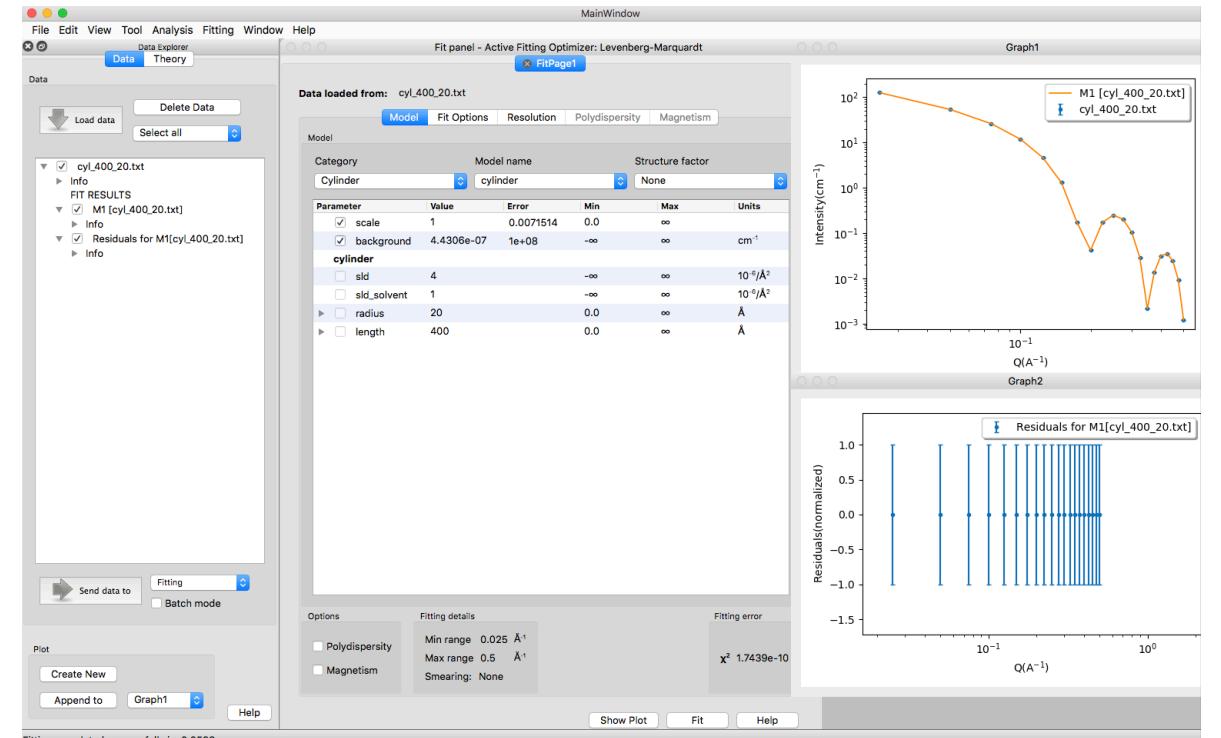


# Data Reduction, Data Analysis and Modelling.



## SasView SANS

- code modularization
- GUI re-design (Qt5)
- code optimization for faster analysis methods using GPUs
- incorporation of model functions from SasFit\*
- Sasmodels marketplace



