The De-Evolution of Informatics

How externalization workflow is impacting traditional data management architectures

Michael H. Elliott

The challenges of the pharmaceutical industry to generate new molecular entities while controlling expenses have been well-documented. With over 75 percent of U.S. drug sales now coming in the form of generic equivalents, new models must be tried to stimulate ideation and increase the candidate pipeline. In the last several years, the desire to externalize research and development has been one of the models in vogue. The belief is that an atmosphere of innovation can be created, while at the same time lowering fixed costs. Partnering with contract organizations, academic institutions and other pharmaceutical companies can "virtualize" discovery and enable global collaboration. As is often the case, the devil is in the details; the information architectures required to allow the seamless flow of data between these multiple entities lag far behind the progression of the business model.

Most of the early work in externalization was in late stage development (e.g., safety assessment) and the clinic, where it was a natural fit to make costs variable with the ebb and

flow of new drug substances. Recently, companies have taken a much more aggressive approach to externalizing drug discovery; some report a target of between 30 to 50 percent of discovery work to be through partner alliances. A full range of arrangements are being pursued, from university collaborations for target identification, to medicinal chemistry with partners in China, to *in vitro* screening shops in India.

Early on, the majority of these arrangements were scattered across a wide number of partners to experiment with new relationships with a clear focus on cost minimization. Over the last two years, we are seeing a movement where there is more work with fewer partners. "Risk sharing" arrangements are trending upward; this is where a partner provides a fuller gamut of horizontal services from synthesis all the way to *in vivo* pharmacokinetic and pharmacology studies. Payments are on achieving milestones rather than purely on the number of assays performed or number of molecules synthesized.

Currently, about 30 percent of biopharmaceutical R&D is spent beyond company boundaries. The overall rate of R&D spending is flat, but outside spending on discovery research is increasing at close to 20 percent per year. If these trends hold, we will see 50 percent of R&D externalized in just a few short years. The race for the business side of these companies to outsource "non-core" activities is coming at a cost: IT plays catch-up and carefully crafted internal data management infrastructures are breaking down. Basically, a "de-evolution" of the informatics landscape.

CHEAPER DOES NOT MEAN FASTER

It all sounds good on paper — triple the number of researchers at the same cost, put new fresh minds on old problems, and allow for program-dependent costing. A major detail habitually overlooked is the most effective way to share and collaborate data and information in this new world. Of the top 20 pharmaceutical companies Atrium Research recently surveyed, only two had a comprehensive externalization data management strategy, though all were engaged in partnerships and collaborations. Eighty percent of these companies are exchanging biology data with partners via e-mail and Excel spreadsheets, despite having robust internal bioassay data management systems, electronic laboratory notebooks (ELN), laboratory information management systems (LIMS) and other very expensive technologies deployed for internal consumption. Two companies with thousands of seats of ELN are sharing chemical reaction designs with partners via paper lab notebooks. Cheaper does not necessarily mean faster, as longer lead times for data analysis slow compound decision making.

It is not just inefficient movement of data that can lead to slower research. Material logistics also comes into play in a virtualized R&D world. Companies have made huge investments in process optimization and workflow systems to quickly turn around biology results from newly synthesized compounds. Internally, a medicinal chemist can hand off material to compound management or directly to biology for plating and *in vitro* assays. Screening results can be generated and posted to databases within two to three days. In an external world, materials must be batched and shipped across the globe resulting in one to two weeks for results. Or, if material is created by a partner and sent to the U.S. for testing, customs comes into play adding additional delays. New IT workflow systems have to be built for requesting, tracking and the creation of customs documentation, adding to the overall costs of the initiative.

A few companies provide access to a select number of internal systems for remote data entry, so why not more? There are a plethora of reasons, not the least of which is that many systems designed for internal consumption seldom play well when exposed externally. There are issues like identity management, compatibility with terminal services, and system performance over a wide area network. Chemical registration databases most likely do not have the row-level security to filter only select compounds to specific partners. A contractor who has many clients may not even want to use your systems or they might have their own technology. Not to mention the security worries of giving contactors — some of which you have never met — access to your internal network. There are also the issues of costs: it takes resources to support the additional users and capital for purchasing new licenses.

If you do not provide access to your systems, then integrating the data generated by a partner becomes the hurdle. When there are common systems across an internal enterprise, IT can enforce a certain set of standards to enable data to be stored in a specific database format. You might be able to get the partner to use your format to enable the ease of data uploading to your systems. Without overall governance, project teams are quite often

left on their own to represent data any way they see fit, leading to a huge level of variability in the Excel spreadsheets and a lack of searching and data comparison. If your contract with a partner is not of sufficient size to warrant customization to your data layout, they will provide data in the layout of their choosing; potentially leading to dozens of different formats. Additionally, even if formats are easily consumed, project teams commonly want an extra level of review of contractor data before it is posted to an internal database, requiring a process for data staging and curation.

I have summarized three different externalization data workflow "types" in Table 1.

- In the Type One scenario, internal systems are externalized and, therefore, used by partners to enter data directly. They may use an ELN to enter chemical reactions, a bioassay database for screening, or chemical inventory for materials management. This is generally a trusted partner relationship where a detailed data review is not required before data upload. The advantage of the Type One dataflow is real-time access, while the downside is the high cost of support.
- A Type Two case is where the partner uses their own systems, but provides data in your approved standard format, allowing for ease of importation. There is no realtime access to monitor workflow, but costs are lower.
- In Type Three data flows, the partner uses their own technology and provides the data in their format. Here, the complexity is increased for the client, as multiple data transformation services will have to be created, but is easiest for the partner. For most large organizations, more than one of the workflows will likely have to be supported, resulting in the need for a tactical plan and informatics architecture to address each scenario.

