

A Collaborative Electronic Laboratory Notebook

JAMES D. MYERS
+1 610-355-0994
Jim.Myers@pnl.gov

ELENA S. MENDOZA
+1 509-376-9252
Elena.Mendoza@pnl.gov

BONNIE HOOPES
+1 407-482-5845
Bonnie.Hoopes@pnl.gov

Pacific Northwest
National Laboratory
PO Box 999
Richland, WA 99352 USA

ABSTRACT

We have developed a secure, collaborative, web-based electronic notebook (EN) designed to provide researchers and students with a means to record and share their primary research notes and data. The Pacific Northwest National Laboratory (PNNL) Electronic Laboratory Notebook (ELN) supports a variety of common multimedia formats and can easily be extended to allow entry and display of custom data types. Based on a combination of Java, JavaScript, and Perl, the ELN provides a hierarchical chapter/page structure for organizing entries and includes features such as search, automated email notification of new entries, and a programmatic interface for automated data submission. The most recent version also includes public-key authentication and digital signatures. The ELN is being used in a variety of scientific research projects at PNNL and is publicly available from our website. Experience at our laboratory and a variety of collaborating institutions suggests that the unique capabilities of the ELN can make it a valuable addition to or a replacement for paper notebooks in a variety of situations.

KEY WORDS: Electronic Notebook, Digital Signatures, Web Applications, Collaborative Systems and Applications.

1. INTRODUCTION

The paper notebook, a reliable staple of the laboratory and classroom, is becoming increasingly obsolete as the volume and complexity of the data we deal with increases. Although paper retains some advantages in portability and in the ease of entering notes, ENs have a variety of compelling advantages. They can eliminate the need for manual transcription of data that already exists in electronic form and can directly display the large, multidimensional, and time-dependent data sets produced in modern experiments. Further, they can be used collaboratively by distributed groups and can have

automated searching and indexing capabilities to aid in knowledge discovery. ENs also have advantages in terms of legal defensibility and as records because of the strength of digital signatures and the low cost of storage on digital media respectively.

Researchers have explored a variety of directions in the development of ENs over the past two decades. Initially, ENs focused on text- and hypertext-based annotation to support individual researchers [1][2][3]. Technological advances, particularly the rise of the Internet and the World Wide Web, led to the development of electronic notebooks that had clear advantages over traditional paper notebooks, particularly in terms of support for multimedia annotations and collaborative use [4][5][6][7][8]. Investigations of pen-and-voice input, and support for lightweight input devices (e.g., personal digital assistants) lessened the ease-of-use advantages of paper. Digital signature technologies, aided by the increasingly recognized advantages of digital storage, have eroded the advantages that paper notebooks traditionally held as a legal record. Other efforts sought to expand the advantages of ENs by bending the notebook analogy. Workflow capabilities showed that electronic notebooks could become an active component in the data analysis process [9][10] while the use of domain schema to provide sophisticated search capabilities moved notebooks into the field of knowledge management [11]. Efforts in industry and government have documented the need to recognize electronic notebooks as primary, legally defensible records at the enterprise scale [12][13][14]. Today, ENs are beginning to see productive use in many communities. Custom solutions have been built and deployed in industry, and commercial notebooks that target structured environments, such as analytical chemistry and pharmaceutical laboratories, have begun to emerge [15][16]. However, very few flexible, general-purpose EN implementations exist.

We have developed a web-based ELN with significant capabilities in most of the areas noted above that we believe provides a new level of general-purpose functionality in combination with powerful customization

and integration mechanisms. The ELN was developed as part of a three-way collaboration between researchers at PNNL, Lawrence Berkeley National Laboratory, and Oak Ridge National Laboratory as part of the U.S. Department of Energy's DOE2000 Collaboratory program [17, 18].

