Workshop on *Data Visualisation, Reduction and Analysis* at Australia's Replacement Research Reactor

Workshop Report and Recommendations

Robert McGreevy (ISIS Pulsed Neutron and Muon Source)
Ray Osborn (Argonne National Laboratory)

High level data analysis and visualisation is an integral part of modern neutron scattering. For any facility to be internationally competitive, it needs to ensure that the software available to users enables a rapid and complete analysis allowing the user to concentrate on scientific interpretation. As data sets become larger, particularly if they are multidimensional, there is a need for sophisticated visualization tools. Indeed, as the speed and complexity of experiments increases, and the proportion of 'non-expert' users rises, the requirement for sophisticated and yet easy-to-use software increases. For the Australian Replacement Research Reactor (RRR), this will be a key factor in ensuring the growth of the national user community and making a significant scientific impact.

However, in common with all other facilities, software development has been left to the end of the priority list and is only considered at a late stage in the design process. It can take up to three man-years to develop data reduction and analysis software from scratch for each instrument. These resources are not available in-house in the time available, so these constraints, to a large extent, determine the recommendations that we make.

Background

The initial instrument suite being built at the RRR is extremely diverse, with each instrument requiring very different forms of data analysis and visualization. There are several approaches to coping with the challenge of this diversity.

1) *Develop specialized analysis software for each instrument, with little or no coordination between different instruments*.

   This has been the traditional development model at neutron scattering sources, where the responsibility has usually rested with the instrument scientists to write the software. The main drawback is that the software becomes difficult to maintain by anyone other than the original author, so that it eventually falls into disuse. It also tends to perform a limited range of analysis determined by the author's interests.

2) *Develop a graphical user interface that allows the visualization of data from a range of instruments, using menu-driven commands*.

   Examples: DAVE, LAMP, ISAW, IGOR

   Commercial packages, such as IDL, Matlab and IGOR, make the development of such interfaces relatively easy, and can provide a high-level scripting language that allows rapid development of sophisticated analyses. Using a menu-driven approach simplifies the number of options available to the novice user (as long as it is well documented), but does make it difficult for the expert to perform non-standard analyses. The users become dependent on the programmers for any changes in functionality, unless the underlying scripting environment is exposed (LAMP and

ISAW provide access to a limited range of scripting commands to allow the development of customized macros). Such packages are usually developed by a team of programmers and are relatively easy to maintain by computer support personnel.

The packages that are currently available are designed to input data in the specific format of the developers' facility, or a range of formats, and therefore would require the cooperation of those developers before use at ANSTO. However, such problems will be minimized by the increasing use of the NeXus data format, assuming that the RRR confirms its decision to use NeXus.

3) *Use a common high-level scripting language to develop analysis packages based on a library of components, with the option of developing graphical interfaces.*

Examples: DANSE, Open Genie

This is the most recent approach and provides the greatest flexibility in building sophisticated analyses; the components can be combined in a large variety of ways, and can incorporate legacy software in a variety of computing languages after wrapping as shareable objects. Once the required scripts have been developed to perform a particular analysis, they can be activated by a graphical user interface in order to help the novice user. In principle, such component-based schemes could use a commercial language, such as IDL or Matlab, which also provide simple GUI builders and visualizating packages. The only problem with this approach is that there is no community consensus about which one to use, and they are too expensive for the user to license a number of them. Alternatively, the components could be built using an open source scripting language, such as Python or Open Genie, which is therefore freely available but which may not have such powerful GUI or visualization options.

DANSE, which will use Python as its scripting language, has the nominal support of all the neutron facilities within the US, and is sufficiently close to European activities in ”eScience” that it could draw on the resources of many groups worldwide. The design has a number of attractive features, including a distributed architecture that will make it easy to perform web-enabled analysis at remote ”supercomputing” centres, and a graphical programming environment. If the neutron scattering community develops a consensus to support the development of DANSE, or a similar framework, then a powerful library of analysis tools could be developed relatively rapidly. However, DANSE is in the very early stages of development, so it is too soon to make a firm recommendation to adopt it as a short-term solution.

