Pragmatic Knowledge Management: An Integrated Architecture Based Approach

August 2002

Tribiosys, Inc.
215 First Street
Cambridge, MA 02142
Phone: (617) 252-0600
Fax: (617) 252-0640
Email: info@tribiosys.com
www.tribiosys.com

Tribiosys

Table of Contents

1 Executive Summary

Recent advances in genomics and rapid adoption of newer technologies like gene expression profiling, high content screening (HCS) and ultra high throughput screening (UHTS) have led to an explosion in the number of potential targets and leads that researchers need to pursue. As a result, biopharmaceutical R&D currently faces significant bottlenecks at the target validation, pre-clinical safety testing and clinical trials stages of drug discovery and development. In addition to the flood of new targets and leads, biopharmaceutical R&D faces a major data integration and management problem due to the large volumes of data generated by advances in screening technologies. Furthermore, collaboration and knowledge sharing across departmental boundaries is not a part of the organizational culture in large R&D organizations leading to redundant or wasted effort.

Effective Knowledge Management (KM) solutions helps minimize risk, cost and time in biopharmaceutical R&D by addressing the productivity bottlenecks in the drug discovery and development lifecycle. KM solutions assist with:

- Early identification of non-viable leads
- Reducing redundancy of effort
- Inefficiencies in the clinical trial process
- Retaining and sharing acquired knowledge

An organization faces numerous challenges in successfully rolling out a KM solution. These challenges range from availability, accessibility and relevance of data to the adoption and use of the KM solution within the organization. Organizations should approach KM from a global, pragmatic perspective – balancing technology, people and processes with their informatics needs to develop a solution that has clearly defined and measurable goals and benefits.

This paper presents a pragmatic approach for implementing KM solutions. The three main principles underlying this approach are:

1. Establish a sound, reusable, scalable and flexible architecture for the KM solution
2. Integrate organizational data so that it can be used across departmental boundaries
3. Implement the KM solution in incremental releases, with the initial release delivering functionality with immediate scientific, business, operational and organizational impact.

Tribiosys is a life science professional services firm that helps biopharmaceutical R&D organizations develop and implement bio-IT solutions, including data integration, management and analysis, knowledge management and regulatory compliance solutions.

2 Drug Discovery and Development is Expensive

In today's highly competitive and regulated environment, biopharmaceutical companies face numerous challenges - from discarding unpromising leads early in the discovery and development process to reducing the project cycle time for each drug that is being tested. Currently, it takes an average of $880M and 15 years to discover and develop a new drug.[1] Of this cost, about 75 percent can be attributed to failures of potential lead compounds along the way. Solutions that help facilitate and expedite decision-making on viable candidates significantly impact the bottom line – a reduction of even a day in the project cycle translates to a calculated increase in revenue of $1M per day to the company.

2.1 Impact of genomics on drug discovery and development

By applying genomics technology, companies could potentially on average realize 35% cost savings and 15% savings in time.[1] However, in the short-term, genomics along with other advances such as high throughput screening (HTS) has led to an increase in R&D costs. There are three main reasons for this increase in R&D costs:

1. Novel Targets
2. Novel Chemistries
3. Uncertain Clinical Development

Genomics has led to an explosion of potential new targets for researchers to pursue. However, researchers do not have adequate biological information about these new targets. This results in poorer quality of target validation and more problems downstream and worsened attrition rates. R&D costs are increased since these poorly characterized targets are only eliminated from consideration after costly clinical work. Secondly, companies are pursuing a much higher proportion of novel targets – these novel targets require novel chemistries. The chemistries of compounds that interact with well-substantiated targets are well understood. New targets require new chemistries that have not been worked out, particularly if the targets are not a member of a well-understood class of proteins. Such targets face the risk of toxicity which may not be apparent until the drug has gone through expensive clinical trials. Finally, the more novel the target and less the information known and available about the target, the more likely it is that the clinical trial process will be prone to expensive trial and error.

