A workflow manager for NMR: objectives, design and implementation

Friday, 21 September 2012 13:30 (22 minutes)

Description of the work

Nuclear Magnetic Resonance (NMR) is an important method for the analysis of macromolecular structure and dynamics. There are a host of different computational processes involved in the interpretation of NMR data. These traditionally make use of different programs, which store data in a wide range of formats. Inter-conversion between these formats is not trivial - the underlying information is extremely complex - and this has limited the ability to chain these processes together in an automated way. Our objective is to develop an application that will provide integrated access to the NMR processing programs presented by the WeNMR portals, allowing them to be combined into workflows in a simple and streamlined manner.

The CcpNmr Workflow Management System (WMS) uses the CCPN data model and API for data storage. Seamless conversion to and from program-specific formats is effected by separate Python modules, using the known input data to disambiguate the program output. For the actual calculations WeNMR web portals that present the relevant NMR analysis programs, are exposed as WSDL web services. These are accessed via a common web-based GUI, with individual protocol interfaces specified by templates. The basic architecture is a GWT client linked to a Java/Hibernate server deployed under Tomcat. The workflow management is carried out using Taverna.

Importantly, WMS supports the systematic tracking and management of the data itself. NMR data analysis involves many different processes with subtly different input data combined either sequentially or as alternatives. WMS keeps a record of all the various data versions and helps scientists keep track of their work - something which a survey of potential users at the start of the project identified as a significant unmet need.

As a case study, we present a workflow that takes a single set of input data, sends it to several different WeNMR structure calculation portals, and integrates the results for subsequent analysis.

Wider impact of this work

WMS is expected to; (i) improve accessibility of the WeNMR services, and (ii) provide a framework for the development of novel workflows to address scientific problems. As structural biology becomes more established, there will be ever increasing demand from non-specialist users for simple, automated tools that provide reliable results. The establishment of standardized user interfaces and seamless data transfer should be very helpful in this regard.

In addition, the WMS should be extremely useful for managing the large amounts of data generated in NMR spectroscopy. We expect spectroscopists to be relatively open to the use of LIMS tools like WMS, given that NMR data processing is already heavily computerized. The direct integration with the programs that are already in common use in the field should make WMS compatible with established work practices without the need for additional data capture steps, thus minimizing the demands on the user and ultimately increasing take-up.

Printable Summary

As part of the WeNMR project, we are developing a workflow and data management system targeted at scientists using nuclear magnetic resonance spectroscopy (NMR). This "end-user platform" will support seamless task execution and data flow across the WeNMR portals for NMR-related calculations, as well as our CcpNmr analysis suite. NMR is a key method for investigating the structure of biological macromolecules, involving both manual and computational steps. For the computational steps there is a growing desire to use distributed computing infrastructures, as the popularity of the WeNMR VO testifies. However, combining grid-based computational tools into more complex workflows presents significant technical challenges, particularly in terms

of format inter-conversion. The prize at the end is a framework for the development of novel, increasingly automated workflows for NMR data analysis. In addition, it will provide improved laboratory data management and simpler access to WeNMR services.

Primary author: Dr FOGH, Rasmus (University of Cambridge)

Co-authors: Prof. LAUE, Ernest (University of Cambridge); Dr IONIDES, Jonn (University of Cambridge); Dr LUNDBORG, Magnus (University of Cambridge); Dr STEVENS, Timothy (University of Cambridge); Dr BOUCHER, Wayne (University of Cambridge)

Presenter: Dr FOGH, Rasmus (University of Cambridge)

Session Classification: Workflow community workshop

Track Classification: Virtual Research Environments (Gergely Sipos: track leader)