2. ELN FUNCTIONALITY

The two basic components of the ELN are an interactive browser-based client and a web-server-based notebook server. In use, the client presents an initial login screen requiring the user's name and password, and then displays a main window containing a table of contents with a user-defined hierarchy of chapters, pages, and notes. The content of the currently selected page appears in a separate browser window (See Figure 1.). Notes added to a page appear chronologically, and it is possible to add a sub-note, which will appear indented and directly after the original note. This provides a capability analogous to "writing in the margin" of a paper notebook. Adding "replies" to notes supports threaded discussions between members of a distributed research team. Since the ELN is a web- and Java-based application, users may access the pages and notes from their lab or their office at their convenience from Macintosh™, MS Windows™, Linux™, and UNIX™ computers.

Notes on a page are created using "entry editors," which are launched from the main window. The notebook currently includes editors to create text (plain, HTML, or

rich text), equations (LaTeX), and whiteboard sketches, to capture screen images, and to upload arbitrary files. A form-based editor allows users to rapidly support new types of experiments by specifying an HTML form that gathers a pre-defined set of parameters and an HTML document that describes how the form data should be displayed on the notebook page (e.g., in an HTML table). Once a note is created, a click on the "submit" button publishes it to the notebook page and makes it immediately available to other authorized users of the notebook. Until submission, entries may be repeatedly edited. Submitted entries appear as part of the page displayed in the browser, tagged with the author's name and the date and time of the entry.

The browser will render individual "notes" (e.g., text, images), using external applications (e.g., launching Microsoft Word™), or by using Java applets or plug-ins (e.g., equations, molecular structures, and VRML). Short "viewer" scripts are used to control how a given type of data is rendered. The current ELN has nearly two-dozen viewers that can display a wide range of notes, sketches, graphs, pictures, biology and chemistry data, and other information. For instance, we developed a short script to specify that protein structures should be displayed using the WebMol Java applet [19]. This allows users of PNNL's Virtual Nuclear Magnetic Resonance Facility (VNMRF) [20] to view the proteins as three-dimensional rotatable models and to display inter-atom distances and angles - enough information to allow quick analysis and comparisons - directly on the shared notebook page.

