

# Outsourcing life science informatics solutions

The life science industry depends on information gathered from new technologies for productivity and competitive advantage in the R&D process. As more information is gathered and the systems become more complex, building and maintaining data management solutions that support rapid and accurate decision-making becomes harder. This article examines ways in which strategic outsourcing of elements of the life science informatics systems can enable companies to apply their resources where they can generate most competitive advantage.

Life science companies have an opportunity to create sustainable competitive advantage for themselves that is unparalleled since the industry attained its current form in the 1930s. That opportunity has been created by the rapid changes overcoming the research, development and portfolio management processes, which in turn have been sparked by the availability of much greater volumes of data from new research technologies.

These new technologies, especially those relating to genomics and microarrays, have led to a transformation of the traditional problem of the industry, namely lack of targets for a disease, which is now largely a problem of the past. The recent glut of targets has solved one old problem, but it has created another new one. The targets that are generated today are much less well characterised than previously as there has been less time for their study in academic and other research groups. It was recently reported that every new target identified today has an average of only eight literature references, compared to 100 ten years ago<sup>1</sup>. This has led to pressure on the downstream activities of target validation, lead identification and lead optimisation, which are struggling to cope with the new volume of poorly qualified targets. With research budgets still increasing and productivity targets being missed<sup>2</sup>, more effective and timely use of information is increasingly seen as the only way to

avoid a significant R&D performance crunch in the next few years.

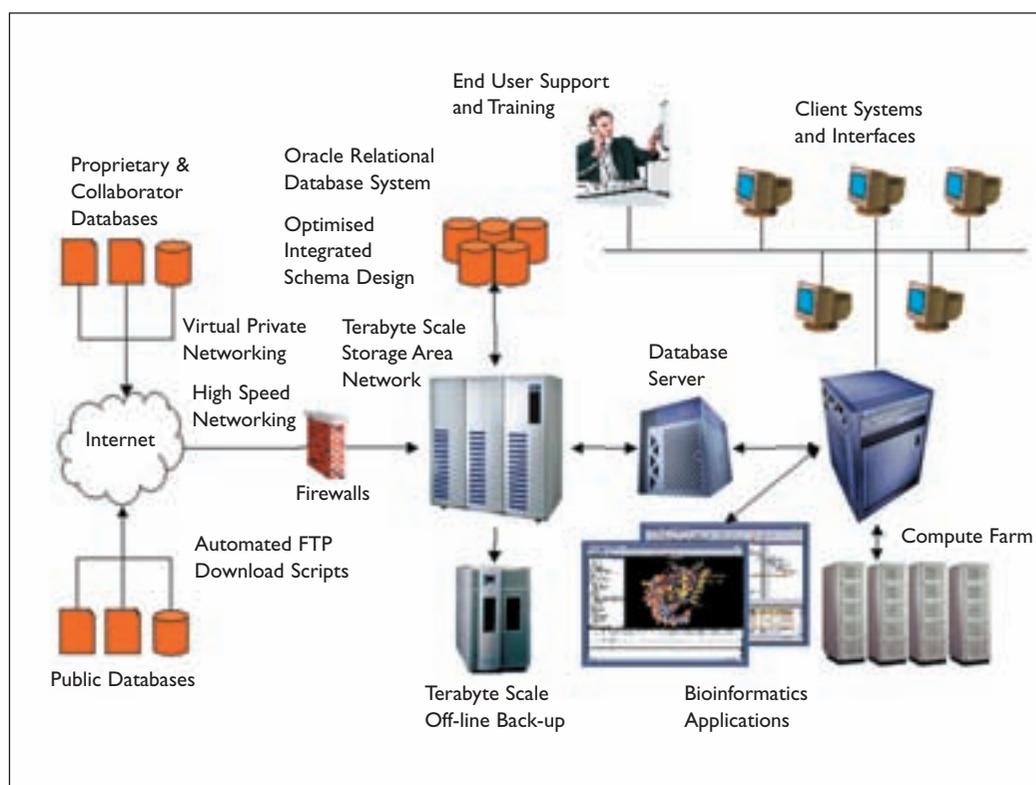
These downstream tasks require that information from a number of diverse sources is correlated and analysed in order to make rapid, but accurate decisions to progress or stop the project. This can only happen if the right information is available at the right time to the scientists and project managers making the decisions. In order to do this effectively all of the available public, collaborator and proprietary data sources must be integrated with the internal data repositories. These data sources have been growing in size exponentially for the last 20 years, and now just to have the basic public bioinformatics databases in a production environment requires at least a terabyte of disk space. Most life science companies have much more storage capacity than this, as storage requirements continue to double every 10 months or so.

