J. Chem. Inf. Model. **2006,** *46,* 1017−1025

Instrument Monitoring, Data Sharing, and Archiving Using Common Instrument Middleware Architecture (CIMA)

Randall Bramley,† Kenneth Chiu,‡ Tharaka Devadithya,† Nisha Gupta,§ Charles Hart,§ John C. Huffman,∥ Kianosh Huffman,*,∥ Yu Ma,† and Donald F. McMullen§

Department of Computer Science, Indiana University, 215 Lindley Hall, 150 S. Woodlawn, Bloomington, Indiana 47405, Department of Computer Science, State University of New York at Binghamton (SUNY), P.O. Box 6000, Binghamton, New York 13902, The Knowledge Acquisition and Projection Lab of The Pervasive Technology Labs at Indiana University, 501 N. Morton St. Bloomington, Indiana 47404, and School of Informatics, Indiana University, 901 E. 10th St., Bloomington, Indiana 47408

Received September 1, 2005

The Common Instrument Middleware Architecture (CIMA) aims at Grid-enabling a wide range of scientific instruments and sensors to enable easy access to and sharing and storage of data produced by these instruments and sensors. This paper describes the implementation of CIMA applied to the field of single-crystal X-ray crystallography. To allow the researchers to easily view the current and past data streams from the instruments or sensors in a laboratory, a crystallography portal and associated portlets were developed for this application. The CIMA-based crystallography system provides an opportunity for anyone with Web access to observe and use crystallographic and other data from laboratories that previously had only limited access.

INTRODUCTION

An aspect of Grid computing[1−3] that has not been well-developed is the integration of scientific instruments into a Grid computing environment. Instruments are still largely “off-line” as far as downstream software analytical components are concerned, and instruments are not at all first class members of the Grid with respect to location, allocation, and access control. This is a serious problem because three issues continue to grow in importance in research: (1) investments in geographically extended (e.g., international) collaborations organized around large shared instrument resources, (2) increasingly “real time” use of instruments by remote researchers both for “first look” activities and pipelined data acquisition and reduction, and (3) sensor networks with hundreds to thousands of nodes being deployed.

Grid computing has proven to be a useful paradigm for organizing and harnessing distributed resources and for providing common services such as scheduling and authentication. Considerable work[4,5] has been done in the areas of computational and storage grids and on combining computing and storage resources. Another important development in scientific computing is the application of component software technologies[6,7] that promise to enhance code modularity, encapsulation, and reuse. At the confluence of component software engineering and Grid computing is Web Services[8] and a Globus[9]-centric implementation, Open Grid Services Architecture (OGSA; note that a list of abbreviations and acronyms is given at the end of this paper).[10,11] Web Services provides a platform-independent, location-independent means to execute code through a remote procedure call or document-oriented interface. Progress has been made[11,12] in developing and evaluating OGSA and in using Web Services as a model for software component interfaces.[13]

The Instrument Middleware Project[14] supported by the National Science Foundation Middleware Initiative[15] is developing tools and methods for representing instruments and sensors as network services. While other Grid projects address allocation and scheduling, the Common Instrument Middleware Architecture (CIMA) is targeting to improve control and access to instruments and to facilitate their integration into the Grid.

Remote access to and control of instrumentation is not a new topic. There are myriad active projects to develop remote access capabilities for specific instruments, each with a unique solution.[16] What distinguishes this work is the goal of developing a general, reusable methodology for remote access. CIMA provides remote instrument control capabilities, but they are not being used in the application described in this paper. Two other general approaches to remote access are discussed in the section below as a basis for understanding the contribution CIMA makes to the field.

NEESdaq. The Network for Earthquake Engineering Simulation (NEES)[17] program is developing remotely accessible shake tables for earthquake simulation in civil engineering. This package is a set of plug-ins supporting NEES-specific hardware for the commercial LabView[18] data acquisition software line from National Instruments. Data transport is via the NASA-developed Data Turbine[19] publish−subscribe software, now commercialized by Creare. As available, this software allows remote users to save data streams from LabView-supported data acquisition boards to their local computers and to plot recently acquired data streams. Grid capability consists of using X.509 certificates for client authorization to access data streams provided by

* Corresponding author phone: +1-(812)-855 6742; e-mail: kihuffma@indiana.edu.
† Department of Computer Science, Indiana University.
‡ Department of Computer Science, SUNY.
§ Pervasive Technology Labs, Indiana University.
∥ School of Informatics, Indiana University.

10.1021/ci050368l CCC: $33.50
Published on Web 03/24/2006

the underlying Data Turbine transport (not data acquisition). NEESdaq makes no claim to general applicability, though the commercial LabView software on which it is based has been used for some time in a wide range of lab and industrial settings.

IEEE P1451.[20] This is a set of proposed standards and evaluation implementations for "smart" sensors, or sensors that have an associated controller and communications interface. These standards cover how a transducer/controller unit should be accessed via point-to-point and bus architectures and how the data from a transducer should be framed for different types of transport (serial, current loop, wireless 802.11 and 802.15.4, Ethernet, etc.). Security in IEEE-1451 consists of defining three roles for the reader, writer, and administrator of a sensor node. In addition to defining "wire" protocols for interacting with sensors and transducers, IEEE-1451 also includes a specification for an "electronic data sheet", which describes sensor output in terms of standard and derived units. Programs that interact with an IEEE-1451 controller can query this memory to determine the units of the data provided by sensors attached to the controller.

The CIMA approach is similar to that of IEEE-1451 except that transport independence is achieved through the Web Services Definition Language (WSDL; see below in Network Transport section) and a richer descriptive capability based on Semantic Web[21] technologies. In addition to industry standard network transport [Web Services via Simple Object Access Protocol (SOAP)[22] over HTTP/s], CIMA leverages the following additional capabilities: location [services can be registered to a Universal Description, Discovery and Integration (UDDI) service directory], line and message security through HyperText Transfer Protocol (HTTP) and SOAP message encryption, an extended descriptive language for both instrument capabilities and signal units, and an open approach to authentication and authorization. Control and data acquisition operations in CIMA are specified in *Parcel* messages sent to and from CIMA implementations. Details of the CIMA Parcel protocol are described elsewhere.[23,24]

The CIMA implementation is currently being evaluated in several settings representing a spectrum of shared instrument applications including X-ray crystallography, astronomy robotic telescopy, and small sensor network nodes based on Berkeley mote wireless sensors. The end product will be a consistent and reusable framework for including shared instrument resources in Globus-based grids.

