Standards for Data Analysis Software (WP6)

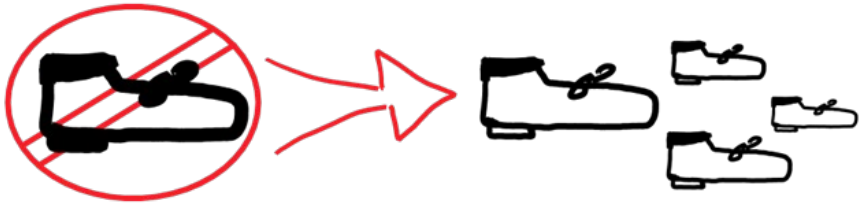
Emmanuel Farhi

Computing for Science, ILL

Partners: ILL, ISIS, PSI, FRM2, JCNS, GKSS, HZB, DTU

<nmi3.eu/about-nmi3/other-collaborations/data-analysis-standards.html>

Scope



Understand how to fight against segmentation of software

Stabilise software (avoid single point of failures) and maintenance

Propose ways to gather efforts: standards

Favour interoperability and re-usability

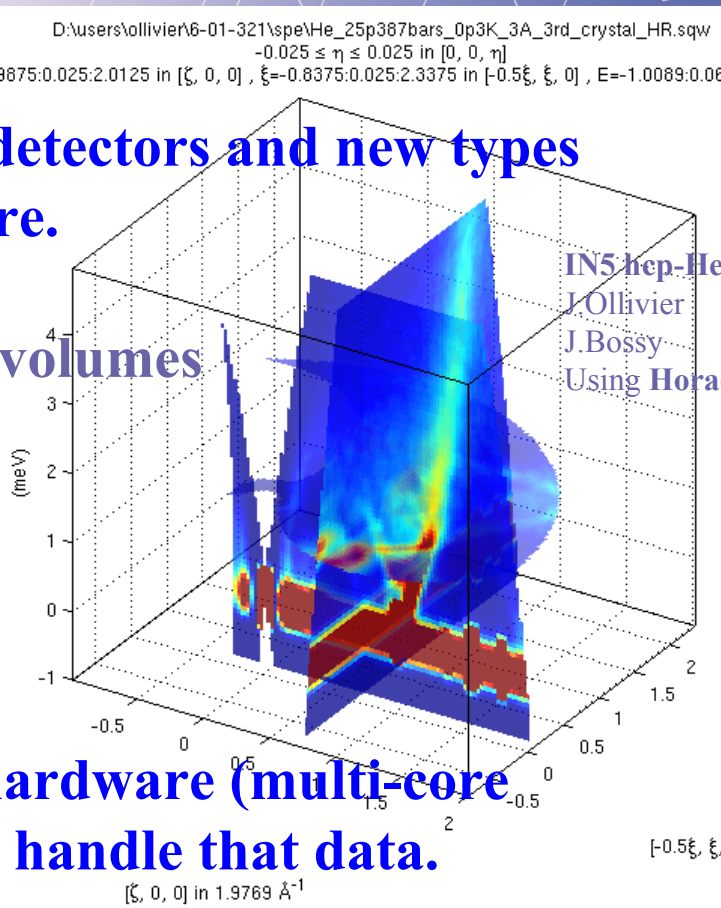
Be flexible and learn from the software history

Bigger machines ?

Shorter experiments, rapid data rates, large detectors and new types of instruments make new demands on software.

BUT: this is just as in the old days, with **larger volumes**

→ require fast processing.



New software (languages and methods) and hardware (multi-core processors, GPU's, etc) offer new solutions to handle that data.

BUT: never forget that the **old codes** are the most valuable and hold the scientific knowledge (stable, fast...)

→ re-use ideas and implementations.

Our tasks and resources

Gather our knowledge and strength for Data Analysis.
Think and evaluate, rather than code.

Our tasks:

Task .1: Review existing data analysis software and practices of software developers
Task .2: Review existing solutions for a common data analysis infrastructure
Task .3: Develop prototype software in chosen solution for representative applications
Task .4: Evaluate prototype software

Our resources:

- 4 months of each of the other participants (that is about 2-3 days per month for each of us).
- 30 months position funding → Sept 2014.
- Our smiles.