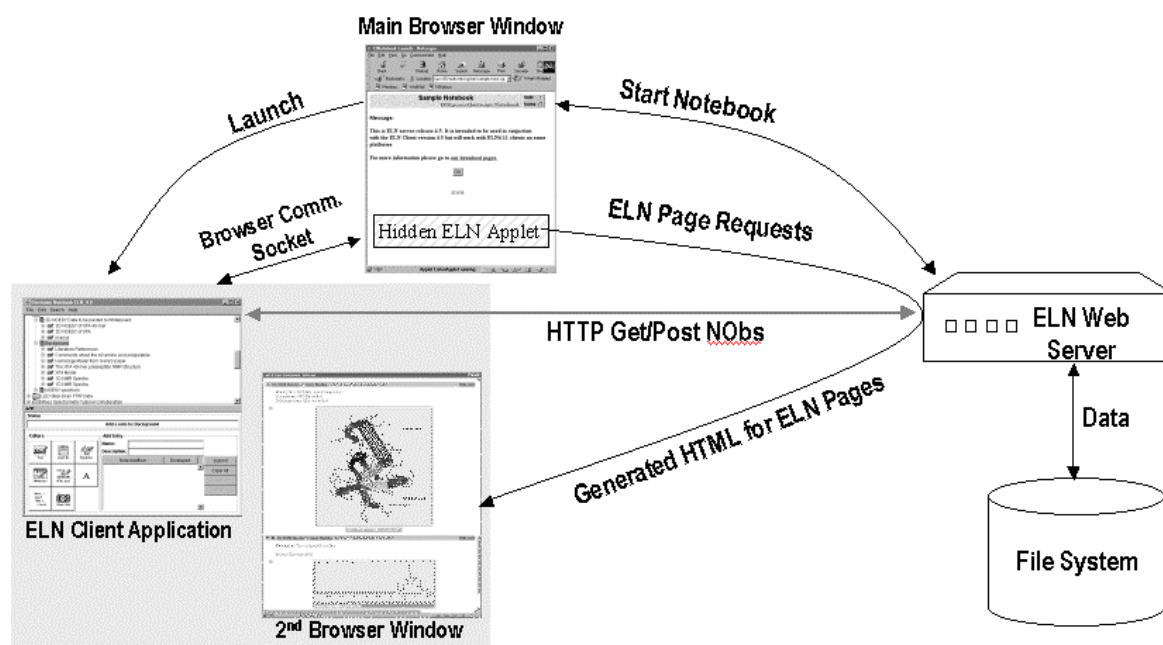


Figure 1. The Electronic Laboratory Notebook architecture showing interactions between the ELN Client application, the user's web browser, and the web server hosting the ELN server CGI scripts. Gray box: The ELN interface consisting of the Table of Contents and palette of editors in one window (left), and a selected page of the notebook showing two entries (right).

Another recent viewer invokes the S-PLUS Graphlet applet to provide dynamic chart displays from S-PLUS files [21]. A notebook administrator, who may also be the end-user, can easily create a customized version of the ELN for a specific discipline/project by specifying which editors and viewers are supported. The ELN client dynamically adds/removes editors from the entry palette based on what the server allows. Documentation of the API for building editors and viewers is available from the project website [22].

The ELN has a number of capabilities beyond those for entry and viewing of multimedia notes. A search capability allows one to quickly find entries based on words in the title and descriptions. An email notification mechanism sends notices about new entries. Users can request email regarding updates to a whole notebook, or specific chapters or pages within a notebook. The ELN Wizard, a component that is callable from within other programs, may be used to automate, or partially automate, the transfer of parameter sets, data files, or screen snapshots immediately to a user-specified chapter and page. If all required parameters (e.g., user name/password, data to store, data placement within the notebook) are programmatically supplied to the ELN Wizard, it completes the submission without displaying a user interface. If one or more parameters are left unspecified, the ELN Wizard will display one or more dialogs to request values from the user. Thus, the ELN Wizard can fully automate the capture of status information and records of off-normal events into the ELN replacing a paper instrument log. It can be used within macros in Microsoft Word™ and Excel™ to create a “Save To Notebook” menu item. Selecting this menu item will pop up a dialog box showing the ELN’s table of contents, analogous to a standard file browser dialog, to allow the user to select where the entry should be created.

The standard ELN is secured through the web server’s basic-authentication mechanism for access control. A separate individual username/password is used to authenticate users and to determine authorship. These mechanisms have obvious limitations when working with sensitive data and in providing legally defensible authorship evidence. A public-key cryptography-based version of the ELN has been developed to address these issues. This version supports SSL-encrypted client-server connections, mutual cryptographic authentication of users and ELN servers, and digital signatures. We are currently investigating options for supporting secure timestamps.

3. DESIGN

The core of the ELN consists of an interactive browser-based client and a notebook server implemented using common gateway interface (CGI) scripts that run on a web server (See Figure 1). The client interface is rendered by a locally installed Java application that

handles user interaction—login, selecting page views, searching, entering notes, signing pages, etc.—and the browser, which serves as a launch point and provides displays of individual pages. A small, signed Java applet within the browser launches the application and maintains communication through a socket. This allows the application to request new page views for display in the browser. The applet uses LiveConnect to invoke a JavaScript method to request the page. The server is written in Perl that generates necessary HTML and JavaScript for displaying the initial notebook entry page and for subsequent notebook page views in the browser. It also responds to HTTP requests from the application that implement the basic functionality of logging in, discovering the notebook contents, submitting new entries, etc. The ELN Wizard is a limited client application that encapsulates the entry submission functionality of the client in a form that can be used by other applications without the need for a browser. Because of the clean separation of browser-related code and the modular design of the ELN client, the ELN Wizard is able to re-use many of the client components. The ELN Wizard does not currently handle retrieval or display of notebook entries beyond displaying the notebook table of contents to allow user selection of where new entries should be placed.