This paper describes the successful application of CIMA in the field of single-crystal X-ray crystallography. For a more detailed description of the CIMA architecture and design considerations, please refer to refs 23 and 24.

The reasons for implementing CIMA in this particular field were mainly that X-ray diffraction instruments are fairly expensive to purchase, the operation of these instruments requires a highly trained operator, and these instruments are frequently only available at larger universities or colleges and the national synchrotron facilities. Hence, improving the accessibility to this type of instrument means that a larger group of users, including researchers and students, especially at smaller colleges and institutions, can utilize data from these instruments.

Additionally, during the past decades, the development of instrumentation in the field of single-crystal X-ray diffraction has significantly increased both the quality and quantity of data, especially in the field of X-ray detector technology. The CCD-based detectors and Imaging Plate detectors[25] are now common in most X-ray laboratories. Although diffraction equipment has become modernized and is capable of generating several gigabytes of raw data daily, the data collection and archival procedures have not advanced to the same extent. Crystallographers are collecting data on these modern instruments using essentially the same routines that they were decades ago by transferring the data off the instrument data acquisition machine to other "postprocessing" facilities and by manually archiving onto other media such as CD or DVD. Therefore, in addition to improving accessibility, a prime target for the CIMA crystallography application was to modernize the data management and storage methodology that is commonly in practice in many X-ray laboratories. By properly designing automated data storage and archiving functions, researchers not only gain a significant savings in time but they also guarantee that no data coming off the instrument can be lost because of accidental mishaps in the manual backup procedure.

Several laboratories are currently working to establish a crystallographic informatics program (Crystal Grid[26]). This international effort is a project to make crystallographic data (including raw data, structural results, and final publications) readily accessible, not only for members of the research community but also for educational and other uses. This will help to accommodate the issues associated with the increasing volume of raw data produced by modern diffraction instruments as well as the increasing number of structures and publications that the crystallographic community is creating. Our team plans to have a significant role in this project, because the goals are critical to a widespread CIMA crystallography implementation. One of the most important items on the Crystal Grid schedule is the establishment of national centers and archives for crystallographic raw data and publications. We are working with Indiana University officials to establish a central archive for CIMA-participating laboratories who would like to store the data from their instruments at a centrally located system. The storage would be readily available via the Internet, so it would be available for sharing to colleagues at other sites and be available for instructional use as well.

CIMA Crystallography Software Architecture. CIMA crystallography application consists of these main components: *Instrument Representative*, *My Manage*r, the *Portal*, and *Data Manager Service* (Figure 1).

At the instrument or laboratory site, there is the *Instrument Representative* component. This component is responsible for representing the instrument or sensor and consists of code to implement a Web Services interface for a group of instruments or sensors at a particular site and protocols for providing individual or multiple readings from a particular data source. Readings are continuously returned by the instrument as a series of SOAP[22] messages after a client registers with the registration service. The client provides a Web Services endpoint to which multiple readings can be returned. The provided Web Services are implemented using gSOAP[27] and use HTTP as the transport protocol.

Each instrument or sensor has a set of zero or more *channels*, shown as dashed lines between *Instrument Representative* and *My Manage*r in Figure 1. The *channels* consist of a set of Web Services definitions that could be

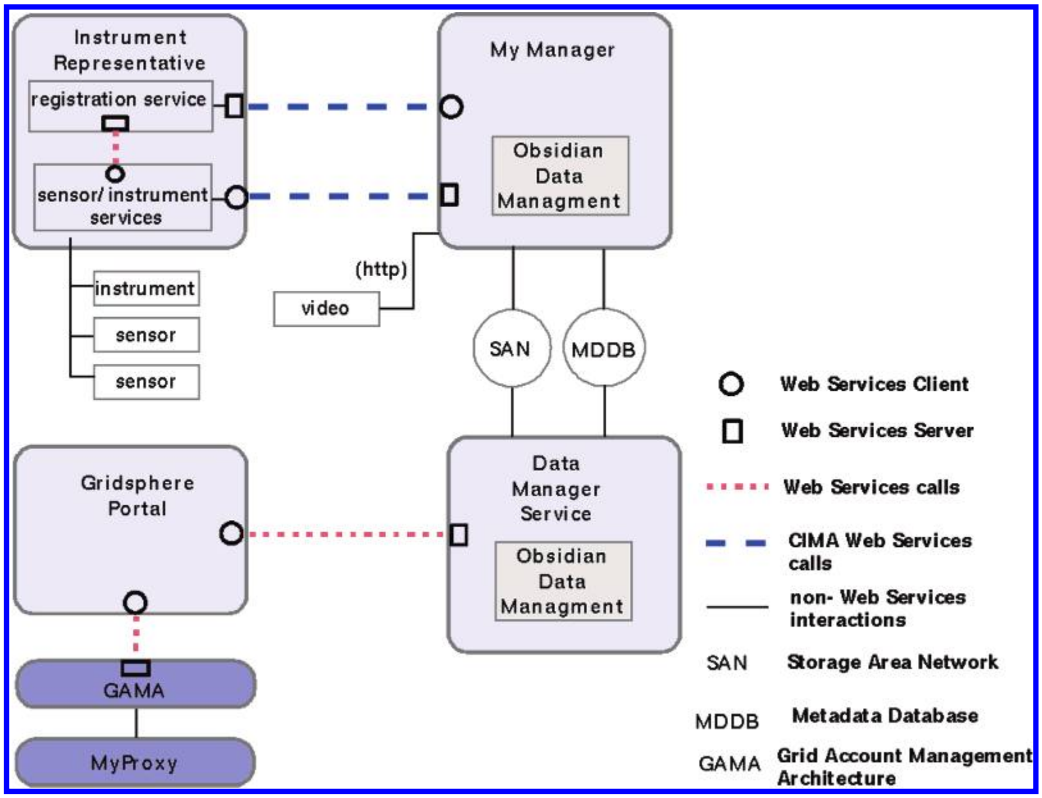

Figure 1. Overview of the main CIMA components.

used independently by other applications. *Channels* provide bidirectional communication between the *Instrument Representative* and user/consumer applications. The messages to the sensor or instrument would be the registration and unregistration requests, while the messages from them would be events or sensor readings. *Channels* support both push and pull mechanisms. The detection of new data from the actual instrument or sensor is monitored by the "*instrument*" or "*sensor*" services components, and their events or readings are communicated from the *channel source* (sensor end) to the *channel sink* (consumer end) through the *channel* and delivered to the *My Manager* component. In addition to the data, the *channel source* can also send status, which tells the *channel sink* the status it should assume, and experiment information. The latter tells the *channel sink* that the data being provided corresponds to a new data collection session and to end the previous data collection. A *channel sink* receives data from the *channel source* using the *push* method as opposed to the *pull* method. The difference is that the push method does not ask for data from the source; instead, it only receives the data whenever the source sends it. The data could be a binary file or a value (for example, a double precision number) of an environment variable and associated timestamp provided by a sensor.