This increase in the number of targets and leads has led to bottlenecks at the target validation, pre-clinical safety testing and clinical trials stages of drug discovery and development [Fig 1].

[1] A Revolution in R&D. HOW GENOMICS AND GENETICS ARE TRANSFORMING THE BIOPHARMACEUTICAL INDUSTRY. THE BOSTON CONSULTING GROUP. NOVEMBER 2001.

Fig 1. Current Bottlenecks in R&D. Increase in the number of targets and leads has led to bottlenecks at the target validation, pre-clinical safety testing and clinical trials stages of drug discovery and development.

2.2 Data explosion

The amount of data being generated from the target identification to regulatory approval stages is increasing exponentially with the rapid adoption of newer technologies like gene expression profiling, high content screening (HCS) and ultra high throughput screening (UHTS). This has resulted in a huge data integration and management challenge for biopharmaceutical organizations. In order for these data to be useful, they need to be:

- Accessible,
- Accurate,
- Meaningful,
- Relevant, and
- Actionable

Finally the information contained in the data needs to be ultimately adopted by the intended recipient. Once it is in this form the data can be used to improve decision-making on potential targets and leads.

2.3 Imperative for organizational change

The situation is further exacerbated in large R&D organizations where collaboration and knowledge sharing across departmental boundaries is not a part of the organizational culture. Further, large organizations are also geographically spread out, which limits the interactivity between human resources. This has led to an inefficient knowledge sharing environment in biopharmaceutical organizations with a resulting loss in productivity. Although management and employees readily admit that the situation is not optimal and that there is plenty of room for improvement, the sheer complexity of the problem has led to a slower adoption of Knowledge Management (KM) solutions.

3 The Business Need for Knowledge Management

Currently, the drug discovery and development pipeline has bottlenecks at the target validation, pre-clinical safety testing and clinical trials stages. Pursuing non-druggable candidates during these stages cost time and money and increase the risk for biopharmaceutical organizations. A number of factors, enabled by KM solutions, can help minimize these costs. These factors include:

- Early identification of non-viable leads
- Reducing redundancy of effort
- Inefficiencies in the clinical trial process
- Retaining and sharing acquired knowledge

3.1 Early identification of non-viable leads

Currently, there appear to be no new scientific technologies on the horizon that are likely to alleviate the bottleneck at the pre-clinical/ADMET stages. As a result of the adoption of HTS, HCS and UHTS it is now possible to screen a very large number of leads. However, most of the leads that pass these screens need to be tested in biological systems of increasing complexity in order to be validated further. Finally, the drug needs to be tested in clinical trials.

Unlike *in vitro* screens, biological screens and clinical trials cannot be accelerated beyond a certain point. The time it takes to complete them depends on fixed biological variables such as cell growth rates, and rodent assays. Further, the complexity of the biological system used in the screen adds to the time and cost of the screen and the unpredictability of the result. Human clinical trials continue to be conducted in a very traditional, painstaking and expensive manner. Thus, the net effect of the recent advances in scientific technology has been rapid generation of a large number of leads that need to be validated – but no quick way to validate them.

A logical way to improve this process would be to only analyze those leads that have the most likelihood of being viable. Emerging technologies such as predictive ADMET and toxicogenomics directly benefit this area. However, these approaches still need to be validated before they are widely accepted in the industry. A major shortcoming of using these technologies is that they rely on relatively small predictor datasets and thus are of limited reliability. This shortcoming may be overcome by using predictive ADMET and toxicogenomic data in conjunction with real-life experimental data previously obtained by an organization. Thus, data for the many leads that have already been investigated by an organization in conjunction with toxciogenomic data for the same leads can be used to build a far more reliable predictor set than those available commercially. This combination of theory with reality is unique to every organization as each organization has its own data set from leads that it has investigated.

