PERSEID
SOFTWARE LTD

The Strategic Value of Web Services for Healthcare & the Life Sciences

Perseid Software Limited
Needham, Massachusetts

March 1, 2003

Table of Contents

1 EXECUTIVE SUMMARY 3

2 INFORMATION COMPLEXITY IN LIFE SCIENCES ENTERPRISES 4

2.1 Introduction—Supporting the Value Chain in the Life Sciences 4

2.2 Information Complexity in the Life Sciences 5

2.3 An Integrated View of Information in the Life Sciences 6

2.4 Traditional Computer and Data Architectures Lack Integration 8

2.5 A Model for Web Services Based Life Sciences Systems Integration 10

3 CONCLUSION 15

4 ABOUT PERSEID SOFTWARE 16

Figures

Figure 1 Life Sciences Enterprises Generate Tiers of Complex Data 6

Figure 2 A Traditional Architecture Lacking Integration 8

Figure 3 A Database Driven Architecture 10

Figure 4 A Model Architecture for Integration of Applications in the Life Sciences Enterprise 12

Figure 5 A Model Architecture for the Enterprise and Local Networks 14

1 Executive Summary

Complexity breeds complexity. As healthcare and the life sciences industry matures and additional automated processes are introduced throughout the entire value chain within healthcare and the pharmaceutical industry, the demand for rapid business process execution, more accurate and timely information, and additional automated information systems increases exponentially.

The pressure to automate the entire value chain within life sciences and healthcare R&D, production, sales, marketing and treatment processes; to make more timely and improve the clinical trials and patient safety processes; to reach and educate consumers, providers and patients and to do so profitably and reliably has resulted in the development of a complex mix of information technology and systems that can extend literally from a patient at home or researcher in a lab to a global manufacturing or sales and marketing application encompassing numerous sites and countries. Because these solutions and their associated IT systems are driven by the same requirements—discovering, validating, producing and selling a pharmaceutical product—they must be more fully integrated and more easily accessible to authorized users, internal and external organizations.

But the rapid growth of pharmaceutical, life sciences, healthcare and regulatory systems has resulted in process and information systems anarchy. Although new telecommunications and data standards driven mostly by the Internet are slowly emerging; in general, each organization within the pharmaceutical value chain has been left to determine for itself how and when to integrate its information systems platforms—computers, operating and storage systems, applications, data and business processes.

Moreover, the complexity of the regulatory and partnership contract "chains" within the life sciences discovery and product development processes have resulted in extraordinary complexity in the security, access and control requirements for participating organizations. Johnson and Johnson, for example, has 200 operating divisions, controlling more than 120,000 users. Wellpoint, a major U.S. healthcare enterprise, has more than nine operating companies. This complexity creates two primary problems: (1) getting and registering the existence of required business and software processes and data in (business) "real-time" and (2) allowing only those who need access to software processes and data to do so selectively and under knowable and auditable circumstances. We call this problem the "Regulated Data Exchange" problem.

The solutions to these problems are (1) integration of computers, operating systems, business and software processes, data, databases and registration information into Internet information systems using "web services" protocols to integrate and communicate (2) creation of global enterprise systems and data integration architectures for registering and sharing enterprise processes and data that will be used within the enterprise and/or leave the contractual boundaries of the enterprise and (3) implementing security, access and control business processes and web services that facilitate what is known as "selective transparency." Selective transparency allows data to become, temporarily or permanently, "visible" and accessible to a known and authorized party for a known purpose for an audited period of time.

This white paper provides a foundation for understanding the anticipated evolution of computer systems architectures to facilitate systems integration processes within the life sciences value chain. It will provide a model for life sciences data and business processes that can facilitate the development of automated business process management (BPM) within the pharmaceutical industry as a whole.

