ELSEVIER

feature

Considerations for implementing an informatics system to support biologics drug discovery

Andrew P. LeBeau

A comprehensive, well-functioning and scientifically aware informatics environment is of the utmost importance in supporting effective drug discovery programs. However, implementing such a system can, without a carefully crafted and enacted strategy, be a disruptive and time-consuming process, fraught with risks of cost overruns and the potential to end up with a system that ultimately does not meet the ever-evolving needs of the organization. In this review, using our experience from the software provider perspective, we discuss how a drug discovery organization should approach the task of selecting an informatics system to support its research discovery programs. We focus on biologics drug discovery because: (i) in general, biologics informatics is less mature than small-molecule discovery; (ii) there is a great deal of activity in biologics drug discovery, from new biologics-only companies to established small-molecule companies extending into biologics programs; and (iii) by their nature, biologics can be more challenging than small molecules to support, owing to the size, complexity and diversity of biologics entities. These factors make decisions about which system(s) to implement even more challenging.

Making the decision to implement an informatics system

Selecting and implementing an informatics system to support discovery research is crucial for ultimate success. However, it also brings significant challenges and uncertainty, and the attendant risks of an unrealized return on investment (ROI) if the wrong choices are made. Compounding the challenge is the problem that those on the customer side who are involved in the selection process often have limited experience with the various vendors and the systems they provide, potentially leading to a stalemate where no affirmative decision is taken. However, inaction is at least as bad as poor choices. The goal of this review article is to provide guidance and key considerations for making an informatics system selection. The key challenges involved are discussed here and suggestions on how to mitigate risk are provided.

When is the right time to adopt an informatics system?

The primary issue we see with organizations selecting a research informatics system is that it is done too late. When a new company is formed, funding is typically tight, and personnel, facilities and equipment are prioritized because they are viewed as essential to the new entity. However, when it comes to research informatics software, there appear to be viable and comfortable alternatives in the form of paper laboratory notebooks and/or the use of general-purpose business software, such as Microsoft Excel, for capturing, managing and visualizing scientific data. All too often, scientific research software can be considered important but something that could be acquired later 'when we have more data and we really need it'. Although understandable, this approach is, with all due respect, wrong. Implementing a proper scientific informatics system from the start means

E-mail address: andrew.lebeau@dotmatics.com.

https://doi.org/10.1016/j.drudis.2018.09.016

all data being captured, managed and utilized are already in a structured, scientifically aware system that can scale easily as more data are generated over time. Informatics systems enforce process and structure on the data, meaning the quality of the data starts high and remains so. ID generation and other nomenclatures are established early, and become documented as part of a standard process, leading to faster onboarding of additional personnel as the company grows. From the first experiment, intellectual property (IP) is generated and this needs to be recorded in a properly implemented and managed system-of-record to ensure that uniqueness and IP rights are established immediately. Given that data volumes in biologics research organizations can start growing exponentially, waiting until there is a realization that there are 'enough data' to start vendor selection and implementation will mean that, by the time the system is in production, the data volume will already be unmanageable.

Implementing a research informatics system early will typically reap cost savings in the short and long term. In the short term, proper data management means data quality is high, avoiding having to redo experiments that fail, and that data are recorded improperly. It also means that the data are accessible to scientists, avoiding inadvertent repetition of experiments because one researcher did not know the experiment had already been done and recorded in a colleague's paper notebook. In the longer term, the cost savings of implementing a scientific informatics system over an *ad hoc* solution are very high. At this stage there will be: (i) more users, who are often resistant to the new systems having grown comfortable with existing processes, and considerable disruption to ongoing research activities; (ii) more data, which are distributed between various file shares, users' computers, paper lab notebooks, among others; (iii) more processes to convert to the new system and these will be less standardized and structured, having evolved without the structure of a proper informatics system; and (iv) a profusion of ID systems and naming schemes or nomenclatures that need to be harmonized. Direct costs as a result of vendor implementation, configuration and training services will scale according to how late the decision is made. When a system is implemented early, those costs can be mostly avoided because *ad hoc* systems were never implemented, and information transfer from early personnel to new employees occurs automatically during onboarding. A properly implemented informatics system increases the likelihood of ultimate company success, by avoiding the impediment of inefficient and ineffective data management systems and by accelerating research activities and overall innovation. The IP of a company will be at risk until the data are correctly captured and secured. For startup companies this might represent the entire value of the company, and can make it hard to secure additional funding, secure partnerships with large pharma or even achieve the company's exit strategy.