3.2 Reducing redundancy of effort

Often times, leads are run through the validation process for one project and the information acquired remains with the project. If the same lead is being tested for another project, there is often no way of easily determining what an organization has previously learned about that lead. This results in the lead being tested again. Likewise, if a molecule very similar to the current lead has been tested before, then everything that was learned about it during the validation process is relevant to the current lead, as it is likely to have similar biological and biochemical properties. Previous information acquired on issues such as bio-availability in different tissue types or behavior during clinical trials is critical for making a "go/no go" decision on the potential viability of a related lead. Thus, even if the lead in question failed during some part of the validation process, the biological, chemical and clinical information captured during the process is very valuable information, and as such represents the "gold-standard" for these kinds of data. This knowledge is certainly superior in quality to a purely *in silico* prediction of how the lead is likely to behave in a biological system or humans.

There are many potential benefits to harnessing this kind of knowledge. For example, if the outcome of a previous project was that there were no "druggable" compounds that could be synthesized for a certain target (regardless for the indication or disease model the project was initiated for), then early research as well as chemistry would benefit from the knowledge of what targets are tractable. Similarly, if a lead was being validated for a particular target but failed because of sub-optimal bioavailability in the target organ, acceptable bioavailability of the compound in other organs may be grounds for reinvestigating it for similar targets in the other organs. This is especially useful for many types of cancer, where the genes involved often belong to the same family. For example, the HER family of oncogenes are involved in many different kinds of cancer – HER-2 is an important gene for breast cancer, while HER-3 has been implicated in pancreatic cancer. Both oncogenes are involved in Epidermal Growth Factor regulation and share similar structural properties and thus are likely to be affected by similar compounds.

3.3 Inefficiencies in the clinical trial process

Clinical data management is an extremely time- and cost-intensive process; pharmaceutical organizations are often unable to effectively manage clinical trials with current data management methods. Advances in human genomics and other development technologies threaten to bring exponential growth in the number of new drug targets and new trials over the next decade. Trial managers have poor visibility into trial progress, they do not have real-time access to critical decision-making data, because of the manual nature of data and analysis. In the past, the FDA has rejected data from trials which have been run incorrectly or which have captured the wrong data – at a cost of millions of dollars and years of time. A collaborative KM environment makes critical decision-making data available to trial managers on a timely basis.

3.4 Retaining and sharing acquired knowledge

Employees are by nature transient resources. However, it is very important that what they learn during their employment remains part of the organization's collective memory, even after they leave the organization. Employee turnover can result in the loss of important corporate assets. In an R&D environment, employees tend to be highly specialized resources that cannot be replaced easily. Due to the cutting-edge nature of their jobs, they acquire a lot of knowledge mostly through experimentation, a trial-and-error process. For example, in companies that develop therapeutic proteins, knowledge gained while developing processes for expression, purification and crystallization of the protein in question is very valuable if the organization is later developing a related family member as a therapeutic protein. This information is beneficial to all the groups involved in the project including biology, chemistry and product development. In such scenarios, knowledge of what did not work is as valuable as knowledge of what did succeed. However, in a typical organization, information about failures is not as well documented as that of the successes. Very few organizations have mechanisms in place that capture such knowledge, as it is a relatively difficult task to accomplish and requires significant employee involvement. The benefits of retaining this knowledge for an organization are immense.

Furthermore, knowing who the experts are for a given area in an organization and having access to them, is an equally important factor in facilitating the R&D process. This is critical in large organizations with multiple geographically isolated centers that perform similar tasks within an organization. This radically decreases the efficient flow of information within the organization as a whole, as one's primary sphere of contacts tends to be the people at a given location.

4 Challenges in Knowledge Management

"Knowledge" for a biopharmaceutical organization spans a number of extremely disparate disciplines. A successful organization depends on biologists, chemists, clinicians, engineers, regulatory specialists, lawyers and marketers to bring a drug to market. These are highly specialized disciplines as biologists may be neurobiologists, immunologists, virologists, or bioinformaticians, while chemists may be synthetic, medicinal or organic chemists, to name a few. Each group of specialists has its own lexicon, diverse knowledge needs, and they produce and consume different kinds of data. However, successful KM requires that relevant information from each specialty be harnessed, appropriately assembled and transformed and delivered to the right person at the right time. Relevant information is not necessarily restricted to individual groups, as something a clinician finds may be useful to a chemist, while a biologist's discovery may be useful to a lawyer. Relevant information is also not restricted to what lies within an organization's boundaries since internal information in conjunction with external information (such as commercially purchased gene expression or pathway data) may be ultimately what is relevant.

