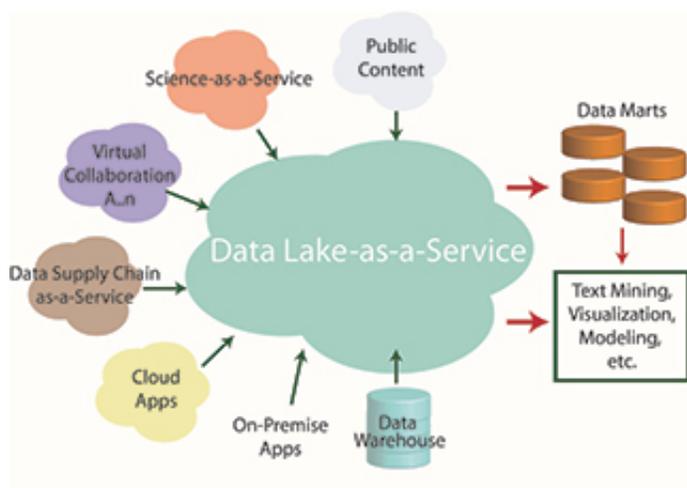# The Future-as-a-Service

Figure 1: Data Lake-as-a-Service

The transformation of conventional R&D operating models is continuing at a rapid pace. The poor track record of preclinical candidates successfully making it to market, rising costs and increasing scientific complexity is forcing biopharmaceutical companies to rethink how they have traditionally executed pipeline projects. It is no longer acceptable to "go it alone" and perform all activities from target identification through clinical trials. Partnering is now in vogue; over 75 percent of biopharmaceutical companies have at least one R&D partner. The expanding ecosystem of research partners is not without substantial challenges, however. Information technology departments in particular are stressed to build effective and agile methodologies to capture, secure, manage and analyze the mounting diversity of data coming from nodes across the collaboration network.

Many have grouped all partner relationships under the generic "externalization" umbrella. It is more nuanced though, as the distribution of research activities is not consistent in drivers and objectives. Depending on internal capabilities, financial requirements and needs, the nature of partnerships can vary widely, even within an organization. Below are just some of the various types of relationships classified by business objective:

- Expense management
  - Outsourcing or "externalization" where tasks considered non-core, such as routine screens and chemical library synthesis, are performed by a contract organization to reduce costs and free internal resources.
- Stimulate innovation
  - **Academic**: Partnerships are formed with multiple universities to research better disease understanding, novel targets, pathways, etcetera.
  - **Crowdsourcing**: Soliciting large numbers of persons and organizations to submit novel compounds.
  - **Scientific expertise**: Partnering with a biotech or specialty pharma that has novel technology that can be applied to in-house programs.
- Risk reduction

- - **Shared risk**: Partner organization is paid for achieving certain milestones to lower the risk of the other.
    - **Co-development**: Two or more organizations work on a program to disperse risk across the participating partners.
  - Functional gap remediation
    - **CRO**: Contract research organization to perform services such as GLP safety assessment and clinical trials for an organization that does not have those internal capabilities.
    - **CMO**: Contract manufacturing for drug substance, drug product, and/or clinical finished goods for a business with no manufacturing capacity.
  - Fill pipeline
    - **In-licensing**: Adding molecular entities, often out of a discovery program of one entity, into the development pipeline of another.
  - Research virtualization
    - A collective of parties working toward particular research objectives. Each has individual expertise and a stake in the success of the others.

The types of relationships are by no means static; companies will continue to explore new operating models. Is it a bit of a shotgun approach? Maybe. Will all models work? Absolutely not. Will some work? Probably. What this vibrant environment does, however, is to create continuous disruption of the functional capabilities required to support partner relationships.

- **Read more**: [Analytical Knowledge Transfer presents a Challenging Landscape in an Externalized World](#)

This chronic change is playing havoc with information technology methodologies and architectures historically designed for internal R&D consumption. IT is often a secondary consideration after partnerships are formed. That is why, even today, the majority of data are transferred in document format, most often by e-mail. In the October 2012 edition of *Scientific Computing* I called this the "De-Evolution of Informatics." In other words, very sophisticated internal systems are being supplanted with PDFs and Excel files, creating ever-increasing stores of "dark data." Data are trapped inside these limited formats and are often so poorly organized (usually in project-specific SharePoint sites), that they are lost for long term preservation and knowledge reuse.

There is slowly appearing light at the end of the tunnel. In the domains that embraced partnering early on, there has been some change since 2012. In safety assessment, the industry standard SEND (Standard for Exchange of Nonclinical Data) is becoming more

commonplace to exchange data, as is CDISC (Clinical Data Interchange Standards Consortium) for clinical trials. The SEND and CDISC formats, though not 100 percent perfect due to nuances between how companies use them, are a model of effective data interchange. Following the same path, there are several other non-profit foundations, such as OpenPHACTS and TranSMART, working on data interchange in pharmacology and translational science. Allotrope Foundation is working on a future framework for laboratory data exchange, and the Pistoia Alliance has several standards initiatives, such as HELM (Hierarchical Editing Language for Macromolecules).