Task 1: Review software

Review existing data analysis software and practices of software developers

Inquire about software usage to estimate if old codes must be maintained, and learn from the past.

Test/analyse software:

- We tested 24 packages (6 months)
- Use e.g. the NMI3 LiveDVD to make-up your mind.

Objective:

Build a table of 'recommended' software.
Identify good, bad and ugly practices.

Action: Wrote the 1st report (see NMI3.eu)

Task 2: Review solutions

Review existing solutions for a common data analysis infrastructure

Common infrastructure

How to work together

- Repositories, documents, communication

Define standards

- Data sets, data formats (NeXus)
- Functionalities, (generic methods, specialized)
- Interfaces (layouts, naming convention)

Action: Report to be produced by end 2013

Task 3: prototyping

Develop prototype software in chosen solution for representative applications

Mantid IS a common infrastructure (ISIS, SNS), but focued on spallation sources

Our contribution actually considers reactor based instruments

Our aim:

Add new type of instruments into Mantid, as a start, such as
Multiplexed TAS
Powder and SX diffractometers

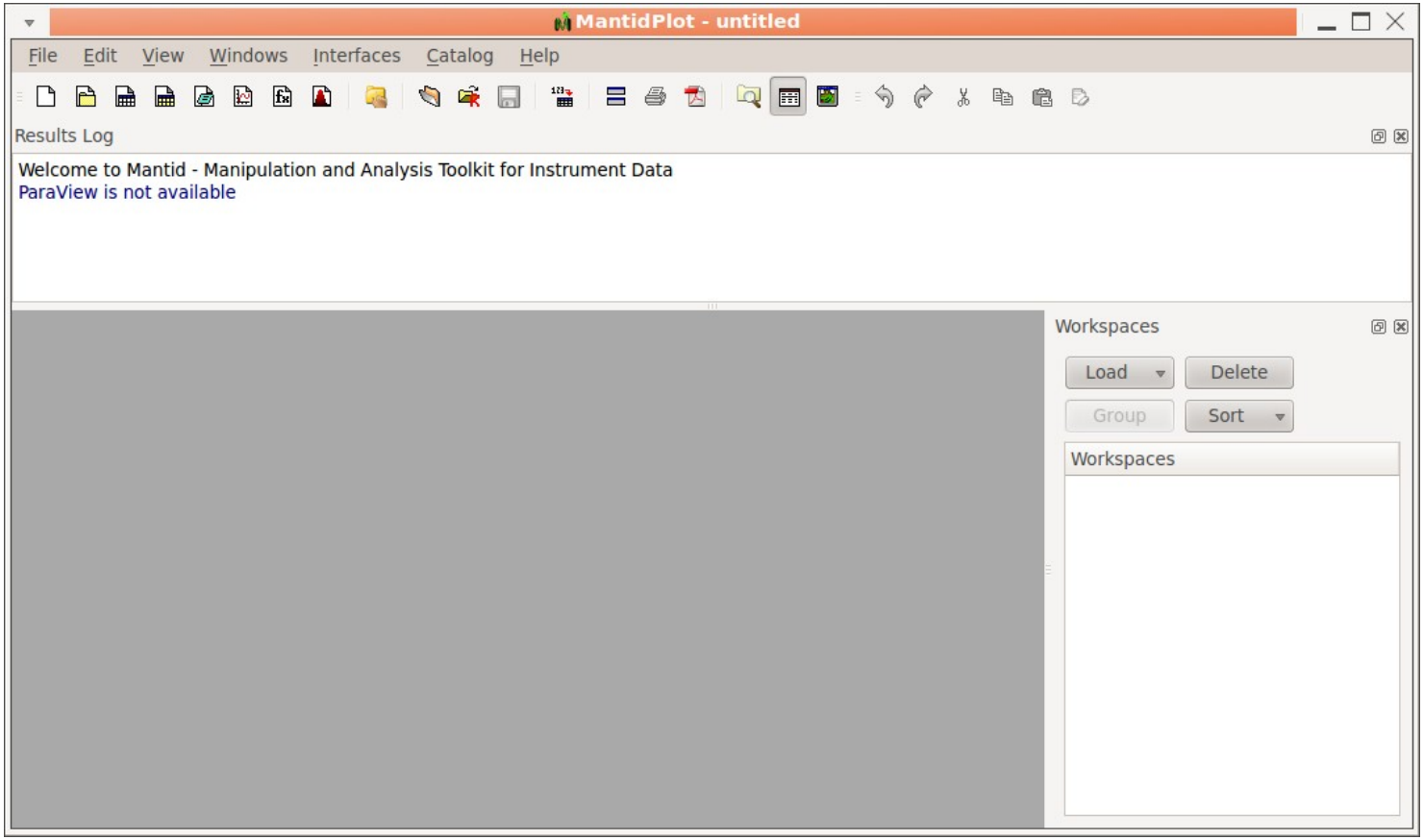
Techno-Geek Collection 4 / 26
ELITE CODER KEYBOARD
[http://www.virtual-trading-cards.com]

