Grenoble, July 18th, 2014

From: E. Farhi, ILL
to: FP7/NMI3-II WP6 members and everybody interested by the topic

Data Analysis Standards (NMI3-II WP6)

4th Meeting Minutes (June 17th 2014, Grenoble)

The meeting was the last organised in the scope of the NMI3-II project, as the available funding for staffing is coming to its end by end of September 2014. The status of the work-package was discussed, as well as its achievements and deliverables. The future of the work-package was also mentioned.

Status of the work-package

Currently, the two first deliverables (D6.1, D6.2) have been produced, and the 3rd deliverable (D6.3) will be released in July 2014. The 4th deliverable is expected by September 2014, so that all planned tasks will have been completed. The work-package production phase will end with the last deliverable.

When comparing the initial plans of the work-package with the actual ones, we notice that Mantid has acquired a central role. This is mostly the result of external agreements between neutron/muon centres to promote common actively developed software tools. As a consequence, no real alternative to Mantid was envisaged during the course of this work-package.

A schematic representation of the work-package production is shown in Illustration 1. It can be seen that the initial learning phase took 6-9 months, including the production of deliverable D6.1 and the Mantid training. Easy Mantid items were completed first (continuous neutron source ToF and SANS raw import). The case of the 'moving' instruments was then studied, but the initial solution making use of groups of workspaces to hold scanned/iterative acquisitions did not work

properly. Effort was then redirected towards SANS reduction, back-scattering (indirect), and other developments (*AllToMantid, McStas import, reductionServer*). Actually, a usable, but not ideal, solution was found for some scanning instruments (e.g. continuous neutron source powder diffractometers) by storing data sets into a single multi-dimensional workspace. Finally, effort was devoted to writing example scripts to be used for ToF, and SANS continuous neutron source based instruments.

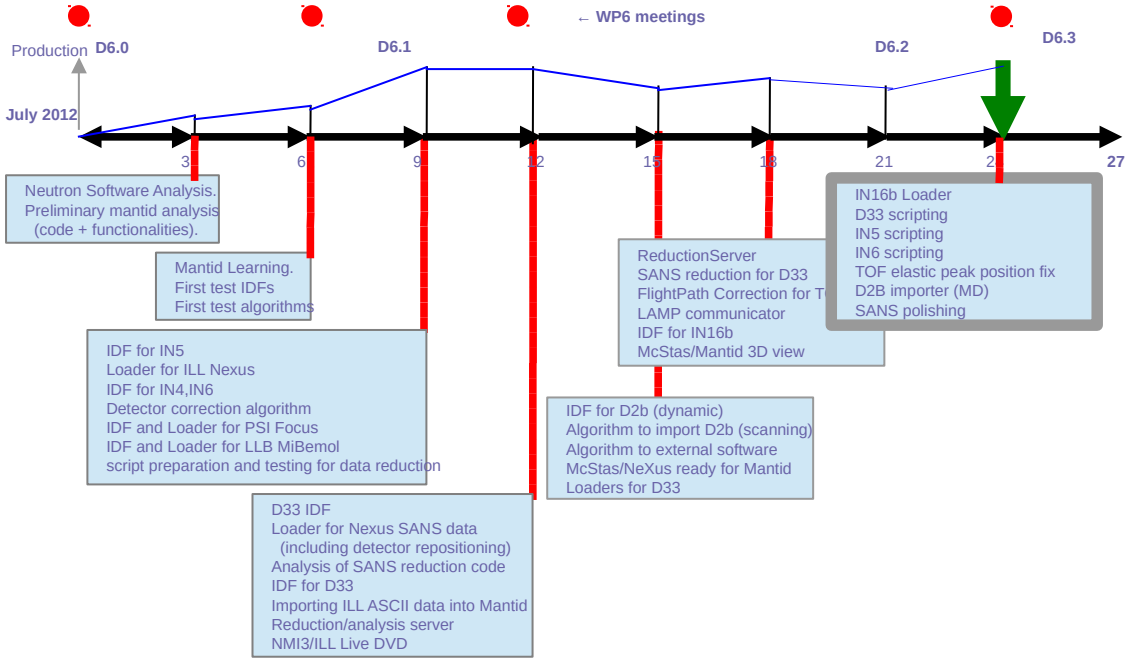

Illustration 1: Schematic representation of the WP6 production. Red dots indicate the meetings, whereas deliverable are indicated as D6.x items.