One or more *My Manager* components can exist on the network. Each can manage multiple instruments and sensors. *My Manager* handles all further processing of incoming data (e.g., where to store data, determining the type of CCD frame format, conversion to other formats, etc.). The *My Manager* component is also responsible for invoking the calls to the Obsidian Data Management Library.[28] Because each instrument can have its own CIMA services provider, scaling is limited only by bandwidth or the CIMA client capabilities.

In the initial stages of CIMA development, we recognized the need for a powerful data management system (DMS) that can easily store and manage the large amount of data and their associated metadata that the instrument's CIMA Web Services would be capable of producing. In addition to storing and organizing data streams, the DMS is necessary so users and other analysis software can easily locate information. For example, X-ray crystallographers would like to know many of the environmental values associated with a CCD frame at the time it was recorded, or over a time range. The Obsidian Data Manager Library handles the data streams from the CIMA Web Services by storing either the actual values or file (e.g., the CCD frame) locations of the data and associated metadata in a database. These values can be retrieved by calls using the Obsidian application program interface. Additionally, the CIMA system integrated with the data manager library can keep track of multiple versions of the same file; for example, if the crystallographer at the instrument deletes or overwrites a file, a copy of the original file is always retained in the CIMA database. Obsidian will be capable of providing an interface for long-term storage solutions, such as the High-Performance Storage System (HPSS[29,30]), to facilitate archiving of the raw CCD data and the associated metadata. Data from the various sources (camera images, environmental values, and CCD frames all with their associated time stamps) are combined into a single package for archiving and easy retrieval of all the data associated with an experiment.

In passing, it should be mentioned that the video and still images collected from the crystal microscope and lab cameras during an experiment are not acquired from the CIMA instance. Video is provided by Web-accessible cameras. Single frames are grabbed periodically using the Linux *wget*

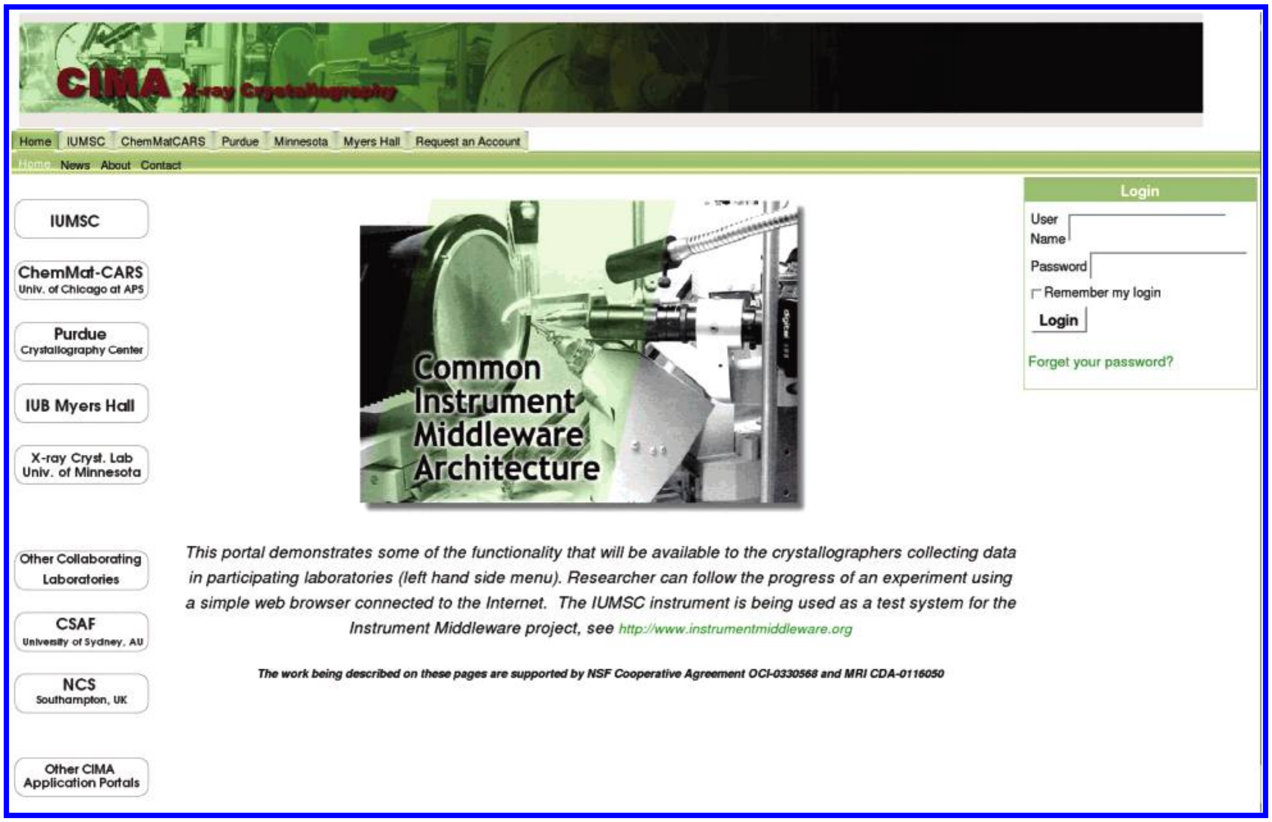

Figure 2. CIMA Crystallography Portal.

command and stored with the other experimental data by the *My Manager* process.