The central data management concept in the ELN is the notebook object (NOB), which consists of an opaque, typed piece of data and key/value pairs of metadata—data about the data. The NOB data may be the text, images, or other data composing an entry, or it may be a list of other NOBs, allowing representation of the potentially infinite hierarchy of chapters, pages, notes, and sub-notes. The metadata includes a core set of key/value pairs whose meaning is understood by the notebook, as well as arbitrary key/value pairs with meaning for the applications that produce or process NOBs. The concept of a NOB, and a core set of keys {authname, datetime, label, description, objectid, datatype, and dateref} have been defined consistently across all three DOE2000 notebook implementations. The dateref key/value allows NOBs to reference externally managed data using standard URL notation in cases where the benefits of not making a copy of data within the ELN outweigh the concern over relying on an external system for data availability, e.g., when the data is very large. Representations of the NOB have been developed in Java and Perl, and a flat, MIME-based format is used during exchanges between the client and server and for archiving notebook entries.

At PNNL, we have extended the core metadata set to include standard key/value pairs representing digital signatures, the purpose of a signature (e.g., “authored by”, “witnessed by”, “approved by”), and the signers’ identities. Editors and external applications that create NOBs are also free to define new metadata keys. We chose to use Entrust’s Java development kit to implement

our public-key authentication, encryption and digital signature capabilities, and leverage an installed Entrust public-key infrastructure at PNNL. All signing and signature verification operations occur on the local desktop. As noted, signatures are submitted to the ELN server as NOB metadata. During verification, the signatures and original data are retrieved from the server.

4. DISCUSSION

The ELN was developed primarily as an investigation of the electronic notebook concept and as a tool for DOE researchers working in distributed teams. As part of the laboratory's operations, we provide notebook customization and administration, user training, and other services upon request to projects involving researchers or equipment at PNNL's Environmental Molecular Sciences Laboratory (EMSL). We also make the ELN publicly available from the project website, along with documentation and developer kits [22]. Over the life of the project, more than 600 people, representing science and engineering research and education projects across academia, government, and industry, have registered to download the software. At PNNL, we now maintain notebooks for more than two-dozen project teams, including users of EMSL's Virtual NMR Facility [20]. In one particular case, involving surface science researchers working in a clean room where paper is a contaminant, the ELN has been the only notebook used for nearly two years.

At PNNL the ELN has been used primarily in the domains of chemistry and biology, and for software development. The ELN allows researchers, from high-school, undergraduate, and graduate students to university professors, to access data and notes generated by remote operation of scientific instruments, to maintain Internet-accessible personal and laboratory journals, and to coordinate work with remote colleagues. These uses span a wide range of emphases on security, laboratory automation, and long-term maintenance of ELNs as project records.

The primary motivations for adopting the ELN have been collaboration between distant members of a project team, and easy capture of multimedia notes (e.g., color images from microscopes). Additional motivations include laboratory automation, the elimination of dust from paper notebooks in clean rooms, availability of data and notes from labs, offices, and homes, and a belief that technologies such as electronic notebooks will be an integral part of research in the future, making it important to introduce them today in the classroom and in laboratory-based training of future scientists.

5. CONCLUSION

The ELN is being used successfully in a variety of research and education settings, as a collaborative tool, and as the primary research record. Its open programming interfaces make it possible to rapidly customize a notebook to support existing project data types and to integrate the ELN into laboratory processes. While the ELN and other ENs are considered robust and valuable tools as is, there are many issues that must be resolved to make electronic notebooks the preferred choice across a majority of industrial, governmental, and academic research. At the user interface level, questions remain about how close the analogy with paper notebooks should be, how to integrate electronic notebooks with the overall scientific workflow and records processes, and how to leverage mobile devices for input and display. At a deeper level, current notebooks are limited in their ability to show non-chronological relationships between entries, to support complex searches, and to interact with other producers, curators, and consumers of annotations such as autonomous feature-detection agents, digital libraries, and data pedigree mechanisms. Issues of long-term data and digital signature management and enterprise-scale management of notebooks have also yet to be addressed beyond documentation of the requirements.

