# ORSO File Formats Session 2

Present in the session were

Andrew Caruana, Thomas Saerbeck, Wei Bu, Mrinal Bera, Robert Dalgliesh, Tim Snow, Jean-Francois Moulin, Gaetano Mangiapia, Chen Shen

The session was tasked with discussing and providing answers to a couple of very specific questions which it largely turned out were fairly generically answered

* proposal: introduce a set of file formats with increasing requirements
  + (e.g. lazy, pragmatic, strict; with lazy as a reduced ASCII format excerpt of the others)
  + discussion outcome
  + probably list of definitions/requirements

## Lazy, pragmatic and strict.

The question was raised as to whether users would actually use the more complex file format.

Chen Shen stated that they already provide 2 different folders of data, one is the reduced data in a simple format and the other is the more complex data. The user then decides which data they wish to take. More advanced users will tend to take the more complex data while less experienced or general users will take the more simplified data.

Thomas Saerbeck then raised the question of the logging of the reduction metadata which may be needed in the future to reanalyse data or may be required by journals. Tim Snow pointed out that the Diamond reduction system already does this and places the history of operations into the Nexus files provided to users. However they also already provide a simplified ASCII file. This would allow the users or instrument scientists to subsequently pick their way through the analysis performed.

The data reduction facilities provided by Mantid, at Diamond and at some of the German sources already allow this sort of processing to happen but the recording of the procedures in the output files still needs some work. The output should also include version number of the software and algorithms used.

The question of whether providing the raw data and a script was discussed but could easily be seen to have potential problems as data set scales increase and major compute resource is increasingly required to perform data reduction. This pushes you away from having access to the raw data and hence having a record of the processing carried out on the reduced data will become more necessary.

The concept of data quality was briefly reviewed but mainly in the context of providing on beam diagnostics that would enable the user/ instrument/ instrument scientist to prevent problems caused by detector saturation or poor signal to noise. The idea of trying to asses useful data information content was not discussed in any great detail as it would require a significant amount of prior information to be provided such that some form of AI could be used. TS (ILL) promised TS (Diamond) a beer if he could make this happen.

The majority of the rest of the discussion centred on other aspects of file content and how to encourage the users to provide sufficient additional detail.

Additional content from in-situ complimentary techniques as well as complex sample environment was clearly something that will be required in any pragmatic or strict data file but the amount of information that will be required was compared to opening a very large can of worms. This is an area that will require considerable further discussion.

Examples included

* Magnet field history from before a particular measurement to enable sample field history to be clearly traced.
* Synchronous measurements from rheometers which are triggered in some complex way where the triggers are important.
* Accurate time stamping of sample environment data relative to the radiation data to enable time filtering
* Recording of instrument positions in appoint by point scan on an X-ray instrument takes a lot more space and needs to be considered

The issue of ensuring users fill in a digital sample log book will be needed to greatly improve the prospect of enabling automated reduction. A carrot and stick approach is highly likely to be required in almost all case to encourage users to do this. The only other thing that is likely to encourage behaviour to change will be increasing requirements from the journals. However it was pointed out that we should get ahead of the curve and this would also be of significant benefit to instrument scientists and users when it came to the possibility of needing to go back to data years after it was originally taken because of e.g. a request form a reviewer.

For the carrot an easing of the rigors of both data reduction and analysis is likely to be the most readily available tool. This will need to be accompanied by suitable script generators that enable data to be easily added. This is done at Petra using a template which the users are warned will make their life more difficult if they do something different.

Another tool to be used would be to ensure that the analysis tools that the users make use of will accept the new file formats and it will give them access to increased functionality. Resolution function fitting is a prime example of this. It is difficult to fit resolution precisely from a 4 column asci file without providing additional capability within the fitting program itself. (see Refl1D and SNS data)