One of the key points which motivated the validation of Mantid was the ability to re-use the (q,ω) 4D data reduction routines, such as VATES (derived from Horace). The instruments which can benefit from these Mantid algorithms are TAS and large detector ToF spectrometers. Progress has been made in the support of continuous neutron source-based ToF spectrometers, but unfortunately there was not enough time to test VATES. We encourage the instrument responsibles of these spectrometers to validate this methodology and check that it can effectively be applied with their data sets.

Achievements: deliverables

In task 1 ***"Review existing data analysis software and practices of software developers"***, we found that there is an urgent necessity to identify code redundancy and potential points of failure in the existing software landscape and propose common low-level shared libraries for e.g. models, algorithms, I/O routines, interface design templates. These libraries should follow adopted standards in order to ease the interoperability. These standards are still to be defined. Old software often perform as well as recent ones in terms of performance/efficiency. Indeed, the underlying

mathematics and physics have not changed, and data volumes have not become problemetic in most cases. Only the programming techniques have evolved, and globally tend to complexify thus increasing the code size, and its maintenance cost. Old software may be used as long as it can be installed on modern architectures, highlighting the issue with unmaintained software and single points of failure. The Mantid project was identified as the largest actively developed software. In order to ease the Task 1, a LiveDVD was produced, which allows to run a Ubuntu system with pre-installed software, especially suited for system deployments, and tutorials/workshops.

In task 2 ***"Review existing solutions for a common data analysis infrastructure"***, we listed items which can help in producing high quality software:

- software repository,
- collaborative work,
- unit and integration testing,
- build servers,
- code review,
- technical and user documentation,
- easy multi platform package distribution,
- allow user contributions,
- use object oriented techniques,
- favour NeXus format for data storage.

Our recommendation is to make use of a community based infrastructure (e.g. on http://www.neutronsources.org, local SVN/Git, Redmine, Jenkins, Deb/RPM package repository) rather than commercial solutions, even apparently free, with a view to keeping full control of our knowledge, intellectual property and dissemination.

In task 3 "*Develop prototype software in chosen solution for representative applications*" we have contributed to the Mantid project with more than 200 commits, and about 17 000 lines of code (LOC). Two other prototypes have been contributed in order to allow the Mantid project to make use of other external contributions, without effectively coding Mantid algorithms. Since facilities tend to be converging on the NeXus format for data storage, all loaders (except one) developed in this work-package focus on that standard. Instruments t ILL, LLB and PSI have been included. The extension to other facilities/instruments should be facilitated when they adopt NeXus.

Contributed Mantid loaders:

- LoadILLAscii
- LoadILLIndirect
- LoadILL
- LoadILLReflectometry
- LoadLLB

- LoadSINQFocus (formerly LoadFOCUS)
- LoadILLSANS

Contributed Mantid algorithms:

- CalculateEfficiency
- ConvertEmptyToTof
- CorrectFlightPaths
- DetectorEfficiencyCorUser
- EQSANSDarkCurrentSubtraction
- EQSANSQ2D
- SANSAzimuthalAverage1D
- SANSBeamFinder
- SANSSensitivityCorrection
- SetupILLD33Reduction
- TransmissionUtils
- IDF_to_PLY (prototype)

Contributed Mantid instrument definitions:

- MIBEMOL_Definition.xml
- FOCUS_Definition.xml
- IN6_Definition.xml
- IN5_Definition.xml
- IN4_Definition.xml
- IN16_Definition.xml
- D2B_Definition.xml
- D17_Definition.xml
- D33_Definition.xml

Mantid Framework contributions:

- Geometry/src/Instrument.cpp
- Geometry/inc/MantidGeometry/Instrument/ParameterMap.h
- Geometry/src/Instrument/CompAssembly.cpp

Last, most algorithms are associated with test procedures, as well as example data files, as part of the Mantid coding standards. Other Python scripts have been written to generate the IDF for new instruments (ILL, LLB, SINQ) and visualise data sets independently of Mantid.

