SMIRP-A Systems Approach to Laboratory Automation

Jean-Claude Bradley,* Benjamin Samuel
Drexel University, Department of Chemistry, 32nd and Chestnut Streets, Philadelphia, PA 19104
*to whom correspondence should be addressed

ABSTRACT

SMIRP (Standard Modular Integrated Research Protocols), a web-based laboratory management system, will be described. SMIRP uses a protocol-experiment-parameter data representation system for all components of laboratory operations, including knowledge acquisition, laboratory data generation and analysis, and document production. Data tractability is inherent to the use of the system, permitting rapid browser-based navigation through linked information. The implementation of SMIRP in an academic research environment is discussed.

INTRODUCTION

Computer database technology, the internet and automation have enabled an unprecedented production of new data and analysis capabilities for recording and managing scientific data. These developments have had a particularly strong impact on the biotechnology sector, where the handling of vast amounts of data has become routine, largely related to the advent of combinatorial techniques for drug discovery. Such an application lends itself very well to automation because compound manipulation and assays can be standardized.

However, the management of discovery-driven scientific processes is not as readily amenable to such automation since experimental designs and results may be highly variable. For such applications, technological advances have mainly taken the form of electronic notebooks. Various approaches to implement electronic notebooks are currently taking shape. Examples include the DOE 2000,[1] ViNE,[2] and ZOO[3] projects. The consequences of shifting to a paperless scientific environment are complex. These issues are being addressed by CENSA, an organization whose mission is to assist with compliance with intellectual property laws and facilitate data-sharing using electronic notebook systems.[4]

One difficulty with the flexibility of most general purpose electronic notebooks is that inherent data tractability is somewhat limited and highly dependent upon the initiative of the researcher to include sufficient information when recording experiments. In other words, there is generally no standard mechanism to trace data related to a particular experiment within an electronic notebook environment. Also, the primary application of electronic notebooks lies in the recording of experiments, with limited attention to other aspects of laboratory management. In order to address these issues, SMIRP (Standard Modular Integrated Research Protocols) was designed in an attempt to develop a general system that can encompass not only laboratory research but also virtually all other components involved in a discovery-driven research operation.

The present paper will briefly outline some of the key features of SMIRP and how information is represented and navigated. Selected applications from an ongoing academic research operation will be used to illustrate the system.

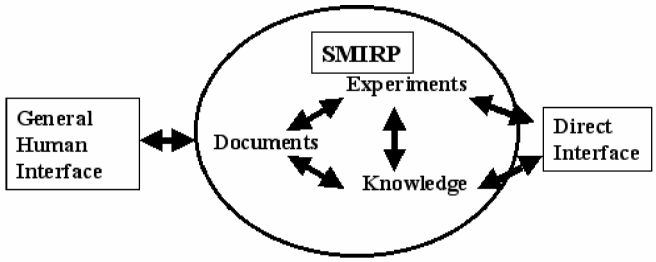


Figure 1 SMIRP General Design

SMIRP ARCHITECTURE

The most common way to communicate scientific information today is still through written documents in the form of research articles. This modality is represented in **Figure 1** as the General Human Interface. However, the ultimate function of the scientific document is to communicate specific *information* from one researcher to another, usually related to a series of experiments. In addition to this fundamental data set, research articles can also provide rich metadata surrounding experiments in the context of other related scientific work. This interpreted data set is represented in Figure 1 as *knowledge*, and may take the form of hypotheses, comparisons, conclusions and recommendations.

A fundamental concern in the design of SMIRP was to offer a standard approach by which it is possible directly to access primary knowledge and experiment components without the need to go through the intermediary of integrated documents such as research articles. The intent is not to bypass peer-reviewed publication but rather to complement this mode of communication and to offer access to data which may not be publishable because it does not fit into a sufficiently integrated format. This type of access may be particularly useful in collaborative research.

