Managing data challenges in the new R&D environment

The discovery and development of drugs is changing. The realm once dominated by 'big pharma' is becoming increasingly fragmented into more agile and flexible consortia. The dramatic change of focus on drug type from small molecule to large molecule, and even large-small molecule conjugates, also brings with it another set of scientific and business challenges. The economics of medicine are shifting too from blockbuster to stratified medicine and diminishing patent portfolios are driving larger organisations to seek new routes for innovation, knowledge augmentation and cost efficiencies. A visible shift has been the rapid increase in research partnerships via collaboration consortia, and more externalisation of those parts of the business deemed as precompetitive. Industry giants are now, in the most part, collaborating with academic institutes, contract research organisations (CROs), contract manufacturing organisations (CMOs) and smaller, focused biotech firms at a scale far greater than ever before.

Historically, many large research and development (R&D) organisations outsourced what they deemed routine work to CROs – ie parts of the process seen to be non-critical to their 'drug innovation' process. Bioanalysis, for example, can be easily outsourced. Other examples included library synthesis, manufacturing and later stage clinical trials. The mindset was that all the knowledge and creative IP should be held internally. Whether this view is right or wrong, the drug pipeline over the past decade has shrunk. The net total of new medicines hitting the market has dropped and the cost of getting a new medicine to market has increased to close to $1 billion. This is unsustainable. The market is simply not able to provide a big enough return on new blockbuster drugs, to allow 'big pharma' to recoup the $1 billion investment and make a profit. Some analysts and the large consultancies have accused the pharma giants of being too big – the economies of scale do not work so well in R&D and being so big these companies are encumbered by big company processes and bureaucracy that limit the kind of innovation and risk-taking seen in smaller, more focused companies and institutes.

The result is that the industry has been forced to accept change – outsourcing and collaboration are the new 'silver bullets'. No part of the R&D process is now deemed too important to outsource or is shielded from efforts to control costs. Big pharma has woken up to the idea that innovation and new ideas can come from anywhere.

By Dr Paul Denny-Gouldson

This is a positive change, but brings a new set of complex opportunities and challenges relating to storing and managing the increasingly distributed and diverse R&D data. This is because companies still need to ensure that what they deliver to market is safe and complies with all the regulatory requirements. Previously, all data and knowledge was stored inside the castle walls. Nothing was allowed out and it was managed centrally – safe, secure, retrievable (to some degree) and protected. IP was secured and defensible, which is particularly important since patents have been the lifeblood of pharmaceuticals for decades. But now, with more stakeholders involved, data and knowledge is originating in multiple locations, multiple organisations and multiple research domains. This represents a new order of collaborative networks charged with delivering results, and highlights the evolving set of challenges facing R&D – a blend of old and new hurdles the industry must overcome.

Data security

Working within a collaborative network means opening up the knowledge of what is going on to many more people. It requires security and access permissions be defined and, more importantly, managed in a simple and dynamic way that does not interfere with the business process. The nature of these collaboration networks means that the old way of working – centrally-controlled security managed through the research informatics or IT groups – is not tenable. Consider the implications of a company's collaborations reaching 200 other

organisations, each with five to 10 people in them. That is 1,000-2,000 users of systems who do not have an internal ID, do not have internal active directory accounts, do not have internal emails and have not signed the internal IP and data security contract.

Furthermore, the nature of these collaborations can be transient and the business needs immediate. When an organisation needs a piece of work done, it needs it now, not in two months once all the contracts and IT infrastructure are put in place. The way collaborators are engaged, contracted and their data gathered must be simple, secure and, above all, managed inside the business by the researchers – not by research informatics or core IT.

This brings a whole set of challenges. Who is allowed to join a collaboration? How does an organisation share the data with them that they need to do the work? This is simple if it is a paid-for testing service where you send a sample; you simply define the test protocol and get the results. But when a relationship is truly collaborative, the sharing of information occurs both ways and there is an iterative definition of what to do next. This fosters the true development of new ideas and concepts but also means that the work environment needs to be able to support this.

IP

Collaboration delivers new knowledge, concepts and ideas. The critical question is that once captured, how are these ideas protected and managed with respect to intellectual property (IP). Who paid for the work, and how are the results to be

shared? It makes the management of IP critical. The collaborative IP must be aggregated with the internal IP so that it can be leveraged and managed. The collaborators will also want to know that any joint IP is protected effectively and can be repatriated back to their organisation, should it be required. The externalised work is worthless if the IP is not captured and managed effectively.