Open Genie is an open source alternative that provides some of the DANSE functionality (although without the distributed architecture, though this is less important in the short term). It has a well-structured scripting language, is able to read NeXus files, is available on a variety of platforms (including Windows, Linux, and Mac OS X) and could be deployed across all the instruments to provide basic visualization and manipulation of the raw data. It is not possible to recommend as a long-term solution, however, because it is maintained by a single facility (ISIS) and future levels of support are uncertain. In the short-term, it could be used to provide a consistent low-level access to RRR data across all the instruments with relatively little work, but it would probably be used for diagnostic purposes by the instrument scientist, rather than for final data reduction by the user.

Recommendations

Given some of the activities that are going on elsewhere in the world, there is a possibility of international collaboration for the development of a range of advanced software tools for neutron scattering, e.g. DANSE project (Caltech), C++ software library (J-PARC), data portal (ISIS), web based user interface portal (SNS). Whether this goes ahead will be known within about a year, since those facilities under construction (SNS, J-PARC) will have to make firm decisions on this time scale. We therefore suggest the combination of a short and long term strategy towards the provision of data analysis and visualization software.

Long Term Strategy

1. If a major new international collaboration on software development does become a reality, then RRR should aim to become involved. This would clearly require putting some resources into the collaboration, though at what level or for what purpose is not yet clear. Such an involvement would require a serious policy decision, so this should be a matter of discussion in the near future. If it is decided to follow this path, then relevant RRR personnel should become involved in the ongoing discussions and planning. (Travel money needed)
2. If international collaboration does not turn out to be feasible, then RRR needs to decide on a uniform platform for basing its future data analysis and visualization. We would not recommend that this is done independently, but in collaboration with another facility that has already made some progress in this direction. Possibilities include OpenGenie (ISIS), ISAW (IPNS), DAVE (NIST).
3. The decision to use NeXus as the standard format for raw data is sensible and should be confirmed. NeXus has been adopted as a standard by several new and existing facilities (e.g. SNS, J-PARC, ISIS, PSI, LANSCE) and use is growing at other facilities. It therefore seems likely to become the de-facto standard within a few years. Furthermore, several external software packages are being adapted to read NeXus files, and therefore become available for the analysis of RRR data with relatively little modification. This allows users to choose which package they prefer based on its functionality, rather than its ability to read a particular format. RRR is already participating in the development of the NeXus format, with membership on the International Advisory Committee, and this involvement should be maintained.
4. While Monte Carlo simulations have been used in the design of instruments for the RRR, they have not yet been considered as a tool for aiding data analysis. We would recommend that RRR maintain some connection with the EU research network MCNSI (Monte Carlo simulation for Neutron Scattering Instrumentation) within which ‘virtual instruments’ are being developed. Some of these could be easily modified for RRR instruments. (Travel money needed)
5. Materials simulations (e.g. molecular dynamics, various types of ab-initio simulations, finite element modelling) will have an increasing importance in relation to the

interpretation of neutron scattering data. Pro-actively instituting collaborations with groups in Australian universities could have a great benefit in the long term.

6. Facilities should be provided to help the user produce realistic simulations of proposed experiments on each instrument, using, for example, simple analytic scattering functions combined with known flux distributions and scattering cross sections. Although the RRR should encourage challenging experiments, the user (and peer reviewers) should have the tools to assess their feasibility in advance.

7. The volume of data produced by the RRR neutron scattering instruments, and the related computational demands, can be predicted to increase more rapidly than Moore's law, i.e. it will no longer be adequate to rely on the increasing power of small workstations and some level of access will be required to larger computing facilities. It would clearly be advisable if RRR were connected to a high-speed network or a relevant GRID system.

8. As much collaboration as possible with the Australian synchrotron is highly recommended. While slightly outside of the remit of this report, we would suggest the implementation of a joint proposal system, user database etc. Future joint software projects should also be considered. Even at this early stage we would suggest the setting up of regular meetings/communications through some form of electronic collaboratory system (also meetings in person).

Short term strategy.

1. The highest short term priority is to ensure that on day 1 of operation each instrument has a basic but complete suite of programs that enables the users to carry out all of the necessary data treatment (i.e. conversion of raw data to scientifically meaningful data).

2. While a uniform 'look-and-feel' of software across instruments would be preferred, it would not be advisable (given the available time and likely available manpower) for RRR to attempt to produce their own software package that does this. Instead the approach should be to leverage as much as possible off existing software from other facilities with similar instruments. At this stage any customization should be kept to the minimum absolutely necessary. Specific recommendations for the different instruments are summarized in Table 1.