In order for this type of representation to be possible, it is necessary to provide a standard mechanism to determine the *context*

of recorded information. SMIRP achieves this by providing a general modular structure where much of the information related to a particular data set can be represented and linked in an intuitive fashion.

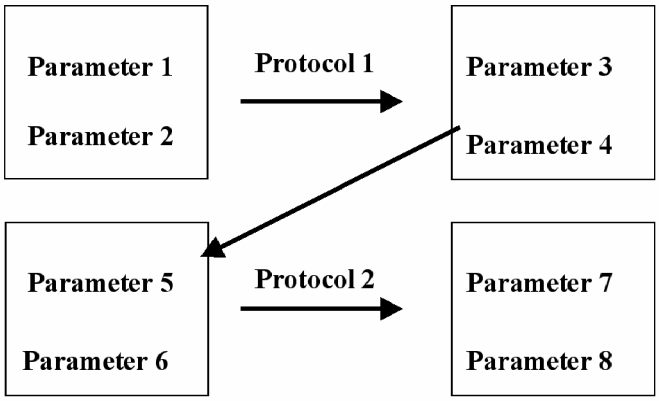


Figure 2 Fundamental Information Representation in SMIRP

The construct used in SMIRP to permit this is the protocol-experiment-parameter representation (**Figure 2**). In the context used here, a *protocol* is a description of the relationship between a set of parameters. Typically this relationship pertains to some kind of action or manipulation of objects, which can be either physical or abstract. Variability in the protocol is handled by *parameters*, generally with input parameters that can be controlled and output parameters that are observed. The execution of a protocol with specific parameter values is defined as an *experiment*. This type of representation is a natural and traditional approach to record laboratory processes. However, other aspects of laboratory management are generally not represented in this manner. By using the protocol-experiment-parameter construct for *all* information recorded in SMIRP, context and data tractability derive naturally.

THE SMIRP USER INTERFACE AND FUNCTIONALITY

A convenient way to illustrate the key features of SMIRP is by using real examples from a research operation where SMRIP has been implemented. An example of the representation of a laboratory experiment in SMIRP is shown in **Figure 3**, taken from actual academic research involving the preparation of palladium catalysts using pulsed electric fields.[5]

Figure 3 shows a screen capture of the General Editor, the default view in SMIRP. This is where data relating to a specific experiment can be entered. In this protocol, parameters such as concentrations of chemicals, electric field frequency and intensity, and duration of the experiment can be entered directly into the window. Additional parameters, including output parameters relating to properties of the resulting catalysts are present within the same frame but are not visible in the screen capture area shown in **Figure 3**. To the right of this frame are the date/time stamps which are automatically updated when any parameter is modified by a user. The identity of the user in a given session is determined by password access.

Navigation from the General Editor is straightforward through the use of two pull-down menus at the top of the screen. The right pull down menu allows rapid navigation to any other experiment within the current protocol. The left pull-down menu will switch to any other protocol within the present SMIRPspace.[6] Between these two pull down menus and the frame containing the parameters is a frame containing the protocol description. This is a detailed description of how to execute the present protocol and is identical for all experiments of a given protocol.

1. EXPERIMENTAL CHAINS

The second parameter in **Figure 3**, "Sprayed by," appears in the form of a drop down menu. This represents an embodiment of the abstraction in **Figure 2**, where a parameter from one protocol can be used as an input parameter in another protocol. The linking of protocols in this manner leads the formation of "experimental

