PERSEID
SOFTWARE LTD

Enabling the Real-Time Life Sciences Enterprise with an IT Infrastructure

Perseid Software Limited
Needham, Massachusetts

February 2002

Table of Contents

Figures

1 Executive Summary

Over the last four decades, information systems professionals have labored to develop the basic technology required to integrate the information required to manage an enterprise precisely and reliably. From disciplines such as Management Information Systems (MIS) crafted in the 1960s to the blossoming of data warehousing and business intelligence approaches in the 1990s, each phase of technology development has resulted in faster and more accurate integration of enterprise data.

Although there have been some spectacular successes in integrating data such as American Express' real-time credit approval business intelligence application and real-time Wall Street trading workstations tied to historical databases of equities prices; many data warehousing and decision support applications have not been deemed stellar successes by senior management. Often over-budget and technically complex, these systems have rarely become the foundation of modern enterprises, particularly in the Life Sciences. By leveraging the positive experiences of other industries, Life Sciences business intelligence and data warehousing professionals can adapt best practices to quickly gain success.

As Life Sciences and healthcare applications merge to create an integrated value chain, the need for analytical information to manage such an extraordinarily data intensive and complex industry will become even more demanding. The next generation of data warehousing and decision support systems will house massive amounts of critical data from bioinformatics, clinical, marketing and financial processes. But, these systems need to evolve to support the immediate management and production demands of the enterprise rather than be viewed as historical and passive "repositories" of retired data. Emphasis must shift from simply "warehousing" information to rapidly transforming raw data into analytical analyses—with the results delivered to the Life Sciences executive in *real-time*. Data warehousing, driven by the enormous expansion of the Internet and inexpensive telecommunications, is evolving into a new discipline to support the fully "wired" enterprise. The era of Real -Time Enterprise Computing (RTEC) based on integrated Life Sciences business intelligence applications has arrived.[1]

RTEC will transform the information technology industry and customer expectations. Having the data will be necessary, but not sufficient. Using derived analytical information to influence management and customers in real-time will become the focus of the Life Sciences enterprise. Those who can get up-to-date information and deploy it effectively will have an enormous competitive advantage as RTEC systems maximize the use of capital, personnel and enterprise resources.

This paper will examine the needs and principles inherent in real-time enterprise computing for next-generation information systems that will be required by Life Sciences enterprises. The foundation technology of real-time enterprise computing—architectures to support integrated enterprise databases—will also be introduced.

[1] Michael S. Malone, "Internet II: Rebooting America" Forbes ASAP, September 10, 2001

2 Data Warehousing and Life Sciences Enterprises

2.1 Creating a Foundation for Real-Time Enterprises

Data Warehousing and Decision Support Systems are decades old disciplines evolving into the foundation for a new generation of high performance enterprises.

Many business intelligence and data warehousing projects fail either because the projects take too long or the management and engineering requirements are not reflected accurately in the design of the solution.[2]

With Bioinformatics being a relatively new discipline based on rapidly evolving science, information systems designers and architects can expect challenges in designing integrated Bioinformatics data warehousing and business intelligence applications. The difficulties will arise from data integration complexities, requirements analysis, data security and trying to meet rapidly changing management needs. Success is imperative, since the very foundation of the integrated Life Sciences enterprise depends on converting, rapidly, data to the knowledge required for new drug discovery and approval. Management needs information in real-time and traditional data warehousing and decision support systems will be redesigned in response to these requirements.

2.2 The ROI for Life Sciences Data Integration

With investigational and new drug development consuming as much as US$800 million and ten years of development,[3] moving the Life Sciences enterprise to real-time may sound overly ambitious, yet it is a critical success factor for Life Sciences companies. Real-time decision-making triggers compression in product development, deployment and post-market analyses of new drugs and discoveries, saving millions of dollars in the process.

Applying market research from McKinsey and Company,[4] each month the R&D cycle for a new drug is compressed out of the development schedule translates into as much as $6 million in savings and adds as much $100 million per month to sales for a "blockbuster" pharmaceutical. At a price/earnings ratio of 30, US$6 million is worth US$180 million in shareholder value for *each month* of R&D time saved. Clearly, time is both money and shareholder value in the Life Sciences.