Current route:

Started with TOF and TAS at continuous sources
Use Mantid as framework, compare with other software (LAMP)

Task 3: Mantid ?

<mantidproject.org> install and try it, make up your mind

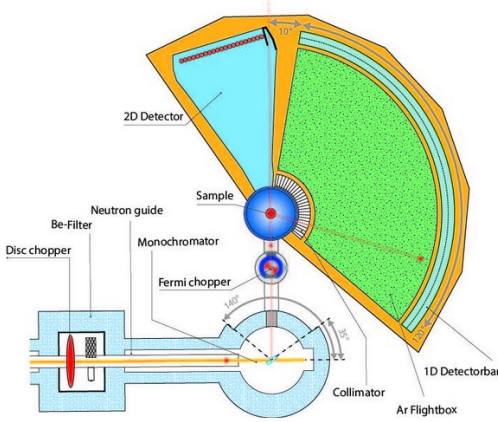


Task 3: prototyping: TOF

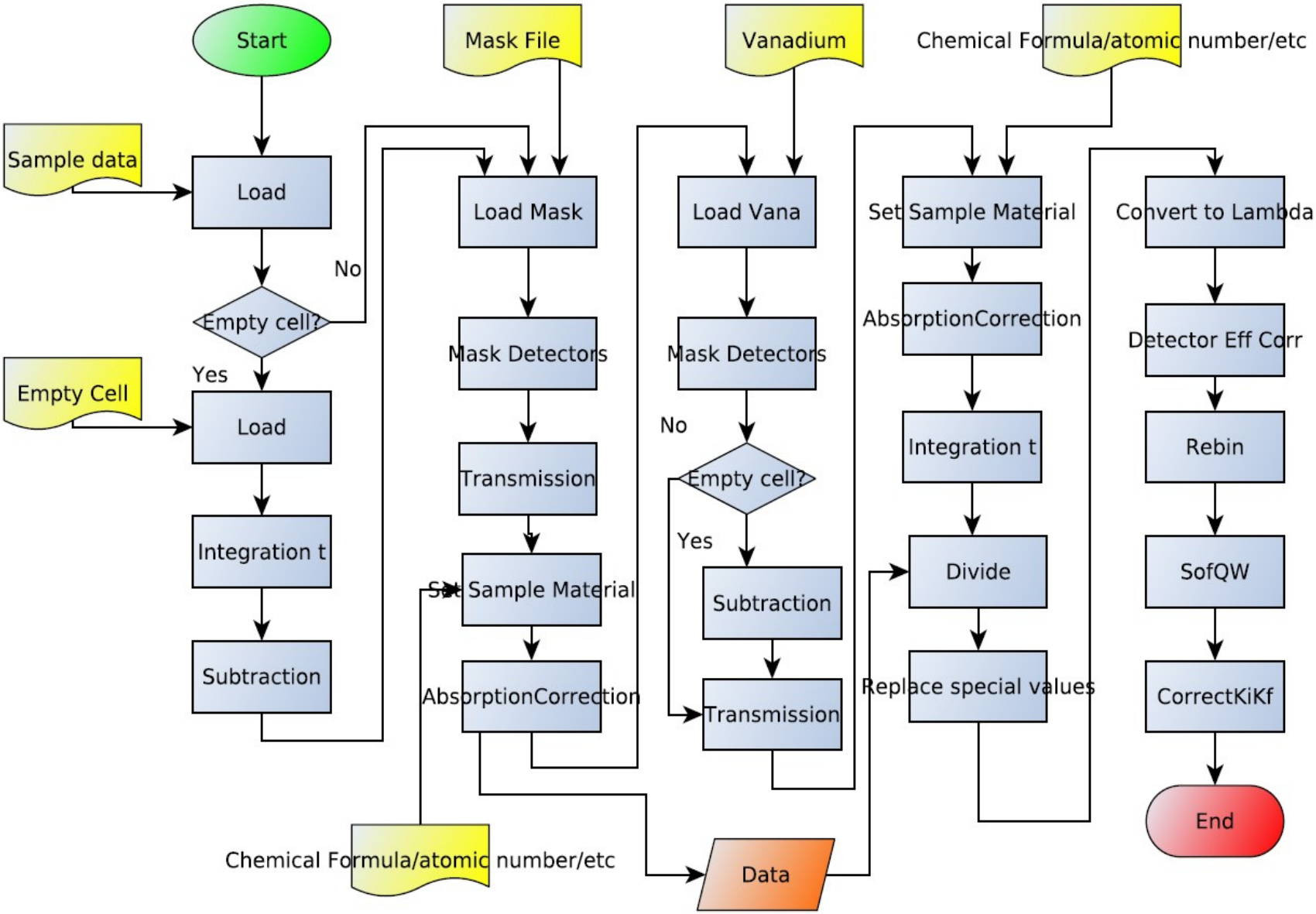
Action: 6 months learning how Mantid works

