An Open Source Informatics System for Combinatorial Materials Research*

Wenhua Zhang, Michael J. Fasolka, Alamgir Karim and Eric J. Amis

Polymers Division, National Institute of Standards and Technology, Gaithersburg MD 20899
NIST Combinatorial Methods Center (NCMC) (www.nist.gov/combi)

INTRODUCTION

Advanced materials development involves the exploration of large, multi-dimensional parameter spaces and the correlation of material/chemical properties and processing parameters derived from disparate sources. To position themselves better in this complex arena, many chemical and materials researchers have turned to combinatorial and high-throughput (C&HT) approaches, which rely on a robust laboratory research informatics system (LRIS). The "nervous system" of a combinatorial research lab, an LRIS integrates an experiment-oriented database, automated fabrication/characterization instruments, and software tools for analysis and data-mining functions. However, an LRIS is complex and difficult to build, and guiding examples for tailoring existing informatics tools for materials studies are sparse. Accordingly, resources that can assist in the development of LRIS infrastructure (and its effective deployment) are necessary for the advancement of C&HT approaches in the materials research arena. In this talk, we describe an open-source LRIS for materials research that is being built at NIST.

THE NCMC LRIS PROJECT

The NIST Combinatorial Methods Center (NCMC, see http://www.nist.gov/combi) aims to lower barriers to the use of C&HT techniques by developing library fabrication devices and high-throughput characterization/analysis tools geared towards chemical and materials research. As part of this mission, NCMC scientists are building a LRIS system in the combinatorial materials research facilities in the Polymers Division at NIST. The NCMC LRIS provides 1) a central database that is specifically geared to accept materials research data from a variety of sources, and that is structured for scientific aims, and 2) a versatile scientific functionality produced by interfacing the central database with instrumentation and data analysis/datamining tools. While this system is useful for NCMC materials research, the primary purpose of the NCMC LRIS is to serve as a model system that demonstrates and tests strategies for informatics infrastructure development. Accordingly, the NCMC LRIS is being built using two guiding principles:

• A *commitment to open-source code.* Open-source software tools offer informatics structures that can be freely used and modified.

• *Transparent Design.* Reasonably trained personnel will be able to easily understand the structure and function of the system.

During its construction, the NCMC informatics system provides working examples of LRIS infrastructure that can be used directly by other parties, easily modified to meet their needs, or (due its transparent structure) compared to existing custom and commercial software. This enables feedback on the system design, and may suggest pertinent avenues for the establishment of informatics standards (interchange formats, etc.) that can smooth LRIS development.[1]

NCMC LRIS DEPLOYMENT

Database System. The core of the NCMC LRIS is a centralized database system that accepts data generated by each instrument and analysis tool, and organizes this data in a way amenable to materials research. Collected information includes fabrication and processing conditions, characterization data (thickness, composition, surface energy etc), measurement data (e.g., micrographs and spectrographs) and results from analysis routines. The NCMC database is built upon the open-source PostgreSQL[2,3], an object-relational database management system that provides suitable performance, speed, and flexibility regarding extension. PostgreSQL accepts most SQL (Structured Query Language) database queries.

User/Instrument and Instrument/Database Interfaces. The NCMC LRIS employs the open-source Python[4] programming language for building user interfaces, instrument/database interfaces, and instrument automation tools. In the NCMC project, the most difficult programming involves the generation of meta-routines to attach instrument control programs and data analysis software to the centralized database. Instrument control and database integration can be difficult, since instrument software rarely includes a suitable database interface, and since most laboratories incorporate both custom devices and instruments supplied from a variety of vendors. During the development of the NCMC LRIS, three basic interface scenarios were encountered:

1) *No instrument software exists.* Custom-made and older instruments often require custom-built control and database interface software. In the NCMC LRIS, these tools are constructed using Python.

2) *Commercial instrument software is available and includes an ActiveX (or dde) component.* In this case, a custom database interface embeds ActiveX controls in its function. For example, optical microscopes in the NCMC labs communicate with the database using Python, while micrograph acquisition is achieved through ActiveX controls provided by the manufacturer.

3) *Commercial control software available, but no ActiveX or dde control is available.* In this case commercial software directs the measurements, while custom-built software is run in tandem. The custom routine checks the commercial software's environment for new files, which are automatically processed and sent to the database when detected. For example, in the case of the NCMC Atomic Force Microscope (AFM), a custom interface continually checks for new files generated during an experiment. When new AFM data becomes available, it is parsed and a representational image (e.g., a .tif file) is generated. Then, the image, the original data and experiment parameters are sent to the database.

Indeed, strategies for meeting these and other challenges are useful to industrial and academic materials researchers engaged in LRIS development.

Next Steps. Now that the basic infrastructure is in place, Phase II of the NCMC LRIS project involves 1) the development of an "expert system" around each piece of laboratory instrumentation and 2) the production of coordinated activity among instrument sub-sets. In the first effort, parameters for instrument operation are collected over time and this data is used to automatically "suggest" best parameters when new applications arise. The second effort will produce a new, more sophisticated level of automation in NCMC laboratories. As with the previous NCMC infrastructure, these tools will be built so they can serve as useful examples for a wide range of LRIS developers.

ACKNOWLEDGEMENTS

The authors thank the members of the NCMC (www.nist.gov/combi) for useful feedback on this project. Thanks also to Barbara Guttman, John Liu (Information Technology Laboratory, NIST) and Mikk Lippma (University of Tokyo) for useful discussions.

REFERENCES

1. See http://www.nist.gov/combi/cmrds for more information.
2. Certain software is identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the software identified is necessarily the best available for the purpose.
3. http://www.postgresql.org
4. http://www.python.org