For the business executive, this white paper provides a means of viewing the value chain in the industry as a series of integrated business and Internet software processes—each supporting a critical business need. If you don't know the data is there, if you can't reliably access it or you can't secure it—for all

intents and purposes, it doesn't exist. Given the extraordinary costs of pharmaceutical R&D and distribution having access to critical data in real-time is not an option—whoever has the best information in discovery, development, production and marketing *wins!*

2 Information Complexity in Life Sciences Enterprises

2.1 Introduction—Supporting the Value Chain in the Life Sciences

The IT infrastructure supporting a life sciences product throughout its life-cycle, particularly "in-real-life" treatment is very complex. Developing a new chemical entity into a commercial pharmaceutical and testing it within the market requires complex partnerships and is based on a regulatory process that can last more than a decade. The information systems supporting these processes must support research, development, clinical trials, manufacturing, marketing, sales, the regulatory process and post-market data capture.

For example, examining the life sciences discovery process within the domain of central nervous system (CNS) product development, there has been an explosion of new devices and databases. In a recent article in *Bio-IT World*,[1] Dr. Stephen Wong of the University of California, asserts that new clinical procedures such as lab tests and neuropsychological exams, new structural imaging techniques such as magnetic resonance imaging (MRI) and angiography, x-ray computed tomography and electron microscopy, functional and metabolic imaging methods such as positron emission tomography, magnetic resonance spectroscopy, functional MRI and optical imaging, while essential, have complicated the systems and data integration problems in CNS discovery. These examination and imaging techniques are also being combined with high throughput genomic techniques such as DNA micro arrays.

Other requirements beyond the discovery process within R&D include documentation and data analysis for clinical research and trials and, most importantly—sharing of data among contractual partners in all aspects of product development, launch and distribution.

As a result, the IT infrastructure shared among partners during development and product introduction can be exceptionally complex and it will be geographically and functionally dispersed among many parties. Computing platforms can literally range from Apple Computers to IBM supercomputers, resulting in numerous disparate storage and computing systems. Hundreds of people at dozens of locations would be expected to simultaneously generate and consume massive amounts of critical business process and research data.

Finally, the data and image storage management and administration requirements in the life sciences have become daunting. Multiple storage types—files, records, databases, images, documents, etc. must be supported. Multi-vendor operating systems must be supported. Multiple computing platforms must be integrated and managed. Disparate storage systems, including network attached (NAS), storage area network (SAN), and enterprise must be integrated. Data systems and structures from simple files to complex relational databases must be supported within these storage architectures, non-stop and in real-time.

This white paper will examine the complexity associated with multi-vendor and -development partner support in healthcare and the life sciences and propose how multi-vendor software and hardware can be

[1] Wong, "Neuro-IT Needs Integrated Infrastructure," Bio-IT World, July 11, 2002. http://www.bio-itworld.com/archive/071102/horizons_neuro.html

integrated to manage a complex IT architecture for discovery, development, production and in-real-life use. The focus shall be on developing a model architecture using "web services" and Business Process Management (BPM) that support local information systems and storage and an enterprise-scale IT architectures that will operate the integrated complex of organizations within the value chain of the life sciences. An emphasis will be placed on the value of BPM and how web services can dramatically improve the management processes and reduce the complexity of systems integration in healthcare and the life sciences.

2.2 Information Complexity in the Life Sciences

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

- Biological — Proteomic, genomic and other biological and chemical information
- Financial — Design, development, pricing, deployment and marketing data
- Payer Data—Claims, membership, medical management, Protected Health Information (PHI)
- Research — Molecular, genetic, proteomic and pharmacological data
- Clinical — Clinical trial, side effect, outcomes and "in-real-life" effects of drugs
- Validation — Strategic planning data on the validation of the drug for the target market
- Market — Marketing information by country and metropolitan area
- Manufacturing — Integrated GMP[2] information on the manufacturing process
- Administrative — Healthcare claims, membership, diagnostic and treatment data
- Metadata — Information about what is contained in the these phases of development

Figure 1 describes the data model for the integration of information across the value chain of the life sciences enterprise that is multi-vendor and multi-source, but one that encourages the definition and sharing of data—within the enterprise and among partners and regulators. Two "stacks" of data are articulated in the figure:

- **Biological and Clinical** — The data and information supporting discovery and R&D of new chemical entities.
- **Administrative and Financial** — The data and information supporting the processes of approval, production, distribution and post-market surveillance.