As it stands, the best case scenario for most organisations sees R&D data captured in systems such as electronic lab notebooks, lab execution systems, product lifecycle management software or a scientific data management system. Much of the data is still captured in databases, home-grown tools built by the scientist, or even Word and Excel. These remain efficient methods for storing and capturing the data generated, and ensuring its provenance is understood. Electronic signatures are used to establish non repudiation and date of creation. However, capturing and storing is only one half of the IP story – the other half is leveraging the IP.

We will continue to see the legal departments of pharmaceutical firms grow, as they take on more ownership and tasks related with IP protection and utilisation. Clear contractual definitions and onboarding of collaborators falls squarely in their field of expertise. Crucially, they must provide frameworks and processes that do not get in the way of the business doing what it needs to do, but still provide enough legal cover to ensure the company's interests are protected.

Data consistency

To derive value, companies must ensure that shared data can be accessed, aggregated and used with other data to drive the decision-making process. Submitting non-conformant data must be flagged at point of entry – not post-entry. Business rules need to be managed and defined to support data collection and delivery tools.

Traditional internal business practices see all members of an organisation use the same project codes, target IDs, active ingredient ID and the same data types and definition names. This centralised approach to ensuring consistency of terminology

and syntax becomes infinitely more complex and difficult to administer and control when most of the people doing the work and capturing data are not internal.

The challenge then becomes one of balancing an organisation's need to have data checked and validated at the point of entry, with a collaborator's need to manage time and cost. Having to qualify their data against another company's set of business rules adds yet another layer of complexity to their operations and, likewise, having to re-check data as it is received stymies the efficiency of collaborative working. At the same time, the data produced as a result of any collaboration needs to be consumable, and an organisation needs to be able to aggregate it with other data for it to be valuable.

For the larger pharmaceutical firms acting as hubs, it is also important to think about where data goes when the project is completed. Where is it archived? Does it get replicated internally? Is it retrievable if the collaboration resumes or is restarted? Does the audit trail get archived or moved?

Communication between parties

The visibility and tracking of which actions are carried out, when and by whom are vital. It is no mean feat internally within a single company, let alone within a distributed collaborative network. When you consider that many collaborators will have multiple different projects ongoing at any one time, the need for tracking and visibility is compounded.

All parties within a collaborative project need an overall view of tasks and progress, to ensure tasks are not duplicated and that all the information is available immediately. Concepts such as alerting, following, project and job boards come into play. 'Social' features such as the ability to comment on work can really help in this type of environment.

It is also worth considering that the work environment will require the scientist to be connected to the project in more ways than just through their laptop. They will want, and in some cases are already demanding, the data and informatics systems to be accessible on all sorts of other devices, from tablets and mobiles to fume hoods and even their safety glasses.

Data and material transfer between parties

This is about who gets what, when and how. It should be easy to transfer data and samples between parties, ensuring the research and development is carried out properly and efficiently. Sample tracking and shipment should be streamlined and tracked across the entire R&D network.

The previous section referenced the importance of communication frameworks; the derivative of that is the content of what you communicate. Data

is one element of this, which means simple data capture and dissemination is critical. Teams must be able to repeat work or make informed decisions – and without all the data and its context, this becomes difficult and risky.

The second element is trickier to manage. Data is relatively easy to transfer between parties, thanks to the internet and cloud. However, much of the work that R&D organisations do is related to samples or materials made: experiments result in the creation of a sample. That sample is then used, of course, in subsequent experiments and projects. This can be managed reasonably well on a single site, and many companies have systems to manage these samples and deal with registration and dissemination. This is something lab information management systems have done for years.

When the collaboration network is distributed across international borders it is another matter altogether. Not only do organisations have all the normal sample labelling, receipt, dispatch issues to contend with, they have the added problems of international compliance mandating – what can be sent by whom and by what methods, not to mention the question of order tracking and delivery. This all needs to be easy to work with and visible to all collaborators who are implicated in the sample or material transfer. The key here is ease of use – a scientist should not have to use five different applications to execute one task. They want one place, connected to the collaboration environment, to do it all – create, label, register, track, send, receive, use and dispose.

Commercial terms, payments and cost

Collaborations must be managed effectively and within a sound commercial framework. This needs to be coupled with simplified deliverable management and payment management. In other words, when the work is completed, delivered and signed off, payment should occur as seamlessly as possible – even automatically.

This automation, or semi-automation, will be important moving forwards. Organisations do not want to spend time processing POs, statements of work and payment requests. The back office admin should be kept to the bare minimum and linked to the project deliverables. This will require integration and linking of lab systems more closely with the procurement and financial systems.