3. Since it is already able to read NeXus files and perform neutron-related unit conversions, Open Genie could be used to provide basic visualization and manipulation of raw data, particularly for diagnostic purposes. It is recommended that it be made available on all the instruments, although other packages may be preferred for routine data reduction and analysis by the users (see Table 1).

4. In-house expertise in other packages that are commonly used at other facilities may be important, based on either open source or commercial software. Given our short-term recommendations, the likely candidates are Open Genie, Igor, and IDL, although Matlab is also used at several facilities. Licenses need to be obtained where necessary. Executable versions of programs such as LAMP and DAVE can be distributed to users who do not have licenses. However this can cause problems if they are only

available for certain platforms (e.g. PC and not Mac OS X or Unix). In the short term this needs to be dealt with on an individual basis.

5. While there is a preference for open source software, no facility is in a position to devote the necessary resources to produce the quality and range of visualization tools that are commercially available. Commercial software should not be avoided, especially if it provides essential functionality that will be costly or time-consuming to reproduce in-house.

Table 1: Short-Term Recommendations for Each RRR Instrument

Quasi-Laue diffractometer	Since this is a commercial image plate device the raw data format will be TIFF, not NeXus. There are a number of software packages for performing crystallographic analysis of image plate data, including those provided through the UK CCP4 collaboration or the ILL. One or more should be installed and used after consultation of the RRR user community. However, the graphical interface and documentation is often difficult to understand by a novice user, so we recommend that local documentation and/or tutorials specific to this instrument are supplied. Software for visualizing crystal structures is obviously required. There are many possible choices, which could be either in the public domain (e.g. POV-Ray) or commercial (e.g. Cerius). However, in many cases, the user just needs structure factors in a convenient format for analysis at their home institution.
High intensity powder diffraction	This instrument is very similar to D20 at ILL. It therefore seems likely that the necessary software within the LAMP program (based on IDL) could be obtained and implemented with little modification.
High resolution powder diffraction	This instrument is very similar to the new Super D2B instrument at ILL. While some software, again based on LAMP, should be available, it is unclear whether this yet deals with the proper integration around Debye-Scherrer cones enabled by the use of PSDs. This should be checked. If the software is not available, then it should be checked on what timescale ILL would plan to implement it. Both powder diffractometers will require Rietveld refinement software. While there seems to be an attachment to the locally developed Rietica code, it is not obvious that this will continue to be supported and developed in the future. In any case, Rietica does not deal with all necessary aspects. A range of codes should be made available, the most obvious being GSAS and Fullprof (particularly for magnetic structure refinement).
SANS	The IGOR package from NIST seems to satisfy all of the requirements for SANS data reduction, particularly if it is adapted over the next two years to use the NeXus format. For analyzing the treated data, there is a very wide range of programs available. Which of these should be implemented would depend on the likely scientific problems of the users. This could be assessed by a short survey. The proposal to make SAXS available on a routine basis as a complementary tool to SANS is strongly supported – the software will obviously need to be able to combine the different data sets.
Reflectometer	The basic data treatment software needs to be appropriate for a TOF instrument with a two-dimensional multidetector. D17, at the ILL, is the obvious candidate, and probably uses LAMP in some form, although some aspects of the NIST reflectivity software may also be applicable.

Residual stress	The SSCANSS package (IDL-based) developed by the Open University for the Engin-X instrument at ISIS is a first rate piece of software and is likely to be used at several facilities (ILL, SNS, ESRF) and possibly become a de-facto standard. It is therefore obvious that this should be implemented.
TAS	TAS data treatment is strongly coupled to the instrument control and data acquisition software. This will be SICS from PSI. While a preference was shown for a data analysis package from LLB, it might make sense to see what PSI have available. We would also strongly suggest that RRR implement a simulation package capable of determining the resolution function for different instrument settings and SCANS. The Restrax package from NPI (implemented at ILL) is the most efficient for TAS instruments and includes both a simulation part and an analysis part.
Polarised Spectrometer	If a LONGPOL type instrument is built, then it will be unique and so will require specially developed software. If a polarized TAS is built by Taiwan, then presumably they will take over responsibility for software provision. However, RRR should insist on the use of the NeXus data format, and harmonize their software with the other TAS as much as possible.