Due to recent technological developments, life science companies have the opportunity to build themselves integrated information and knowledge systems that can support their R&D decision-making processes. Building such a system will allow the researchers across the discovery process to exploit fully the information available throughout the company. Providing this integrated information infrastructure to their scientists and using it as a primary Discovery tool can make the whole R&D process more efficient. Being able to reproducibly identify successful and interesting projects

**By Dr Steve Gardner**

## Informatics

**Figure 1**  
The components of an enterprise informatics system



to strengthen the corporate portfolio reduces R&D costs and gives long-term competitive advantage. This is why industry leaders such as Merck and GlaxoSmithKline have invested hundreds of millions of dollars developing their R&D informatics systems.

### Components of a life science informatics solution

Integrated informatics systems are by their very nature complex and large, although they contain many off-the-shelf components, as shown above in Figure 1.

Building a modern informatics capability internally is therefore a major, long-term IT project that requires much more than a computer-literate biologist with a PC in the corner running a couple of sequence similarity searches across the web. There are many sophisticated components (shown in Table 1) and many different IT skills required to build and manage this system. Lehman Brothers recently estimated that the minimum investment needed in informatics technologies to guarantee returns from a genomics programme was in the order of \$15-20 million<sup>1</sup>. This is in line with our experience of implementing a number of these types of systems internally in major pharmaceutical companies.

### Levelling the playing field with strategic outsourcing

Obviously this would suggest that deploying an effective informatics system is the prerogative only of the mega-companies and that they will be able to use this a major competitive advantage over the smaller pharma and biotech companies.

In fact this is not necessarily true; in reality the situation may be exactly the reverse with smaller companies holding the upper hand. Just as developing countries with no wire-based telecoms infrastructure such as China, Israel and Finland have embraced the newer, better and cheaper wireless mobile phone technologies, so smaller companies who have not yet made significant investments in informatics can avoid many of the pitfalls of the larger companies. They can also develop systems that are more agile and flexible to the changing needs of their research as they avoid the huge inertia of internally developed systems.

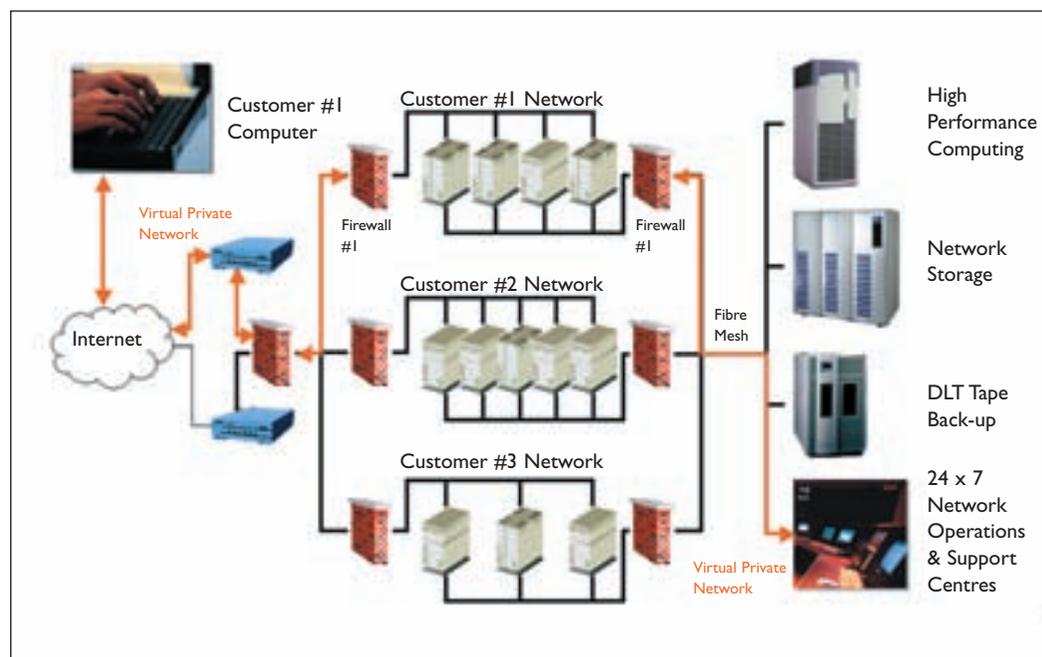
By strategic outsourcing of elements of the design, implementation and management of their informatics systems, smaller companies can expect to significantly reduce their costs, get access to more experienced staff than they (or in many cases their larger competitors) could recruit, and design and build better systems than they could by themselves. This is such a compelling proposition that it

