

# What You Should Know Before Selecting an ELN

Fri, 05/15/2009 - 11:22am by Michael H. Elliott

## What You Should Know Before Selecting an ELN

Electronic laboratory notebooks have evolved into four distinct types of systems



*First-generation ELN system*

click to enlarge

Figure 1: With a first-generation ELN system, scientists continue to make use of applications such as Microsoft Word, Excel, PowerPoint, JMP, Spotfire and other technologies for data aggregation, analysis and reporting outside of the ELN.

In my last article on electronic laboratory notebooks (ELN), I noted that the technology had migrated from use only by visionary users to the broader market in just a few short years.<sup>1</sup> I also stated that there are many different

flavors of products available, from general-purpose systems to those focused on a particular application area. In this article, I will discuss the progression of systems from the late 1990s to today, describing four generations of ELN. Each of these generations is available on the market as a commercial product; the determination of which product is correct for your project is dependent on your organization's ELN vision and strategy.

### **Type 1: Basic data capture**

First-generation systems are relatively non-invasive to a laboratory's operation. Designed for basic data capture and intellectual property retention, they are the closest to a pure paper replacement. These early products are said to be generic or non-specific<sup>2</sup> because they do not provide any particular domain functionality; nor do they support any specific laboratory workflow.

As illustrated by Figure 1, scientists continue to make use of applications such as Microsoft Word, Excel, PowerPoint, JMP, Spotfire and other technologies for data aggregation, analysis and reporting outside of the ELN. Entering data into the ELN is analogous to cutting reports and gluing them into a bound laboratory notebook; data is pasted to an Office application and uploaded to the ELN — or printed to portable document format (PDF) and then imported. The information contained in the ELN is static, encapsulated in binary objects, making it difficult for cross-experimental data analysis. Also, there is a limited amount of metadata for each experiment, e.g., project number, compound id, notebook number, to facilitate searching and record retrieval.

The primary objective of this generation is to satisfy the organization's legal counsel, since a consistent mode of intellectual property (IP) storage, clarity of notebook content and the assurance of record retrieval simplify the chore of patent filing and the risks associated with electronic discovery. In this configuration

- Assay data may still be uploaded to a data warehouse for project-level analysis and mining.

- Study reports and presentations are manually created and stored in an enterprise collaboration system for access by a project team.
- Raw data could be stored in a scientific data management system (SDMS), but flat pictorial representations of chromatograms or mass spectra are pasted into the ELN document or presentation.
- Workflow and tracking could be supported by a laboratory information management systems (LIMS).
- The fundamental capabilities of generation one are listed in Table 1.

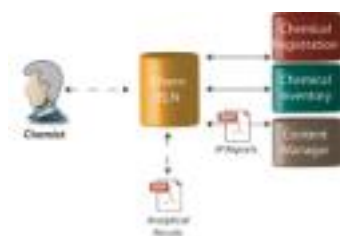
There are pros and cons with this generation. To some, the non-intrusive nature of the technology into the laboratory workflow is a major plus. Implementation is straightforward, with little change to current data management practices. The corporate attorneys are content, and the costs of dealing with bound notebooks are reduced. Life goes on as normal.

A contrary argument is made that the system does not change how data are managed. Practices that may have been inefficient in the first place remain, like personalized formats, nomenclatures or processes. The drudgery of copy/gluing into a paper notebook has been replaced with the labor of copy/pasting into the ELN. Not much was really advanced in terms of efficiency or productivity. Additionally, data management is oriented toward storage of information in a document format, restricting data reuse and repurposing.

The net result is that generation one systems never really caught on broadly in large biopharmaceutical companies, the largest consumers of ELN. These systems rarely answered the “What’s in it for me?” question from scientists and engineers, resulting in several supplier exits from the market. There also are a wide number of options available to manage electronic records in this manner. For example, Microsoft SharePoint is increasingly being turned to for those desiring to implement an ELN with no need for scientific domain-specificity.

## Type 2: Specific solutions

Around 2001, a few years after the first-generation technology was introduced, systems from companies such as VelQuest, MDL, Synthematrix and Intellichem appeared on the market. These solutions did not start out to be ELNs at all — VelQuest was designed for GMP/GLP procedure execution, Synthematrix and Elsevier MDL products were reaction planning tools for synthetic chemistry research, and Intellichem targeted process research. (Note: Synthematrix, Intellichem and MDL have since been acquired by Symyx Technologies.) In fact, most of these companies tried quite intensely to avoid using the ELN moniker!