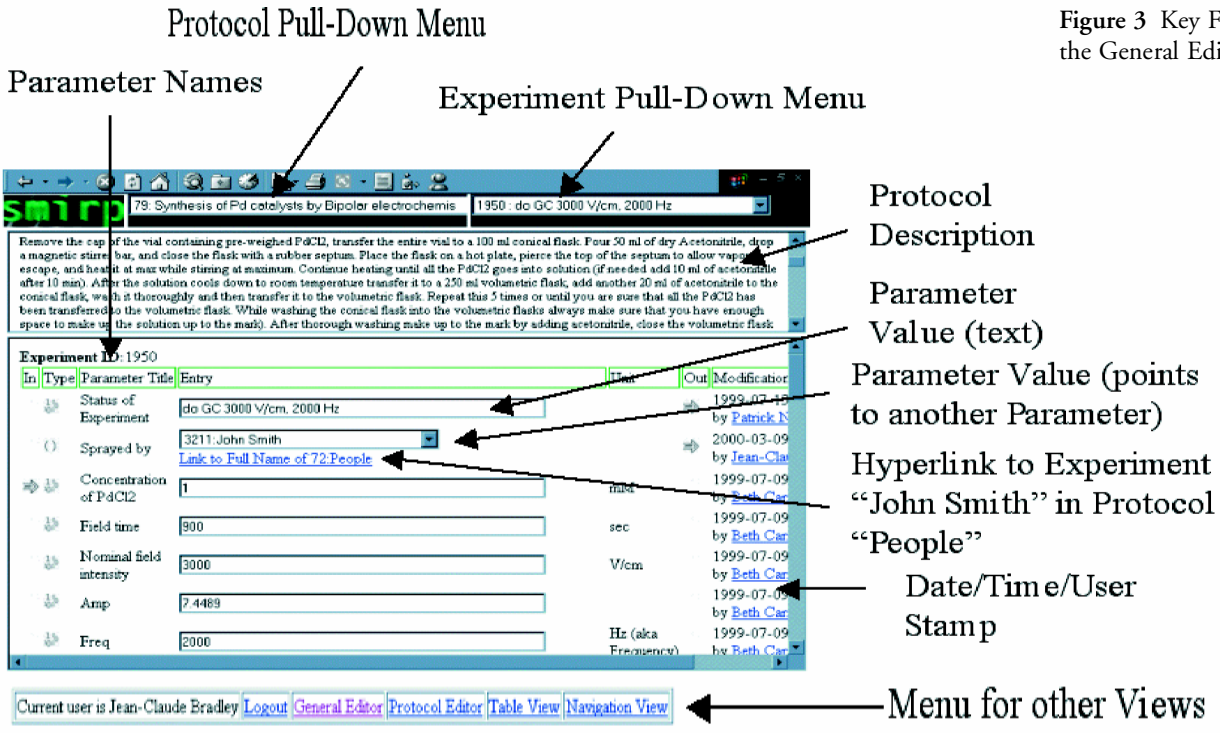


Figure 3 Key Features of the General Editor View

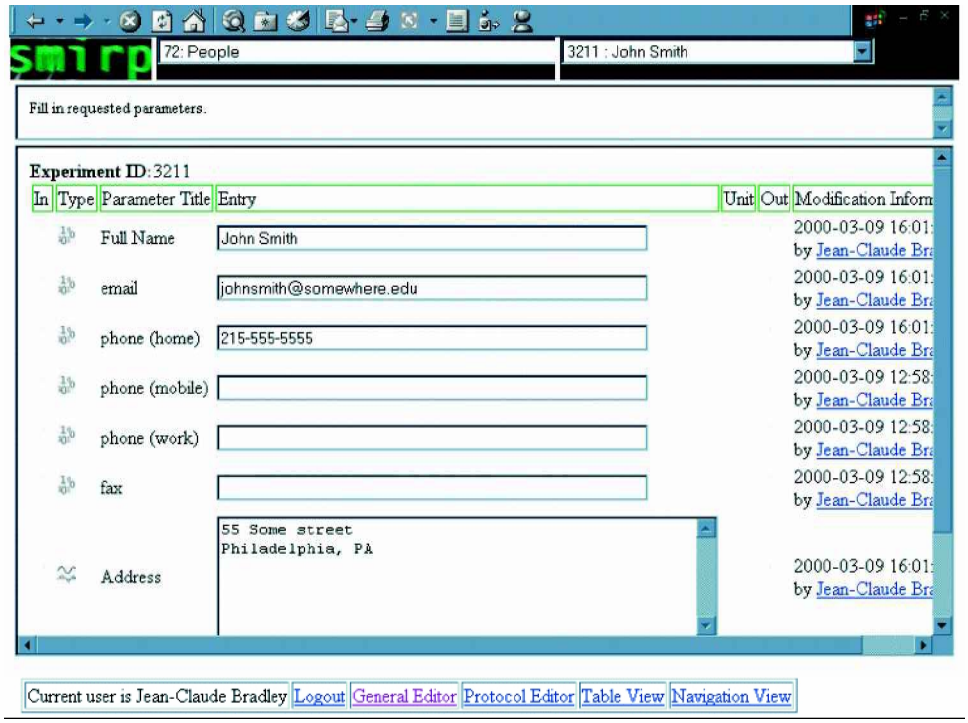
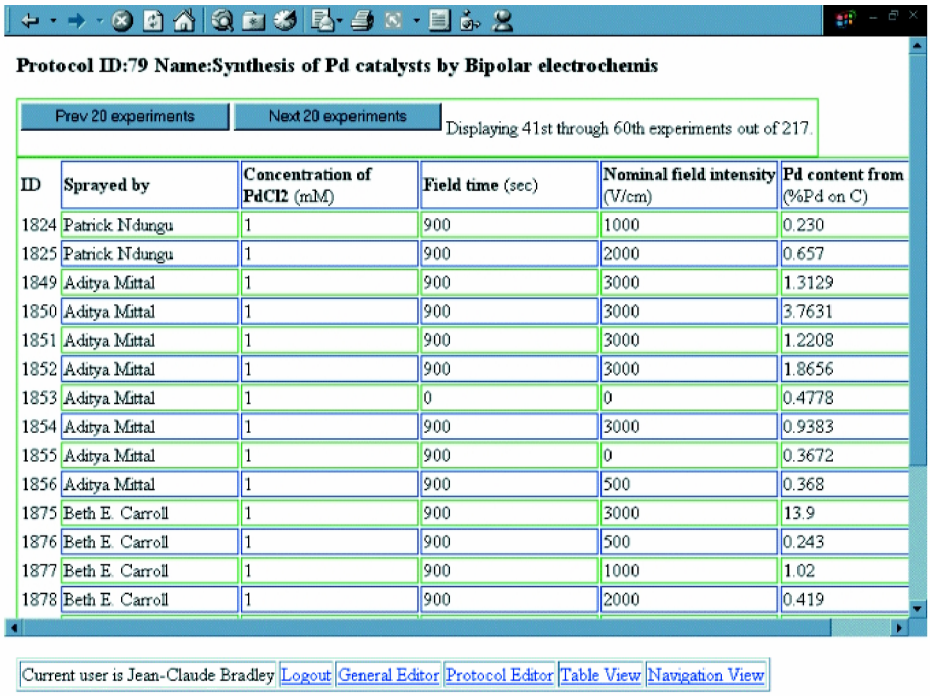


Figure 4 Link obtained upon clicking on "John Smith" hyperlinked parameter value



ID	Sprayed by	Concentration of PdCl2 (mM)	Field time (sec)	Nominal field intensity (V/cm)	Pd content from (%Pd on C)
1824	Patrick Ndungu	1	900	1000	0.230
1825	Patrick Ndungu	1	900	2000	0.657
1849	Aditya Mittal	1	900	3000	1.3129
1850	Aditya Mittal	1	900	3000	3.7631
1851	Aditya Mittal	1	900	3000	1.2208
1852	Aditya Mittal	1	900	3000	1.8656
1853	Aditya Mittal	1	0	0	0.4778
1854	Aditya Mittal	1	900	3000	0.9383
1855	Aditya Mittal	1	900	0	0.3672
1856	Aditya Mittal	1	900	500	0.368
1875	Beth E. Carroll	1	900	3000	13.9
1876	Beth E. Carroll	1	900	500	0.243
1877	Beth E. Carroll	1	900	1000	1.02
1878	Beth E. Carroll	1	900	2000	0.419