Network Transport. CIMA uses Web Services as a means of communicating with instruments and sensors. The Web Services specification does not provide or even recommend a network transport, instead leaving the choice to the implementor. WSDL provides a means of indicating which transport or transports are bound to a service at run time. Applications which intend to use a service can query the WSDL for that service to find the location and the transport as a *Uniform Resource Identifier* (URI) needed to connect to the service. The typical protocol used for Web Services is HTTP over the Transmission Control Protocol (TCP), but other types can be used as well. For example, SOAP messages can be sent by the Simple Mail Transport Protocol (SMTP) used for e-mail and by publish/subscribe systems such as Java Message Service (JMS)[31] and NaradaBrokering.[32] As a relatively extreme example of transport neutrality, we have implemented SOAP using the Antelope[33] system, a network ring buffer used in the seismology community to aggregate and transport data from ground motion sensors.

Line-level security for messages between *Instrument Representative* and *My Manager* can be provided through SSL (HTTPS; SSL = Secure Socket Layer). Authentication is not supported directly by the instrument in this implementation but is handled by authentication and authorization services provided by the portal component described below. Direct authorization and authentication by the CIMA instance at the instrument is in development. Because procedures in the lab are driven by the availability of technical and scientific support, scheduling of the instruments is "manual". Users can be notified when their samples are mounted and can, at that point, observe the data acquisition remotely and interact with on-site personnel using video or audio conferencing. Our current implementations of the *Instrument Representative* and *My Manager* components are written in C++ and use the gSOAP[27] library.

Probably the most visible component of CIMA crystallography application is the Crystallography Portal[34] component, Figure 2. At the time of writing, we have completed a transition between a Jetspeed-1[35]-based portal framework to a Gridsphere[36−38] portal framework. The CIMA Crystallography Portal and associated portlets were developed to allow the crystallographers to easily view the current and past data streams from the instrument. The portal allows researchers at each CIMA-participating laboratory to observe real-time data coming off the instrument and the sensors in the laboratory, as well as previously collected data. Each instrument portal can access data stored by multiple *My Managers*.

An active area of research is the integration of security services provided by Open Grid Computing Environments Collaboratory (OGCE)[39] portals with applications that run through these portals and with instruments or sensors that users may want to interact with through the portal.

Future Work. Another goal of the CIMA crystallography project is to acquire and provide data in the proposed Image Crystallographic Information File/Crystallographic Binary File (img-CIF/CBF) standard[40] format for crystallographic raw data. Within the CIMA-enabled crystallography system, we not only store the raw data in the native instrument format (various manufacturers' CCD or Imaging Plate format) but we also plan to provide a mechanism to convert raw data formats to the img-CIF/CBF standard format. This facilitates the use of data because users can read the raw data frames independent of their native format. By providing a standard

Table 1. CIMA Data Channels for the Indiana University Molecular Structure Center Laboratory

instrument/sensor	model	manufacturer	output quantity	derived quantity	frequency
CCD detector	SMART 6000	Bruker AXS Inc.	Bruker frame format	Bruker frame format	every 1−120 s
CCD chip temp.	from Bruker frame header	Bruker AXS Inc.	Celsius × 100	Celsius	every 1−120 s
crystal temp.	iSeries controller, built-in HTTP server	Omega[47]	HTML	Celsius	every 60 s
Bay1 room temp. & humidity	probe EI-1050	LabJack	Celsius & rel. humidity	Celsius & rel. humidity	every 60 s
DE temp. & humidity	probe EI-1050	LabJack	Celsius & rel. humidity	Celsius & rel. humidity	every 60 s
campus chill water (in)	probe EI-1022	Labjack	voltage	Celsius	every 60 s
lab chill water (in)	probe EI-1022	Labjack	voltage	Celsius	every 60 s
lab chill water (out)	probe EI-1022	Labjack	voltage	Celsius	every 60 s
liquid nitrogen scale	thinner	source: Bed, Bath & Beyond	analogue output from elec. scale connected into LabJack (voltage)	level in %	every 60 s

conversion interface, such as an img-CIF/CBF interface, the user is no longer obligated to use a particular instrument manufacturer's proprietary software. This standard interface also facilitates collaborations between groups that have different instruments or versions of software.

Instrument Description. One of the operations on a CIMA instance returns a description of the instrument as a Resource Description Framework (RDF) and Extensible Markup Language (XML). The descriptions are based on an OWL-DL,[41] a sublanguage of OWL (Web Ontology Language), ontology and provide information about the observables and their units and other characteristics of an instrument available through a CIMA instance. Although this function is not active in the implementation described here, it will be included in future versions. This is primarily because much of the relevant metadata for an experiment is included in the headers of frames output by the area detector and must be extracted from the frames after they are acquired.

CIMA Crystallography Test Site. The Indiana University Molecular Structure Center[42] (IUMSC) is a well-known service and research facility located in the Indiana University Department of Chemistry. Established in 1974, the laboratory specializes in the determination of the structures of crystalline materials via the techniques of X-ray diffraction and scattering. The primary emphasis is on "small molecules" that can be studied at an atomic resolution; however, the IUSMC is also developing a capability to examine macromolecular compounds and other materials in the crystalline state, as well as a variety of X-ray scattering techniques. Specialized techniques have been developed to readily handle materials that are unstable under ambient conditions. Nearly all experimental data are collected at ca. 100 K using a locally developed cryostat. IUMSC has three full-time employed crystallographers, each with many years of experience in the fields of chemistry and crystallography. The majority of the samples submitted to IUMSC are from the researchers in Department of Chemistry; although, frequently, the IUMSC scientific staff collaborate with and assist researchers from other departments on the Indiana University Bloomington campus as well as laboratories outside the I.U. Bloomington campus. The X-ray diffraction equipment available in IUMSC consists of several different types of single-crystal diffractometers and one powder diffractometer.

The major instrument in this laboratory is a Bruker-AXS SMART 6000 CCD area detector system[43] with a molybdenum (Mo) sealed X-ray tube, purchased and installed in 1997. The Bruker instrument is capable of collecting a complete data set in less than 1 day, with each set being as much as 4 gigabytes of raw data.

Besides the X-ray diffraction equipment in the laboratory, the IUMSC also has several types of microscopes, crystal mounting equipment, audio and video collaboration equipment, Web-accessible cameras, and access to advanced computing facilities such as a Beowulf cluster for rendering high-resolution ray-traced images and the Indiana University Analysis and Visualization of Instrument-Driven Data (AVIDD) supercomputing facility.[44] Through the Service Crystallography at Advanced Photon Source (SCraps)[45] project, the IUMSC staff can also utilize the ChemMatCARS' beamline[46] in the Advanced Photon Source, Argonne National Laboratory, for samples that require the high-brilliance X-rays which cannot be obtained by using the in-house X-ray generators.

