Table 1.3 d1: Work package description for Networking activity or Joint research activity

For each work package:

Work package number		**Start date or starting event:**					
Work package title	Standards for data analysis software						
Activity Type[1]	COORD (or RTD)						
Participant number							
Participant short name	ILL	ISIS	PSI	FRM2	JCNS	GKSS	HZB
Person-months per participant[2]**:**	4+24	4	4	4	4	4	4

Objectives

To determine how to develop, deploy and operate a common data analysis software infrastructure to facilitate joint software development in the neutron scattering community.

Description of work (possibly broken down into tasks), and role of participants

The scientific output of neutron scattering facilities is determined by a production line which starts from the neutron source and finishes with published results. There are many links in this chain like neutron guides, detectors and sample environment, many of which have been or are supported by NMI3 activities. Any weak link jeopardises the whole production line. Facilities rarely have the resources to adopt a holistic approach to the production of scientific results and effort tends to be committed sequentially along the production line leaving data analysis, as the ultimate step before publication, under-funded. Instrument scientists are therefore left to find specific software solutions, as has traditionally been the case, leading to a pluralistic, heterogeneous provision of software. This ad hoc solution has worked in the past but new, high resolution detectors, higher data rates make, etc make new demands on software that require professional software solutions. In particular, new instruments constitute a considerable challenge for software provision and failure to address this issue leads to a delay in the scientific impact of new investment.

The importance of data analysis is now being realised and is becoming a focal point of efforts to optimise scientific production. In facilities which have scientific computing groups with responsibility for data analysis (less than half), a more rational approach to data analysis has been pursued leading to multi-functional, software platforms like LAMP [1] at the Institut Laue Langevin (France), DAVE [2] at NIST Centre for Neutron Research (USA) and, most recently, Mantid [3] at ISIS (the Rutherford Appleton Laboratory, UK). But at a European and world-wide level, these software solutions still constitute a duplication of effort, using a range of methods, practices and languages, for what are mainly standard data analysis tasks.

Broader, multi-centre approaches are therefore being pursued. The construction and now operation

[1] Please indicate one activity per work package:
MGT = Management of the consortium; COORD = Networking activity; RTD = Joint research activity.

[2] Except human effort already included in the calculation of the access costs.

of the Spallation Neutron Source (SNS) in the USA prompted the DANSE project [4], which has collected together and wrapped a wide range of software, accessible via a web portal, to facilitate science peformed with neutrons. This software initiative has not however addressed the need for primary data analysis directly on the instruments and SNS is now collaborating with ISIS in the development of Mantid, both facilities operating similar types of instruments.

In Europe, more than ten central facilities for neutron and photon (X-ray) science are starting to join forces within the FP7-funded, PaNdata initiative [5] to address the management and flow of information from users and proposals through to publications. Data analysis is one of the major workpackages in this support activity, the goal being to determine a strategy for sharing software. The model being pursued is a repository of mainly existing, inter-operable software, which takes into account the diversity of software at a wide range of existing and new facilities. Centralising software will facilitate access for all users, encourage best software practices and allow tracking of software use.

Finally, within the High Data Rate Initiative (HDRI) project in Germany, all Helmholtz centres contributing to the Photons, Neutrons and Ions Research Programme (PNI [6]; i.e. HZB, FZJ, HZG for neutrons; DESY, KIT, GSI) are working together to make substantial progress in the efficiency of use of new sources. This includes agreement on common tools for data management, the development of software and hardware for real-time data processing and the provision of user software for further data analysis and simulation.

In this context, the support activity proposed here will explore “standards for data analysis” to prepare the next step in collaborative software provision within the neutron scattering community, the goal being to define the requirements of a common data analysis platform. While the PaNdata workpackage on data analysis aims to bring together a wide range of neutron and photon software in a central repository without any attempt to rationalise content, in this project we want to make important steps towards optimising software development within a unique software infrastructure, which will also give users of neutron scattering facilities a common look-and-feel as they move between instruments and facilities. The two projects are however complementary in that the tracking of software use via a centralised repository will define the most useful neutron scattering related codes that should ultimately be integrated in a common software environment.

Defining a new, unique software environment for data analysis would be a formidable task, beyond the scope of the support activity proposed here. However, existing facility-based software, mentioned above, constitutes an advanced starting point and possibly a solution. In particular, Mantid is the most recent project, it is based on open source software (C++ and Python) and it is already a multi-facility collaboration (ISIS and SNS), involving also an external company (Tessella [7]) for software engineering. Mantid is therefore a possible solution, the challenge being to extend the existing functionality for instruments at spallation sources to those at reactor sources, which constitute the majority of the partners in this project.

Mantid is, however, not necessarily the only solution. The project will therefore begin with two review tasks: review existing data analysis software and practices of software developers and review existing solutions for a common data analysis infrastructure. Based on the reports from these tasks, the key phase of the project will be to develop a small number of prototype software solutions (probably 2) in areas chosen by the partners. The goal will be to deploy functionality that exists at some facilities in the common framework thereby making it accessible to and optimised for all partners. The goal will NOT be to develop new functionality. The final task will be to evaluate and report on the prototypes, thereby defining the software solution for integrating existing software

into a common infrastructure and developing new software on an efficient, collaborative basis in the future.

Ten facilities will participate in this support activity, each contributing 1 man-month per year to the reviewing and evaluation tasks. The majority of the resources will be used to employ a software scientist for two years at the lead facility to investigate prototype software solutions.

[1]http://www.ill.eu/computing-for-science/cs-software/all-software/lamp/the-lamp-book/, [2]http://www.ncnr.nist.gov/dave/, [3]http://www.mantidproject.org/, [4]http://wiki.cacr.caltech.edu/danse/, [5]http://pan-data.nd.rl.ac.uk, [6]http://www.helmholtz.de/en/research/structure_of_matter/research_with_photons_neutrons_and_ions_pni/, [7]http://www.tessella.com

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

Deliverables (brief description and month of delivery)

D.1: Report on current software and practices.

D.2: Report on solutions for developing a common software infrastructure.

D.3: Prototype software in chosen solution.

D.4: Report on evaluation of prototype software.