Then implemented loaders for TOF@continuous source

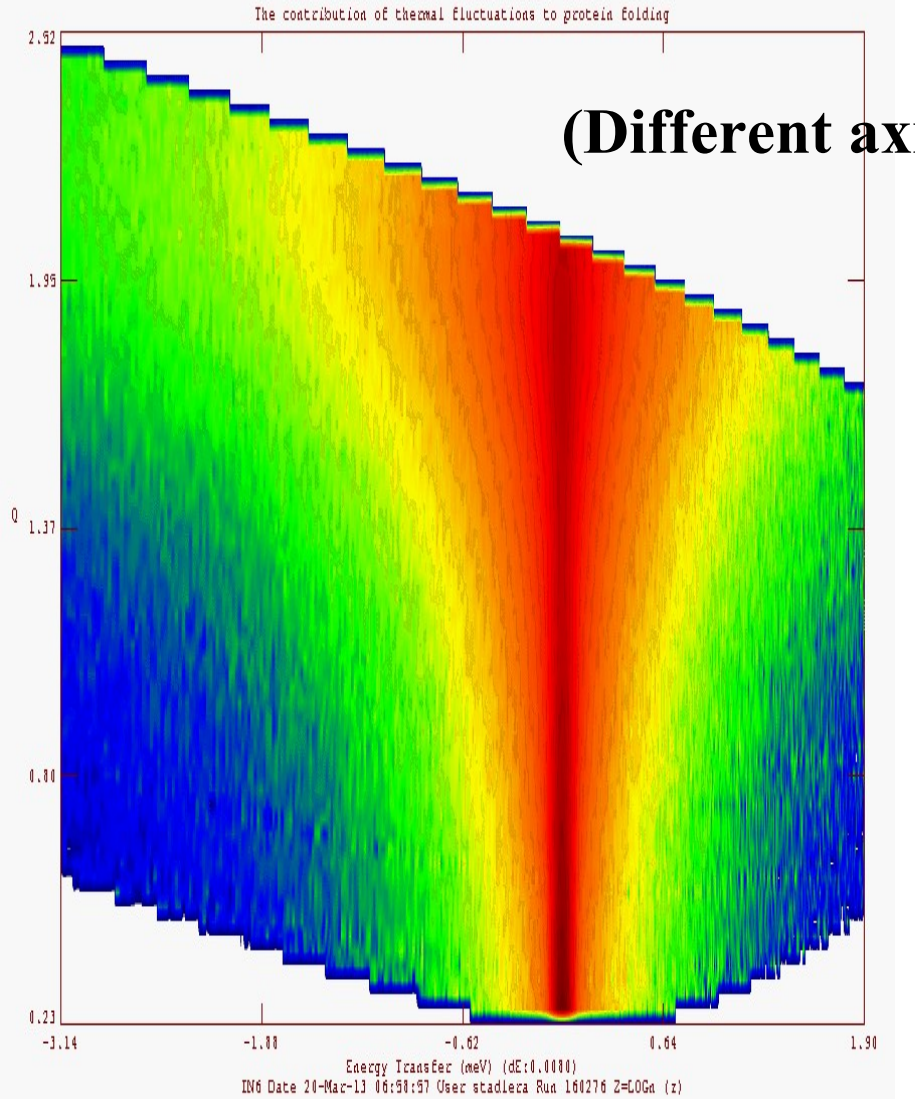
- IN4, 5 and 6 @ILL
- Focus @PSI
- MiBemol @LLB
- TofTof @FRM-II (on going)