[2] Bernard P. Wess, Jr., "Building Mission Critical Document Management Solutions for Global Pharmaceutical Companies, EMC White Paper, Perseid Software, June, 2001.
http://www.emc.com/vertical/pdfs/life_sciences/interstitial_3.jsp

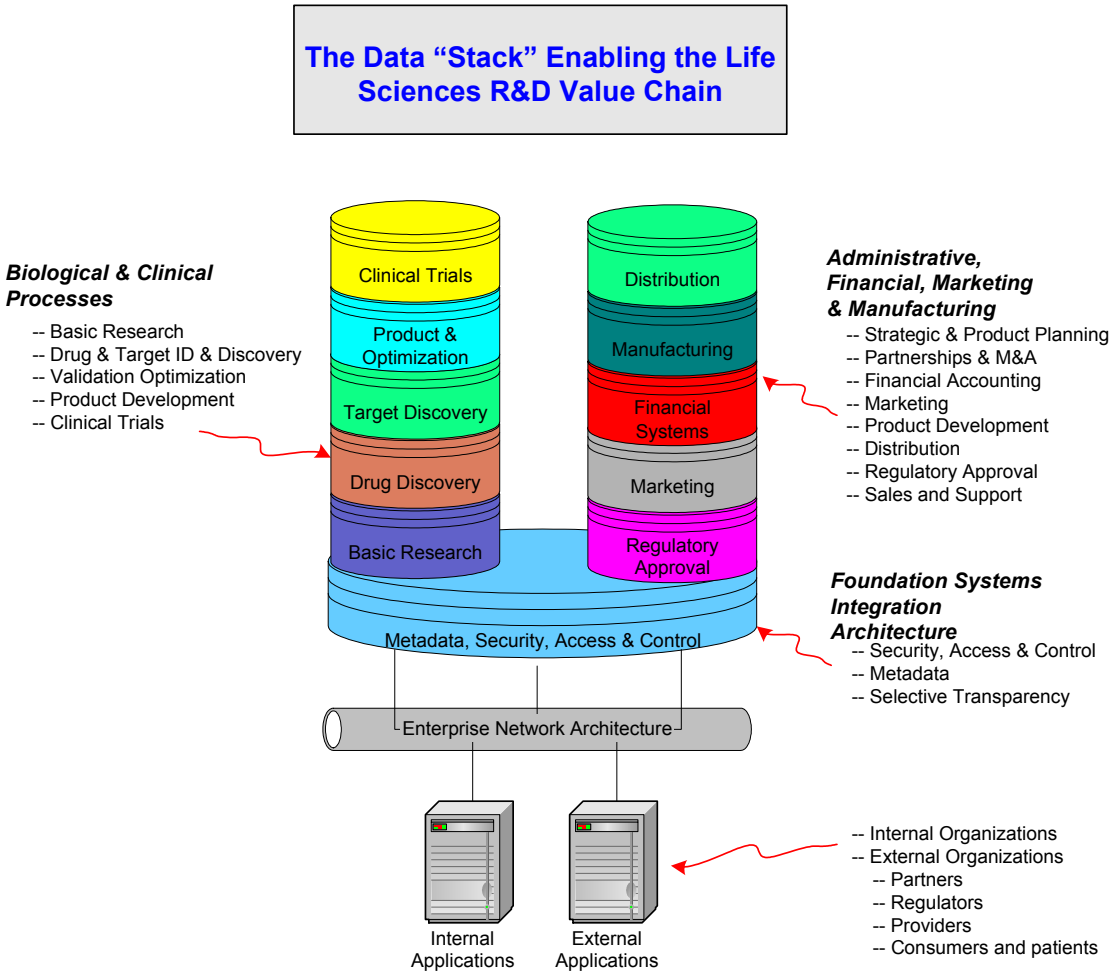

Figure 1 Life Sciences Enterprises Generate Tiers of Complex Data

2.3 An Integrated View of Information in the Life Sciences

These segments of the value chain in the life sciences *require* integration. Speed-to-market and efficacy of new pharmaceuticals is about having the right information at the right time for the right person in development and production processes. Moreover, many partners must share, *selectively*, data. Discovery partners need access to internal systems of the pharmaceutical company. Clinical trials partners, payers and treating clinicians may need to use current and past clinical data. Regulators may need access to a variety of clinical and market data in a confidential manner for an extended period of time and in multiple countries.