*Second-generation chemistry-specific configuration*

click to enlarge

Figure 2: Typical second-generation chemistry-specific configuration

Driven by the perceptions of customers as notebook replacement systems, these products began to be recognized as specific ELN solutions. Their domain functionality enhanced users' efficiency by providing new tools previously unavailable. In the case of the chemistry products, features like a searchable reaction database, automated stoichiometry and reaction planning answered chemists' "What's in it for me?" question. In the case of VelQuest, it was the automation of standard procedures for cGxP, lowering company's costs of compliance and, where possible, the automation of manual tasks. However, users did not want to upload a finished report or experiment record into yet another system like a

generation one ELN. They wanted these systems to have the same features of security, data management, search and electronic signatures. Lacking another category to define them, these products were considered "ELNs" despite limitations in their ability to be leveraged across other workflows like biology.

A typical chemistry-specific configuration is shown in Figure 2. It should be noted that

- The user works within the ELN for reaction searching, planning and procedure documentation.
- Forms automatically calculate and update stoichiometry data.
- Through systems integration, users register a product against a corporate registration system, returning a compound ID number to the ELN.
- An inventory system is also linked to update reagents in the reaction.
- Reports may be printed as a PDF and sent to a content manager or printed to paper and attested with a "wet" signature.
- In the early part of the decade, company lawyers required users to print research records to paper in this hybrid configuration, using the system for laboratory efficiency, but not for IP documentation retention.

Versus the first-generation technology, chemical data are dynamic, as structures and reactions can be searched and reused. Retrosynthetic pathways can be dynamically built showing relationships between multiple reactions and their products. In the case of procedure execution systems, regulatory reports can be automated and data posted to LIMS. These powerful tools did much to change the efficiency of the laboratory information process.

Because they provided tangible benefits to the end user, these early systems are responsible for the rise and acceptance of ELN as an informatics technology. Early adopters could envision an electronic laboratory environment that scientists could buy into. Pragmatists understood the advantages as soon as they started to use

the product.

There were downsides to this approach. As companies desired to expand utilization of the ELN concept beyond just one domain, they were hampered by the lack of a generic component or application specificity beyond the installed area. This limited their enterprise reach.

### **Type 3: Expanded capabilities**

Third-generation systems soon emerged to enable the sophisticated capabilities of the domain-specific solutions, protect IP like generation one, and allow expansion into other domains. Pioneered by CambridgeSoft and their E-Notebook ELN, research was targeted with explicit chemistry functionality in combination with a generic component. Synthetic chemists work exactly as they would with the second-generation chemistry systems, and others, such as biologists, use existing tools and import files into the ELN. This flexibility enabled CambridgeSoft to quickly establish itself as the overall market in units and sales.



*Fourth-generation ELN system*

click to enlarge

Figure 3: Fourth-generation ELN systems feature LIMS-type capabilities provided for workflow automation, data analysis and visualization.

---

Based on user input, the generic capabilities were expanded much beyond generation one. Clients did not want to constantly write up experiments in Microsoft Word and upload them to the ELN. They desired to work within the ELN to simplify the user experience. In response,

wrappers were placed around MS Office so data entry could be made directly into Word or Excel and existing templates could be reused. Or, vendors built rich text format (RTF) editors so users could document and annotate research from within the system without the need for a locally installed version of MS Office. Tools to create reusable forms and templates provided a consistent mode of data entry, resulting in expansion of metadata tagging and searching. As the ELN moved across departments, integration expanded to include other technologies like SDMS.

These architectures achieved project goals of productivity enhancement and IP protection. Patent preparation was expedited, collaboration was enhanced, and institutional knowledge was preserved. Users documented efficiency gains of over 20 percent.<sup>3</sup> The changes to the U.S. Federal Rules of Civil Procedure (FRCP) in 2006 lessened worries about the acceptance of electronic signatures and IP storage preservation in the courts.<sup>4</sup> These changes allowing electronic discovery lowered ELN resistance barriers and made the hybrid deployment a thing of the past.

Nevertheless, some organizations felt there were downsides to this approach.