Figure 5 Table View of selected experiments and parameters from a protocol

chains" that can be navigated through hyperlinks in both the forward and reverse direction. This is an important component of automatic context generation. By following hyperlinks it is generally possible to obtain details of the input parameters and to discover how the objects or information were used in subsequent steps by following output parameters. Coupled with the protocol description frame, the navigation of experimental chains can be a powerful method to extract the context of a particular experiment.

For example, clicking on the "Sprayed by" parameter from the General Editor view takes us to the protocol named "People," specifically to the experiment named "John Smith," as shown in **Figure 4**. Using an identical representation as for the previous laboratory protocol, we can quickly find other parameters such as telephone numbers, fax, email and address information for that

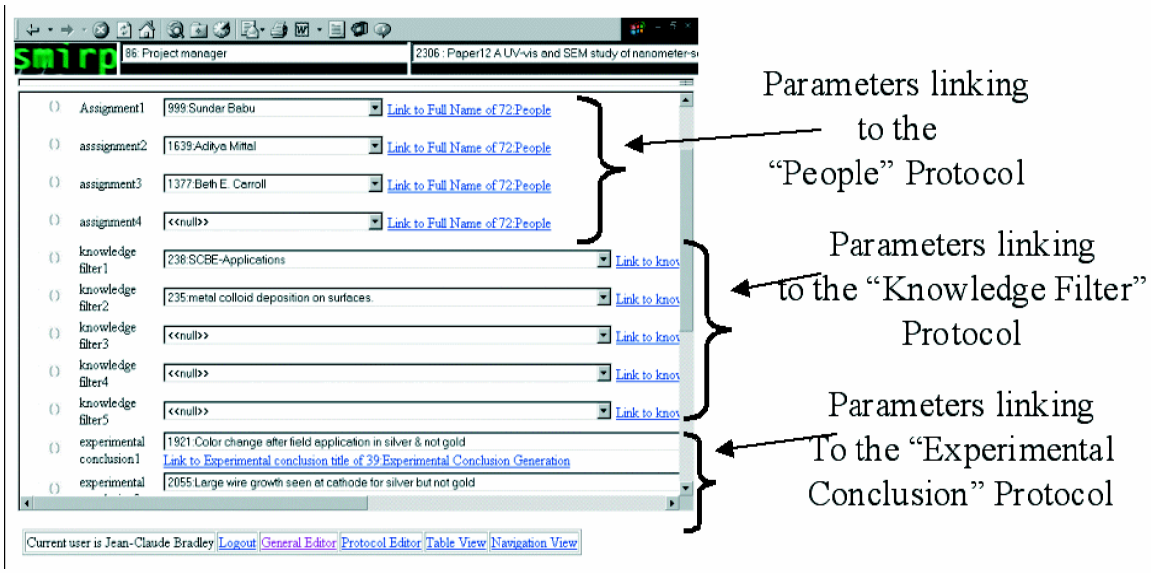


Figure 6 Partial View of an experiment in a Project Management Protocol

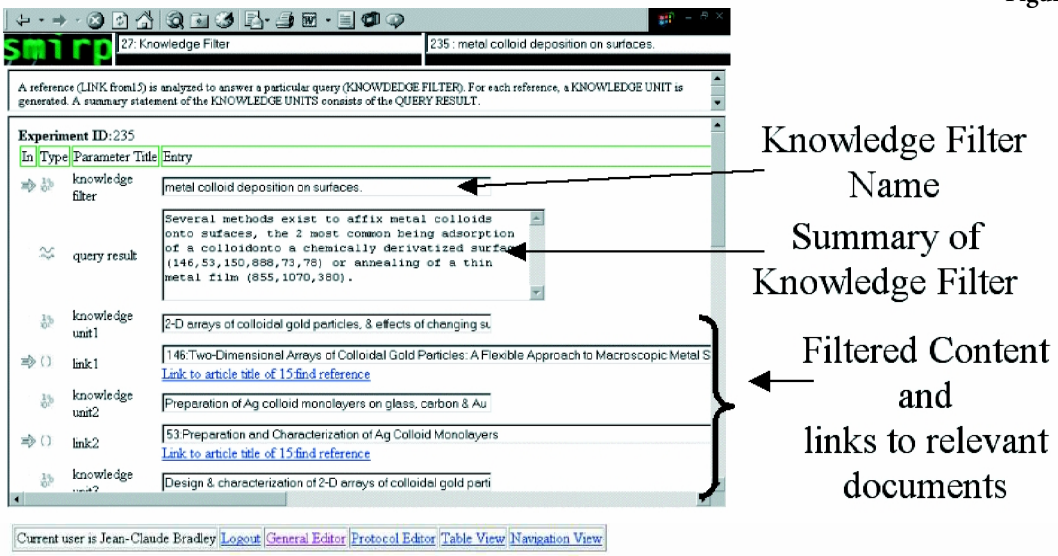


Figure 7 Knowledge Filter Protocol

particular individual. The execution of this protocol consists simply of filling in the available data for a given Name. Any time the parameter "Name" from the protocol "People" is used as an input parameter within a protocol, it will automatically appear as a drop down menu and create a hyperlink to information relating to that particular individual.

2. TABLE VIEW

At the bottom of the screen in **Figure 3** are a series of hyperlinks that permit different views of the information represented in SMIRP. The Table View creates a summary view of all experiments in the protocol currently being viewed in the form of a table. User selected parameters appear as columns whereas individual experiments appear as rows. In order to aid in navigation, the rows are hyperlinked to the General Editor view for any given experiment being clicked. An example of this is shown in **Figure 5**, representing a Table View of the experimental protocol "Synthesis of Pd Catalysts…" shown in **Figure 3**. This represents a convenient mechanism to scan for parameter values that have been used within a given protocol, regardless of who performed the experiments. For other types of protocols, the summarization feature provided by the Table View represents a convenient access to certain types of information not typically associated with experimental data stored in electronic notebook systems. For example, the Table View from the "People" protocol in effect yields an address book of all individuals associated with the research operation.

