Appendix 1: R4L Scenario

A chemistry researcher proposes a strategy for synthesising a new chemical compound and devises an experimental plan that complies with COSHH regulations. Using the SmartTea system the researcher outlines the synthetic strategy, including ratios of all reagents and solvents involved and the methodology to be used and submits it to the system, along with metadata relating to proposed identifiers and workers involved. When the researcher is ready to commence the synthesis the plan is recalled on the laboratory tablet PC and the system prompts the laboratory worker to measure out the required reagents and solvents in the predetermined order. The measurements are performed in the Smart lab environment, where actual values for the amounts of reagents employed are seamlessly recorded and archived for future retrieval. On completion of the reaction, once separation and purification processes have been performed, the volume or mass of product are recorded and if necessary the sample is crystallised into a solid form, the mass of which is recorded.

For both publication purposes and investigation of the properties of the new material a thorough characterisation must be performed. The research team decide to perform infra-red spectroscopy, mass spectrometry, single crystal diffraction and theoretical quantum mechanical calculations. An account for the sample is generated in the R4L laboratory data management and archival software and metadata core to all processes is generated, captured and deposited. This metadata principally comprises a (proposed) chemical identifier and the research workers involved in the study (which sets access permissions).

Infra-red spectroscopy is a technique that can easily be performed by any researcher (after brief basic training) on a desktop instrument controlled by a PC running proprietary software in a matter of minutes with virtually no post collection data correction or workup involved. The researcher prepares the sample and initiates the R4L software on the controller PC, opens the account for the sample in question and selects infra-red spectroscopy as the analytical experiment about to be undertaken. At this point, time-stamping metadata are generated and submitted to the prior assertion service. Metadata regarding the experiment are then captured and primarily include instrument (manufacturer and model) and software (including version) used and the researcher conducting the experiment. The sample is then loaded on the instrument and the proprietary software initiated. A spectrum is acquired and saved in native software format. The spectrum is then 'saved as' a file in a common exchange format (plain text / XML) and a file containing metadata on the operational parameters of the instrument during the course of the measurement is generated. The data capture service is then invoked from within the R4L software and the files pertaining to the experiment are deposited, along with the necessary metadata, in the laboratory repository.

Mass spectrometry analytical experiments are performed as a service for many researchers and the sample is to be submitted to such a facility. The mass spectrometry experiment requires a trained technician to perform decision making prior to the experiment,

which may then be easily and rapidly performed with little or no post collection correction or work up of the data. A sample of the material is submitted to the service, labelled with the name of the originator and the chemical identifier. Metadata regarding the sample are also provided to enable the service to decide on the most appropriate technique for the analysis. In the R4L system the originator of the sample delegates responsibility for the mass spectrometry measurement to the service. When the sample is scheduled for measurement the service technician initiates the R4L software and selects mass spectrometry as the experiment to be undertaken. At this point, time-stamping metadata are generated and submitted to the prior assertion service. Metadata regarding the experiment are then captured and primarily include instrument (manufacturer and model) and software (including version) used and the technician conducting the experiment. The sample is then loaded on the instrument and the proprietary software initiated. A spectrum is acquired and saved in native software format. The spectrum is then 'saved as' a file in a common exchange format (plain text / XML) and a file containing metadata on the operational parameters of the instrument during the course of the measurement is generated. The data capture service is then invoked from within the R4L software and the files pertaining to the experiment are deposited, along with the necessary metadata, in the laboratory repository.

Single crystal diffraction analysis requires decision making at numerous stages during the experiment and a lengthy data collection process, with detailed post collection data correction and work-up. The process must be performed by personnel trained in the field over a significant period of time. When the sample is scheduled for measurement the service technician initiates the R4L software and selects single crystal diffraction as the experiment to be undertaken. At this point, time-stamping metadata are generated and submitted to the prior assertion service. A suitable specimen is selected from the sample and digital images of both the sample and specimen are recorded. The specimen is then loaded on the instrument and the proprietary software initiated. Preliminary scans (binary format files) are recorded to assess the quality and suitability of the specimen. Further scans are then recorded, decisions made and parameters calculated for the scan strategy of the data collection. Data are collected, corrected, processed and reduced to a format suitable for the researcher to download and work up. The researcher downloads the reduced data onto an office PC and performs the process of working it up (solution, refinement and report preparation). When data work up is complete the data capture service is then invoked from within the R4L software and the files pertaining to the experiment are deposited, along with the necessary metadata, in the laboratory repository, whilst the raw data (binary files) are sent to an off site magnetic tape store for archival and curation.

The researcher also wishes to perform in-silico theoretical calculations to determine some of the properties of the structure of the compound. The researcher initiates the R4L software and selects theoretical calculation as the study to be undertaken and at this point, time-stamping metadata are generated and submitted to the prior assertion service. The result of the single crystal structure determination is used as the initial starting point for the

study and the process of geometry optimisation and then property calculation is undertaken. When complete, the data capture service is then invoked from within the R4L software and the files pertaining to the experiment are deposited, along with the necessary metadata, in the laboratory repository.

When all investigations are complete the report generation tool is initiated and the researcher selects the identifier for the compound under study from the list of 'active' compounds in their R4L account. The researcher is presented with a list of all studies performed for this particular identifier / compound and may select which are to be included in the report. The researcher selects, synthesis, infra-red spectroscopy, mass spectrometry, single crystal diffraction and theoretical calculations and the R4L interface presents the results for each different dataset in the study in turn. For each technique the researcher is presented with the data via an interactive interface and may select the components to be included in the final report. After processing each selected dataset for publication the R4L software automatically generates a full report in standard journal format. When the interpretations of the full study are to be submitted as a paper to a learned society journal the experimental data report is deposited in an institutional repository from the R4L software. Links to the underlying datasets in the laboratory repository are generated and enabled from the IR.