**Table I**  
**Components of the enterprise system**

<b>COMPONENT</b>	<b>DESCRIPTION</b>
<b>Data Storage</b>	The entry level for a basic research informatics system based largely on public data is about one terabyte. Most mid-size biotech companies now have 3-10TB, top 20 pharma companies store 10-40TB, and the genomics data providers have hundreds of TB of data.
<b>Database Updates</b>	The major public databases are updated nightly, and these database updates need to be made available to internal staff as soon as they are available. This requires downloading, checking, reformatting, validation and deposition to occur overnight in an unsupervised fashion.
<b>High Speed Networks</b>	Database updates for the larger databases often amount to several GB per night. This volume can swamp even dedicated T1 (1.5 Mbps) lines, so T3 or higher connectivity may be required.
<b>Virtual Private Networks</b>	When accessing proprietary data remotely security and privacy are paramount. This requires a Virtual Private Network or encrypted connection to be established between the client and the server.
<b>Security</b>	Companies must protect themselves from the outside world using layered firewalls and other security technologies.
<b>Compute Farm</b>	Processing and searching terabytes of data requires significant compute power. This is often separate from the main application server, and takes the form of a large supercomputer box, a dedicated accelerator or a farm or cluster of smaller computers. Optimising performance of the compute farm is critical for acceptable response of the system.
<b>Bioinformatics/ Chemoinformatics Applications</b>	There are many Bioinformatics and Chemoinformatics applications that perform subsets of the total functionality required by a research group and almost always a company will implement a number of different systems from different vendors. These systems must be integrated in order to avoid major bottlenecks.
<b>Integrated Informatics Schema Design</b>	In order to allow data from different sources to be correlated effectively they must be stored in some form of integrated database schema. This typically requires a common view of the vocabulary used in the domain to be applied to a wide range of data and is one of the hardest elements to achieve.
<b>Decision Support Systems</b>	End-users need to ask questions correlating data from multiple sources in a simple fashion, and have the results presented to them in a simple and visually obvious way. This requires decision support systems to structure and present information back to the users.
<b>End User Support</b>	End-users faced with a new system will require support for all of the applications, databases, systems and networks. This means having both a systems administration group to ensure that the systems are always available, but also ensuring that scientists can get help on applications when needed.
<b>HIPAA Compliance</b>	The new Health Information Portability and Accountability Act (HIPAA) legislation seeks to guarantee the security and privacy of health information. Companies must establish systems and policies by 2003 to ensure that network access, intrusion detection and reporting, and wide scale public key encryption can be enforced. This will have a major impact on hardware, software, networks and security inside the enterprise.

## Informatics

**Figure 2**  
Example of a solutions hosting architecture



was recently estimated that the volume of research informatics services outsourced will increase by 1,800% between 2000 and 2005<sup>3</sup>.

### Different outsourced services

Outsourcing takes many different forms. From well understood services such as business process consulting, systems design and contract engineering to more recent developments such as Application Service Providers (ASPs) and remote hosting. Recent developments in network and security technologies have made it entirely possible for whole corporate infrastructures to be hosted remotely, providing a range of new options for smaller companies.