6. ACKNOWLEDGEMENT

The author would like to acknowledge the efforts of developers including William T. Valdez, Wyllona Harris, and Eric Bradley in the development of earlier versions of the ELN. We also would like to acknowledge the design and code contributions of Al Geist (Oak Ridge National Laboratory), Noel Nachtigal (Sandia National Laboratories) and Sonia Sachs (Lawrence Berkeley National Laboratory). This work was supported by the U. S. Department of Energy through the DOE2000 program sponsored by the Mathematical, Information and Computational Sciences Division of the Office of Science. PNNL is operated by Battelle for the U. S. Department of Energy. The W. R. Wiley Environmental Molecular Sciences Laboratory (EMSL) is a national scientific user facility sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research and located at PNNL.

REFERENCES

1. H. Kaplan, An Electronic Notebook, *Popular Computing*, 3, 1884, 174-&.
2. T. D. Costello, RS/1, An Electronic Notebook, *Journal of Coatings Technology* 58, 1986, 75-80.

3. J. Figueras, An Electronic Notebook for chemists, *Acs. Symposium Series*, 341, 1987, 37-47.
4. D. C. Edelson, R. D. Pea, and L. M. Gomez, The Collaboratory Notebook, *Communications of the Acm.*, 39, 1996, 32-33.
5. J. Hong, G. Toye, and L. J. Leifer, Engineering design notebook for sharing and reuse. *Computers in Industry*, 29, 1996, 27-35.
6. D. Strauss, K. Meserve, T. Hess, R. Meyers, M. Kersey, J. Reilly, G. Thornton, G. Atwater, C. Jones, P. Graf, and J. Jenkins. 1996. Expressive electronic notebook environment, *Abstracts of Papers of the American Chemical Society* 212, 1996, 5-CINF.
7. E. S. Mendoza, W. T. Valdez, W. M. Harris, P. Auman, E. Gage, J. D. Myers, EMSL's Electronic Laboratory Notebook, In *Proceedings of the WebNet '98 World Conference*, Orlando, Florida, November 7-12, 1998.
8. J. D. Myers, C. Fox-Dobbs, J. Laird, D. Le, D. Reich, and T. Curtz, Electronic Laboratory Notebooks for Collaborative Research, In *Proceedings of the IEEE Fifth Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, Stanford University, California, June 1996.
9. A. D. Malony, J. L. Skidmore, and M. J. Sottile, Computational experiments using distributed tools in a web-based electronic notebook environment, *High-Performance Computing and Networking, Proceedings*, 1999, 1593:381-390.
10. A. D. Malony, J. E. Cuny, J. L. Skidmore, and M. J. Sottile, Computational experiments using distributed tools in a web-based electronic notebook environment, *Future Generation Computer Systems* 16, 2000, 453-464.
11. LBNL DOE2000 Homepage Available URL: <http://www.lbl.gov/CS/DOE2000.html>
12. R. Lysakowski, Electronic notebook system technologies and architecture base-line requirements, *Abstracts of Papers of the American Chemical Society*, 1993, 206:103-COMP.
13. Records Management Document Available URL: <http://www.ornl.gov/records/notebooks.pdf>
14. FDA Compliance Available URL: http://www.fda.gov/ora/compliance_ref/part11/
15. Labbook Available URL: <http://www.labbook.com/>
16. IceBreaker Available URL: <http://www.icebreaker.com/>
17. DOE2000 Notebook Project Homepage. Available URL: <http://www.csm.ornl.gov/enote/>
18. DOE2000 Project Homepage. Available URL: <http://www.mcs.anl.gov/DOE2000/>
19. D. Walther, WebMol - a Java based PDB viewer, *Trends Biochem Sci.*, 22, 1997, 274-5.
20. K. Keating, J. Myers, J. Pelton, R. Bair, D. Wemmer, and P. Ellis, Development and use of a Virtual NMR Facility, *J. Mag. Res.* 143, 2000, 172-183.
21. S-PLUS Graphlet. See URL: <http://www.insightful.com/graphlets/default.html>
22. EMSL Collaboratory Project Homepage. Available URL: <http://collaboratory.pnl.gov/>