- In R&D, departments beyond chemistry started to demand application additions to provide the same level domain-specificity.
- Users wanted tools for personal productivity.
- Managers wanted improved laboratory efficiency and cycle time.
- In some cases, the result was a heavily customized system with one-off modules.

#### **Type 4: Converging functionalities**

Third-generation architectures faced another challenge: a segment of users craved a single solution to manage structured data and automation of the laboratory's information workflow as additions to ELN. The desired capabilities augmented the ELN's unstructured data

management, crossing the line into what has been traditionally considered LIMS territory. Resultantly, a new generation of ELN emerged, converging functionalities seen in ELN, LIMS and even SDMS.

LIMS in discovery research has never been particularly successful, except when applied to standardized processes like high throughput screening. Few commercial LIMS products have the flexibility research users need to adapt to frequent changes in protocols and study designs. Even fewer products include the functionality for experiment design and documentation found in an ELN. Nevertheless, LIMS manages structured data, tracking and automation tasks very well.

Third-generation ELNs have the inherent flexibility users require, but they handle structured data and task automation poorly in terms of cross-experimental analysis and data mining outside of synthetic chemistry. This design has strong appeal to those who view ELN in terms of a system for only experiment documentation, but less so for those confronted with growing mounds of structured data. Building complex data models or leveraging semantic relationships is quite tasking if data are encased in Excel binary objects. In particular, biology research domains suggested new approaches; particularly those with structured workflows like drug metabolism/pharmacokinetics (DMPK), pharmacology or toxicology.

To address these needs, fourth-generation systems appeared on the market. A fourth-generation ELN is a framework incorporating generic functionality with pluggable domain-specific modules and extensions built using reusable components or services to facilitate unstructured and structured data management. LIMS-type capabilities are provided for workflow automation, data analysis and visualization (Figure 3). Results data from instrumentation can be parsed and uploaded in addition to raw data capture and storage, analogous to an SDMS. These additional functionalities can be provided by the ELN



supplier or through incorporating services from third parties. The concept is that modules can be mixed and matched to adapt to domain-specific workflow automation and data management, creating what Geoffrey Moore in his book on technology adoption Crossing the Chasm calls the “whole product” to increase appeal to “main street” pragmatic users.<sup>5</sup>

This generation was pioneered by Rescentris with the Semantic Web architecture of their CERF ELN and by IDBS with their BioBook and ChemBook extensions utilizing the Quantrix multi-dimensional data modeling and analytics platform. Most of the major ELN suppliers now are heading in this same direction, including Symyx taking advantage of the Isentris framework obtained through their MDL acquisition and Agilent exploiting Accelrys’ Pipeline Pilot for data process automation. Recognizing the market interest in converged solutions for unstructured and structure data management, STARLIMS announced an ELN product at Pittcon this year, and Thermo Fisher recently publicized a new relationship with Symyx, reselling that company’s ELN in conjunction with their LIMS products.

There are benefits to both end user and suppliers of the fourth generation approach. The user experience can be simplified, as end users can benefit from a single platform to execute laboratory data processes. Workflow can be simplified and tools are provided to expedite data analysis and reporting, both increasing laboratory efficiency. From the vendor perspective, it is a less complicated architecture to support and reusable components expedite module delivery.

But, as with anything else, there are downsides.

Specifically

- These additional modules and new concepts in data reuse can radically change work habits.
- To effectively modify processes to adjust to the enhancement opportunity, an increase in business analysis time is required.
- Users must accept a common suite of software rather

than using what they personally feel is the best product.

- There is a sizable portion of the market that is highly resistant to change, which can impact adoption.
- Another major challenge is that the “best of breed” for a particular capability may be a product other than the embedded application, necessitating integration.

### **View to the future**

Suppliers will find it increasingly difficult to be all-encompassing, necessitating more partnerships and off-the-shelf product assimilations in the future. There also are increasing demands to bridge information silos across departments. As this trend continues, it could be said that the next generation of ELN will be even more open and thinner, akin to a portal with the ability to transpose data between best-in-class commercial applications in a framework we refer to as the “scientist’s desktop.”

Conceptually, this holistic approach intends to broaden ELN into a view of the overall enterprise scientific information architecture, lowering barriers between internal departments and outside collaborators. Microsoft has prototyped this model with their “Scientist Workbench,” demonstrating a perspective to be watched closely over the next few years.