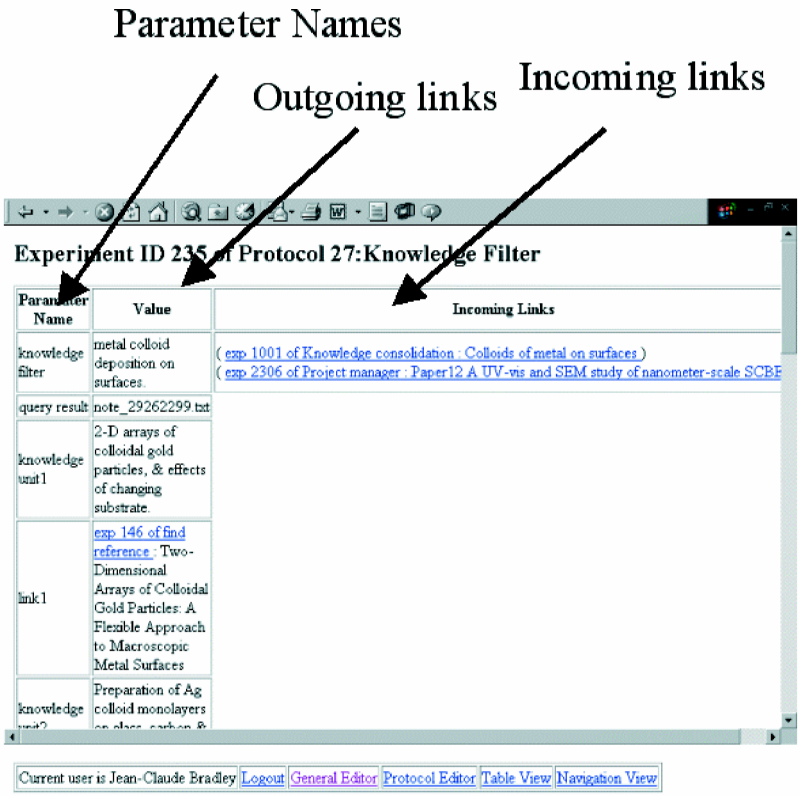


Figure 8 Navagation View of an experiment in Protocol "Knowledge Filter"

3. PROJECT MANAGEMENT

Figure 6 shows a screen capture of a portion of the General Editor view of a "Project Management" protocol used to coordinate laboratory activities and information related to specific projects. The topmost parameter set links to the "People" protocol, used in this context to signify individuals assigned to this particular project. The next parameter set links to experiments from a "Knowledge Filter" protocol, which are subsequently linked to references and documents relevant to the project. The third set of parameters links to an "Experimental Conclusion" protocol that eventually links to individual experiments, such as those described in the "Preparation of Catalysts…" protocol described above.

Clicking on one of the Knowledge Filter parameter links will lead to the screen shown in **Figure** 7. This protocol derives its name from its function as a type of conceptual filter when applied to the collection of documents present in this SMIRPspace. Only those documents that yield an answer to the stated query should be "filtered" through and become collected within a specific experiment. Answers appear in text boxes next to each document link and the summary of all these answers is included in a text box at the top of the page. Each of the links shown here will lead directly to the "Find Reference" protocol that contains as a file the actual document being referenced. In this way it is possible to quickly drill-down to specific information. In contrast to the "Preparation of Catalysts..." protocol above, the "Knowledge Filter" protocol involves only the transformation of information without physical manipulation. However, its representation within SMIRP is completely identical.

4. NAVIGATION VIEW

Thus far, navigation through the General Editor or the Table view has been unidirectional. We have also seen that modules can be reused. For example, the "People" protocol was pointed to within two different contexts, one to indicate who performed a specific step in a laboratory protocol and another to indicate who is part of a project under the "Project Management" protocol. Clearly it would be useful to navigate in the opposite direction, for example to ask for all instances where a particular experiment is being referenced. The Navigation view can be used for this and can be accessed at any time by clicking the corresponding link at the bottom of the page. **Figure 8** shows an example of the Navigation view for an experiment in the "Knowledge Filter" protocol. For each parameter instance, both incoming and outgoing links appear as hyperlinks in separate columns.

5. PROTOCOL EDITOR VIEW

The protocols described thus far are used for illustrative purposes and represent only *one* way to use SMIRP to link data. The only information representation that is hard-coded within SMIRP is the protocol-parameter-experiment construct. When a SMIRPspace is created, it is completely blank. Protocols must first be created and for this the Protocol Editor View is used. The construction or modification of a SMIRPspace is also accomplished over the internet through the same browser-based interface that is used for recording data.[7] Since protocols and parameters can be modified on the fly, a SMIRPspace is a very dynamic and flexible system that can adapt to environmental changes. For this reason, the ability to associate context with data is necessary to minimize confusion.

OTHER APLICATIONS

The flexibility of SMIRP has been exploited in a variety of contexts. The following represent some of these applications.

1. AUTOMATIC EXECUTION OF LABORATORY EXPERIMENTS

The standard representation of SMIRP with well defined input and output parameters offers the possibility to substitute, in a modular fashion, components of experiments that are currently done manually with automatic systems. In our laboratory this has proven to be a useful feature to enable the acquisition of large amounts of data. During the course of a particular experiment current and voltage sampling is required. There exist additional buttons in the General Editor view that allow passing of parameters between SMIRP and Labview to both execute an experiment and to acquire the data from an experimental run and load the resulting files and data into the SMIRP experiment. Since there is no difference in the way manual and automated processes use and store information in SMIRP, this greatly facilitates a stepwise modular approach to automation.

