Data Reduction

Can we agree on

1. basic data reduction steps?
2. necessary amount of metadata?
3. common vocabulary and units?

Thomas Saerbeck

What is data reduction?

Data reduction is the transformation of [..] information [..] into a corrected, ordered, and simplified form. The basic concept is the reduction of multitudinous amounts of data down to the meaningful parts. [Wikipedia]

Using metadata

- detector efficiency, normalization, distances, wavelength(s), user input…

and transformation into agreed set of coordinates

- e.g. Q, ΔQ, I, ΔI

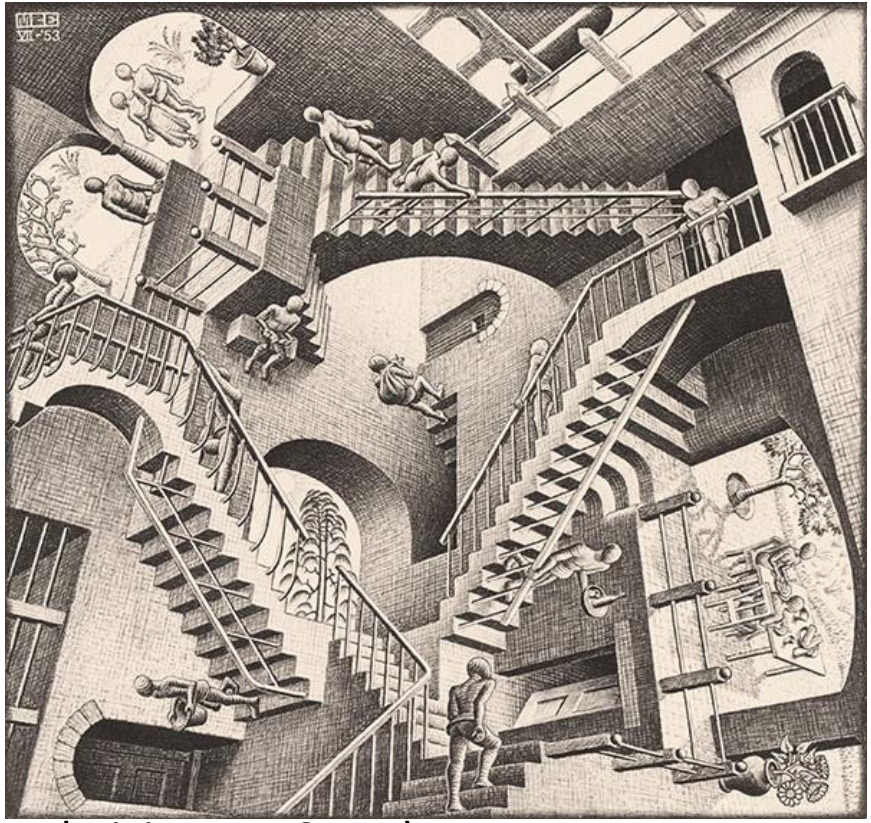
data reduction should lead to

- Separation of instrumental artefacts from physics
- Maximum usability of data
- Maximum comparability of data

Where does reduction start and end?