In summary, there is a wide range of options now available to potential users of ELN. The path you take depends on an analysis of your operations, desired future direction, and how you wish to best make use of your data and information. For those wishing only to retain intellectual property and/or document experiments without giving up current tools for data aggregation and analysis, there are a wide number of choices, from ELN suppliers to generally available content management systems, such as Microsoft SharePoint or Open Text ECM. For those needing an integrated solution encompassing domain-specific modules, the choices are more complex. The newest generation of products crossing over the historical demarcations of ELN, LIMS and SDMS offer interesting new perspectives to laboratory data management. As suppliers try different approaches to complex problems,

careful attention should be paid to your use cases and how you can best exploit data to meet your strategic goals and objectives.

## References

1. Michael H. Elliott, "Electronic Laboratory Notebooks Enter Mainstream Informatics," *Scientific Computing*, November 2008
2. Michael H. Elliott, "Are ELNs Really Notebooks?" *Scientific Computing*, July 2004
3. Michael H. Elliott, "It's Not About the Paper — The Benefits of Electronic Laboratory Notebooks," *Scientific Computing*, September 2004
4. Michael H. Elliott, "The Rules have Changed: The Management of Electronic Research Records is More Important Than Ever," *Scientific Computing*, July 2007
5. Geoffrey Moore, *Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers*, Harper Business Essentials, 1991 revised 1999

*Michael Elliott is CEO of Atrium Research & Consulting. He may be reached at [editor@ScientificComputing.com](mailto:editor@ScientificComputing.com).*

## Table 1: Fundamental Capabilities of Generation One ELNs

Data management and organization are supported by a relational database management system.

- A simple data model enables storage, metadata tagging and an organizational hierarchy (i.e., taxonomy).
- Data are most often stored as encrypted binary objects in database tables.
- Through a function like a hash digest, multiple record versions are tracked in the database.
- A database audit trail maintains a history of record access, entry, modification and deletion.

Security functions exist to protect and restrict access to records.

- There are utilities to assign access rights (e.g., read, write, modify, delete) to individual experiments or projects and restrict system functions.

- There also might be components for record authentication.

Search features facilitate record access through a query.

- There can be conditional queries with Boolean operators to search metadata tags assigned to experiment records.
- A full text index provides a “Google-like” experience, finding keywords or strings within documents.
- A basic report utility enables individual experiments to be downloaded in a format like PDF or printed to paper.
- E-signatures are an electronic means of author and witness attestation.
- There are a number of possible options, from a simple electronic signature with a database level identifier to sophisticated digital signatures using certificates and authorities.

### Acronyms

**DMPK** Drug Metabolism/Pharmacokinetics | **ELN** Electronic Laboratory Notebook | **FRCP** U.S. Federal Rules of Civil Procedure | **IP** Intellectual Property | **LIMS** Laboratory Information Management System | **PDF** Portable Document Format | **RTF** Rich Text Format | **SDMS** Scientific Data Management System

Data Mining

## RELATED READS



Cycle Computing Debuts Newest Version of Groundbreaking

Cyclecloud

A Spatiotemporal Model to Unlock The Secrets of the



Human Heart



Unusual  
Quantum Liquid  
on Crystal  
Surface Could

Inspire Future Electronics



Programmable  
Network  
Routers

0 Comments

Scientific Computing

1 Login ▾

♥ Recommend

🔗 Share

Sort by Best ▾



Start the discussion...

Be the first to comment.

ALSO ON SCIENTIFIC COMPUTING

### Construction of World's Most Sensitive Dark Detector Moves Forward

2 comments • 3 months ago •



**Mark Lund** — This is a very exciting endeavor, certainly as important as gravity waves in the way that it will affect the future

### New Class of Materials Could Realize Quantum Computers

2 comments • 2 months ago •



**sparcboy** — "...but no for the first time are experimentally identified." "no"? Maybe, "now". No worries, it's not a quantum

### Wireless 'Data Center on a Chip' Aims to Cut Energy Use

1 comment • 3 months ago •



**wagnertinatlanta** — Often requiring air conditioners to cool their many processors, data centers consumed about 91 billion

### Giant Machine Shows How a Computer Works

1 comment • a month ago •



**TheOldMan** — Where are the squirrels?