\* <https://kur.web.psi.ch/sans1/SANSSoft/sasfit.html>

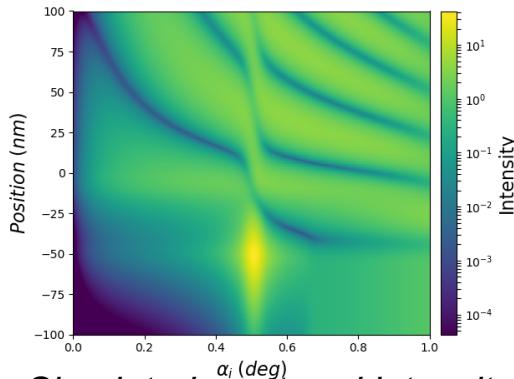
<http://www.sasview.org>

# Data Reduction, Data Analysis and Modelling.

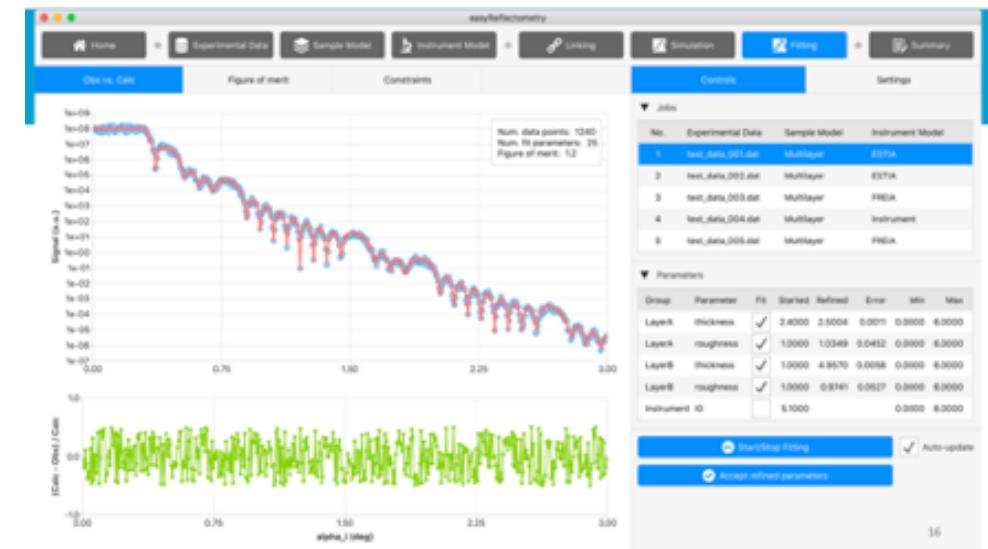


## BornAgain reflectometry

- basic functionality for reflectometry simulations
- simulation functionality → GUI
- usability check on experimental data
- depth probe simulation for experimental setup (see figure)



Simulated scattered intensity  
from Ti/Pt resonator with beam  
divergence



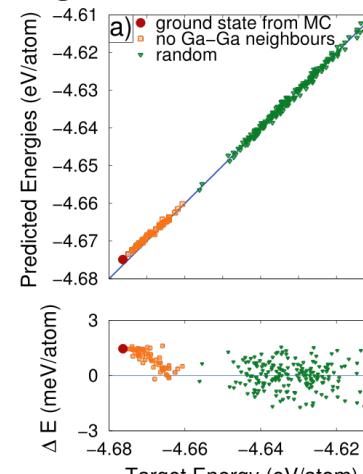
# Data Reduction, Data Analysis and Modelling.

## Icet diffraction

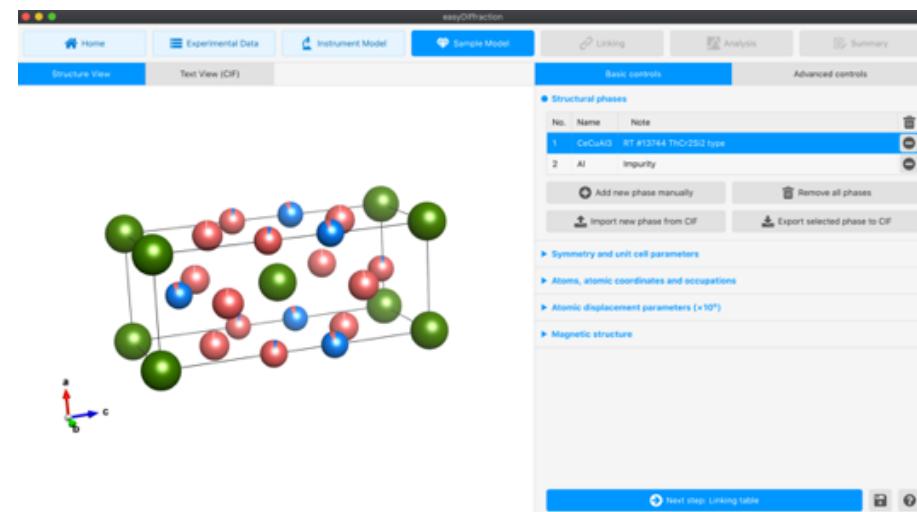
- simulating ordering and segregation in multi-component systems
- DFT calculation
- construction and sampling of cluster expansions