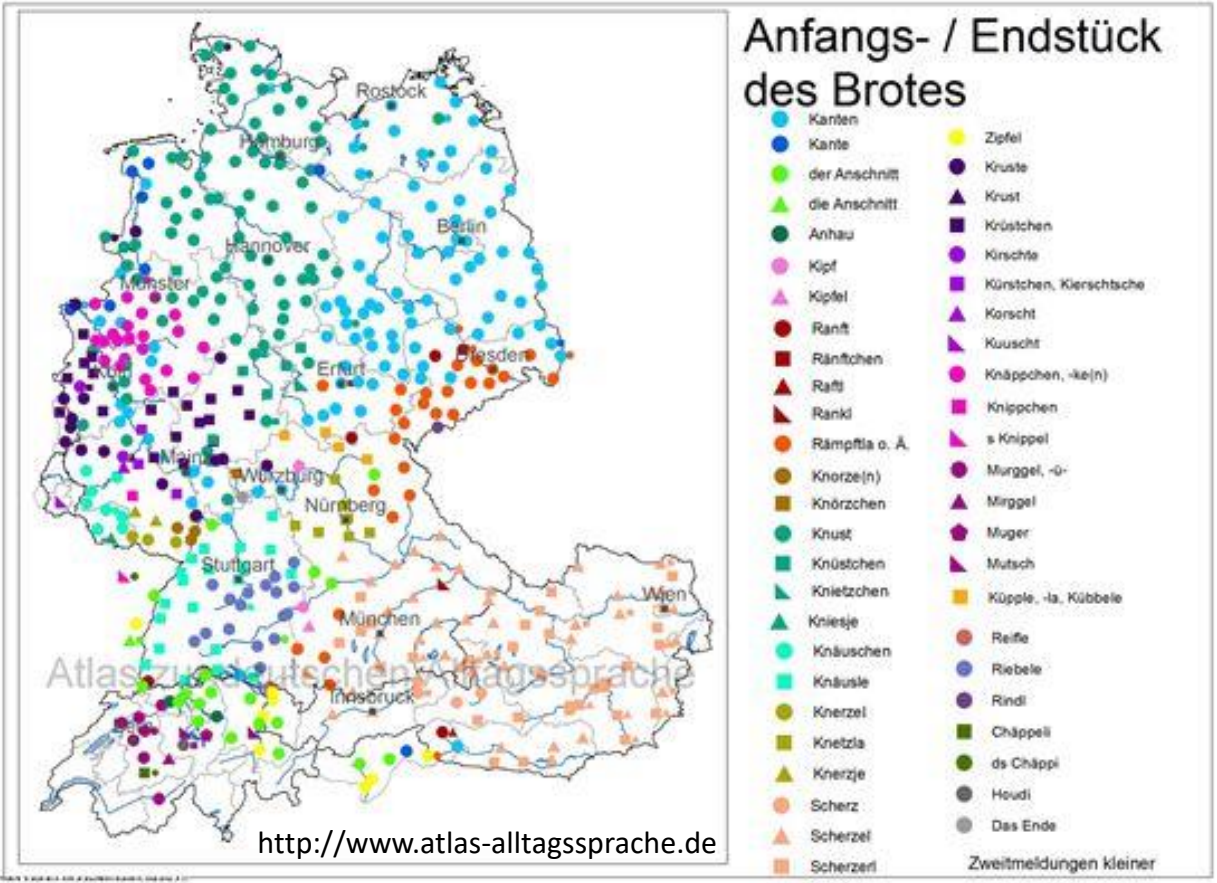
1. In the instrumental hardware
2. During the experiment
3. After the experiment
4. During analysis
5. In publication

- Data reduction must include all steps that are not accounted for in instrument hardware OR not included in the analysis software.



Relativity ; M. C. Escher

Common Definitions:



Germany has over 200 words for the end-crust of a bread…

Data reduction must use a common language and units

Compromise

- σ vs. FWHM
- Phi vs. San vs. Ω vs. θ…
- Binning vs. grouping
- Flood vs. waterrun
- Direct beam vs. transmission

Current approach:

- Each instrument has their own data reduction procedure
- Many instruments are seen incompatible (x-ray + neutron; mono + ToF)
- Every facility designs their own pieces of code (even with Mantid)

⇒ Every reduction is different, leading to different formats, reproducibility questions and difficulties in comparison for multi-institute users.