What do I do first?

So, the decision is made to adopt a scientific informatics system. Where should an organization start? The answer will depend on the current situation and what is already in place. Vendors are often approached by prospective customers asking about specific software applications, such as an electronic laboratory notebook (ELN) or a registration system. The most common mindset is that it is better, simpler and/or easier to start small and build up, thus deciding on the most needed software application, implementing it and building a system incrementally from there. Given that, priority typically follows workflow dynamics; so, because the standard workflow is to create an experiment, run the experiment, and then analyze and review the results, start by implementing an ELN, then an assay data management system and finally get an analysis and visualization application(s). A registration system fits in along the way. In other words, capture the data first and worry about accessing and using it after that. In our experience, this approach is flawed. The drug discovery industry has traditionally been much better at capturing data than making it easily accessible to scientists. Informatics vendors share much of the blame for this – a common complaint from customers is that it is difficult to access data held in system-of-record databases such as ELNs and registration systems. The folly of this approach is clear when considering the discovery research cycle (design > make > test > analyze & report; Fig. 1), which occurs throughout discovery research and particularly during the crucial lead optimization phase. This is a cycle, not a linear process, with each step depending on the one before. What the savvy customer does is build a strategic plan that clearly defines the end-goal, to ensure that the needs of the full cycle will be fulfilled, irrespective of which applications are implemented first.

Within the vision of the end-goal, focus should be placed on the analyze & report > design step, which is where the collective knowledge, experience and intuition of scientists are primarily utilized, and where most innovation occurs. Failure to provide informatics tools that deliver the necessary information in a timely, complete and consumable fashion will lead to frustrated researchers, undermine innovation and slow project progress. Not coincidentally, this is also the phase where innovation in scientific applications is the greatest. There is greater diversity across vendor

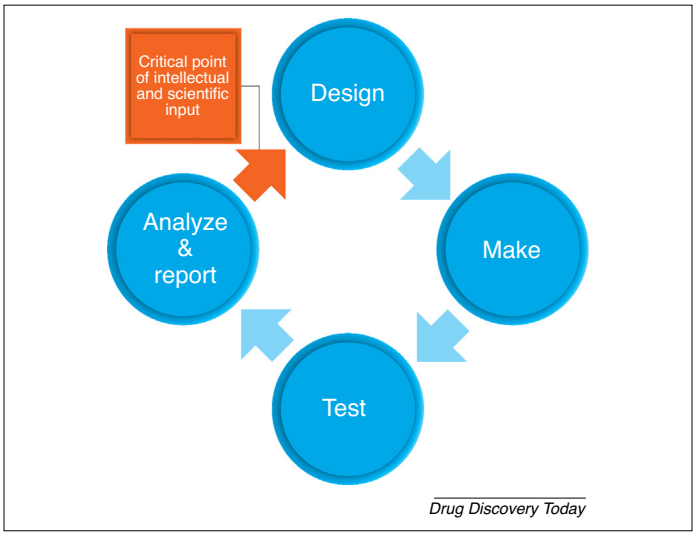

FIGURE 1

Drug discovery workflow cycle. Most informatics implementations focus initially on implementing an electronic laboratory notebook (ELN) to support the design > make step. But the greatest scientific input, including innovation and ideation, occurs at the analyze & report > design phase. Strategic decisions regarding solutions to support this step should factor highly when considering adoption of a scientific informatics system.

applications in how they present data in a consumable format, their ability to tailor views and reports to individual users' and teams' needs, and the ability of embedded analysis and visualization tools to support advanced and emerging computational methods. As a result, discovery organizations should be very careful regarding selection of applications to support this crucial stage of the discovery research cycle.