<https://www.materialsmodeling.org/>



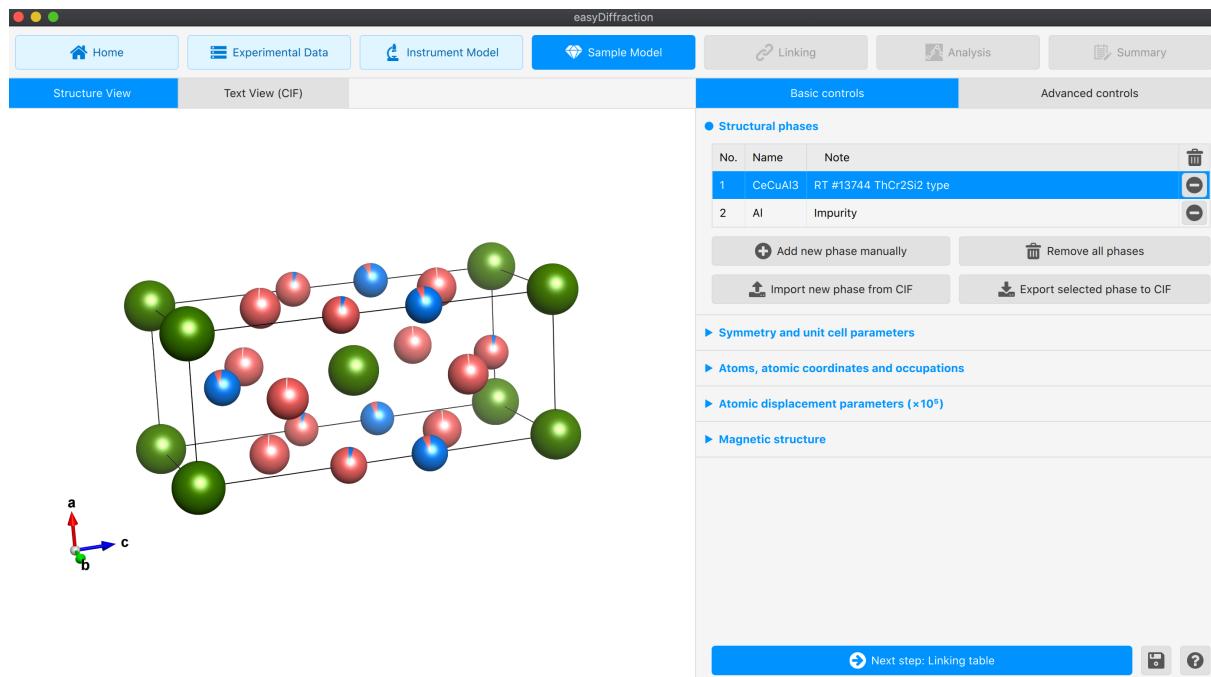
Ångqvist & Erhart, *Chem. Mat.*



- Collaborative development of the FullProf Suite with ILL

# User experience.

- Sustainable software development for Neutron Scattering data analysis.
- Enable the community with easy to use feature rich software



The screenshot shows the easyDiffraction software interface. On the left, there is a 3D visualization of a crystal structure composed of red and green spheres connected by lines. A coordinate system with axes labeled 'a', 'b', and 'c' is shown at the bottom left. On the right, there is a table titled "Structural phases" with two entries:

No.	Name	Note
1	CeCuAl3	RT #13744 ThCr <sub>2</sub> Si <sub>2</sub> type
2	Al	Impurity

Below the table are several sections: "Basic controls", "Advanced controls", "Symmetry and unit cell parameters", "Atoms, atomic coordinates and occupations", "Atomic displacement parameters ( $\times 10^3$ )", and "Magnetic structure". At the bottom, there is a button labeled "Next step: Linking table".