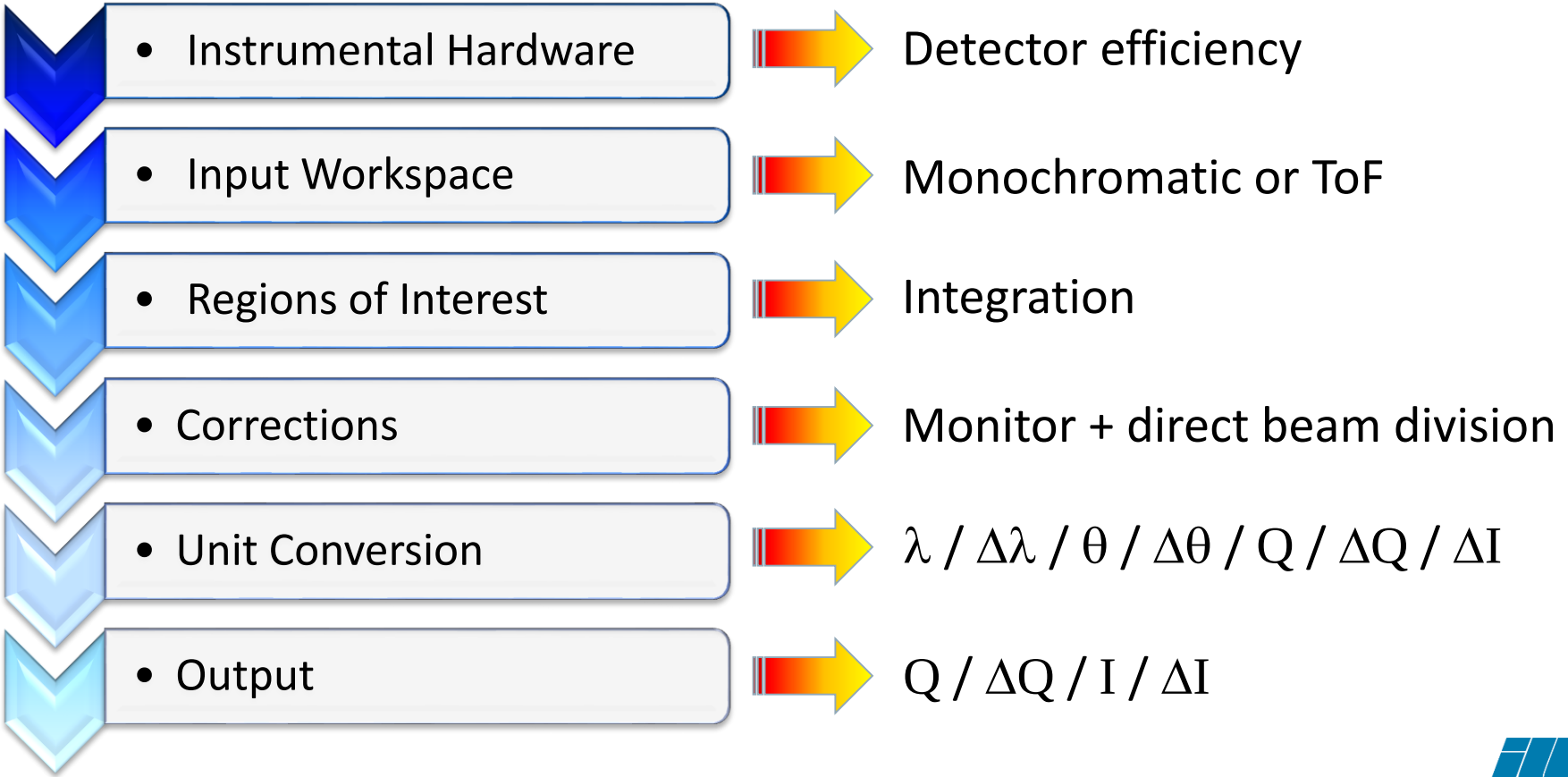
Aim:

We will probably not agree on the appropriateness and precise order of each step, but we might agree on what are the **minimum necessary steps**, which parts are **"voluntary"**, what to include in the **metadata** and a **defined vocabulary**.

Workflow?

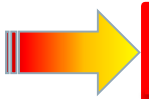
- Instrumental Hardware
- Input Workspace
- Regions of Interest
- Corrections
- Unit Conversion
- Output

Minimum Steps?