An organization faces numerous challenges in successfully rolling out and reaping the intended benefits from a KM solution. These challenges include:

1. Accessibility of the data: Often the centers of data production and data consumption are distributed within the organization. The challenge lies in identifying the sources and consumers of information.
2. Accuracy of the data: Valuable raw data generated by a particular group within an organization may need to be validated before being transformed into normalized or consistent representations and content.
3. Meaning of the data: Information derived to data generated by one group may need to be mapping to a standard context in order to be meaningful to someone else in the organization.
4. Relevance of the data: Does the data support and truly answer the question being asked by the user. This requires the appropriate meta data about the data to be held in the KM solution.
5. "Actionability" or ability of the data to support/deny hypotheses: Does the information truly support decision-making? Does the KM solution include a statistical or rule-based model for the workflow within which the question is being asked?
6. Adoption of KM solution: Does the organization culture foster and support the voluntary usage of the KM solution?

5 A Pragmatic Approach for Knowledge Management

Biopharmaceutical organizations should approach KM from a global, pragmatic perspective – balancing the mix of technology, people and processes with their informatics needs to develop a solution that has clearly defined and measurable goals and benefits. This approach is based on three underlying principles:

1. Establish a sound, reusable, scalable and flexible architecture for the KM solution. This architecture will enable integration with current and future products and technologies. It also provides the greatest flexibility, scalability and extensibility to the biopharmaceutical organization to respond to future product, technology and business changes.
2. Integrate organizational data so that it can be used across departmental boundaries. Organizations can minimize a significant amount of wasted or redundant effort by making integrated biological, chemical and other experimental data accessible to all relevant stakeholders in a timely fashion.
3. Implement the KM solution in incremental releases, with the initial release delivering functionality with immediate scientific, business, operational and organizational impact. This approach allows organizations to refine the architecture and continually deliver high value business functionality to the end users.

5.1 The knowledge framework

After establishing the KM goals, requirements and priorities, organizations should develop an inventory of data sources and processes and use the results of the analysis to design a robust, scalable, standards and component-based KM architecture – the knowledge framework - the framework for institutional learning. The knowledge framework [Fig 2] enables the following features:

- Portal for one stop shopping of discovery information
- Personalized information services for alerting and access, based on roles
- Knowledge base of people (experts and communities) and information assets
- Modeling and simulation / data analysis and statistics / linking knowledge and interpretation
- Standardized data repository to link information to data
- Integration to operational and external data sources