The need for "selective transparency"[3] of information systems first proposed by W. Roy Dunbar, CIO of Eli Lilly drives the need for integration of systems and data. Selective transparency allows an employee,

[3] Mark D. Euhling, "The Pharma Prophets", Bio-IT World, April 7, 2002. http://www.bio-itworld.com/archive/040702/boston_it_pharma.html

research partner, regulator, or any authorized entity or person access to aggregate data or the ability to execute a transaction, program or report—based on role and security, access and control criteria.

But the key to selective transparency is having:

1. An integrated information systems architecture that allows for and encourages real-time access to real-time information and raw data.
2. Secure, timely and selected access to integrated data that can be controlled by business processes
3. The ability to access data and processes anywhere, anytime
4. The ability to introduce software to control business process management and improve workflow and to reduce the costs associated with business process execution
5. Simple and effective enterprise application interfaces to enable application integration

Integrated computer architectures facilitate business and operational integration by encouraging employees, partners and others to quickly register and share data. When in place, these architectures facilitate sharing data with regulators and partners who are outside the literal boundaries of the life sciences enterprise. Given that life sciences discovery and commercialization is a national or even a global exercise, the need for registering data and sharing it effectively is profound—both in terms of costs and managerial value.

The dual "stacks" of enterprise information in delineate the data requirements within the critical management functions within the life sciences value chain. These functions include the chemical, biological, clinical and validation processes associated with developing a pharmaceutical solution and the administrative, financial, marketing, manufacturing and distribution needs once a product requires preparation for the market. Note that the foundation architecture for selective transparency is implemented as a core systems integration function. Data is registered by form and type (as metadata) and then security, access and control constraints are applied to the data to register its presence in the enterprise and to provide selective access to interested and approved parties. A key aspect of effective registration and distribution is to register a data asset only *once* as a unique resource. This is a primary role of the Enterprise (ESA) and Network (NSA) Storage Architectures. These architectures avoid the proliferation of multiple versions of records, files and databases which is a classic means of destroying effective data integration in an enterprise. On a global scale, these systems are supplied by, for example, IBM and EMC.

The metadata of the ESN and NSA identifies who has access to data and the circumstance and means by which the data may be accessed. Thus, the systems integration process becomes very much a process of data integration—driven by appropriate security, access and control needs and software.

The foundation of the data integration architecture is one based on security, access, control and metadata resources that facilitate the overall enterprise systems integration processes. New systems and data are registered in the enterprise and data access is provided through controls that allow departments, partners and others selective access to data depending on their validated needs and rights. Data is secured, integrated and available for re-use at all times. Thus data integration encourages Internet-based systems and applications integration. The complexity arises from the fact that the data is regulated by the FDA and the Center for Medicaid and Medicare under HIPAA. This requires that a full audit trail of access and use be maintained.

2.4 Traditional Computer and Data Architectures Lack Integration

Figure 2 highlights the "traditional" systems architecture of life sciences information systems. The computer and storage systems are not integrated and each tends towards duplication of data and systems resources. Each information system utilizes its own processing resources and storage architecture. The duplication of the resources is expensive and encourages a lack of data integration in enterprise and poor shared security, access and control processes.

In the traditional enterprise systems model, business process management is performed by individuals, not software. Business processes are managed using project plans and email, for example.

Communications among applications is accomplished through database extracts and transfers and "hard-coded" custom protocols among applications programming interfaces (APIs). On average this model involves expenses for simple application or module integration of between $60,000 and $1,000,000 *per* interface.[4] [5] The resulting enterprise application and systems integration methods are complex, unique to the enterprise and very expensive to manage, validate for the FDA and improve.