2.3 Merging Market, Administrative, Financial and Clinical Data

No industry integrates more complex information than the Life Sciences. From units of organization ranging from the molecular, document, family unit and, ultimately, country and government, Life Sciences companies generate enormous quantities of documents and data. Market information, regulatory documentation, disease management, clinical trials, detailed research and development, medical records, claims data and documents must all be integrated to improve the clinical efficacy and financial performance of a new pharmaceutical for a specific target market and patient

---

[2] Riggle, Mark, "Breaking the Cycle of Failure", *Intelligent Enterprise*, August 10, 2001
[3] "Tufts Center for the Study of Drug Development Pegs Cost of a New Prescription Medicine at $802 Million," Tufts Center For The Study Of Drug Development, Press Release, November 30, 2001
[4] Bhandari, Garg, Glassman, Ma and Zemmel, "A genetic revolution in health care", *The McKinsey Quarterly*, 1999 Number 4, pp.58–67

population. As indicated in Figure 1, the scale and layers of information in a modern Life Sciences Bioinformatics repository proceed from the general (the market) to the specific (the individual). With the fully burdened costs of new pharmaceutical development spiraling upwards, data and document integration impacts speed-to-market and shareholder value.

According to PricewaterhouseCoopers, full data integration, from passive repositories to active repositories of data and documents will facilitate:[5]

- Better compliance and faster regulatory approval
- Creating a management environment of flexible and adaptive decision making by utilizing scenario planning
- Using predictive market modeling and simulation to develop smaller market segments, new pharmaceuticals and devices
- Monitoring a global "e-environment" of markets and research and development
- Tracking and documenting emerging trends and competitive threats
- Better allocations of internal and partner resources
- Adaptive experimentation in marketing, i.e., electronic CRM for patients
- Real-time management review and iteration of decisions

Figure 1 highlights tiers of global management, financial, marketing and clinical data and documents. The data and documents are stored to support an integrated view of the pharmaceutical enterprise. Market segment research can be conducted at the population level and tracked over a period of years. Integrated clinical, biological, financial, regulatory and efficacy documentation and data are tracked at all levels of the enterprise. Because of the massive size of the central repository, the "layers" of the data pyramid may be implemented as multiple physical databases.

Tiers of Bioinformatics Data

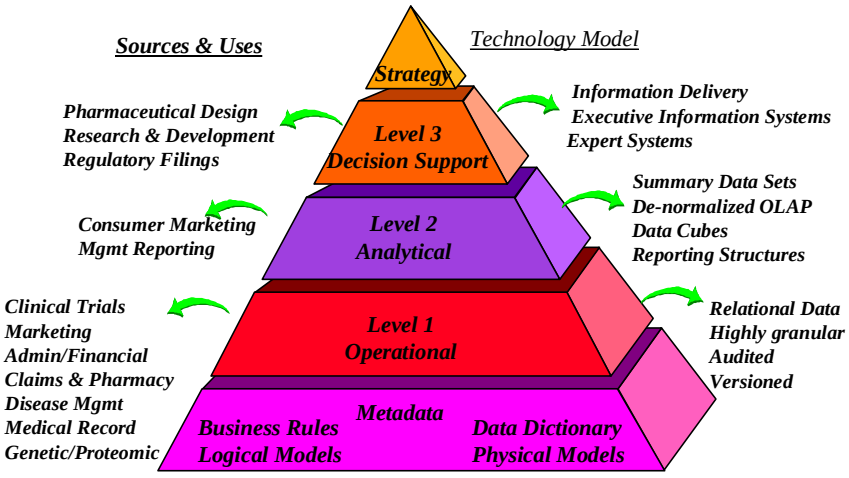

Figure 1 Life Sciences Enterprises Generate Tiers of Complex Data

[5] Pharma 2005, "Marketing to the Individual", PricewaterhouseCoopers, 2001

2.3.1 Bioinformatics — The Ultimate in Data Integration

To design, test, deploy and manage during new drug discovery, or to re-market an older pharmaceutical, data must be integrated, managed and deployed to serve multiple simultaneous purposes. To manage in real-time requires the integration of complex enterprise information:

- Market — Marketing information by country and metropolitan area
- Financial — Design, development, deployment and marketing data
- Research — Molecular, genetic, proteomic and pharmacological data
- Clinical — Clinical trial, outcomes and post-market effects of drugs
- Biological — Proteomic, genomic and other biological and chemical information
- Administrative — Healthcare claims, membership, diagnostic and treatment data
- Metadata — Information about what is contained in the data warehouse

2.4 "Active" Repositories Emerging as Real -Time Management Tools

Historically, data warehousing and business intelligence applications have been viewed as "back room" batch-oriented activities beyond the pale of serious operational executives with immediate needs. But increasingly, Life Sciences enterprises need Bioinformatics data converted to knowledge that can be acted upon immediately to maintain a competitive edge. For example, gaining the real-time knowledge of mutations in the CYP3A family of cytochrome P450 and the cardiac effects of antihistamine Seldane on a very small class of patients could have saved Hoechst Marion Roussel a US$600 million a year franchise.[6] Not only are new genetic tests and procedures required, but also clinical information must to be integrated and analyzed continuously—to facilitate identifying adverse reactions and to immediately detect promising new drug applications once a product is introduced into the market.