For the CIMA crystallography application, our team chose to integrate the Bruker SMART 6000 diffractometer instrument and the various meta data (Table 1) associated with this instrument into our system. The requirements for CIMA data channels in IUMSC were identified to be as follows (Table 1):

• Bruker SMART6000 CCD frames were collected every 1−120 s depending on the crystalline sample being examined.

• Environmental observations (ambient temperature and humidity, instrument enclosure temperature and humidity, and X-ray generator chill water in/out temperature) were taken every 60 s.

• Technical observations such as liquid nitrogen level and CCD back plate temperature were taken every 60 s and every 1−120 s, respectively.

The Bruker SMART 6000 CCD instrument control and data acquisition software (SMART v 5.628) is proprietary software from Bruker AXS Inc. The captured CCD image from each diffraction observation is written to a disk as a proprietary Bruker SMART format frame. The number of frames collected during a typical diffraction experiment is 2000−4000 frames, each ranging from ~512 KB (for 512 × 512 pixels) up to ~8 MB (2048 × 2048 pixels) depending on the resolution setting of the instrument. An important design criterion for the CIMA project was to not modify the computer system supplied with the Bruker diffractometer instrument. To be able to access the CCD frames collected on the Bruker SMART 6000, we mirrored the file system

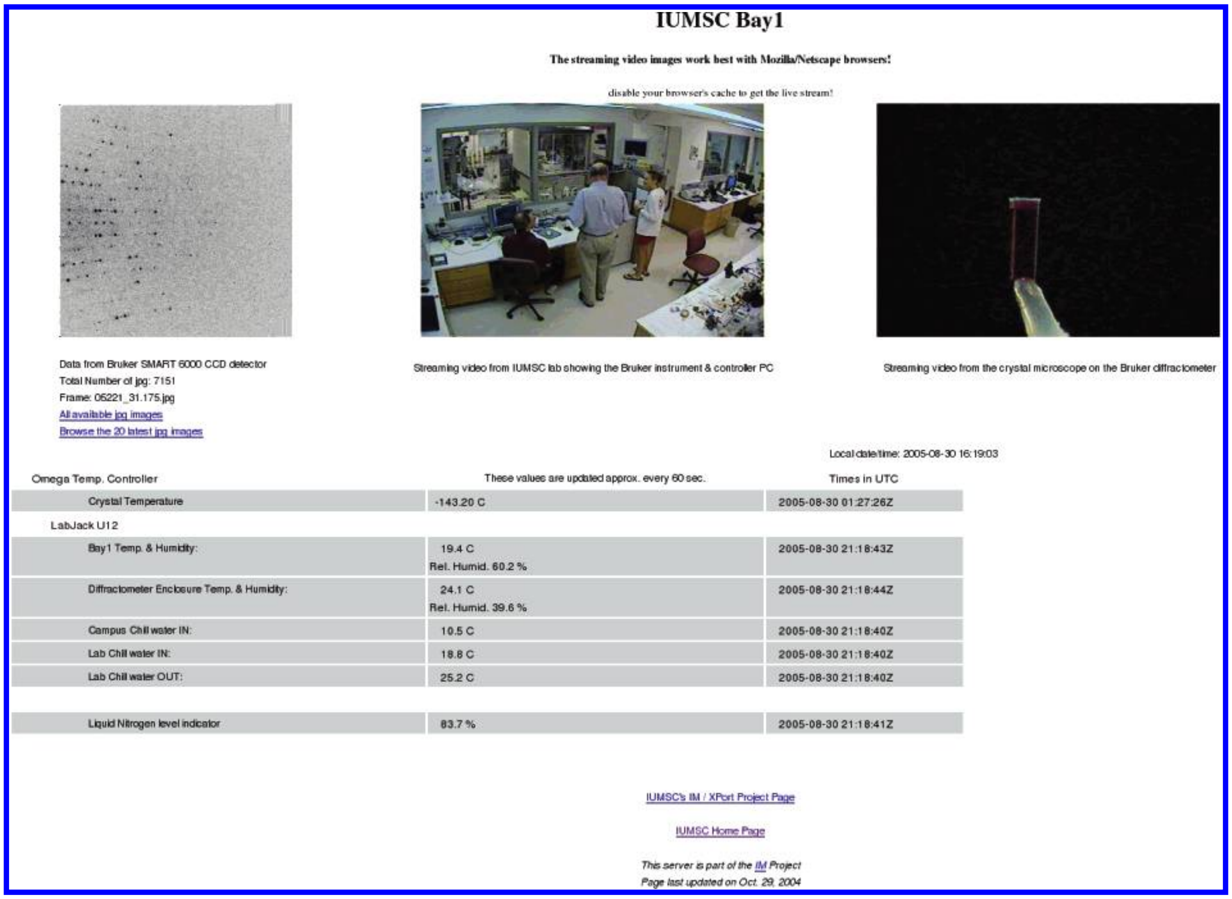

Figure 3. Simple Web-based application showing the latest CCD data and parameters from IUMSC.

of the control system (PC using Windows 2000) onto a Linux "CIMA proxy box", using the networking file system client Common Internet File System−Virtual File System (CIFS−VFS).[48]

An inexpensive USB-based data acquisition device, LabJack[49] U12 with thermocouple temperature probes and digital humidity probes, was employed. This device is able to provide the environmental data in digital format. The IUMSC laboratory staff also requested that the LabJack device be used to monitor the level of liquid nitrogen (LN2) in the cryo cooling system (used to cool the crystal sample to LN2 temperatures). An inexpensive household electronic bathroom scale was modified[50] to feed its strain gauge output signal into one of the analogue inputs of the LabJack U12 device. After calibration, we were able to correlate the output from the strain gauge to the amount of liquid nitrogen present in the Dewar. Using these hardware systems, the CIMA Web Services software provides all the laboratory temperature/humidity and liquid-nitrogen levels/values from the LabJack device. A simple Web application (Figure 3) demonstrating online viewing of the above parameters coming off the instruments has proven to be very popular by the IUMSC crystallographers because it allows them, with a quick glance, to monitor the current status of various parameters in the laboratory and also get an image of the lab/instrument operator and the crystal being examined. With this encouraging approval by the IUMSC staff, we continued the Web Services implementation of the above parameters.

