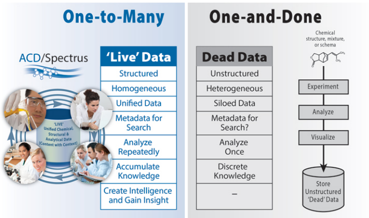
Creating Intelligence to Gain Insight

Ryan Sasaki, Director of Global Strategy; Advanced Chemistry Development Inc., Toronto, Canada



click to enlarge Figure 1: With the ULI intelligence-from-information live cycle, organizations can benefit from the one-to-many relationships between the data and scientist interaction (left). Currently the one-and-done chemical data life cycle dominates laboratory environments based on large amounts of dead data.

Despite the emergence and adoption of software technologies such as electronic laboratory notebooks (ELN) and laboratory information management systems (LIMS), the collection and usage of data in a paperless environment has yet to optimize specific laboratory processes. A 2011 survey of R&D professionals revealed "88% of R&D organizations lack adequate systems to automatically collect data for reporting, analysis, and decision-making."[1] To understand the results of this survey one must investigate the existing informatics landscape to try to understand why significant investments made in different technology areas are not helping to address this gap.

A 2010 report from Accenture, Shaping the Future of Pharmaceutical R&D: Executing a Vision of High Performance, presents a vision of how R&D informatics can move well beyond data management. "Today, most of the focus of R&D informatics lies in executing traditional and noncore activities, including maintaining current systems, enhancing existing systems, filling technology gaps, and implementing technologies that will supplement existing processes and create further efficiencies." Most of this current work deals with the management of data and information, and in some cases it results in the creation of knowledge. Yet very little of it rises to the level of generating insight. By contrast, there is tremendous value to be gained by focusing on the D2i [Data to Insights] capabilities that enable organizations to generate valuable insights from existing and new sources of data.[2]

Unified laboratory intelligence (ULI) is a technology framework focused exclusively on improving chemical identification, characterization, and optimization for life science, chemical, energy, consumer, industrial, and technology industries. With this technology comes the ability to apply chemical context (why) to the vast amounts of analytical content (what) generated on a daily basis in the laboratory. The ability to collect and unify chemical, structural, and analytical data creates the foundation of an intelligence-from-information "live" cycle. This enables scientists to search, retrieve, and reapply live data throughout the research and development process and to help address the gap acknowledged in the aforementioned survey.

No more one-and-done

While the drive towards better productivity through better access is the one of the themes being used to justify the implementation of new informatics capabilities, more and more often researchers are claiming that it's the application of the data, not the data itself, which needs to be leveraged more within the R&D process. Jay Galeota, Senior Vice-President, Strategy and Business Development, Merck & Co., states, "The most important thing is what you actually do with the data. It's one thing to have interesting information, but it's the insights that are important to guide smarter, better decisions."[3]

ULI combines software, algorithmic tools, and databases to create an opportunity for one-to-many data insights throughout the drug discovery and development process. Chemical, structural, and analytical information is homogenized and unified as live data. Intelligent algorithmic loops learn—which improves future performance—as they analyze new and existing information and convert it to knowledge. Finally, this knowledge can be retrieved and reanalyzed by scientists across the discovery and development process to test new hypotheses without having to conduct additional laboratory experiments. The end result is a one-to-many relationship between the data generated in the lab and the scientists; and forms the foundation for the ULI intelligence-from-information live cycle (Figure 1).

To fully comprehend the value of the ULI live cycle, contrast it with the typical one-and-done chemical data life cycle (Figure 1). Beginning with a chemical structure or schema, initial information is generated from a specific experiment and is analyzed to provide specific knowledge about a specific characteristic of the structure or schema. This data is then usually converted into an unstructured "dead" format, such as a .pdf, tagged with relevant metadata, or a raw data instrument file, and stored in an electronic record or archive. This process creates a one-and-done, information-to-knowledge life cycle, using data once and then effectively freezing it in an unstructured format. This typically ends the life cycle, and the data is no longer useful beyond its specific initial purpose. If additional hypotheses need to be investigated, additional, and often repetitive, experiments are conducted which represent costly, time-consuming, and very inefficient activities. This approach can add significant cost and time over the course of the drug discovery and development cycle. And while experiment repetition can generate multiple instances of discrete knowledge, the one-to-one information-to-knowledge life cycle rarely, if ever, generates insights.