Adding Content

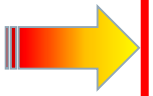
- Instrumental Hardware



- Source instrument
- Flood normalization
- Gamma discrimination
- Dead time
- Instrument background
- Detector resolution
- Distance conversion
- Wavelength(s) + angles
- Appropriate units
- Sum / Group similar datasets

Adding Complexity

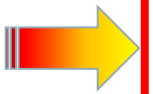
- Instrumental Hardware
- Input Workspace



- Monochromatic or ToF
- Static, kinetic, event mode
- Polarized data?
- Data space (3D, 2D, 1D, 0D)
- Integrate unresolved dimension
- Instrument metadata
- Scan + data axis

Adding Flexibility

- Instrumental Hardware
- Input Workspace
- Regions of Interest

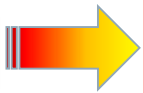


- Data binning (3D / 2D)
- Peak finder
- Peak analysis (X_C, FWHM)
- Foreground
- Foreground shape
- Background
- Wavelength band
- Integration methods

(see selene, divergent, bent, prism, birefringent)

Accounting for Instrument Specifics

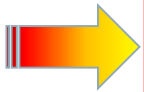
- Instrumental Hardware
- Input Workspace
- Regions of Interest
- Corrections



- 1D binning
- Monitor normalization
- Direct beam normalization
- Normalize to slits
- Background subtraction
- Polarization correction
- Gravity correction
- Over-illumination factors

Provide Proper Units

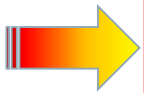
- Instrumental Hardware
- Input Workspace
- Regions of Interest
- Corrections
- Unit Conversion
-



- Output dimensions?
- Calculate λ, θ, Q, p_i, p_f,
- Calculate $\Delta\lambda$, $\Delta\theta$, ΔQ
- Binning / grouping to resolution
- Propagate errors

Match Data Requirements

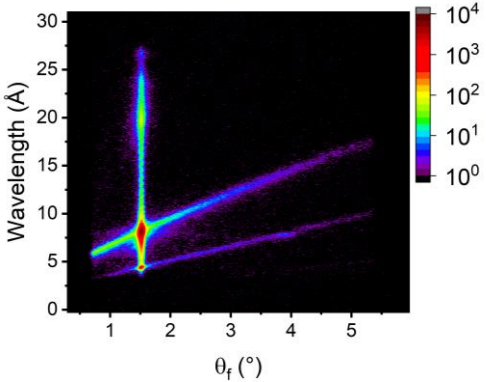
- Instrumental Hardware
- Input Workspace
- Regions of Interest
- Corrections
- Unit Conversion
- Output



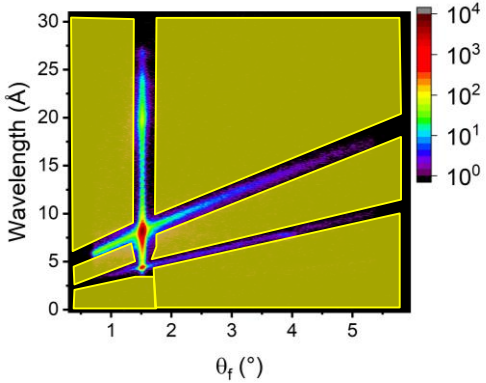
- Join / stitch data
- # of files to write
- Off-specular data
- Number of subsets to include (Q / ΔQ / I / ΔI, [Q_X, Q_Z, λ, θ])
- Calculate auxiliary data (spin-asymmetry)
- Loop?
- Display data