From a commercial perspective, another factor here is cost. Many years ago we saw the pharmaceutical giants build their own bespoke systems on their own budgets, because the money was flowing. The technology was custom-built, or heavily customised, based on a "we do it differently here" mentality. In reality, much of the mechanics of what these organisations do is the same. This is having repercussions now on the informatics environment, coupled with massive cost pressures. A change in fiscal dynamics will see a perceived need for perfection fade away, and be replaced by a more measured and cost-conscious attitude. As a result, organisations may increasingly turn to off-the-shelf systems that are domain workflow centric, as well as technology centric – but offer far less scope for customisation.

System management and integration

It is crucial that businesses can turn systems off and on easily, and manage the informatics infrastructure rather than relying on the IT function to create and manage users and projects. The ideal scenario here is that a scientist can send an email to the person they want to bring on board and set their security privilege to one of three types: 'view', 'contribute' or 'administrator'. The onus is then on the collaborator to set their password, and removing them from a project is simply a case of turning off their privileges. Other options would include linking users to other applications that deal with security and user identification, such as LinkedIn and Facebook. The linked application then provides more information about the user such as name, age and location, which can all be used to help sign data electronically.

When thinking about the crossover of data and identity, we should also note how teams go about finding a suitable collaborator. There are examples of 'science discipline brokers' already, who act as a single point of contact for organisations providing a service of a similar type. The natural extension to this would be not only finding organisations, but also individuals.

Another challenge to note here lies in systems integration. Until now the approach taken by organisations has typically been bespoke, and linked to legacy thinking of "this only has to work internally". This approach is no longer fit to support today's R&D environment. Given the inevitable move to the cloud and distributed collaborators, the integration paradigm will also change. Systems will need to integrate out of the box. We see this with other industries and in our home lives – you can use Facebook to log in to many other applications, for example. We are moving towards an environment dominated by 'as a service' offerings, public APIs, data exchange and cloud hosting – radically different from today's landscape of custom, behind-the-firewall technology.

Domain support

As organisations ramp up their pursuit of innovation, the creation of multidisciplinary teams is increasingly common. We now see many organisations opening up project teams and diversifying their teams. Gone are the days of having 20 medicinal chemistry scientists sitting in a room, making project decisions – these are now cross discipline teams with scientists from all areas taking part. The natural extension to this is that the team is virtual and spread across multiple organisations, and meets in an e-room instead.

When reflecting on collaboration networks, the importance of supporting all the scientific disciplines with tools and data environments that cater to each of their needs becomes clear. Chemistry, biology, analytical, translational and simulation scientists also have their own requirements that must be supported within one environment as they work together on projects.

Each domain has distinct and specific needs, not only with regards to data types but also the tools they use to generate data. It therefore becomes more and more important for tools to be easily integrated into the collaboration environment. It may not be long before we see scientists using apps on demand, as and when needed. This of course brings up an array of questions for the application providers themselves, as they balance meeting the business and scientific needs of the end users.

SaaS, PaaS, XaaS

The future of all these systems is on the cloud – in some form or another. This brings with it another set of issues – not from a collaboration standpoint, but from a change in organisational mentality. It is clear that all R&D systems will be managed and delivered 'as a service' (Xaas) in the future. What is less clear is how this transition will occur and how long it will take for companies to acclimatise. Current trends suggest this will be sooner that we might think, with many companies adopting a cloud-first approach for all new informatics solutions.

Clearly then, this change of focus is diversifying the R&D environment. Larger companies are refining the way they operate to think and act more like 'research hubs', but they need to up their game. The barriers outlined earlier must be overcome, and results delivered, in a way that ensures there is simple on-boarding, simple management and a single point of knowledge – a single source of truth. There must be consumable, contextually rich data that is secure and aids decision making. Technology will need to serve diverse scientific domains, help commercial aspects to be managed with minimum fuss, and of course, give scientists easy access to data from almost any device.

It could be argued that this is what informatics, supporting an R&D environment, should do now – and that most aspects are catered for with the current landscape of informatics tools. However, as organisations continue to break down their external walls, the sheer complexity of communications, data security and IP challenges make the next decade in R&D an intriguing prospect. These change factors offer opportunities for disruptive technologies, new ideas and new solutions to come in and fill the gaps. All things considered, we might just be seeing the rebirth of R&D informatics into the 21st century.   DDW

*As Vice-President of Strategic Solutions, **Dr Paul Denny-Gouldson** heads the overall strategic planning for the various market verticals and scientific domains in which IDBS works. He joined IDBS in 2005 as part of the acquisition of his ELN company and has spearheaded the drive to make E WorkBook the market leader. Paul obtained his PhD in Computational Biology from Essex University in 1996, and has authored more than 25 scientific papers and book chapters.*