3 Real-time Enterprise Computing in the Life Sciences

3.1 Designing Real-time Enterprises Computing Solutions

Success evolving data warehousing and decision support systems into "active" repositories supporting RTEC requires careful planning and attention to detail. This basic approach will facilitate success:

- **Focus on Requirements Management** — Match the needs of operational and senior management carefully during the design and architectural development phases.
- **Realistically Schedule —** Rapidly prototype a solution and develop schedules based on incremental delivery of new solutions—ship solutions early and do so often.
- **Think "Real-Time" —** Immediately use available data that the organization has generated. Rapidly show operational and strategic management visible and useful solutions. Think of the enterprise as a real-time organization and plan the design and data acquisition processes accordingly. Depend on the Internet for speed of distribution.
- **Use State-of-the-Art Technology —** Use experts in research, healthcare, finance and the biosciences with experience in large-scale database and information systems design. Buy

[6] Manish Bhandari, Rajesh Garg, Robert Glassman, Philip C. Ma, and Rodney W. Zemmel, "A genetic revolution in health care," *The McKinsey Quarterly*, 1999, Number 4, pp. 58–67.

reliability and availability as each day a pharmaceutical is in the market may be worth millions of dollars.

- **Focus on Solutions —** Speed solutions to operational and senior management and products to markets.

3.2 The Internet is the Foundation for Real-Time Management

The Internet, modern data base management systems and non-stop computing architectures enable the foundation for RTEC. Although computational costs have fallen linearly 30% annually since the 1970s, communications costs in the last five years have fallen exponentially.[1] Telecommunications bandwidth has also increased exponentially. Transmission speeds now exceed one million bits per second, up from 19,000 five years ago. This is a 50-fold increase in performance in five years.

Improved connectivity means management cooperation and new methods of planning are possible for the Life Sciences enterprise in ways that were unimaginable ten years ago. Thus, real-time enterprise computing means that operational and research personnel can ask questions today of their decision support systems that can influence the *immediate* behavior of the enterprise in areas such as:

- Operational Control
- Research & Development
- Deployment
- Market Research
- Patient Outcomes
- Consumer Perceptions and Behavior

One example of these improvements is that pharmaceutical companies can make more immediate and better use of information they already possess to produce safer drugs. This is not just a case of re-using their own data from clinical trials on related drugs to reduce the number of trials and refine the clinical development process; it is really about tapping into the flow of healthcare data, generated by providers of care.

For example, although politically and legally complex, PricewaterhouseCoopers suggests pharmaceutical companies should merge non-competitive data on toxicity and serious adverse events, using independent intellectual property exchanges.[7] The authors point out—"However, any organization that wants to capitalize on this trend will need a much greater degree of connectivity, including integrated systems that enable it to collect information on *individual* patients and distribute it to scientists and marketers across the entire enterprise."

3.3 Accessing Real-Time Information in a Global Enterprise

Life Sciences companies are global organizations leveraging global bioinformatics research and development processes. Research organizations, clinical trial organizations and markets are global in nature, but operate within the purview of a sovereign national government. This means that market information; local clinical variations and outcomes should be centrally stored, but locally available.

