Integrating the Drug Discovery Laboratory

Published in Scientific Computing – May 2006

Michael H Elliott

The Need for Integration

One of the many challenges biopharmaceutical companies face is improving the throughput of drug discovery. Despite vast sums of investment, the US FDA in 2005 approved a total of 20[i] new molecular entities (NME), the same number as in 1974[ii] . In 2004, R&D spending by members of the Pharmaceutical Research and Manufacturers Association (PhRMA) was US$ 38.8 billion, an increase of 12.5% over 2003 and a sizable leap from the inflation-adjusted US$ 4.1 billion spent in 1974.[iii] [iv]

Based on a study in 2004[v], approximately 20% of discovery chemists' time is spent in nonproductive activities such as manual data transcription, merging data from various sources for analysis, and writing reports and notebook entries. In essence, the scientist has become the "data integrator," performing tasks that take time from designing and performing experiments. There are a number of reasons for this inefficient use of valuable resources, but the large number of autonomous informatics systems and paper laboratory notebooks are chiefly among them.

It is ironic that information technology has created many of the problems that information technology can solve. Driven by the need for discrete solutions to specific problems, stand-alone applications and databases have proliferated in R&D. These systems address chemical inventory, registration, bioassay data management, data analysis, sample tracking, chemical reaction planning, structure activity relationships, toxicity and so forth. Additionally, external databases of publications, abstracts, reactions, genomics and proteomics data increase the number of systems a scientist must navigate. Unfortunately, most of these have different security architectures, user interfaces, and data formats, each requiring discrete training, administration and management.

These information silos provide solutions to their individual challenges. As the number of systems increases, so does the aggravation level of both the user and the IT professional. In our latest survey[vi], the number one response from discovery scientists to the question: "What do you see as your primary information / data management challenge?" was "I have to manually integrate data from

multiple sources." Not only are there many systems for the scientists to use, but IT departments have to maintain an increasing number of systems with fewer resources. A head of advanced technology at a major pharmaceutical company says they have "120 active systems throughout drug discovery, with a goal to reduce that number significantly," due solely to the costs associated with system maintenance.
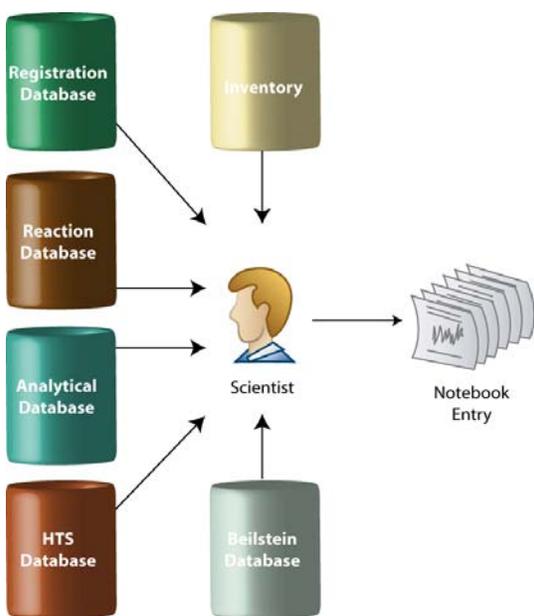
The scientist as the data integrator

A trend in discovery informatics is to provide different levels of system and data integration to streamline workflow and operations. "Integration" is the latest market buzzword and means different things to different people. The scientist's perspective is often one of obtaining the right information when it is needed and logging data where it is required – all with minimal effort. More time should be spent on science and less on trying to

remember how to navigate through the maze of multiple software applications.

The ELN as Integrator

Over 80% of discovery groups still use a paper laboratory notebook as their final data repository[vi]. For these users, they must collect data from many sources, manually transcribe it into a notebook and/or literally cut and paste software printouts onto a page. This is rapidly changing, as Electronic Laboratory Notebook technology is maturing to not only replace its paper equivalent, but to act as the central focal point of scientific activity integrating various systems in the enterprise through a common interface. The goal of many ELN projects is for the scientist to not even know (or care), where information is stored. They should be able to access relevant information in real time when necessary – simply and easily.

An early trail-blazer using the ELN as central integration tool for the scientists, Novo Nordisk (Bagsværd, Denmark) custom developed their ELN almost six years ago. Their ELN was programmed internally due to the lack of mature commercially available solutions at the time. "Our development philosophy was "write once, update everywhere", says Gorm Kruse, department manager of scientific computing for Novo Nordisk. "We wanted to have a single environment for the user to update multiple previously non-integrated databases. We

eliminated many of the various systems the technician and scientist had to interact with."

Built on a two-tier client/server architecture, the Novo Nordisk ELN started in medicinal chemistry and now has over 300 users across high throughput screening, large molecule biology, and pharmacology. Their ELN interface provides integration to various databases in use at the company. When entering experiment details or performing database lookups – for example a compound registration number – the scientist has only one interface to navigate. All databases requiring updates are made contemporaneously, negating the user having to log in and update other systems in the enterprise.

Kruse says that they "were surprised by the broad acceptance of the ELN." He relates one example where a technician "said she would go on pension before she used the system." Once she started using the ELN, however, she immediately became an advocate. "She realized the system made it much easier for her to do her job, since she didn't have to work with so many other databases." Scientists are now aware of what others are working on, which reinforces the Novo Nordisk culture of collaboration and information sharing. The system provides many other unforeseen benefits. Unnecessary trips to the stockroom, for example, can be eliminated by a simple search for chemicals used by nearby scientists.

The company has now deployed the system in 80% of targeted discovery functions and will expand the system to smaller geographically dispersed laboratories. A future goal is to become completely electronic, eliminating the paper records that are now used for the support of patents.