Example: Background Definition

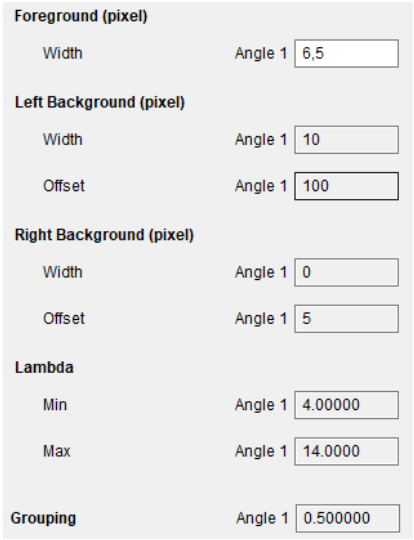
Detector Image

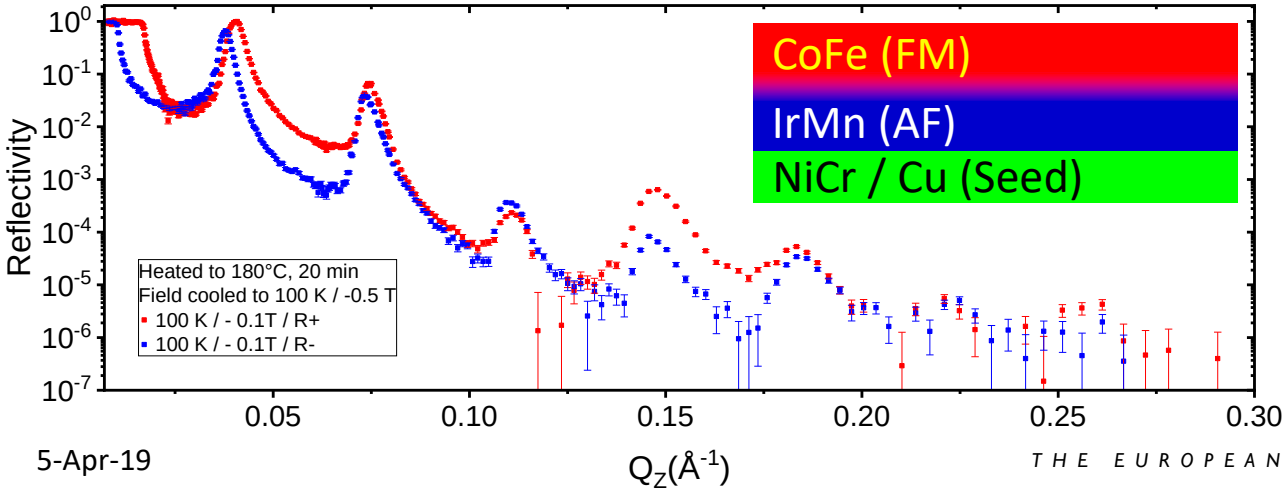


Background Regions



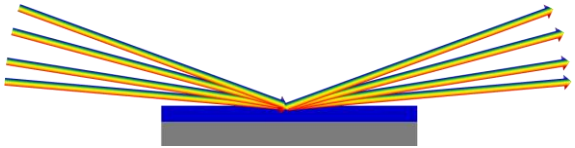
Currently:
Only rectangular background

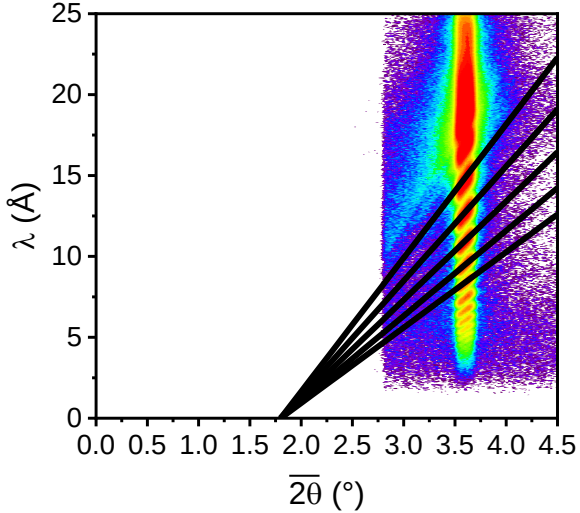




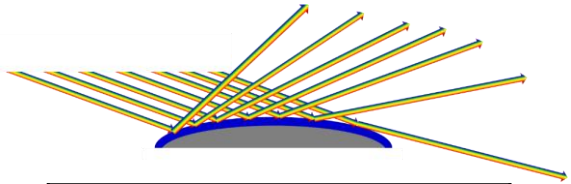
Example: Integration Methods

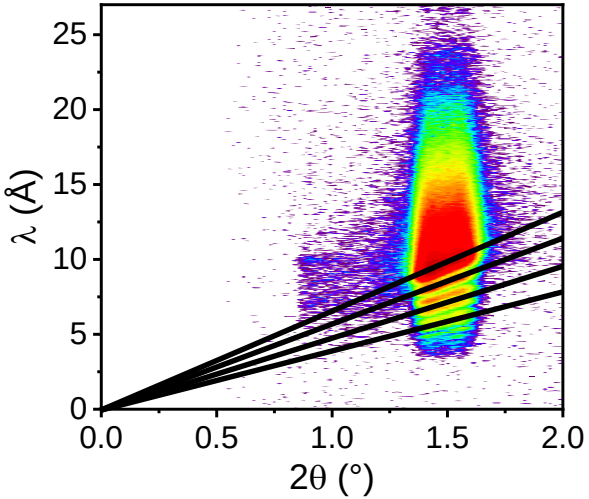
Divergent beam





Curved sample