These contributions allow to test Mantid with a reduced but comprehensive set of continuous neutron source instruments, e.g. for ToF, Back-Scattering, SANS, Reflectometer, and partially diffractometer.

Other contributions:
In addition, a generic *AllToMantid* Algorithm was written to execute an external program, send it a Mantid workspace, process computing tasks, and retrieve the modified workspace back into Mantid. This prototype was tested with LAMP <http://www.ill.eu/?id=3463> and iFit <http://ifit.mccode.org/>, and proved to be effective.

As a continuation of this mechanism, a "reduction" server was initiated, which allows to request computational tasks. These are distributed to a number of computational resources (e.g. Mantid, LAMP, iFit, DAVE, …), and results are sent back upon completion to the requester. This prototype is simple to configure and maintain. It could be used to run data analysis from instrument control software allowing on-line data reduction and analysis.

Achievements: software

The software produced during the work-package is now available on <http://nmi3.eu/about-nmi3/networking/data-analysis-standards.html>. It comprises

1. the *LiveDVD* software evaluation system, with pre-installed software ;
2. the software repository <http://packages.mccode.org> which provides Debian packages for 25 neutron/muon software, as an example of solution aiming to ease the installation and distribution ;
3. the *reduction server* allowing to request data analysis tasks distantly, with any external computational code (5400 LOC) ;
4. the *All to Mantid* algorithm, which allows to transfer Mantid workspaces to other external software, e.g. LAMP and iFit, with associated scripts, and retrieve the result so that it appears as a regular Mantid algorithm (700 LOC);
5. the *Mantid loader algorithms* for continuous neutron source based instruments at ILL, PSI, and LLB. 29 Mantid tickets were treated during the workpackage, resulting in about 200 commits (17 kLOC).

Achievements: knowledge

In addition to the produced software, knowledge has been gained regarding the Mantid software, and its links with other projects.

In the scope of this work-package, we organised in June 2014 a training session on the Mantid software, for continuous neutron source scientists. The training presented both the current software, and focused towards algorithms usable for continuous neutron source based instruments.

Future of the data analysis standards work-package

Developing the Mantid project follows pragmatic constraints: the project exists, it is actively developed and maintained, and it already involves major neutron/muon facilities. In this context, most neutron/muon facilities in Europe have agreed to get involved in that project. This approach follows the same criteria as adopting NeXus for file formats, that is it gathers resources around a common infrastructure. This choice does not guaranty the best technical solution in all cases, but its ensures a wide spread dissemination of the knowledge, a common set of standards and understanding.

The future of the Data Analysis work-package is simple: the effort should be continued with a small developer team for a reduced cost. Indeed, the potential issue with Mantid is the overall cost related to the size of the project. A way to go further could be, for continuous neutron sources, to build a limited task force of three coders funded by e.g. ILL, FRM2 and PSI. This team would have to remain grouped, to build knowledge and expertise, but in order to serve these facilities, the team could be moved on a yearly basis around EU facilities. The task of the team would be to develop and maintain data importers for Mantid, as well as identify the Mantid algorithms to be used consecutively to ensure a continuous data work-flow.

Concerning NeXus, any new instrument adopting this format must use existing templates and nomenclature, and should certainly not build new data structures.

In the scope of this work-package, we plan to organise, within year, a workshop centred around data analysis. This workshop would present recent developments in the Mantid project, in a broader context of scientific software projects for data reduction and analysis.