---

[7] Simon Hughes and Sam Barnett, "Getting Wired for a World of Individuals**",** PricewaterhouseCoopers Pharmaceutical Practice, In Vivo - November 2000

Molecular/genetic variations of a drug must be known quickly in order to produce slightly different formulations to affect a new molecule that has a slightly different clinical outcome, thereby removing side effects for a sub-population of patients. Data warehousing evolves in this scenario from a passive "back-end" research or reporting tool to become an active component in the simulation, design, construction, deployment and evaluation phases of new product development. Rapid response becomes the mantra of the Bioinformatics professional as he/she seeks to convert the enterprise to real-time. The potential benefits are numerous:

- Shareholder value is increased as days, weeks or months are removed from design, development and marketing
- Medical errors are reduced as negative side-effects are identified earlier
- Software simulation refines the process of drug development improving the "targeting" ability of a new drug formulation
- Market research databases become focused on the individual and patient sub-populations for which drug variations are more effective and safer
- Behavioral change occurs along the pharmaceutical value chain resulting in an operational orientation where "cheaper, better, faster" new drug investigation and development is the norm, not the exception

4 Enabling Real-time Enterprise Computing

4.1 Evolving from Separate Databases to Integrated Real-Time Repositories

Current estimates from PwC suggest that the volume of biological data alone is doubling every six to nine months, which is twice the rate at which computer processing power is expanding. As a result, the Bioinformatics profession will have to manage a 100-fold increase in data over the next few years.[8] Much of this data will be focused on "fine tuning" pharmaceutical products for better outcomes and fewer side effects for smaller and smaller market segments. For example, business intelligence applications in marketing are evolving from a focus on "pushing" drugs into a market without much *a priori* information on individual consumers to a focus on the consumer "pulling" a product into personal use through their physician based on immediate real-time clinical needs.

Figure 2 describes the traditional model of data warehouse design, that is, separate databases each serving a specialized reporting function, but without common data integration. As a result, this architecture makes it difficult to reliably merge clinical, financial, marketing and claims data into an accurate central reporting system. What is needed for real-time enterprise computing is the integration of the databases into a common repository, using shared "metadata."[9] This improves data integration, removes redundancy and encourages a focus on the individual consumer or patient.

[8] PricewaterhouseCoopers, "An Industrial Revolution in R&D," Pharma 2005

[9] "Metadata" is information in the Data Warehouse that identifies the contents of the warehouse itself—its fields, tables, reliability and validity, for example.

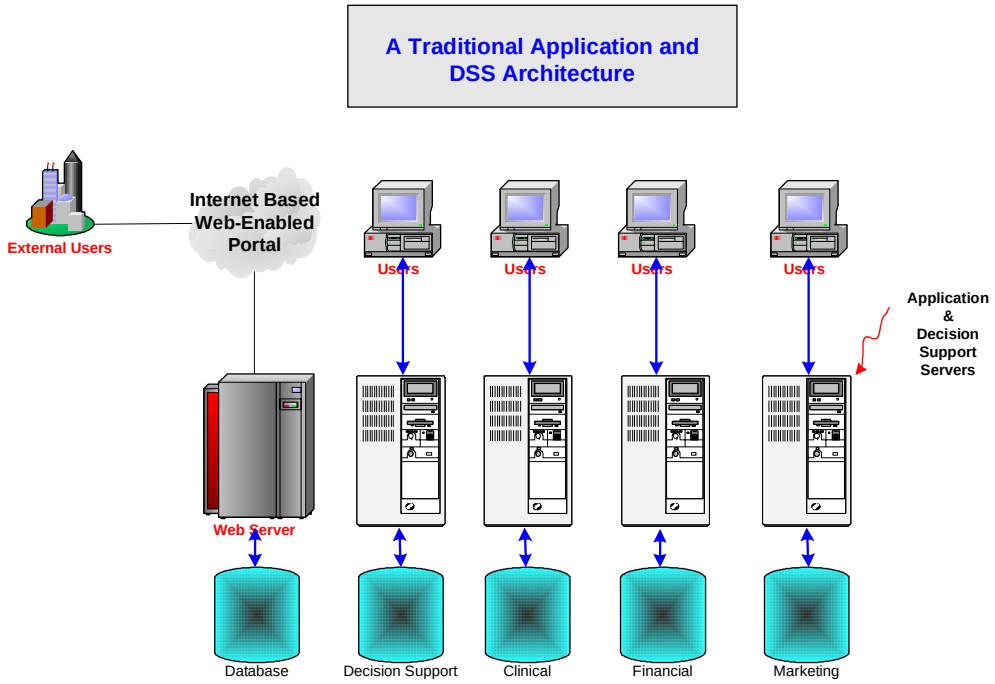

Figure 2 A Traditional Architecture Lacking Integration of Databases

4.2 A Model Real-Time Life Sciences "Active" Repository

With extreme degrees of regulatory supervision and liability exposure, Bioinformatics data warehousing and business intelligence applications must meet or exceed current state-of-the-art standards for validity, reliability and security.