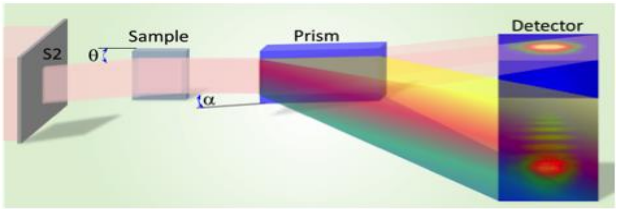




Further considerations:

Selene

Rainbows (?)



R. Cubitt, T. Saerbeck, R. Campbell, R. Barker, P. Gutfreund, J. Appl. Cryst. **48**, 2006 (2015)

Example Contour Plot Coordinates

Instrumental

$$[\theta_f, \lambda]$$

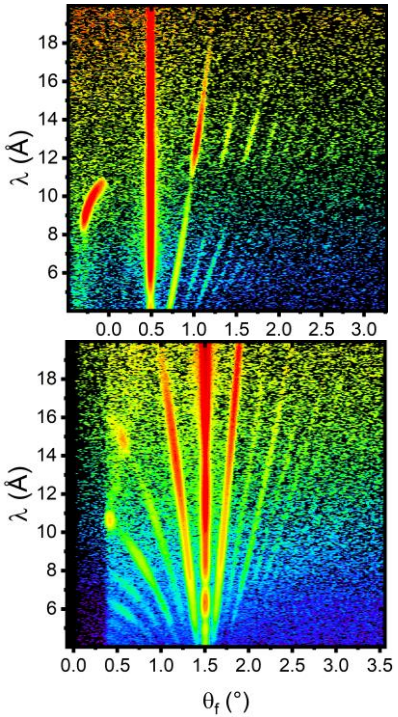
Intermediate

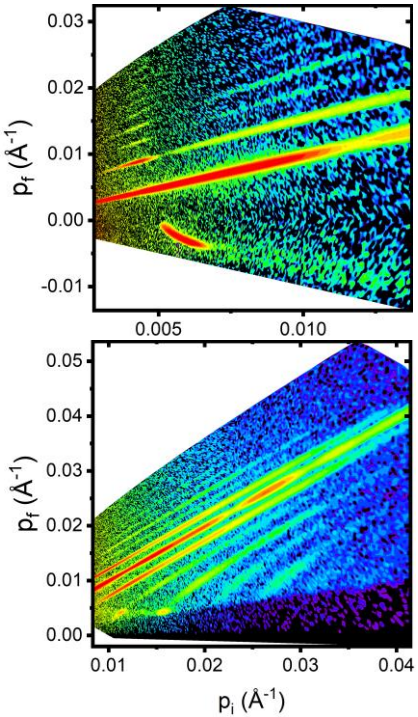
$$p_i = \frac{2\pi}{\lambda}\sin\theta_i$$
$$p_f = \frac{2\pi}{\lambda}\sin\theta_f$$

