Tribiosys

Informatics in BioPharmaceutical Research and Development: Challenges and Opportunities

The medical world is looking to new technologies to provide advanced understanding of disease. Effective drugs for treatment and prevention are needed for many disease areas, including cardio-vascular disease, cancer, neurological disorders, infectious diseases, endocrinology, and inflammatory and chronic degenerative diseases. There is excitement about the potential biological revolution that will emerge with understanding the human genome. The challenge for pharmaceutical research is to unravel the pathophysiology of human diseases and thus, make it possible to identify targets. Now that the first set of complete human genome data has been reported, the focus of pharmaceutical and biotech companies needs to shift from sequencing the genome to understanding the relationships of genome to diseases and finding new, innovative drug molecules. Parallel developments of new biological technologies, such as transcript profiling, allow scientists to examine almost any biological system in high molecular resolution. Contemporary drug discovery research is now focusing on the identification and validation of pharmaceutical targets in the molecular pathways/systems embedded in this information. Novel therapeutic interventions are being developed and evaluated as a result of this research, and these are expected to be the basis of innovative pharmaceuticals of the future.

Data Explosion

New R&D technologies have dramatically increased data volume and complexity. The mapping of the human genome, and the increased use of combinatorial chemistry, high throughput screening, and other technologies have dramatically increased the amount of data available to drug discovery organizations. The amount of public and proprietary data being generated today is several orders of magnitude larger than that of the pre-genomics era. Vast amounts of biological data are being accumulated in the areas of gene sequencing, gene expression, protein structure, cellular metabolic processes, cell-signaling pathways, and cell system analysis, and the volume and complexity of this new data defies efforts to transform it into information and to integrate it into existing discovery processes. New technologies have driven a 10,000-fold increase in the number of compounds a company can investigate during the discovery phase, yet the difficulty of managing this unprecedented volume of data has become one of the biggest bottlenecks in the drug discovery process.

Data come from different, disconnected sources, and must be integrated manually

New drug discovery from genomic-based targets requires connecting and integrating information and

knowledge from genetics/genomics, medical and chemical libraries and combinatorial chemistry and screening. However, these sources of information are rarely resident within one company and tend to be disconnected silos. Researchers can access public research initiatives such as the Human Genome Project and libraries such as GENBANK and SWISS PROT, as well as a host of genomic databases and libraries provided by private companies. However, the data-intensive task of deriving useful information from this varied data (cross-comparing gene-sequence information and gene expression information, for example) tends to fall to the individual researcher.

Strategic and Operational Challenges

Many discovery organizations are not equipped to deal with the strategic and operational challenges that accompany the transformation of this new data into useful and actionable information.

Existing data is inaccessible

Many pharmaceutical and biotechnology discovery organizations collect, store, and analyze their research data in a wide variety of off-the-shelf and internally-developed spreadsheets and databases. These data sources are often disconnected: either by geography, corporate structure, or underlying data format and annotation. As pharmaceutical research requires an increasingly integrated approach, gathering information from these disparate and disconnected systems creates enormous data management and research workflow challenges. Because data and intellectual property are not effectively managed within most drug discovery organizations, individual researchers spend valuable time repeating others' work, or worse, making expensive decisions without the benefit of additional analysis. In this environment, internal collaboration and knowledge sharing is difficult and the output of "druggable" compounds is significantly below potential levels.

Research workflow processes are resistant to change

Although it is widely accepted that the new data available represents an enormous opportunity for pharmaceutical research, there is a inertial resistance on the part of many individual researchers to change the underlying drug discovery process. With multi-million dollar R&D projects at stake, the move to an integrated informatics data environment must be managed in a deliberate and thoughtful fashion.

Inadequate internal knowledge management systems and processes

Possessing a large compound collection has long been a significant competitive advantage for discovery organizations. A standard practice for internal research is to evaluate in-house compounds to try against new targets, but a common situation is that these experiments can only reference compounds that have

already been assayed and filed, effectively excluding the thousands of compounds presently undergoing research elsewhere in the pipeline. As a result, discoveries that cannot be developed internally frequently languish, because there is no market for a discovery that has not reached a recognized state of development. By not having access to up-to-date information on the progress and subject of ongoing experiments elsewhere in the discovery organization, researchers are unable to effectively mine previous or concurrent research.

Decisions are made with insufficient information

Researchers in drug discovery organizations spend much of their time searching for data; time that could be better spent extracting information from that data and making valuable research decisions. It takes a significant amount of time to arrive at a go/no-go decision for a specific candidate compound, and often, multi-million dollar investment and attrition decisions are made based upon value-weighted best-guesses of success or failure.