With petabytes of data being used in real-time, production transaction and business intelligence applications for marketing, clinical trials and research and development cannot ordinarily be "backed-up" to magnetic tape. Mirrored 24x7 database management systems must be put in place to ensure the continuity of these real-time enterprise information systems. The evolution of the poorly integrated databases and applications in Figure 2 to the next generation real-time architecture is shown in Figure 3.

Figure 3 depicts a fully integrated foundation for real-time enterprise computing in the Life Sciences. This example is based on EMC's "E-Infostructure" which is composed of physical, connectivity and functional layers of hardware and software supporting one or more integrated database management systems with a common logical database design architecture and integrated metadata.

The physical layer includes Symmetrix™ and CLARiiON™ enterprise storage systems that provide the basic foundation for performance, capacity, availability and other physical requirements of the central repository and database management systems. The Enterprise Storage Network ("ESN") is the connectivity layer and through two information connections—Connectrix™ and Celerra™—it provides a means of using *all* primary and secondary operating systems to connect into the enterprise application and data management platforms. These systems could include IBM operating systems, Unix operating systems, including Linux and those from Compaq, Sun, HP, and Microsoft operating systems.

The storage management layer is composed of the Enterprise Storage Network ("ESN") and the Database Management System(s). EMC Symmetrix™ and enterprise storage management software are used to integrate all operating systems and database storage into a uniform central repository. EMC Connectrix ™ switches are used to handle multiple connections to the servers. EMC TimeFinder™ software provides each researcher with local copy of certain data to analyze, thereby increasing researcher productivity. TimeFinder can be used to refresh data warehouses with timely information without disrupting production systems. The remote mirroring capabilities of EMC's SRDF™ (Symmetrix Remote Data Facility) software can protect enterprise databases and other critical data to avoid costly interruptions in speed to market. EMC TimeFinder is used for non-disruptive backup and data warehouse loading with EMC SRDF used for disaster recovery and information mobility. The EMC solution supports continuous availability of reporting, decision support and web access to the enterprise repository and continuous availability of databases at the core of the federated data base management system.