Technical considerations

Once organizations have determined the general strategy for selecting an informatics system, there are many technical considerations to be assessed. We focus here on some of the most important.

Single vendor or mixed vendor environment?

Should you go with a single vendor for your informatics needs or should you look for the best tools across a range of vendors? There is no right or wrong answer, and obviously it is ideal if the best tools for your organization happen to come from the same vendor, but this generally won't be the case. Single vendor solutions are typically more integrated (you must check) but come with the risk of a weak link(s) within the suite. Concerns about 'vendor lock-in' might be voiced but, with developing integration conventions (cloud computing, among others), switching costs are typically less than they were in the past. Mixed vendor environments run the risk of poorly integrated apps and challenges with data fidelity and security across the full workflow.

One key dynamic is the increasing role of collaboration within and across organizational boundaries in drug discovery. Collaboration within organizations is supported optimally with a unified platform approach, where applications are built on an underlying architecture that has a common data model, manages core services (users, security, etc.) and is extensible through application programming interfaces (APIs) built on industry conventions such as RESTful web services [1]. Collaboration across boundaries often means data and metadata exchange between different systems. The use of an informatics platform with RESTful APIs is an advantage. An alternative approach to cross-organization collaboration is to use a cloud-deployed system that collaborators can all interface with. Although there is no single right answer, in our experience the most successful implementations are those in which a single vendor, with a platform, fulfills most informatics needs. Additional apps fill specialist needs or gaps in the primary system.

Cloud

Adoption of cloud computing has increased significantly in the past 2–3 years, after many years of resistance in the drug discovery industry [2]. Organizations should ensure that a cloud-deployed system is an option from any prospective vendor. Even if an organization does not plan to immediately deploy to the cloud, it is likely that they will in the future, because the advantages of cloud deployments will ultimately win out over (largely unfounded) data security concerns. If organizations are bound by data sovereignty regulations (e.g., The EU General Data Protection Regulation, GDPR [3]) then the ability to deploy to a regional instance is necessary. The software system should be identical whether it is deployed in the cloud or on the premises. This ensures maximum flexibility should the organization wish to change its deployment methodology in the future, or use a hybrid model, where a cloud deployment supports collaborations with partner organizations. For small companies, total cost of ownership is often the most important reason for cloud compared with hiring IT experts buying hardware and infrastructure licenses. Another advantage of cloud deployment is that it is the easiest and quickest way for vendors to have access to the customer's system to provide support, such as adding or modifying configurations and troubleshooting issues.

APIs and integration

Clunky, click-ey software requiring the user to traverse a multitude of interfaces to complete a workflow is perhaps the biggest barrier to user adoption of an informatics system. Modern IT has solved this problem by embedding services, such as registration, into common user interfaces (e.g., an ELN), so a vendor who has not adopted this approach wholeheartedly does not deserve your business. You should look for a system in which openness and integrability are fundamental tenets of the architectural design. An open system means a better system today, and one that can scale and evolve over time to meet new demands.

Performance

Software performance is another crucial factor that can derail user acceptance of an informatics system. This is particularly true at the crucial analyze & report > design phase of the research cycle, because this is where the scientists are applying the greatest level of their intellectual input, experience and intuition. If the software is slow to load and display data, and analytical computations take extended periods to complete, users will lose focus, disrupting their intellectual flow and undermining higher-level thought processes and innovation. You should ensure that technical presentations from vendors clearly demonstrate that the performance of the system will meet your needs based on the nature and volume of data you routinely work with. If you need to work with millions of sequences with thousands of columns of data you should ensure the demo you are given reflects this.