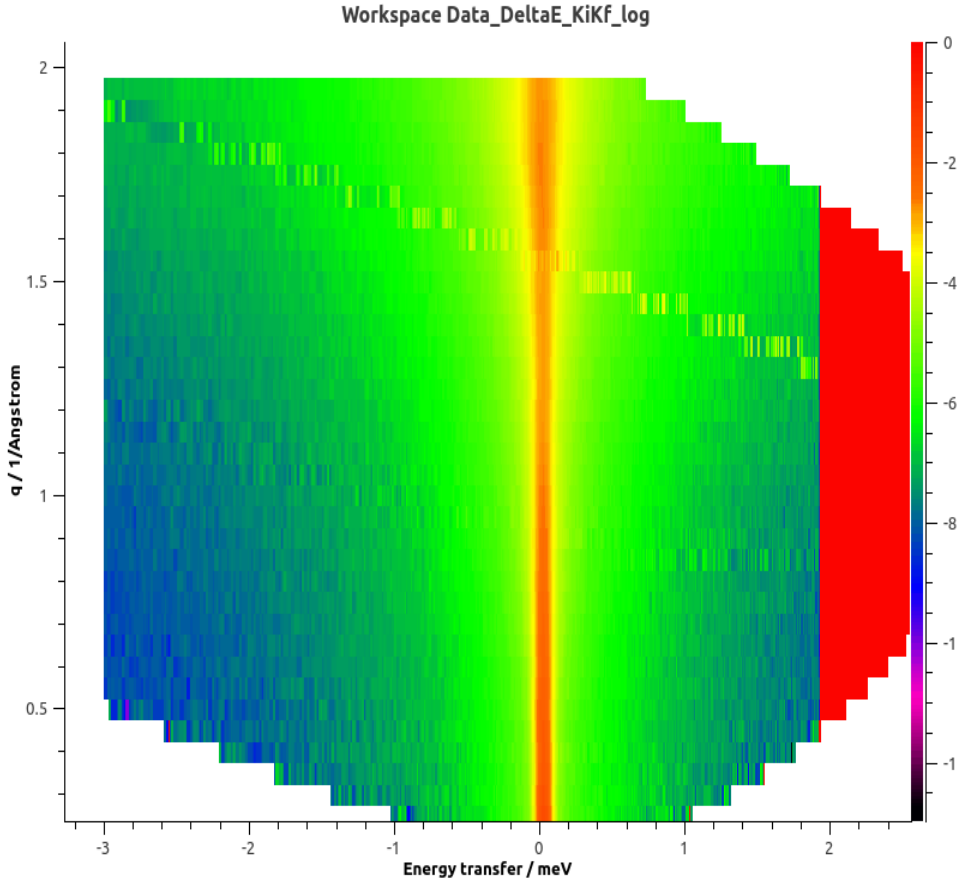
Identify algorithms which fulfil basic operations
Correction, coordinate transformation, reduction


Start
Mask File
Vanadium
Chemical Formula/atomic number/etc
Sample data
Load
Load Mask
Load Vana
Set Sample Material
Convert to Lambda
No
Empty cell?
Mask Detectors
Mask Detectors
AbsorptionCorrection
Detector Eff Corr
Empty Cell
Yes
Load
Transmission
No
Empty cell?
Integration t
Rebin
Integration t
Yes
Subtraction
Divide
SofQW
Set Sample Material
Subtraction
Replace special values
CorrectKiKf
AbsorptionCorrection
Transmission
End
Chemical Formula/atomic number/etc
Data