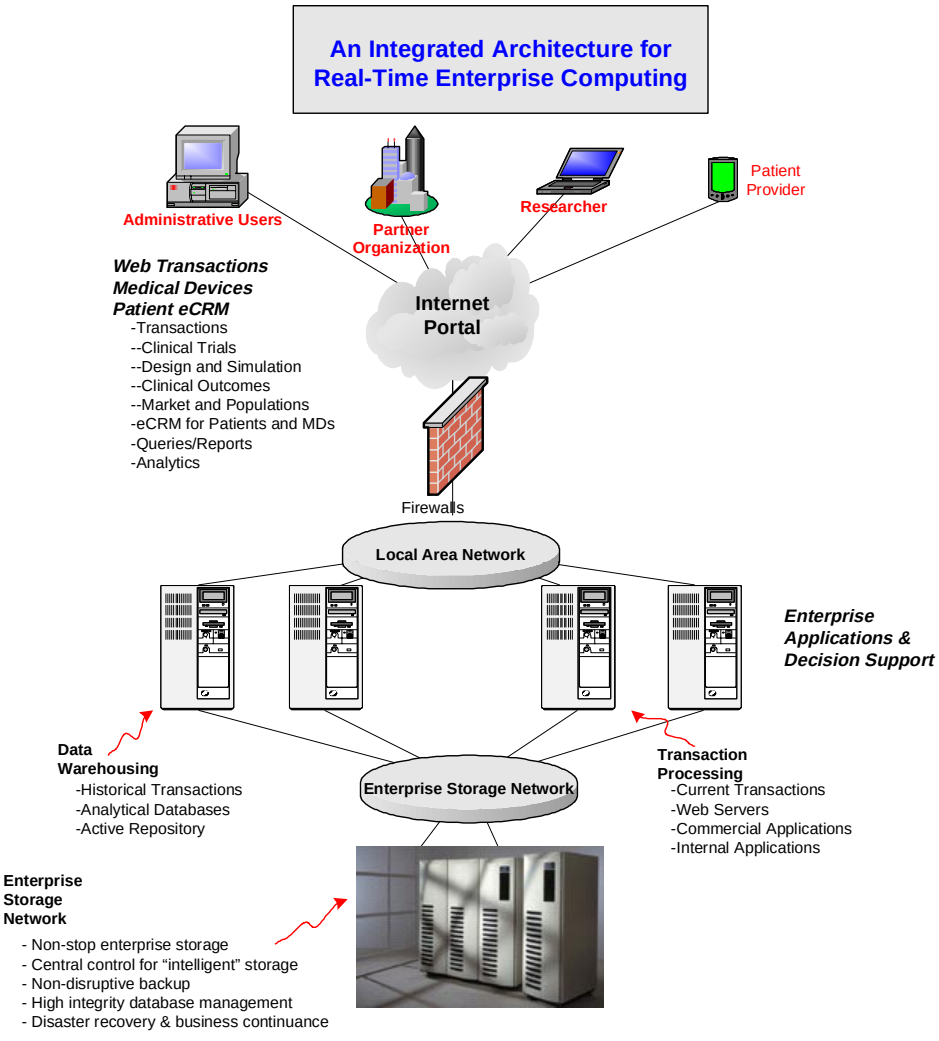

Figure 3 A Model Architecture for the Integrated the Real-Time Enterprise

The applications and databases are divided among production RTEC transaction processing and business intelligence applications. New data may arrive from numerous transaction processing systems as depicted in the telecommunications "cloud" of Figure 3 or from other databases—from partners, research organizations or third-party data vendors. As new data arrives, the production databases must be updated and reprocessed. New clinical trial data, for example, may arrive and require that analytical results to be amended. Market research and personal medical information, eventually flowing from the patient/consumer directly, will form the foundation of a personalized CRM application that allows a pharmaceutical to be crafted for a *subpopulation* of patients. This level of marketing precision will place heavy emphasis on timeliness, accuracy and reliability of the databases within the enterprise storage network. However, once sufficient data integration has been achieved, the real-time Life Sciences enterprise will have been born.

5 Conclusion

Decision Support and Data Warehousing applications for the Life Sciences are evolving rapidly, driven by the enormous amounts of data generated in clinical trials and new drug discovery processes. This data serves as the intellectual capital source and foundation of the modern Life Sciences enterprise.

Traditional data warehousing and business intelligence applications have been based on the long-term accumulation of data, using a passive style of information management. The goal in the past has been to archive information for later use. Rarely has the process of building business intelligence applications been focused on data integration with an intensity that would facilitate using the data to influence the core business processes of the Life Sciences firm in real-time.

Focused on using metadata defining integrated clinical, administrative, marketing and bioinformatics data, the new strategy for the real-time enterprise is about rapid integration of data and rapid application development. To achieve real-time enterprise computing, independent databases must be merged into a foundation repository using enterprise application integration tools. A common enterprise information storage architecture can then be put in place to rapidly develop business intelligence and decision support applications. Using this approach, the emphasis for the Life Sciences enterprise gradually shifts from a focus on data integration, that is, simply getting the data into a useful format and stored for later use, to a focus on the rapid production of business intelligence solutions that influence the enterprise in real-time.

To be successful, a particular approach is required for real-time enterprise computing. Understanding the requirements for each incremental solution is imperative. Responsive information systems departments must clearly understand each business intelligence need and have the tools to rapidly craft a solution. The technical architecture must be enterprise-wide and industrial strength. Data must be both reliable and valid. Solutions must be constructed using rapid application development techniques and flow in a predictable and continuous manner

Real-time enterprise computing for the Life Sciences promises to introduce to the Life Sciences, management techniques and integrated business intelligence applications that are seen far more often in traditional high-technology manufacturing than in the pharmaceutical industry. But with speed-to-market a critical driving force in the Life Sciences, real-time information systems and management processes that have been successful in other disciplines and industries will begin to appear in Life Sciences enterprises. An enormous competitive advantage awaits the enlightened and motivated Life Sciences enterprise that leads the way.

6 About Perseid Software

Perseid Software is engaged in providing strategic consulting and information technology design services to healthcare and life sciences enterprises. For more than 30 years, the principals of Perseid Software have been engaged in the development of mission-critical information systems and in the analysis of healthcare, disability and pharmaceutical data.

Perseid Software is not merely a strategic consulting firm. It is an engineering management and design firm focusing on database design and implementation of very large and complex life sciences and healthcare information systems. Perseid's clients include or have included some of the largest and most progressive computer, healthcare and manufacturing companies in the world.

Contact:

Bernard P. Wess, Jr., President
Perseid Software Limited
Needham, MA
Direct Dial:(781) 453-2351
bwess@perseidsoftware.com
www.perseidsoftware.com