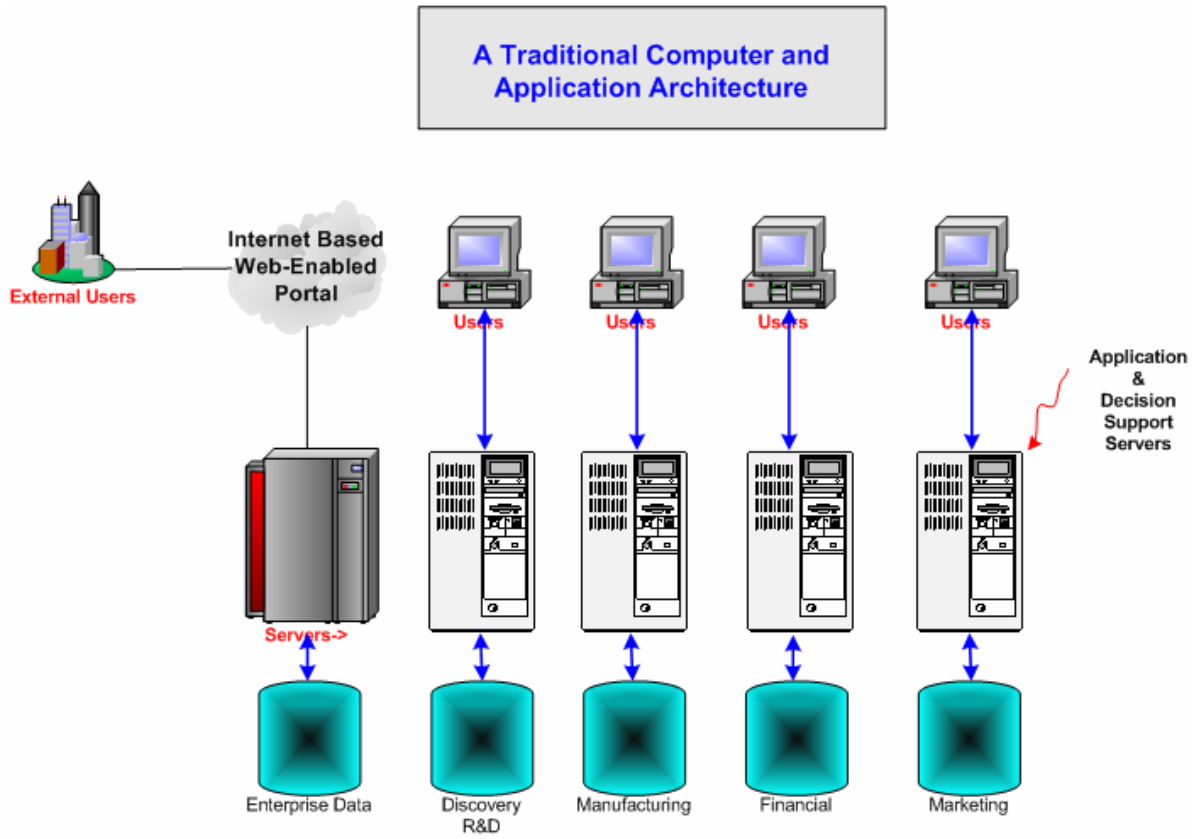

Figure 2 A Traditional Architecture Lacking Integration

[4] John Hagel III, "Edging into Web Services", *McKinsey Quarterly*, 2002, Number 4, Technology http://www.mckinseyquarterly.com/article_abstract.asp?ar=1245&L2=21&L3=37&srid=21&gp=1

[5] These expenses are larger than those in the article to account for validation and verification expenses in the life sciences.

Figure 3 presents an evolutionary view of systems and data integration, focusing on using central databases of management, financial, marketing and clinical data and documents. The data and documents are stored to support an integrated view of the pharmaceutical enterprise and its product development processes within a database. 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 each tracked at multiple levels of the enterprise and with external partners using file transfer of database extracts or database-database updates.

Because of the massive size of the central repositories, the databases in the figure may be implemented as multiple physical databases. This data architecture makes it easier to reliably merge clinical, financial, marketing and patient/consumer data into accurate reporting systems.

What is needed to support this form of real-time enterprise computing[6] is the integration of the data into repositories, using shared "metadata."[7] This improves data integration, removes redundancy and encourages regulatory and partnership data sharing. However, because the business processes are still uncoupled from the software processes, these complex database-driven applications are still managed by people using traditional project management systems and documentation.