AstraZeneca R&D Mölndal (Mölndal, Sweden) was another early adopter of ELN technology for discovery, working with Elsevier MDL's (San Leandro, California) consulting group in the late 1990s to develop their system based on the company's LabJournal ELN. LabJournal was developed by MDL's professional services as a starting point for custom ELN projects. AstraZeneca R&D Mölndal went live with the system in 1999 and now has over 120 scientists who use it on a daily basis.

"Our goal for our project was to have a searchable reaction database," said Ingrid Hansson, project manager for the ELN installation. To make historical reaction information available to all the users, AstraZeneca R&D Mölndal inserted data into the system from paper notebooks as far back as 1950. These paper notebooks, though diligently cataloged, were rarely accessed – as is the case in most companies - due to the effort involved in retrieval and handwriting interpretation.

To facilitate the entry of historical information, Hansson hired summer interns from a local college. These interns entered data into the ELN from the vast collection of past notebooks, including electronically re-creating hand drawn structures using MDL's ISIS Draw. A double entry and QA review process was used to ensure the integrity of the data. "The quality of the data was critical. We wanted to make sure our chemists saw immediate value from using the system," says Hansson. "This was not an easy task, as it took almost five years to log in over 55,000 journal records, but it was well worth it." Hansson cited several instances where accessing past experiments was crucial to streamlining synthesis design. For example, one of the chemists could not find a reaction from any available public source that produced a yield greater than 60% for a compound he was assigned to synthesize. By searching the newly electronic historical records, he discovered a reaction for the same compound from the 1960s that produced a yield greater than 98%, saving considerable trial and error.

In addition to the reaction database, Hansson's system has interfaces to chemical inventory and compound registration services. From the ELN, the user can register a compound with a single mouse click. This will return the assigned compound number which is logged with the ELN experiment record. When designing the reaction in the ELN, the user can select the appropriate compounds or reagents from the inventory system. The approved inventory names and numbers will be attached to the experiment record, ensuring consistency between scientists. Researchers can then easily gather stockroom materials identified by their shelf location.

AstraZeneca R&D Mölndal has recently worked with MDL to upgrade their custom ELN to the Isentris framework. Isentris is MDL's enterprise backbone for integrating databases and systems in drug discovery. The objective of Isentris is to provide central access to manage data, users and applications into a coordinated workflow. Between the researcher and any integrated databases, Isentris provides a translation layer for data consolidation and a common user interface. This shields the researcher from having to reformat complex data from the various sources so it can be easily accessed and queried. A single query using a chemical structure, for example, could retrieve the related reports from EMC's Documentum enterprise content management system and activity data from a bioassay database.

MDL recently announced the initial shipments of MDL Notebook, which is the company's "off-the-shelf" ELN targeting drug discovery built on the Isentris foundation. The software has functionality for discrete or parallel chemical synthesis,

IUPAC name generation, collaboration and experiment documentation.

Simplifying Data Analysis

Data analysis is another area where companies are turning to technology to simplify data codification. In the past few years, process automation software targeting the life sciences has appeared from companies like Inforsense (London, U.K.), Incogen (Williamsburg, VA), Teranode (Seattle, WA), and White Carbon (Melbourn, U.K.). These service-oriented systems, built with workflow engines at their core, allow data to be integrated and analyzed from multiple sources. Through a graphical "drag and drop" interface, workflows can be created for data mining and assembly, laboratory process automation, analysis, simulation, and report creation.

Curis, a Cambridge, MA biotechnology company, has been using Inforsense's Knowledge Discovery Environment ("KDE") software for almost three years. Curis' discovery operation uses a variety of informatics tools and databases developed internally and from MDL, Cellomics, Chemaxon and other suppliers. According to Dr. Douglas Barker, Curis' senior director of informatics, the use of process automation software offers his team "great flexibility" to combine data from diverse systems.

Barker uses the tool for a number of applications. For Curis' bioluminescent luciferase screening assays, Barker uses KDE to link up with their chemical registration system to acquire compound identifiers to merge with screening data stored in a bioassay database. The data are then sorted and filtered, and the related chemical structures are reported using the integrated link with MDL's chemical searching cartridge. In another area, data from whole-cell fluorescence assays are acquired and stored in Oracle using Cellomic's ArrayScan HCS system. Percent activity or inhibition data are extracted and analyzed by KDE from the Cellomics database, with the results imported automatically into the bioassay database, saving hours over manual methods.

Barker says that "the flexibility of the system reminds me of Excel. Once you are familiar with how to use it, it's the first tool you turn to integrate and analyze data." Curis now uses KDE as a base architecture for all integration projects and has a number of ideas for other applications. Barker does have advice for those considering technology of this type, "Focus on how it will integrate with what you already have." He warns prospective users not to try and "reinvent everything." In other words, the baby should not be thrown out with the bath water!

Conclusion

The future of the discovery laboratory is unknown, but changes are occurring to remove the barriers between the various silos

of information that have propagated in the last decade. The application of a rational approach to workflow automation can increase the time for science and alleviate much of the inefficient use of expensive scientific resources. The hope is that this time can now be used to discover new medications to improve the human condition.

References

[i] Berenson, Alex, *Drugs in '05 Much Promise, Little Payoff*, New York Times, January 11, 2005

[ii] Source: US FDA

[iii] Source: Pharmaceutical Research and Manufacturers Association

[iv] Source: InflationData.com

[v] *Electronic Laboratory Notebooks: A Foundation for Scientific Knowledge Management*, Atrium Research, Wilton CT

[vi] *2006 ELN Survey*, Atrium Research, Wilton CT

Michael H Elliott is president of Atrium Research, a market research company focused on scientific informatics. He is a member of the editorial review board of Scientific Computing. He can be reached at info@atriumresearch.com www.atriumresearch.com