

THE COLLABORATIVE ELECTRONIC NOTEBOOK SYSTEMS CONSORTIUM

Rich Lysakowski, Ph.D.

The Collaborative Electronic Notebook Systems Consortium

800 West Cummings Park, Suite 5400,

Woburn, MA Massachusetts 01801

Phone 617-935-9600; Fax 617-935-3113.

Abstract

Electronic laboratory notebook systems have the potential to provide increased access to scientific data. The Collaborative Electronic Notebook Systems Association (CENSA) has been formed through the support of eleven major pharmaceutical and chemical companies. The consortium is working with scientific software and hardware vendors to facilitate the design of a system that will meet the needs of the consortium members, including compliance with recordkeeping rules by the various regulatory authorities and patent offices.

Introduction

Starting in October 1994, researchers at TeamScience, a science oriented software research and engineering company located in Woburn, Massachusetts, initiated a research study called "the R&D Team Computing Study"[1,2]. The study was commissioned by nine multinational corporations in the pharmaceutical, chemical, biotechnology, and related industries. The study detailed the vision, the business requirements, and the software functional requirements for companies to be effective using the new class of systems called "R&D Team Computing Systems." A purpose of the study was to evaluate potential electronic notebook solutions. The study evaluated Digital Equipment Corporation's "LinkWorks", Documentum's "Enterprise Document Management System," ForeFront Group's "Virtual Notebook System," Helix Systems and "ResearchStation," Lotus Development Corporation "Notes." The study uncovered major areas of improvements required

in these products. It also measured significant process improvements possible with the proper application of electronic notebooks. In December 1995, in response to the R&D Team Computing Study results, several global chemical and pharmaceutical companies decided to enter into a consortium to specify and facilitate the creation of robust electronic notebook solutions designed to meet member needs.

Results

The Consortium has started the development of a technical architecture that will build on top of already successful base platforms. It will place value-added services for electronic recordkeeping and records management lifecycle support on the servers. It will build special modules and add-in components to facilitate the work of chemists, biologists, materials scientists, and other technical professionals working on collaborative teams. Finally, the architecture will build specialized and standardized integration hooks into these platforms and provide graphical application integration tools to take advantage of these hooks. There will be both desktop hooks and collaboration infrastructure hooks. The Consortium development process is designed to leverage the core competencies of both large and small "best-of-breed" vendors. The technology emphasis is on CENSA's open, modular, client/server component architectures that can leverage current and near future infrastructure investments, and result in mass-marketed but customizable products at reasonable prices. The tools developed by vendors are required to take advantage of popular base computing platforms, and utilize successful integration and middleware

tools and emerging standards such SQL3, CORBA, Java, OLE, X.400, X.500, DCE, http/html, PDF, SSL, DNS, and others. As of this writing the technical architecture design is in process.

Discussion

R&D records possess several unique characteristics: First, R&D activities are always "project-oriented," meaning that they comprise a closed set of activities with a fixed endpoint. Electronic recordkeeping and records management are a core subsystem of an R&D Team Computing System. R&D team activities nearly always result in specific deliverables. Very often those deliverables are materials, equipment, or reports, although often they are simply conclusions or presentations of results [3]. While the project orientation of R&D is not particularly unusual, since many businesses are run using projects, R&D project structures are extremely variable and ad-hoc, and change or spawn new projects quickly.

The second important and unique characteristic of R&D is that projects are frequently interrelated. Project data may be similar in one set of dimensions yet very different along others. This implicitly links projects across several dimensions. In computer terminology such project data are "hyperlinked." Yet there is now no way for people to easily drill backward or drill forward from one project to another, to browse any data, and create reports from them. Project data are very often locked up in paper in notebooks or recordkeeping systems.

Because projects are rapidly spawned from one another, it is easy for a project portfolio to become unmanageable. Thus project data tracking and management capabilities are crucial in R&D. For example, a medium-sized chemical company may have hundreds to thousands of projects open at once. Sometimes thousands of people scattered across the globe must participate in the generation, organization, review, summary, and secure release of project information. Companies must have powerful ways of simplifying the inherent complexity of their tasks. Sophisticated software tools for managing all data from projects are badly needed by scientists.

Another unique characteristic of R&D is that laboratory record types are among the most varied anywhere. Typical data types encountered include text, images, graphics, structures (molecular, protein, DNA), spectral data, protocols, audio, video, and others—each with perhaps tens of variations of data formats which change with successive releases of software. This makes automation of lab data management a major challenge because there is so little consistency or standardization.

Conclusion

R&D data from laboratory notebooks is not always accessible to scientists. CENSA has been formed to facilitate the creation of collaborative electronic notebook systems designed to store experimental data. The consortium is designing a technical architecture that will build upon existing products and standards to solve many of the needs for scientific information. The consortium will be involving national laboratories, vendors and various other companies to build components for member companies. The ultimate output of the consortium will be hardware and software systems that facilitates collaboration of scientific and technical teams, the construction of R&D data and document warehouses, and specialized record repositories. Knowledge deliverables from the consortium will form the basis for standards for fully electronic recordkeeping procedures and systems that meet all the legal and regulatory requirements. Eventually the group's mission is to move the storage of R&D intellectual property away from "paper as a necessity" to "paper merely as a convenience" with fully electronic recordkeeping for the default storage mode. This will enable much greater utilization of R&D and corporate intellectual property, and allow much higher scientific productivity.

References

- [1] R. Lysakowski, S. Obrien, and J. Garden, "The Team Computing Study Report," TeamScience Inc., October 1996.
- [2] R. Lysakowski, "Moving to Electronic Records and Electronic Lab Notebooks, While Meeting

Legal, Regulatory, and Quality Standards," TeamScience Inc., June 1996.

[3] R. Lysakowski and S. Schmidt, "Automating 21st Century Science: The Legal, Regulatory, Technical, and Social Aspects of Electronic Lab Notebooks and Collaborative Computing Systems in R&D." TeamScience Publishing Company, Sudbury, Massachusetts, *in press.*, July 1997.