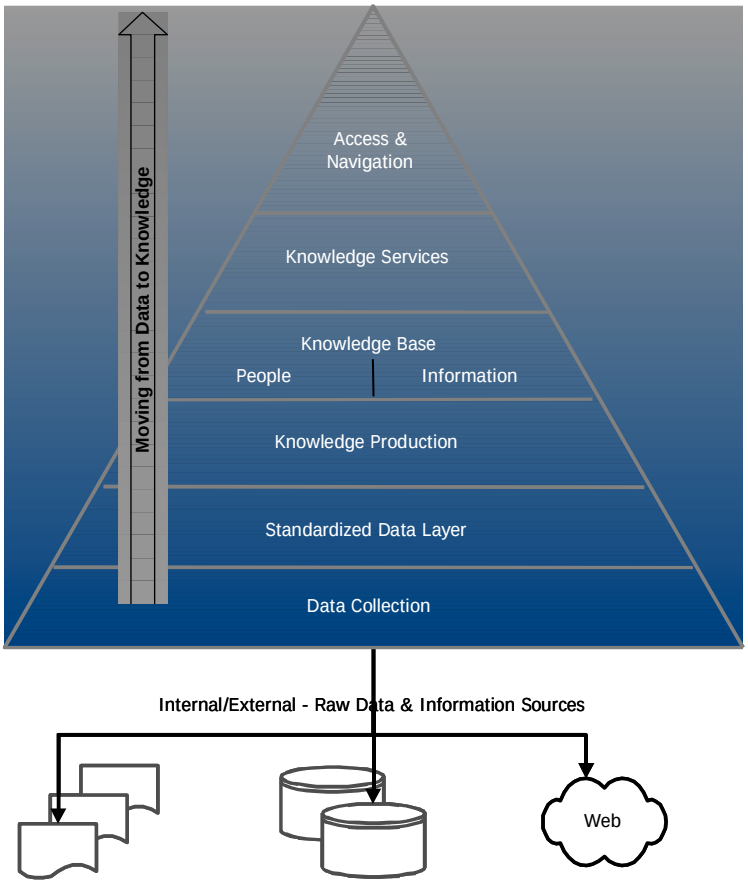

Fig 2. Knowledge Framework. The knowledge framework is a conceptual, multi-tier architectural model of a KM solution.

The knowledge framework is designed to integrate decision support into daily user workflow. The knowledge framework is rolled out to the organization via a phased approach that delivers business functionality via incremental releases. The initial release of the knowledge framework should select an area with immediate scientific, operational and organizational impact. The first release implements the complete end-to-end knowledge framework, designed to maximize correct data use and user-driven functionality. The knowledge framework is implemented using Tribiosys Solutions Methodology that includes the following phases:

- KM Strategy – The goal of this phase is to achieve consensus on data integration, informatics and KM priorities amongst the key stakeholders and users in the organization. An implementation roadmap is developed based on the organizational priorities.
- User Requirements – This phase elaborates on the user, functional and architectural requirements of the KM solution based on priorities from the KM Strategy phase.

- Architecture Design – This phase results in the design of the technical infrastructure and application architecture of the KM solution. The architecture is a pragmatic combination of commercial off-the-shelf components and custom functionality. The architecture design creates a scalable, flexible and extensible model for integrating R&D data across the organization.
- Data Mapping – The goal of this phase is to create an integrated data model of the information in the KM solution and map the entities in the data model to the sources of data from operational systems.
- Application Design – This phase results in the functional design of the KM solution. It specifies the application modules, application interfaces, the user interface, design of the business rules, physical data model etc.
- Application Development and Testing – This phase results in the development and testing of the first release of the KM solution. During this phase application components are integrated and developed, as necessary. The entire application is then rigorously tested via Unit, System, Integration and Acceptance testing.
- Application Rollout – The KM application is rolled out to the end users in a planned manner.

Once the first release of the KM application has been rolled out to the R&D organization, subsequent releases and enhancements are developed based on the integrated, architecture design.

5.2 Case study

Tribiosys is currently implementing a KM solution for a privately held biotechnology company developing novel therapeutics by using its proprietary knowledge of critical pathways for pathogenesis. In this organization experimental data and results were held in a silo-ed information environment and the scientific workflows were non-collaborative and often times redundant. With the turnover of certain critical scientific resources, the company was faced with the loss of key intellectual property. The lack of access to integrated internal experimental biological and chemical data and external biological and cheminformatics data was hampering productivity and quick decision-making within the company.

Tribiosys developed a KM solution that integrates biological and chemical data and activities around the drug R&D processes. The KM solution integrates relevant data while supporting day-to-day operational use. The solution is a combined build and buy model that maintains proprietary, differentiated, intellectual capital. The solution establishes a reusable, scalable, and flexible underlying data architecture for the biopharmaceutical company. [Fig 3].

The company has adopted a phased approach to the implementation of the knowledge framework with the first release delivering researchers access to integrated biological and chemical data, enabling researchers to compare and analyze information across experiments. Tribiosys is using its Solutions Methodology to architect, design, develop, test and rollout the system.