In the remote hosting model, customer machines are built to their specifications inside one of the hosting company data centres. The systems are isolated from other customer systems via front and back firewalls to ensure security, and accessed by the users via a Virtual Private Network connection. Large-scale items such as storage, back-up and recovery and high performance computing are typically provided on an as-needed basis. The hosting company is usually responsible for the systems management, monitoring and updating of all the systems, and may also provide customer support.

Using one or more outsourcing provider, companies can expect to realise a number of important benefits:

- Significantly lower the risk of failing to build, deploy and manage a successful enterprise system

- Get access to specialised expertise in all relevant IT and Bioinformatics domains
- Build competitive enterprise systems much quicker than they could internally
- Build more secure and flexible systems than they could internally
- Guaranteed performance and service levels often well in excess of their own companies capabilities
- Lower the total cost of ownership of their systems.

### Lowering the risk of development

As we saw earlier, life science informatics systems are complex IT systems with multiple interacting components. Getting systems to work well together and optimising their performance takes a great deal of experience. As such they are inherently difficult to build, deploy and manage. Experience is invaluable in particular in ensuring that the architecture, specifications and project plans are realistic and achievable. Without significant expertise in the domain it is almost impossible to build a successful system at the first attempt, as shown by the First Timer column in **Figure 3**. Even experienced staff with two or three systems under their belt will struggle to avoid all the pitfalls involved in building such a complex system.

By using an outsourcing provider who specialises in the domain (often called a Vertical or Full Service Provider), companies can expect to increase their probability of a successful implementation of a system to close to 100%. Any remaining problems can be managed through the imposition of

## Informatics

appropriate service level agreements, which guarantee the availability, performance and support of the system.

### Accessing specialised expertise

Although internal IT staff are often confident of their ability to design, deploy and manage these systems, experience would tell us that there are many more failed projects than successes, and there remains a high rate of disillusionment with informatics. By choosing the correct marriage of internal knowledge of the domain and specific company research process with external knowledge of designing, building and managing life science informatics the probability of successful implementation rises sharply.

Outsourcing providers may also be able to bring optimised components, platforms and vendor relationships to the design of a customer's system. In particular because the staff within an outsourcing company will typically have designed and deployed a number of similar systems in different environments, they can simultaneously accelerate the deployment while reducing the failure risk.

### Security

The most persistent argument levelled against outsourcing relates to services remotely hosted outside of a company's firewalls. It is a comfortable accepted truism that data are intrinsically safer inside a company's firewall than outside it. Experience and statistics tell a different story however. More than 90% of all security breaches occur inside the organisation.

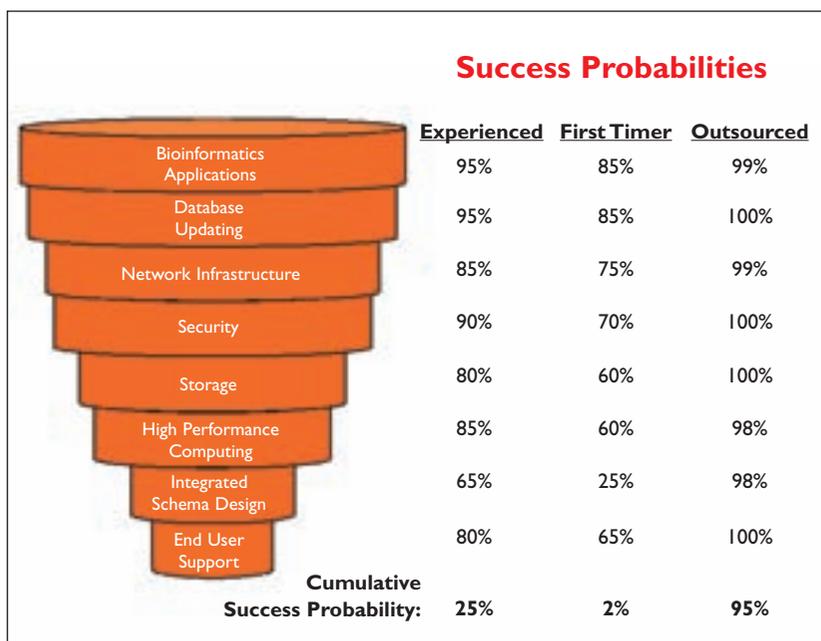
Life science companies typically have a large number of collaborators, contractors and staff from other sites working on their systems and interacting with their data. With the ability to plug a laptop PC with a 20GB hard disk directly on to the corporate network, most users can download significant quantities of data on to a highly portable and lovable device, often after entering only a single password. Even without any deliberate intent to steal data, users using their laptop or even PDAs to store and analyse corporate data can unknowingly leak critical information. Corporate data can be lost if the device is stolen, lost or even just plugged into a hostile network (which we define as anything that isn't behind your own firewall).

