Summary IUCr Workshop

Selina Storm

What are raw data good for?

- hidden treasures (data forgotten / not properly processed)
- training / education
- machine learning
- software development (there might be more in the data that you can currently get out, but: software ages faster than data)
- prevention against fraud

Challenges

- struggle to keep up with the amount of data
- as data storage is expensive, solutions are based on the availability of funding, resulting in different solutions
  - reduce data early on (SSX, DESY)
  - don’t offer the possibility to store data (Australian Synchrotron)
  - let the users decide which 10% of the data they want to store (European X-FEL)
  - keeping everything up to now
  - discussions of making the users pay for data storage
- different solutions based on the availability of funding
- users want to get answers quickly (more advanced experiments, *e.g.* in small-molecule crystallography)
- new techniques to solve the same structures (electron diffraction)

FAIR data

- data need to be findable (DOI solution ESRF)

- accessible: data policies: data should be kept for three years, but it's not clear what happens then
  - data management plan, organised storage big issue: interoperability (medium rare data: unmerged intensities and extensive meta data)

- reusable: meta data
  - DAPHNE: need to do this from the very beginning and keep records, *e.g.* Biosync
  - documentation of the samples (powder diffraction) / use of persistent identifier

- reproducible → need to think about that much more