Also in this model, the application interfaces are still either entirely supplied by the software vendor, for example, SAP, Siebel, Oracle, etc., or they are typically modified for the enterprise using custom-coded APIs—again, at great time and expense.

[6] Bernard P. Wess, Jr., "Enabling the Real-Time Life Sciences Enterprise with an IT Infrastructure," EMC White Paper, Perseid Software, February 2002.
http://www.emc.com/vertical/pdfs/life_sciences/interstitial_data_warehouse.jsp

[7] "Metadata" is information in the data systems that identifies the contents of the data itself—its fields, tables, reliability and validity, for example.

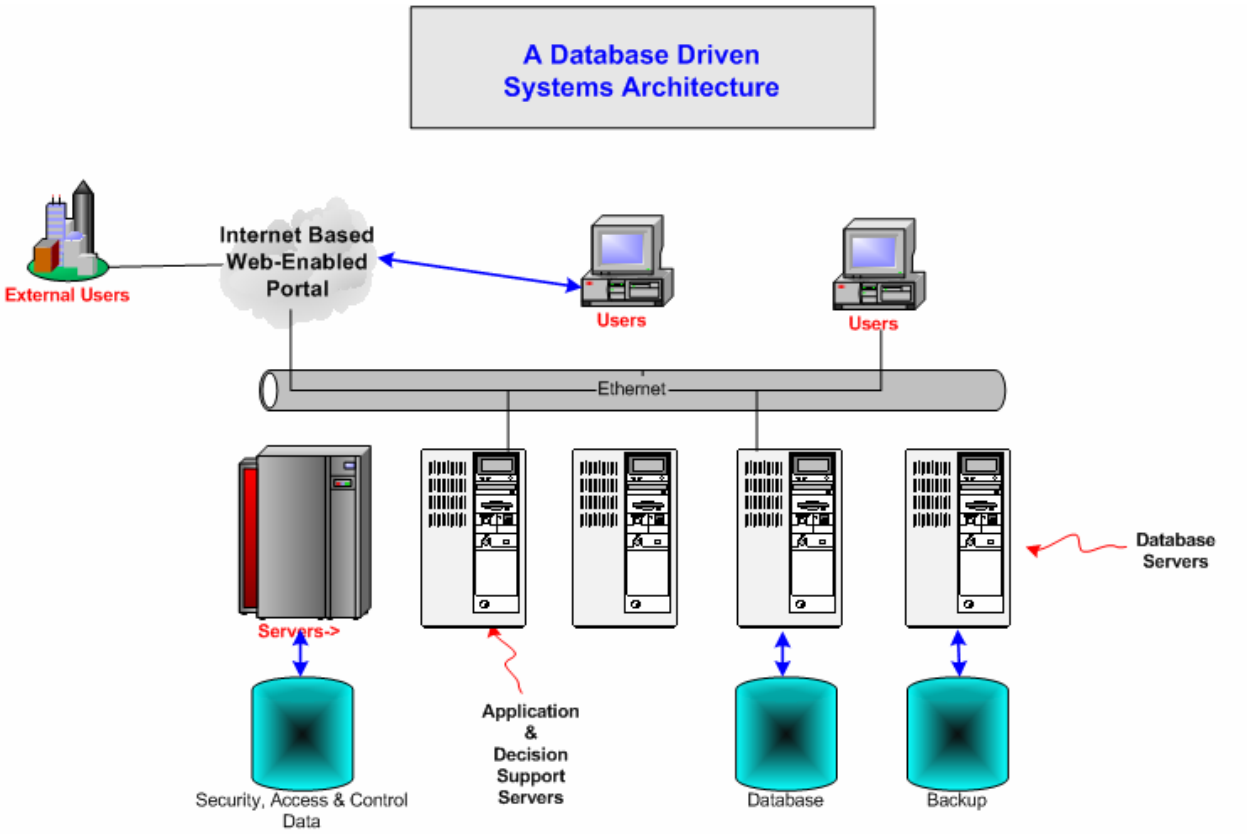

Figure 3 A Database Driven Architecture

2.5 A Model for Web Services Based Life Sciences Systems Integration

Internet (web-) based application and systems integration is based on the ability of Internet-enabled applications to exchange in (business) real-time messages through a remote procedure call (RPC) using web software services that register applications, software services, security, access and control information as enterprise data in various directories available in a secure manner within the enterprise. An RPC is a function or "subroutine call" between two cooperating and communicating software processes, wherein the call takes place over the Internet, not necessarily within the same computer.[8]