Task 3: Mantid vs Lamp: S(q,w)



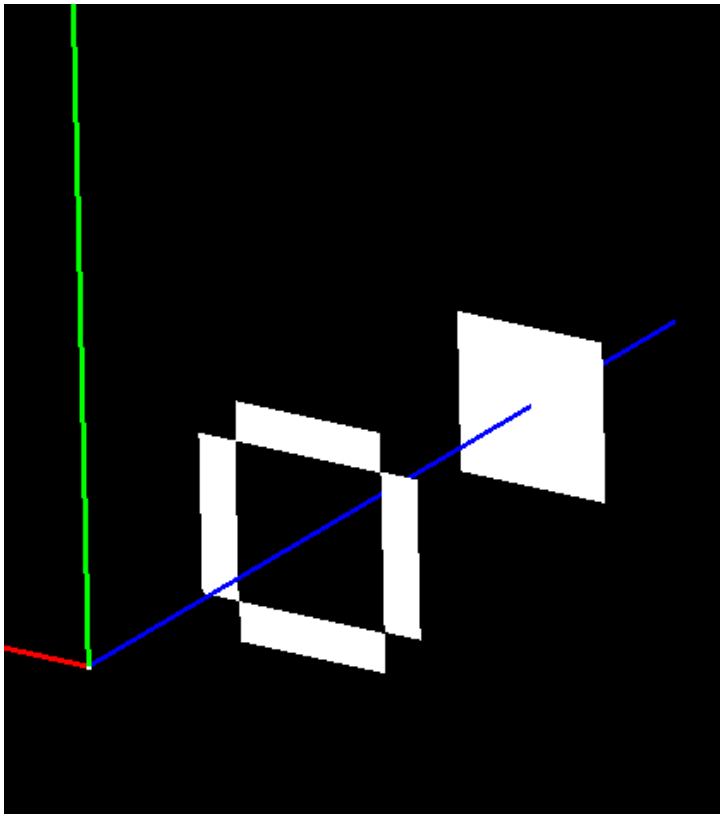
(Different axis/color scales)



Task 3: SANS : D33

Mantid has built in GUIs for SANS@spallation

We study how SANS/TOF can benefit from Mantid algorithms



D33@ILL In Mantid
Similar rendering exist in other packages

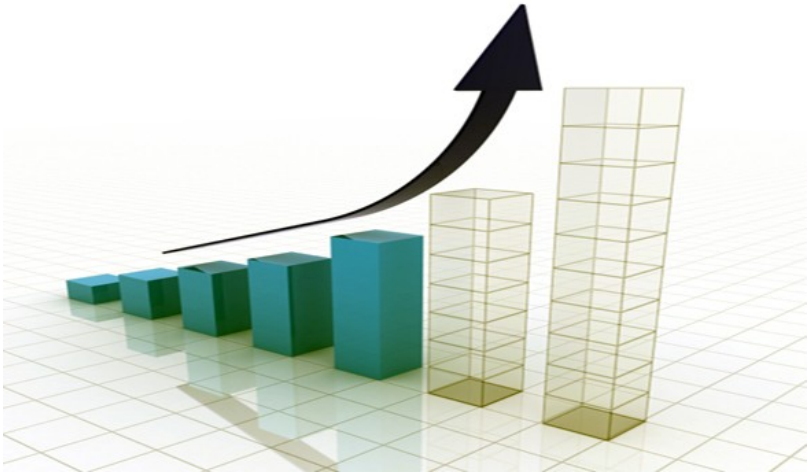
Task 3: performance

Main tasks in Mantid:

- File read (4 s)
- Convert units ([t,θ] → [ω,q], few minutes)
- SofQW (10 min for IN5, 1 min for others)
- Requires 10 GB of memory
- Data sets currently require time information for most algorithms

Lamp:

- File read: 15sec
- Convert to energy: < 0.5 sec
- S(q,ω): 1 min (maximum)



Task 3: outlook

We will proceed with Mantid evaluation

- More TOF machines
- Multiplexed TAS (no time axis)
- Powder and SX diffractometers (no time axis)

Benchmarking with other software is essential (LAMP, DAVE,...)

Evaluate interconnection between different packages