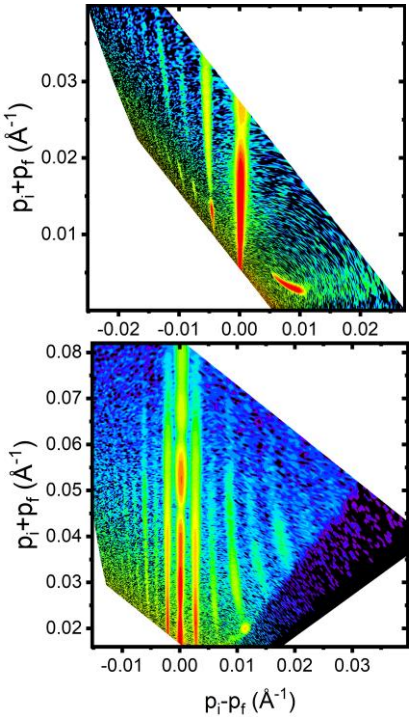
$$[p_i - p_f, p_i + p_f]$$

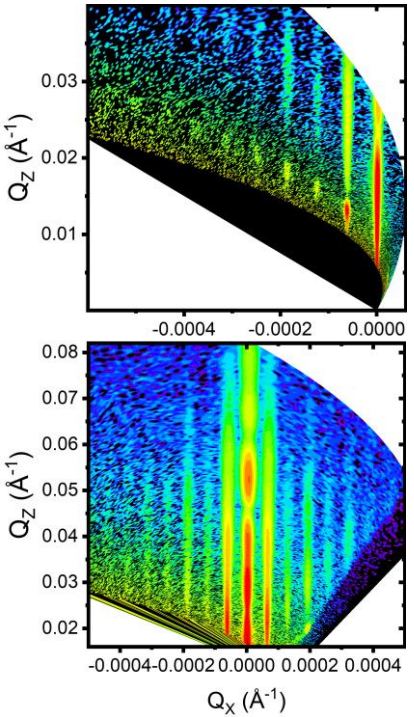
Reciprocal

$$Q_Z = \frac{2\pi}{\lambda}(\sin(\alpha_i) + \sin(\alpha_f))$$
$$Q_X = \frac{2\pi}{\lambda}(\cos(\alpha_f) - \cos(\alpha_i))$$









Discussion Topic 1

Explore the possibilities for unification across techniques and facilities of the following reduction steps and agree on a defined vocabulary and units. Decide where the step should take place (instrument hardware, reduction software, analysis software). Define a set of metadata necessary for traceability.
The list below is not exhaustive, so please add your own ideas.

- Detector efficiencies and background subtraction

  Detector flood, background in metadata; instrument or sample background; …
- Binning in 2D and 1D

  Constant/log/fraction of resolution steps in Q; stitching of data; difference between binning, grouping and summing; …
- Conversion of instrumental parameters and resolution function

  Wavelengths, distances and angles; σ vs. FWHM; include $\Delta\lambda$, $\Delta\theta$, ΔQ; …

Discussion Topic 2

Explore the possibilities for unification across techniques and facilities of the following reduction steps and agree on a defined vocabulary and units. Decide where the step should take place (instrument hardware, reduction software, analysis software). Define a set of metadata necessary for traceability.
The list below is not exhaustive, so please add your own ideas.

- Polarization and geometrical data correction

  Polarization efficiencies; which correction algorithm; provide spin asymmetry; gravity; over-illumination; normalization and scaling; …
- 2D maps and off-specular data

  Q_X/Q_Y or p_i/p_f or $\alpha_i,\alpha_f,\lambda$; data format (column or matrix); 2D binning and errors; …
- Identification and Integration of data

  Divergent beams, bent samples, selene; peak finder and summing in $\alpha_i,\alpha_f,\lambda$ or Q; ...

INSTITUT LAUE LANGEVIN

NEUTRONS FOR SOCIETY