Non-redundant value

The ULI one-to-many live cycle has significant advantages over the typical one-and-done life cycle in its ability to address opportunities for R&D productivity improvement. Regardless of the potential improvement in productivity, it's unlikely any R&D organization would adopt a system that requires wholesale changes in existing hardware, software, and methods. ULI is not a replacement for other R&D informatics products, such as a laboratory information management system (LIMS) or an electronic laboratory notebook (ELN). Rather it can complement both while providing one-to-many chemical and data insights that neither a LIMS nor an ELN can deliver.

A LIMS can manage samples, experiments, and results as structured data in an analytical lab. Most LIMS are deployed for a specific analytical lab (e.g., chromatography or spectrometry), though some can be deployed across multiple labs. LIMS usually do not incorporate secondary analysis tools; instead the structured data from the LIMS is ported to a separate software tool for analysis. An analytical lab with regulatory requirements for strict sample tracking and reporting may benefit from a LIMS that automates the completely end-to-end laboratory workflow. In this case, a ULI-enabling technology can accept the structured data from the LIMS and homogenize and unify it with information from other analytical labs. In analytical labs without a LIMS, ULI can acquire data directly from different instruments for the same purpose.

An ELN is a software application that was originally designed to replace the paper notebook used by most laboratory scientists. As such, it is used primarily to document research, experiments, and procedures in unstructured data formats, including text documents, spreadsheets, and images, and facilitate more efficient reporting. Many ELNs are optimized for a specific discipline (e.g., synthetic chemistry, analytical chemistry, drug metabolism, and pharmacokinetics, etc.), while some provide general cross-discipline capabilities with discipline specific modules.

While a LIMS can manage structured analytical information, and an ELN can manage a broad range of unstructured information and knowledge, neither provides the capabilities of ULI to collectively enable one-to-many chemical and data insights. This includes the ability to:

· Homogenize and unify chemical, structural, and analytical information from a variety of heterogeneous instruments across multiple labs.
· Analyze information and convert it to knowledge with intelligent algorithms that continually learn and improve performance.
· Visualize knowledge graphically and dynamically with contextual information.
· Aggregate and integrate new and existing ‘live’ data, which can be retrieved and reanalyzed throughout the discovery and development process.

ULI provides a platform for one-to-many chemical and data insights that cannot be achieved by other R&D informatics software, such as LIMS and ELN. However, ULI can be used with LIMS and ELN software as part of an integrated informatics environment to generate data-driven chemical insight throughout the R&D process.

Improving R&D productivity
According to a 2010 report published by McKinsey & Company, The Road to Positive Returns, a major pharmaceutical company’s internal rate of return (IRR) for a typical small-molecule drug was about 12% between 1997 and 2001.[4] Over the decade, the overall probability of success (POS) for small molecules has declined by five percentage points, the time required for R&D has increased by 12 to 18 months, and R&D costs have risen by about 8% annually. As a result, only 30% of drugs

launched earn an acceptable rate of return. The average present IRR is 7.5% for a small-molecule drug, and the net present value (NPV) is minus $65M.

“The trend has been to view diminishing return as a science problem. But while scientific innovation is certainly part of the solution, management should not overlook other, more familiar means of value creation. Increased attention to costs, speed of development, and decision making could increase the internal rate of return (IRR) of an average small molecule from around 7.5%—which is less than the industry’s cost of capital—to 13%.”

The McKinsey & Company report concludes that, “Although the current environment is tougher, managers are not yet fully exploiting the value-creation levers described here [attention to costs, speed of development, and decision making], and moderate improvements can substantially increase returns.” Data cultures in life sciences are very heterogeneous, and no single approach will suit the needs of everyone. “The most successful strategies are likely to be those that address needs in the context of sub-disciplines.”

ULI can generate data-driven chemical insight, and a successful strategy to achieve moderate improvements in overall R&D productivity can substantially increase returns for small-molecule drugs.

References
1. R&D Informatics: Are you ready for 2012?, http://blog.idbs.com/2011/10/18/rd-informatics- [1]–-are-you-ready-for-2012/, IDBS (October 18, 2011).
2. Shaping the Future of Pharmaceutical R&D: Executing to a Vision of High Performance, Accenture, (2010), 7–8.
3. Progressions, Building Pharma 3.0. Ernst & Young, Global Pharmaceutical Industry Report 2011
4. Eric David, Tony Tramontin and Rodney Zemmel, The Road to Positive Returns, Invention Reinvented: McKinsey perspectives on pharmaceutical R&D, McKinsey & Company (2010), 1.