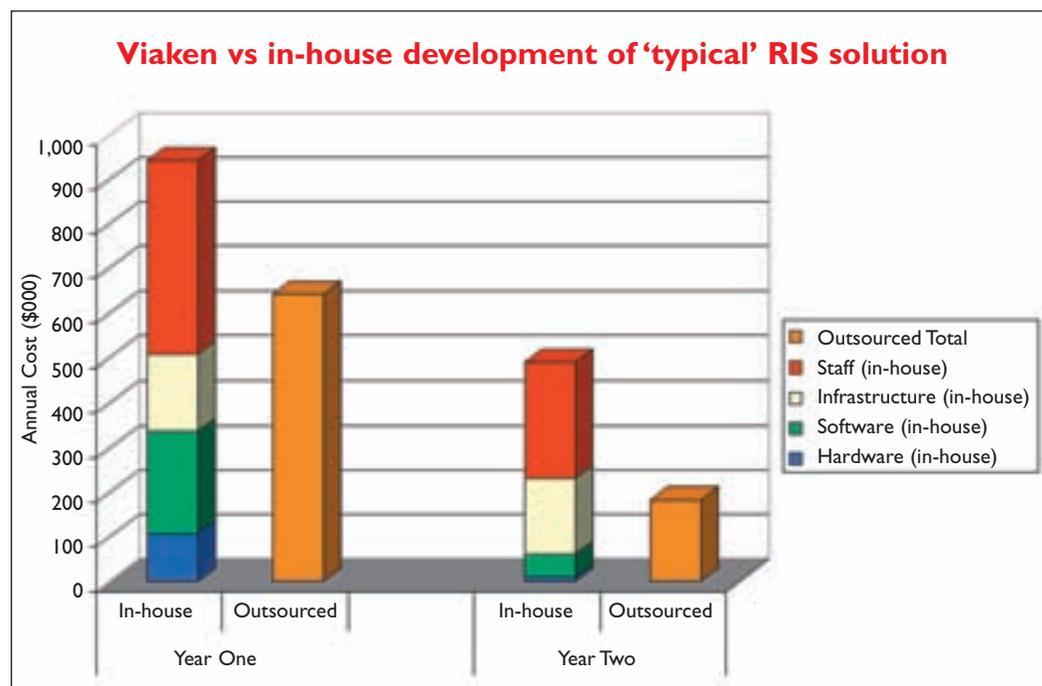
Well established Application Service Providers will typically have a much higher security service level than can typically be guaranteed by a life science company's IT organisation, although this is often not realised or acknowledged. This is achieved partly by strong user authentication, Virtual Private Networks, strengthened hosts and the mechanisms used to host applications. Two-factor user authentication is typically used so that a user has to simultaneously enter a password and a code from a smart code device that regularly updates an encrypted cipher. This combination requires the user to enter something that they know and something that they have, and avoids any possibility of the wrong person getting access to their data. This also ensures that individual and group privileges to access specific resources can be applied stringently and tracked.

Many companies require remote access to their data for researchers who are on the road, for collaborators or when they are visiting other company sites. Remotely hosted applications are ideal for this, as they provide a mechanism for a researcher to work on the same hosted machine whether they are at their desk or travelling. Further, many of these systems transmit only an encrypted picture of the screen of the user's hosted machine, rather than allowing them to run the application locally on their PC. This means that the user never has any corporate data on their hard disk, and it cannot therefore be lost if the laptop is stolen.