THE ROLE OF ELN

Is ELN the solution? Since ELN only handles a portion of an organization's data management needs, the answer, for the most part, is no. ELN can play a significant role and does for many companies. It was one of the first categories of systems in discovery to be externalized, deployed initially to support medicinal chemistry outsourcing (a Type One dataflow). There are now hundreds of external-facing ELN clients across major biopharmaceutical companies. Implementations are mainly via Citrix

Table 1: Data Logistics — Externalization Dataflow Types

Type	Description	Pro	Con
One	Trusted partner	Real-time access to workflow	High degree of complexity for partner
	Externalize internal applications	Use of common technology	Systems may not fully support externalization
	No data staging	No data conversion	Global support
			Security risk
Two	Partner uses their own systems or a cloud-based system	No real-time access to workflow if not cloud	Must have import utilities
	Will comply with client data formats	Lower costs of support	Must have process for review/ staging
	May require review before uploading	No data conversion	Medium complexity for partner and client
		Medium complexity for partner and client	
Three	Partner uses their own systems	No real-time access to workflow	Must have data transformation services
	Data output is in their format	Lowest complexity for partner	Must have process for review/ staging
	May require review before uploading		Highest complexity for client

terminal services to reduce costs and eliminate the need for local client installs. But, project managers report that system performance and training are the two biggest issues with externalized ELN. According to Atrium Research's surveys, the number one complaint of ELN users overall is system performance. Combine that with deployment under Citrix, the poor latency of virtual private networks, and organizational firewalls (and countries, in the case of China), and the performance can be rather pitiful.

Type Three data flows can be extremely challenging for ELN, as there are no standards for data exchange between vendor systems. Many initiatives to develop an interchange format have started and stopped, the most recent example of that being the Pistoia Alliance's ELN working group. Companies resort to importing PDF printouts of experiments into their internal ELN. Data are stored, but searching reactions and products via chemical structure is not possible. As contractors grow in capabilities, their desire to use clients' ELNs diminishes and interest in having their own solution escalates. Therefore, over time, Type Three workflows will increase in proportion, creating considerable data integration difficulties for clients.

In any case, identity management is a serious concern for intellectual property and ELN. If one is sued in the U.S., the admittance of records to support your case is essential. If the person who created the records (or attested to them) is not available, then admittance under the Business Records Exception (BRE) of the Federal Rules of Evidence (FRE) is necessary.[1] For BRE, you must have established policies for lab notebooks and prove they are being followed. It is likely you will not be able to compel your contractor in China who signed off on an experiment years ago to appear in court. There are no legal methods of doing so. Therefore, auditing partners for compliance with your notebook (electronic or not) policies is necessary. As with IT systems, this is too often an oversight, with records management personnel also playing catch-up to the expansion of outsourced work. This can place your IP at risk.

Cloud ELN is a natural fit in this environment. True cloud solutions tend to be Web-based, eliminating the need for remote thick clients, and hosting can be distributed to any location for performance optimization. On-boarding a new contractor — or turning off an existing one — is relatively painless. Collaboration is enhanced, as a shared environment enables real-time access and data annotation. I noted two years ago in this publication the resistance to ELN in the cloud.[2] In 2010, Atrium's survey indicated 29 percent will not consider ELN in the cloud while only 20 percent would (remainder was "maybe"). Unfortunately, in 2012,[3] this has worsened. Now, only 13 percent say they would consider using a cloud ELN while those indicating they would not increased to 34 percent. IT is the most resistant — only a tiny percentage of respondents indicated they would consider. The security fears and doubts surrounding the cloud do not seem to be going away anytime soon in the laboratory data management market.

Is a proper cloud solution any less secure than allowing contractors from all over the planet access to your internal network and systems? We feel that global collaboration will make the cloud inevitable across a number of different technology categories. This is already happening with document management solutions in the clinical space. With IT budgetary constraints increasing every year and the expansion of externalization, one or two large companies will be forced to make the leap. And, as what happened with rapid adoption of fully electronic environments (dumping the hybrid ELN model where experiments were printed and hand-signed), a bandwagon effect will drive change throughout the industry.

Those embarking on externalization or the expansion of existing initiatives should carefully consider how the workflow will be enacted in reality and how to support it in the most efficient manner. Business development, labs, IT, legal and records management must be working together to address the short- and long-term strategic needs of the business, including looking at new process models. Without it, the rush to externalize will actually slow down research and put valuable intellectual property — your data — at risk. If the scientific processes need new models, so does informatics. Doing things the same way they were done internally will not work in an externalized world. SC

REFERENCES

1. Elliott, M. H.; ELN Authentication: Navigating a Sea of Options; *Scientific Computing;* August 2011
2. Elliott, M. H.; ELN in the Cloud - Marketing Buzz or Reality?; *Scientific Computing*; June 2010
3. Atrium Research & Consulting LLC, 2012 ELN Survey; Wilton, CT; September 2012

Michael Elliott is CEO of Atrium Research & Consulting. He may be reached at editor@ScientificComputing.com