Business, personal and personnel considerations

Successful adoption and maintenance of a drug discovery informatics system means an ongoing relationship with a vendor(s). Successful implementations occur when a strategic relationship is formed and, as with any relationship, there are going to be ups and downs. Interpersonal relationships play a key part in the ongoing success of the system, because assistance from the vendor(s) is necessary to ensure the system operates optimally as it evolves over time. Below are some key elements to consider before selecting a vendor and informatics system to implement.

Past presentations

Check the vendor's website to view recorded webinars and previous presentations. Did the presenter genuinely raise issues and challenges, and demonstrate how its informatics system answers those, or was it just a high level, jargon-filled review of current industry trends, disconnected from the product line-up? You want to know that the vendor is in tune with, and can meet, your current, real-world needs.

Technical presentations

After initial discussions, a technical presentation is an important step in the assessment process. You should expect a presentation that clearly demonstrates that the vendor understands your business needs and has factored these into the presentation. You should be prepared to provide guidance on this including sample data for the vendor to use. This commences the relationship and gives you an opportunity to assess the people as well as the software before you commit to a purchase. The technical presentation should communicate a story that reflects the problems and challenges you face and present how the system can solve those problems. The demonstration should not simply be a traverse of the menu options, showing each feature in excruciating detail.

Vendor personnel

You should check the backgrounds of key vendor personnel, who have often worked across companies within the space. Experience counts for a lot and some due diligence in LinkedIn (https://www.linkedin.com) can provide insight into which vendors are on an upswing. Also, check financial performance, stability, background and tenure of senior management. Industry analysts that cover the life sciences, such as from Gartner (https://www.gartner.com), Frost and Sullivan (https://ww2.frost.com/), among others, can be a good source of information.

User group meetings (UGMs) and roadmaps

Speaking with existing users at UGMs will help you understand the strengths and weaknesses of a given vendor. At these events, or in dedicated meetings with vendors, you should be able to review product roadmaps. Roadmaps provide information about future plans and should be discussed in a wider context of how market and user information are factored into future software development plans. Plans for the next 6–12 months should be mostly firm, whereas plans beyond a year should be directional and demonstrate that the vendor understands the trajectory of the market and emerging technologies that need to be supported. Ask what the process is for providing input on future product development. Is this structured, *ad hoc*, driven by the loudest voice (i.e., the largest customer)? Ask about the release timetable and cadence, how many versions are supported, and the complexity of upgrading (this is a good question to ask existing users).

Support organization

No matter how carefully you vet potential vendors, all software has bugs, and there are always issues that arise during and after implementation. To address these, you will be interacting with the vendor's technical support organization. Good or bad support can make or break a successful informatics implementation. Typically, interactions with the support team occur after a purchase is made, but there is no need to wait to test the support. As part of your due diligence you should contact the vendor's support organization directly. You generally won't have a valid issue to be addressed but you can still get a sense of how quickly your call or email is responded to, and what the personal interaction is like. If you are testing the software out with a trial license, these interactions can be more involved and give an even better sense of the support performance.

Concluding remarks

Implementing a scientific informatics system might seem like a challenging and disruptive task but it doesn't have to be. Making the decision and adopting a system as soon as possible brings huge productivity and cost benefits. Having committed to a system, careful assessment of what is required compared to the available solutions on the market, and due diligence of prospective vendors, should lead to an efficient and successful implementation. This will result in ongoing benefits in scientific innovation as well as increased productivity in the long term.

References

1 Fielding, R.T. Representative State Transfer (REST). Architectural styles and the design of network-based software architectures. Available at: http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm
2 Accelerating Drug Development with Cloud-Based Computing. Available at: https://pharma-life-sciences.cioreview.com/cxoinsight/accelerating-drug-development-with-cloudbased-computing–nid-12895-cid-36.html
3 GDPR Portal. Site overview. Available at: https://www.eugdpr.org/

Andrew P. LeBeau

Dotmatics Senior Manager of Biologics Marketing, 6050 Santo Road, Suite 270, San Diego, CA 92124, USA