In a recent article in *Bio-IT World,* Salvatore Salamone provides the technology foundation for understanding the internals of web services.[9] We are going to focus on how to strategically apply web services to dramatically reduce costs and complexity of applications.

The evolution of the poorly integrated systems, data and applications in Figure 2 to a real-time web services-based architecture is shown in Figures 4 and 5. All users pass through a security, access and control (SAC) application, itself implemented as a web service application that provides for common verification, validation and access to production applications. This process gives users and software processes access to the Internet to send XML messages among applications. Since XML is a simple text

[8] RPC were initially designed for use within the same computer or local computer network.
[9] Salvatore Salamone, "Divide and Distribute*", Bio-IT World*, November 12, 2002
http://www.bio-itworld.com/archive/111202/divide.html

based “markup” language using standard Internet TCP/IP transport protocols, there is no “hard-wired” API to deal with among applications. Nor is there a hard-wired telecommunications network to maintain.

Essentially, the following protocol is used by every application and every service:

1. A software service, for example, access to a particular database, is registered as a web service in a directory on the Internet or within an intranet.
2. All users software processes (services) are registered with security, access and control rights.
3. Application interfaces are coded as messages sent over RPC TCP/IP protocols sending and receiving XML messages. An application need only know the XML message and how to respond.

The network becomes decoupled from hard-wired database and API applications. Protocols can be coded for a fraction of the cost of coding APIs. Geographical and functional dispersion is inherently enabled since all messages utilize the Internet, thus providing ubiquitous and global instant access to each web service. Figure 4 describes the XML based web services architecture and the intended end state is a global enterprise architecture depicted in Figure 5.

For the organization, several financial and functional benefits are realized:

- The internal and external divisions, partnerships and regulatory agency relationships can be realistically automated for the first time, since access is defined by a web service.
- Systems integration costs are dramatically reduced and interfaces are standardized, by as much as an order of magnitude.
- Data integration is facilitated as database proliferation ceases.
- Most importantly, the use of business process management software is enabled so that business processes, software and business process latency and workflow associated with clinical, administrative and financial software and human-based systems can be monitored.
- R&D capital can be re-deployed to return maximum value to the enterprise and not consumed by IT systems integration costs.
- A community of payers, providers, suppliers and researchers is effectively and economically enabled.
- Security is pervasive and adjustable by operating division or region

Johnson and Johnson is developing a security, access and control system to enable web-based access to more than 200 operating divisions covering more than 120,000 application end-users. This “regulated data exchange” enterprise application is enabled using PKI and a web services based LDAP directory to enable secure, authenticated access to regulated applications and web services. By building the RDE application as a web service, J&J is able, with its systems integration partner, Northrop Grumman Mission Systems (formerly TRW Healthcare) to rapidly create XML “wrappers” around enterprise applications like Oracle and SAP and internal custom applications. Web services are dramatically reducing systems integration expense and improving systems reliability and maintainability for J&J.