2. SELF-REFERENCE

When a SMIRP user logs on and modifies information within a SMIRPspace, that activity is logged by user and is date/time stamped at the parameter entry level. The highly atomistic recording of activity within the SMIRP environment is rich in information content, with respect to the transformation and creation of materials, data and concepts. In order to analyze this dataset it is convenient simply to port the event data from one SMIRPspace into another SMIRPspace for analysis and data mining. In a sense this represents a form of self-reference because SMIRP is used to represent the analysis of its own behavior.

3. AS AN EDUCATIONAL TOOL

The flexibility of SMIRP to represent information has also been used in a classroom environment to study the way in which students process information.[8] In this application, graduate students from a Neuroscience class were assigned a series of questions to be answered by a certain date. All answers were required to be supported by specific references, with the only limitation that two students could not use the same references. Since all students had access to the answers and references from the entire class, this permitted a type collaborative knowledge sharing approach to learning. Within the SMIRP architecture, the system was represented through interlinked protocols such as “Questions”, “Answers”, “References”, “Students” and “Grades”.

D. CONCLUSIONS AND FURTHER CONSIDERATION

Preliminary data obtained so far with SMIRP indicate that it offers an additional perspective for the managing of discovery-driven research operations. New technological advances offer an opportunity to re-examine the modalities that scientists use to communicate with each other and with automated systems. It is unlikely that a single approach will serve all the requirements of future scientific advances and communication. Given the inherent plasticity in the way in which SMIRP organizes information, it is anticipated that further applications, especially within collaborative environments, will shed light on the processes of scientific advancement and learning.

ACKNOWLEDGEMENTS

Partial support for this project through NSF grant CHE-9875855 is gratefully acknowledged. Helpful and stimulating discussions with John Bielec, Ken Blackney, Jan Biros, Beth Hazany, Claudia Ruiz-Shipe, Martha Ankeny, Donald McEachron, Diedre Harper, Joanne Sparks and Sundar Babu are also acknowledged.

REFERENCES

1. A demonstration of the DOE2000 project is available at http://www.epm.ornl.gov/enote
2. A tutorial and demonstration of the ViNE project is available at http://www.csi.uoregon.edu/nacse/vine
3. Ioannidis, Y. E., Livny, M., Ailamaki, A., Narayanan, A., Therber, A. Proceedings of the ACM SIGMOD international conference on Management of data 1997, 126, 580; Also see http://www.cs.wisc.edu/ZOO.
4. Information on CENSA (Collaborative Electronic Notebook Systems Association) can be found at http://www.censa.org.
5. For a more detailed description of this work see Bradley, J.C.; Ma, Z. Angew. *Chem. Int. Eng. Ed.* 1999, 38, 1663.
6. A SMIRPspace is a separate instance of SMIRP that is fully self-contained in terms of available protocols and authorized users.
7. Currently only Internet Explorer is supported, with optimization for version 5.0.
8. Bradley, J.C.; McEachron, D., manuscript in preparation.

BIOGRAPHY

Jean-Claude Bradley is an Assistant Professor of Chemistry at Drexel University. He received his Ph.D. from the University of Ottawa in 1993 followed by postdoctoral appointments with Michael Pirrung at Duke University and Jean-Marie Lehn at the College the France in Paris. His interests lie in both the execution and understanding of multidisciplinary science and he has published in various fields including synthetic organic chemistry, DNA-hybridization technology, gene therapy, electrochemistry, materials science and neuroscience. Contact information: Department of Chemistry, Drexel University, 32nd and Chestnut streets, Philadelphia, PA 19104; bradlejc@drexel.edu. ■