Once the above-mentioned hardware modification and installations were complete, the CIMA Web Services source interface running on the "Proxy box" was able to successfully retrieve all the Bruker CCD frame data, crystal temperatures, laboratory temperatures and humidities, and liquid-nitrogen levels from the LabJack device and make these available as Web Services. A Web Services client (e.g., *My Manager*) that registers for that specific service (e.g., laboratory humidity) can then easily receive the desired values and metadata, such as time, associated with that value.

DISCUSSIONS AND CONCLUSION

There is a great demand in the instrument and sensor community for a "universal interface" that will allow sensors and instruments to be connected seamlessly to individual researchers' analytical processes. The concomitant requirements for data management, security, and access control must be either transparent to the end user or extremely straightforward to use. The CIMA software is currently installed and operational in the IUMSC test site and is being implemented in several other X-ray crystallography laboratories as a useful method of remotely accessing and sharing data emerged from these laboratories. One of the most interesting observations we have made is the recognition of the impact of adequately presented auxiliary data to the researchers.

The ability to create a "time line" of the experiment that includes not only the raw instrument data but also auxiliary information such as the environmental conditions, images of the sample, and so forth has proven popular (Figure 4). The time-line presentation has also proven to be an excellent teaching aid.

Researchers using commercial instruments are usually satisfied with the data presented by the manufacturer in

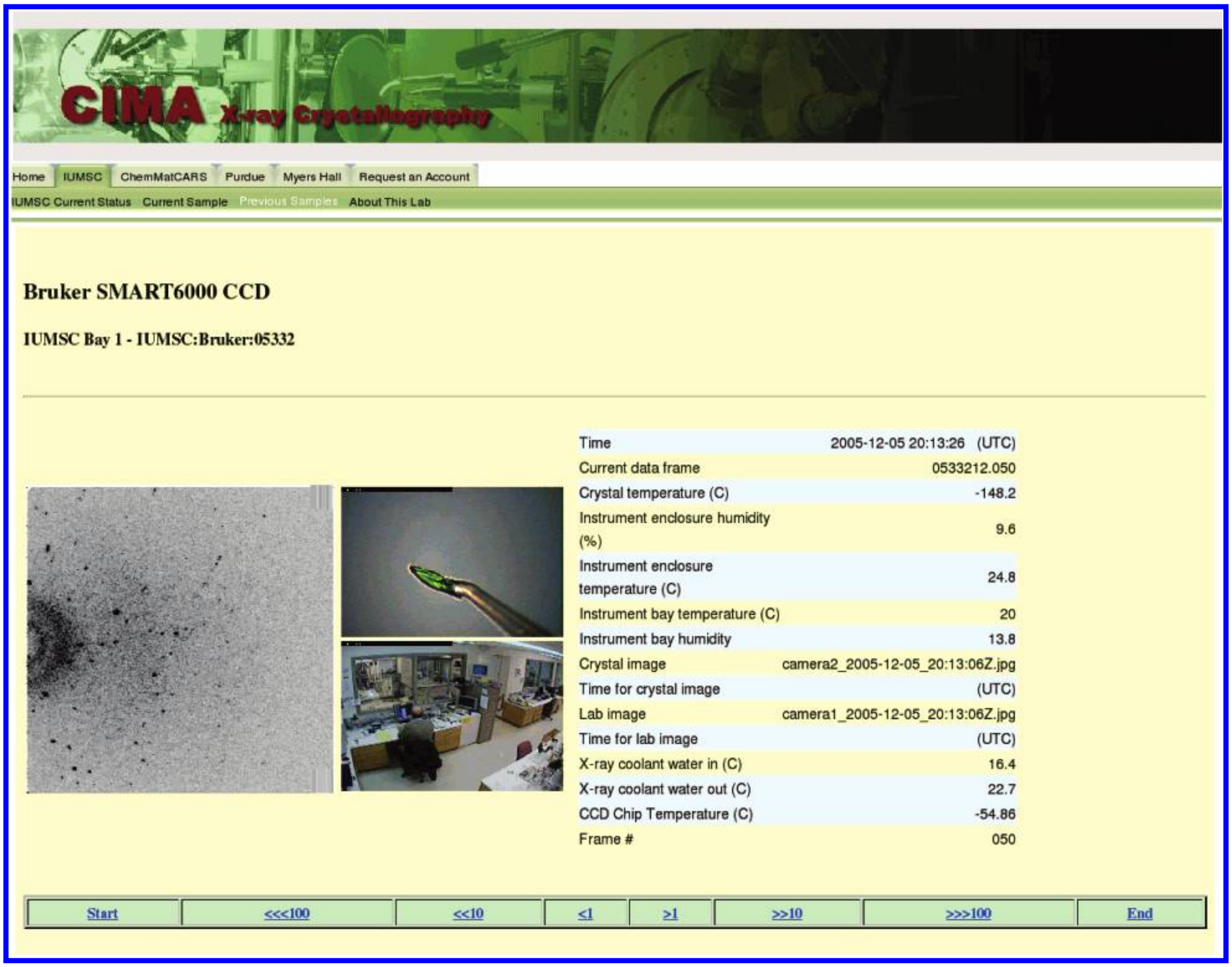

Figure 4. Time-line view of an entire experiment made available through the CIMA Crystallography Portal.

whatever form it is provided. We have discovered, however, that researchers are interested in other auxiliary data that is easily captured but is not provided by the commercial instrument. The ability to easily add these additional data streams to the CIMA portal has proven invaluable.

A major goal of the Crystallography Portal is to allow researchers to develop a system that allows remote access to Grid-enabled instruments and to provide a collaboration environment that would allow the remote user to interact with on-site researchers, as well as provide near real-time access to the data as it is being collected. In addition to these functions, the portal also facilitates postprocessing and archiving of raw data and metadata.

The portal is expected to have a significant impact on access to highly specialized facilities such as instruments located at synchrotron sites. It will also be possible for others to observe details of the experiment and potentially reduce training time and costs for new users unfamiliar with the instruments.