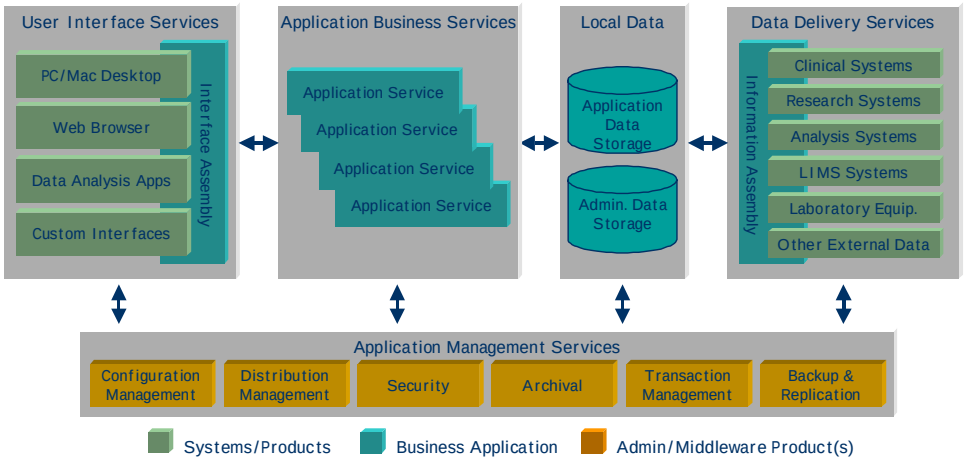

Fig 3: Multi Tier Architecture for the KM solution. The various services within the architecture collaborate to deliver the KM functionality.

The KM project resulted in a reusable, scalable, flexible enterprise architecture that established integration of biological and chemical data across the drug R&D processes. The architecture was also compliant from a regulatory point of view.

As a first step in the implementation, Tribiosys gathered user requirements to identify, document, and prioritize the business needs of the KM solution. The result from this phase was a set of business and scientific requirements from the various R&D groups within the company and corresponding functional requirements that the architecture solution must support. The functional requirements were prioritized as a part of the implementation plan with regards to their criticality to the business and scientific priorities. At the conclusion of the user requirements phase, the current operational systems and other data sources were in order to understand the underlying software architecture, data model, functionality, and integration requirements. Following the Operational Assessment phase, Tribiosys conducted a Data Mapping phase to define a preliminary identification mapping structure to tie biological and chemical data sets from various operational data sources.

During the High Level Architecture Design phase, Tribiosys conducted an in-depth review of current hardware, software, and application environment in order to provide a high level architecture and design plan. Within this design, the component layers of the architecture were refined and specified. The data extraction/data integration layer was specified to support the operational systems identified in the previous phases. The architecture design activities resulted in a specification of the requirements for each layer along with recommendations for off-the-shelf software products to support each architectural layer.

The architecture design realizes the knowledge framework for the company and the KM initiative will deliver the complete business functionality via incremental application releases.

6 About Tribiosys, Inc.

Tribiosys is a life science professional services firm that helps biopharmaceutical R&D organizations develop and implement bio-IT solutions, including data integration, management and analysis, KM and regulatory compliance solutions. Tribiosys bio-IT solutions include Better Experimental Design, Comprehensive Gene to Drug Data Integration, Lab Optimization, Regulatory Assessment and Remediation, Bioinformatics and Cheminformatics Analysis and Application Development and Integration Services. Tribiosys' primary customers are pharmaceutical, biotechnology and medical device companies worldwide.

Tribiosys resources work in multi-disciplinary teams that combine life sciences and IT expertise. Tribiosys designs KM solutions with multi-tier architectures that conform to the following principles:

- Build only open systems
- Leverage existing investments when possible
- Integrate best of breed commercial off-the-shelf solutions as needed
- Build in adaptability for the future

Tribiosys employs proven and disciplined methodology and best practices around quality and project management, based on an experienced staff that has developed and refined these practices over many years and complex projects.