In discovery chemistry, it is now customary for companies to deploy cloud-based Electronic Laboratory Notebooks (ELN) to capture reaction experiments. In the last four years, the number of commercial cloud-ELN products has risen to over a dozen; driven in large part to the increasing demand for externalization. A major challenge remains in that there is no standard for the exchange of ELN records. Companies may have several ELNs they use between different partners, but integration between systems remains a manual process. Additionally, related analytical data is often not considered other than static images copied and pasted to the ELN page. Live data for reactions, but dark data for analytical results.

Nevertheless, the rise in the number of commercial products and standards bodies does not address the majority of challenges organizations face when trying to develop a collaborative informatics ecosystem. Point software solutions and standards only address a piece of the overall puzzle, focusing on the results of the science and leaving out the business and process capabilities. The movement of data between technologies, operational workflow, sample logistics, payments, vendor management and project management are often missing, resulting in a significant level of manual intervention. In addition, the rapid scientific advancement in areas such as next generation sequencing (NGS), precision medicine and next-generation biologics is rapidly increasing the volume, variety and complexity of datasets on a seemingly weekly basis.

Data security classification is proving to be especially problematic. It is not unheard of for a major company to have over 100 R&D collaborators across the entire R&D spectrum. Certain data have to be limited to specific people, some to specific companies, while other data have to be constrained to persons within a specific list of countries. Other agreements stipulate not only who can see data, but also who expressly cannot have access (e.g., a competitor). Restrictions also can change over the life of a project. These new security considerations go beyond modifying traditional data warehouses, since data can be spread by type across a wide number of platforms such as NGS, bioassay, LIMS, ELN, SDMS and document management.

The dynamics in the types of partner relationships, collaboration-specific capabilities and rapid changes in the science creates the prerequisite for agile and adaptive informatics architectures. What you design for today may not be what is essential in the future. It is no wonder IT teams are having a difficult time keeping up.

In an effort to establish a simple solution to a complex problem, a few pharma companies are considering going back in time to project-specific workspaces. Twenty years ago, chemists and biologists worked in project spaces supported by technology such as MDL's ISIS. This was not deemed useful from a strategic perspective, so data warehouses were built to make data broadly accessible, at least internally. The new thinking is that project-specific spaces — with tailored informatics support — might be a way to address the nuances of the collaboration and contain data until such time as they could be released to a central repository.

Future-as-a-Service

Collaboration spaces, whether project-specific or not, are increasingly being supported through cloud services. This is a rather dramatic change from a few years ago where our surveys showed little willingness to manage intellectual property (IP) in the cloud. The accelerated movement toward virtualization along with the belt tightening of IT departments overcame much of the resistance levels. Today, the vast majority of those who are investing in the cloud do so based on private cloud deployments, rather than multitenant software-as-a-service. Nevertheless, this is slowly changing as well, particularly for small biotechs and academic institutions who gravitate to the lower costs of a public cloud solution.

In the future, agility and flexibility will come through cloud-based services. It is inevitable that rigid on-premise applications will not be able to address the ever-changing virtual research environment. Additionally, the budgets for IT departments will be pressured for many years; professionals must look beyond their own walls to bring diverse groups of partners together under one virtual canopy. Collaborators should be on-boarded quickly; new connections must be made within hours, not days or weeks.

The arrival of science-as-a-service is adding to the diversity of cloud solutions to enhance traditional research. These businesses provided a narrow range of technical and scientific expertise for hire. Their numbers are growing strongly, particularly in genomics. They combine analytics-as-a-service (e.g., sequence analysis) with Data-as-a-Service (i.e., data management, collaboration and security), connecting corporate, government and academic entities via the cloud. Other Science-as-a-Service providers are emerging

to address activities as diverse as imaging, protein characterization and *in silico* drug design. Distinction is not from providing assay results at the lowest cost, but through bridging scientific knowledge and data services.

The expanding network of science-as-a-service providers, vendor cloud-based software, and partners who have their own informatics platforms, generates further headaches for informatics departments. There is a need to connect the various systems and their related data formats and security classifications into virtual project spaces. The concept of a data supply chain-as-a-service framework in the future will help to manage the workflow of projects, access the sundry data in a single user experience, and manage projects joining tasks to data. This cloud environment will manage security on a data container basis and enable movement to points in the ecosystem based on its contents. Due to the high variability in formats and terminologies, it may be necessary for machine learning and semantic technologies to provide data interoperability.

Data analytics is necessary to discover new insights across all the virtual project spaces. Moving beyond the paradigm of the structured data warehouse, the concept of a data lake-as-a-service (Figure 1) will enable organizations to have a flexible object-based repository combining internal, partner and public data without restriction to volume or variety. Content in the lake is stored in its native format, combining both structured and unstructured content. Today, data scientists are needed to extract the maximum value out of a data lake. In the future, there will be a plethora of tools for self-service and automated analytics supporting the interactive exploration of data.

As we look to the future, informatics must play a major role in helping organizations collaborate in new and mutually beneficial ways. Information technology can be an enabler of these new business models not only in pharma, but across all R&D environments. To do this, architectures must be more adaptive than they are today. In the years ahead, flexible cloud services will play a larger role in facilitating the flow of data between partners. IT organizations should make the strategic decisions necessary to get ahead of these new operating paradigms, rather than being left to play catch-up.

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