An obvious use of the crystallography portal is to provide students and faculty located in smaller universities and colleges the opportunity to use remote facilities in a similar manner. All chemistry students use molecular structures often in their studies without recognizing their origin. By being able to remotely observe the data collection procedure and participate in the analysis of data that provides a structure, they achieve a solid understanding of where the molecular structures come from and also the limitations of this analytical method. Many research universities provide crystallographic services to institutions that cannot justify the acquisition and operational costs of an X-ray diffractometer. The IUMSC, for example, has ongoing collaborations with faculty from six smaller institutions in Indiana. Recently, a group of six faculty and about 25 students from Ball State University (Muncie, IN) used the remote access facilities to observe the data collection procedures and, later, the structure solution for a compound synthesized as part of an undergraduate research project. The initial test involved having a classroom equipped with several workstations with Internet browsers and inexpensive video conferencing equipment. The remote observers were able to view the process from crystal selection to data collection and discuss the analysis and data processing steps with the crystallographers located in the IUMSC. The response was overwhelmingly positive, and we are in the process of trying to decide the best way to incorporate the experience into their curriculum on a regular basis. We are also in the process of getting local faculty members within the Chemistry Department to be involved. This will allow further collaboration with the participating schools and will provide an excellent recruiting tool for talented undergraduates. The CIMA-based crystallography system provides an opportunity for anyone with Web access to observe and use crystallographic and other data from laboratories that previously had only limited access.

Abbrevations. AVIDD, Analysis and Visualization of Instrument-Driven Data; CIFS−VFS, Common Internet File System−Virtual File System; CIMA, Common Instrument Middleware Architecture; DMS, Data Management System; HPSS, High-Performance Storage System; HTTP, HyperText Transfer Protocol; IEEE, Institute of Electrical and Electronics Engineers, Inc.; img-CIF/CBF, Image Crystallographic Information File/Crystallographic Binary File; IUMSC, Indiana University Molecular Structure Center; JMS, Java Message Service; NASA, National Aeronautics and Space Administration; NEES, Network for Earthquake Engineering Simulation; OGCE, Open Grid Computing Environments Collaboratory; OGSA, Open Grid Services Architecture; OWL-DL, Sublanguage of OWL (Web Ontology Language); RDF, Resource Description Framework; SCraps, Service Crysrallography at Advanced Photon Source; SMTP, Simple Mail Transport Protocol; SOAP, Simple Object Access Protocol; SSL, Secure Socket Layer; UDDI, Universal Description, Discovery, and Integration; URI, Uniform Resource Identifier; WSDL, Web Services Definition Language; X.509, a standard for public key infrastructure (PKI); XML, Extensible Markup Language.

ACKNOWLEDGMENT

This work was supported by National Science Foundation grants SCI 0330568 and MRI CDA-0116050. We thank Drs. Maren Pink and John Bollinger (IUMSC), who provided us with invaluable user feedback, and Dr. Marlon Pierce, Mehmet Nacar, and Yin Hao from the OGCE group for assistance with the Crystallography Portal development. We thank Ji Young Chong of the I.U. Department of Telecommunications for the graphical design of the Crystallography Portal. We thank Ashwini Prakash Athavale for much of the initial programming in the CIMA crystallography application. We thank John Pohlman and Andy Alexander of the Indiana University Department of Chemistry's Electronic Instrument Services group for assistance with some of the electronic sensors. We thank John N. Huffman for assistance throughout the project and, last but not least, all the crystallographers at CIMA-participating laboratories for invaluable help and feedback.