Many smaller companies who do not have good internal informatics systems submit their database searches to public services such as the European Bioinformatics Institute (EBI) or National Center for Biological Information (NCBI). This is a significant intellectual property risk, as the act of using a public server may constitute publication of data, regardless of the encryption levels employed. In addition, public services also have many visiting

**Figure 3**  
Estimates of likely implementation success for enterprise system developed internally or outsourced





**Figure 4**  
Total cost of ownership comparison

scientists and are a regular target for crackers seeking interesting resources and data. Use of these services with sensitive data exposes potential intellectual property assets to unacceptable and unnecessary risk.

### Enabling collaborations and alliances

Many smaller companies are dependent on the formation of alliances and out-licensing deals with larger pharma companies who are either interested in their technology or who wish to co-develop a particular molecule. The problem for the smaller partner is three-fold. Firstly, how can they effectively demonstrate their technology without tipping off their competitors, as they would if they put it up on their website? Secondly, when they are working with a partner how can they make sure that they only disclose those pieces of information that they wish to without putting the whole set of corporate data at risk by letting external researchers on to their network? Thirdly, how can they track the information that have been examined by their collaborators and create an audit trail of their studies?

Externally-hosted collaborative environments can provide a safe location on to which the data to be shared can be placed. This is a remotely-hosted workstation containing the data under strict access control along with all of the relevant analysis and visualisation tools to allow research and communication. Through the hosting security mechanisms,

all such environments automatically have strong security and individual user authentication and can therefore be used to share and track disclosures within an environment that facilitates collaborative research.

### Lowering the total cost of ownership

Outsourcing companies can typically use their internal resources more efficiently than their customers. This is because, for example, an Oracle DBA can be allocated on to three customer projects in a year, rather than just a single internal project. Outsourcing companies also typically negotiate reseller arrangements with their suppliers that reduce their costs in a way that can be passed on to their customers.

A case study of the adoption of an enterprise bioinformatics system in a pharma company is shown in **Figure 4**. The company achieved cost savings of 32% in Year 1 and 62% in Year 2, with total savings of more than \$500,000.

### New opportunities created

Outsourcing of the design, deployment and management of key elements of the complex informatics systems provides smaller pharma and biotech companies with a chance to leapfrog their larger competitors. As well as avoiding the huge investments currently made by top 20 pharma companies, outsourcing provides flexible systems that can remain closer to the state-of-the-art than large

## Informatics

### References

- 1 The Fruits of Genomics, Lehman Brothers, 2001.
- 2 High Performance Drug Discovery: An Operating Model for a New Era, Accenture, 2001.
- 3 Life Science Informatics: The Next Quantum Leap in Drug Discovery, 2001.

internally driven projects. This creates many competitive opportunities to adopt and integrate the information from new technologies more quickly and reallocate the internal resources saved to use it to better advantage.

Companies making use of services such as remotely-hosted collaborative environments may also be able to advertise their technology and intellectual property better to larger pharma companies looking for in-licensing deals. Collaboration environments show that an organisation that has considered its own and its partners' informatics needs and invested in meeting them. This is an important differentiator when a pharma company is evaluating the likely ease of working with a smaller partner.

When faced with sweeping new legislation such as HIPAA, companies will have to take very seriously the impact of informatics on their core business competitiveness. Life science informatics has very quickly evolved into a highly complex large scale IT problem. For many companies being able to take advantage of other's expertise will be the

only way that they can manage the transition. For others making strategic use of outsourcing will bring them the key competitive advantage that they need to differentiate themselves from their competitors and grow more quickly. **DDW**

---

*Dr Steve Gardner has more than 15 years' experience in the field of bioinformatics, specialising in protein structure databases. After his PhD from Birkbeck College, London, UK, he worked for Oxford Molecular as a Senior Product Manager. He was then the founding Director of Astra's Bioinformatics Centre serving 3,500 researchers at eight sites worldwide. He then founded and was CEO of Synomics Ltd, a life science systems integration business, before joining Viaken Systems as VP and CTO.*