Collaboration and outsourcing is difficult

R&D researchers have little ability to share data with external partners or even other internal research teams. As more and more pharmaceutical companies partner with smaller biotechnology firms, the ability to share information in a secure setting will become increasingly important.

Shareholders demand continual earnings growth

With a sharp eye on the pharmaceutical industry's historic track record of successfully bringing drugs to market and growing revenues, shareholders of the major companies continue to set progressively higher expectations. Over the next decade, therefore, meeting this relentless demand for double-digit financial growth will remain a critical management objective. For R&D leaders, it will mean rising pressure to deliver more, higher-value compounds for commercialization. An Accenture study found that the industry must increase the number of new molecular entities (NME)-the first step to creating a new drug-by 50 percent to meet 10-year growth projections.

Pharmaceutical IT organizations are not equipped to address large, data architecture redesign

Pharmaceutical organizations are late adopters of enterprise information technology, and their IT departments are significantly less capable than their counterparts in other industries. A significant amount of data management within pharmaceutical firms is managed with paper-based processes. Large biotechnology companies tend to be the most IT savvy, although some large pharmas have begun to make investments in this area.

New research directions will require even more capability

Even as pharmaceutical and biotech companies are grappling with the aggregation and analysis of genetic information, the leading edge has shifted from gene identification to the discovery of the functions and relationships of these genes and the proteins they direct. With the exponential increase in volume of data predicted for proteomic analysis (estimated to be 100x that of genomics data), an even greater need will arise for data management resources that can help interpret this volume of information.

Role of Life Sciences Informatics

Life sciences informatics, or the use of computers and sophisticated algorithms to store, analyze and interpret large volumes of life sciences data (bio-, cheminformatics), is essential in order to capture value from this growing pool of data. By integrating data sources, development organizations can integrate and analyze genomic information from multiple sources to discover genes that may represent the basis for new biological targets, therapeutic proteins, or diagnostic products. Through curation and annotation (the checking and re-tagging of data within a database) even dissimilar information can be integrated, including gene expression, gene sequence information, SNP data, functional genomics studies, preclinical pharmacology and toxicology, and results from clinical development studies.

Informatics Strategy

In order to benefit fully from the range of new technologies available to the life sciences industry, discovery organizations need to develop and implement a cohesive informatics strategy for managing corporate research and discovery processes. Management and research personnel need to cooperate to develop a comprehensive informatics platform that will meet the organization's present and future data requirements. Although most development organizations realize the potential of life sciences informatics to improve research productivity, few understand how they might integrate informatics into their current research architecture. Most also have concerns regarding how the necessary process change, implementation, and training would be realized.

Informatics Solutions

Few pharmaceutical and biotechnology firms have the experienced systems IT personnel necessary to architect and implement an enterprise-wide data infrastructure. However, many of the large IT systems consultants provide only a standard solution, or a solution based around their own product, to the detriment of the client. Each discovery organization's data management practices are unique, with its own proven processes and methodologies for gathering, mining, and storing research data. There is an opportunity

to help management and research personnel of R&D organizations to implement an informatics strategy that complements this legacy process and hardware infrastructure. There is also an opportunity to migrate legacy software applications into multi-tier client/server applications, allowing customers to preserve the core functionality and other benefits of their legacy applications while eliminating the constraints of legacy system architectures.

Tribiosys, Inc.

Tribiosys provides Life Sciences Professional services for BioPharmaceutical R&D organizations. We help these organizations design and rapidly implement quality informatics solutions to facilitate drug discovery and development. Our services encompass data and process management including:

- Data Annotation, Curation, Modeling and Mining
- Data Integration, Migration and Warehousing
- Process Automation and Workflow
- Knowledge Management and Collaboration

Our Ph.D level scientists and informaticians with expertise in molecular and computational biology, bio, chemical, process and clinical informatics collaborate with your staff to design informatics approaches and strategies. Our highly skilled technical staff, including architects and software engineers work in tightly integrated teams with your staff and Tribiosys informaticians to architect and develop rapid, cost-effective solutions that realize the informatics strategy. Our staff has experience in fields such as architecture design, data warehousing, enterprise data integration, and technology standards. The Tribiosys team has deep bio-pharmaceutical and information technology expertise. In the course of their careers, the Tribiosys team has successfully delivered mission-critical, enterprise-wide solutions to over a hundred Fortune 500 companies. Tribiosys has offices in Cambridge, Massachusetts and Mumbai, India.