REFERENCES AND NOTES

(1) Foster, I.; Kesselman, C. *The Grid: Blueprint for a New Computing Infrastructure*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, 1999.
(2) Foster, I.; Kesselman, C. *The Grid 2: Blueprint for a New Computing Infrastructure*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, 2003.
(3) Berman, F.; Fox, G.; Hey, T. *Grid Computing*−*Making the Global Infrastructure a Reality*; John Wiley & Sons Ltd.: Chichester, U. K., 2003.
(4) Foster, I.; Kesselman, C.; Tuecke, S. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *Int. J. Supercomput. Appl.* **2001**, *15*, 200−222.
(5) Grimshaw, A. S.; Wulf, W. A. The Legion Vision of a Worldwide Computer. *Commun. ACM* **1997**, *40*, 39−45.
(6) Krishnan, S.; Bramley, R.; Gannon, D.; Ananthakrishnan, R.; Govindaraju, M.; Slominski, A.; Simmhan, Y.; Alameda, J.; Alkire, R.; Drews, T.; Webb, E. The XCAT Science Portal. *Sci. Prog.* **2002**, *10*, 303−317.
(7) Gannon, D.; Bramley, R.; Fox, G.; Smallen, S.; Rossi, A.; Ananthakrishnan, R.; Bertrand, F.; Chiu, K.; Farrellee, M.; Govindaraju, M.; Krishnan, S.; Ramakrishnan, L.; Simmhan, Y.; Slominski, A.; Ma, Y.; Olariu, C.; Rey-Cenvaz, N. Programming the Grid: Distributed Software Components, P2P and Grid Web Services for Scientific Applications. *Cluster Comput.* **2002**, *5*, 325−336.
(8) Booth, D.; Haas, H.; McCabe, F.; Newcomer, E.; Champion, M.; Ferris, C.; Orchard, D. Web Services Architecture. http://www.w3.org/TR/ws-arch (accessed 12/15/05).
(9) Globus Toolkit Web site. http://www.globus.org/ (accessed 12/15/05).
(10) Foster, I.; Kishimoto, H.; Savva, A.; Berry, D.; Djaoui, A.; Grimshaw, A.; Horn, B.; Maciel, F.; Siebenlist, F.; Subramaniam, R.; Treadwell, J.; Von Reich, J. The Open Grid Services Architecture, version 1.0. http://www.ggf.org/documents/GFD.30.pdf (accessed 12/15/05).
(11) Foster, I.; Kesselman, C.; Nick, J. M.; Tuecke, S. Grid Services for Distributed System Integration. *IEEE Comput.* **2002**, *35*, 37−46.
(12) He, H.; Haas, H.; Orchard, D. Web Services Architecture Usage Scenarios. http://www.w3.org/TR/ws-arch-scenarios (accessed 12/15/05).
(13) Foster, I.; Gannon, D., Kishimoto, H.; Von Reich, J. J. Open Grid Services Architecture Use Cases. http://www.ggf.org/documents/GFD.29.pdf (accessed 12/15/05).
(14) The Instrument Middleware Project Web site. http://www.instrument-middleware.org (accessed 12/15/05).
(15) NSF Middleware Initiative. http://www.nsf-middleware.org (accessed 12/15/05).
(16) (a) Frey, J.; Peppe, S.; Surridge, M.; Meacham, K.; Coles, S. J.; Hursthouse, M.; Light, M. E.; Mills, H. R.; De Roure, D.; Zaluska, E. Grid-Enabling an Existing Instrument-Based National Service. In *Proceedings of the First IEEE International Conference on e-Science and Grid Computing (e-Science 2005)*, Melbourne, Australia, December 5−8, 2005. http://doi.ieeecomputersociety.org/10.1109/E-SCIENCE.2005.49 (accessed 3/15/06). (b) Remote Instrumentations. http://science.internet2.edu/remote.html (accessed 3/15/06).
(17) NEESit. http://it.nees.org/about/index.php (accessed 12/15/05).
(18) National Instruments − Test and Measurement. http://www.ni.com/ (accessed 12/15/05).
(19) RBNB DataTurbine. http://outlet.creare.com/rbnb/ (accessed 12/15/05).
(20) NIST IEEE-P1451 Draft Standards Home Page. http://ieee1451.nist.gov (accessed 12/15/05).
(21) Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web. *Sci. Am.* **2001**, *284*, 34−44.
(22) Simple Object Access Protocol. http://www.w3.org/TR/soap (accessed 12/15/05).
(23) McMullen, D. F.; Devadithya, T.; Chiu, K. Integrating Instruments and Sensors into the Grid with CIMA Web Services. In *Proceedings of the Third APAC Conference and Exhibitionon Advanced Computing*, The APAC Conference and Exhibition on Advanced Computing, Grid Applications and e-Research (APAC05), Gold Coast, Australia, Sept 26−30, 2005. http://www.apac.edu.au/apac05/index.html (accessed 12/15/05).
(24) Devadithya, T.; Chiu, K.; Huffman, K.; McMullen, D. F. The Common Instrument Middleware Architecture: Overview of Goals and Implementation. In *Proceedings of the First IEEE International Conference on e-Science and Grid Computing (e-Science 2005)*, Melbourne, Australia, Dec 5−8, 2005. http://doi.ieeecomputersociety.org/10.1109/E-SCIENCE.2005.77 (accessed 3/15/06).
(25) Giacovazzo, C. *Fundamentals of Crystallography*; Oxford University Press: New York, 1992; pp 281−284.
(26) Crystal Grid Collaboratory Web site. http://www.crystalgrid.org (accessed 12/15/05).
(27) van Engelen, R.; Gallivan, K. The *gSOAP* Toolkit for Web Services and Peer-To-Peer Computing Networks. In *Second IEEE Int'l Symposium on Cluster Computing and the Grid*, Berlin, Germany, May 22−24, 2002.
(28) Ma, Y.; Bramley, R. Obsidian, A Composable Data Management Architecture for Scientific Applications. In *Proceedings of Challenges of Large Applications in Distributed Environments*, Research Triangle, NC, July 24, 2005.
(29) High Performance Storage System (HPSS). http://www.hpss-collaboration.org/hpss/index.jsp (accessed 12/15/05).
(30) HPSS at Indiana University. http://storage.iu.edu/geog.html (accessed 12/15/05).
(31) Hapner, M. *Java Message Service*; Technical Report; Sun Microsystems, April 11, 2002.
(32) Pallickara, S.; Fox, G. NaradaBrokering: A Middleware Framework and Architecture for Enabling Durable Peer-to-Peer Grids. In *Proceedings of ACM/IFIP/USENIX International Middleware Conference*; Endler, M., Schmidt, D., Eds.; Springer: Germany, 2003; Lecture Notes in Computer Science, Vol. 2672, pp 41−61.
(33) Boulder Real Time Technologies, Inc. http://www.brtt.com (accessed 12/15/05).
(34) McMullen, D. F.; Huffman, K. Connecting Users to Instruments and Sensors: Portals as Multi-User GUIs for Instrument and Sensor Facilities. In *Proceedings of GCE 2005: Workshop on Grid Computing Portals held in conjunction with SC05*, Seattle, WA, Nov 18, 2005. Submitted to the Journal of Concurrency and Computation: Practice and Experience.

(35) Jetspeed-1. http://portals.apache.org/jetspeed-1/index.html (accessed 12/15/05).
(36) GridSphere Portal Framework. http://www.gridsphere.org (accessed 12/15/05).
(37) Novotny, J.; Russell, M.; Wehrens, O. GridSphere: An Advanced Portal Framework. http://www.gridsphere.org/gridsphere/wp-4/Documents/France/gridsphere.pdf (accessed 12/15/05).
(38) Novotny, J. Developing Grid Portlets Using the GridSphere Portal Framework. http://www-128.ibm.com/developerworks/grid/library/gr-portlets (accessed 12/15/05).
(39) Open Grid Computing Environments Collaboratory (OGCE). http://www.ogce.org.
(40) imgCIF/CBF. http://www.iucr.org/iucr-top/cif/mmcif/ndb/cbf/index.html (accessed 12/15/05).
(41) Smith, M.; Welty, C.; McGuinness, D. L. OWL Web Ontology Language Guide. http://www.w3.org/TR/owl-guide/ (accessed 12/15/05).
(42) Indiana University Molecular Structure Center. http://www.iumsc.indiana.edu (accessed 12/15/05).
(43) Bruker-AXS Model SMART 6000. http://www.bruker-axs.de/index.php?id=smart6000&L=0 (accessed 12/15/05).
(44) Analysis and Visualization of Instrument-Driven Data (AVIDD) Supercomputing Facility. http://www.indiana.edu/~uits/rac/avidd (accessed 12/15/05).
(45) Service Crystallography at Advanced Photon Source (SCraps). http://www.iumsc.indiana.edu/projects/SCrAPS/index.html (accessed 12/15/05).
(46) Consortium for Advanced Radiation Sources (CARS). http://cars9.uchicago.edu (accessed 12/15/05).
(47) OMEGA iSeries Temperature Controller. http://www.omega.com/temperature/iseriesindex.html (accessed 12/15/05).
(48) CIFS−VFS filesystem. http://linux-cifs.samba.org(accessed 12/15/05).
(49) LabJack. http://www.labjack.com(accessed 12/15/05).
(50) Huffman, J. C. Department of Chemistry, Indiana University, Personal Communication, 2005.

CI050368L