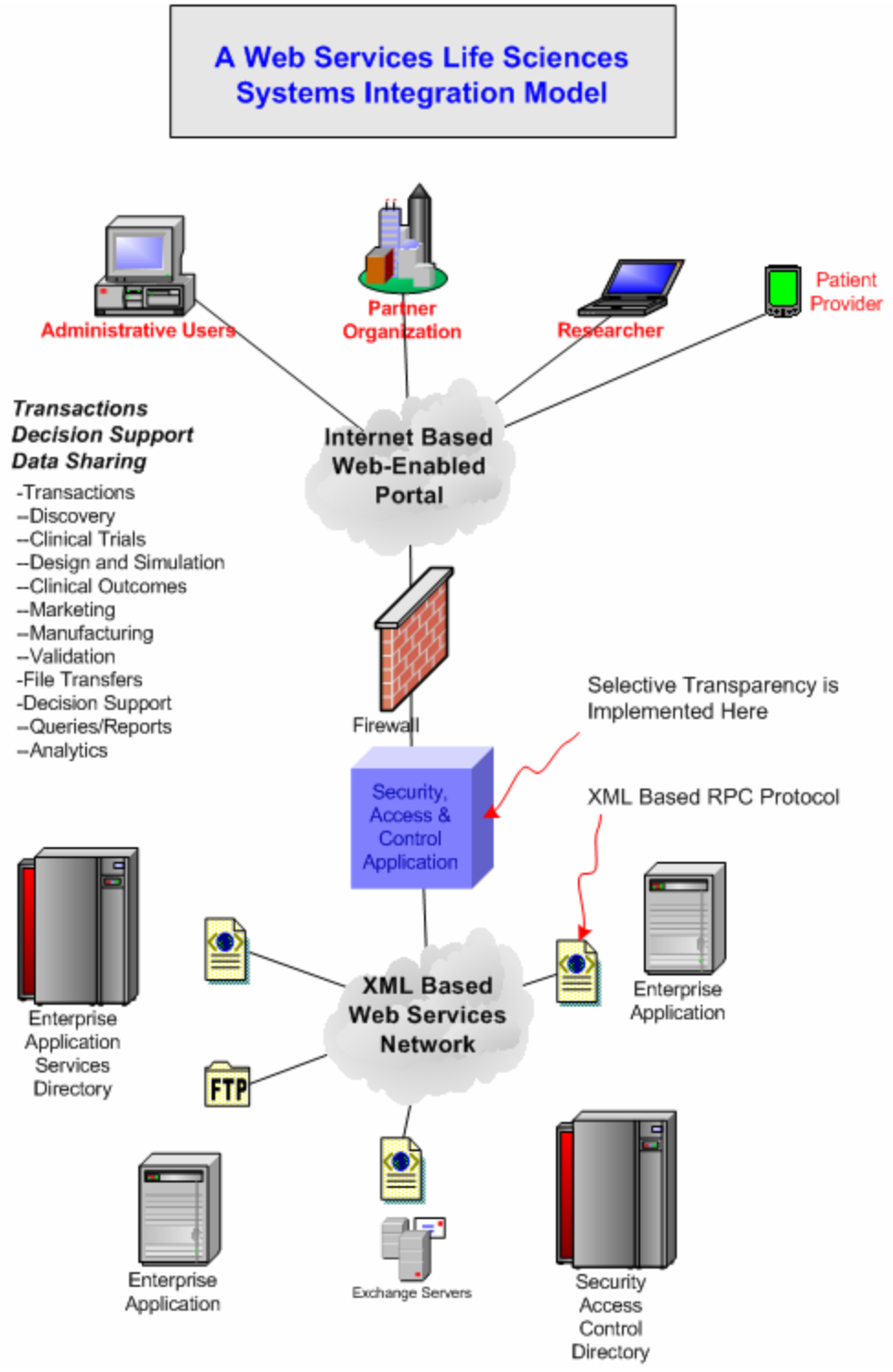

Figure 4 A Model Architecture for Integration of Applications in the Life Sciences Enterprise

The final state of the global enterprise network is shown in Figure 5. Applications, files, persistent data objects (images, annotations) and databases are divided among production transaction processing and business intelligence applications and most importantly, enterprise departments, external agents and partners.

New data or requests for data arrive from transaction processing systems through web service requests—from partners, research organizations or third-party data vendors. The systems integration architecture is responsible for ensuring that local applications and databases can transfer data to the central enterprise storage network (ESN) for registration and integration. Using enterprise-level security, access and control solutions, regulated data is integrated in the ESN or local NSA for use by others. This ensures that the data is registered, secured and validated so that it can be accessed by all authorized local and global users, either locally in the NSA or globally through the ESN. It enables effective regulated data exchange in a secure manner with dramatic reduction in systems integration costs.

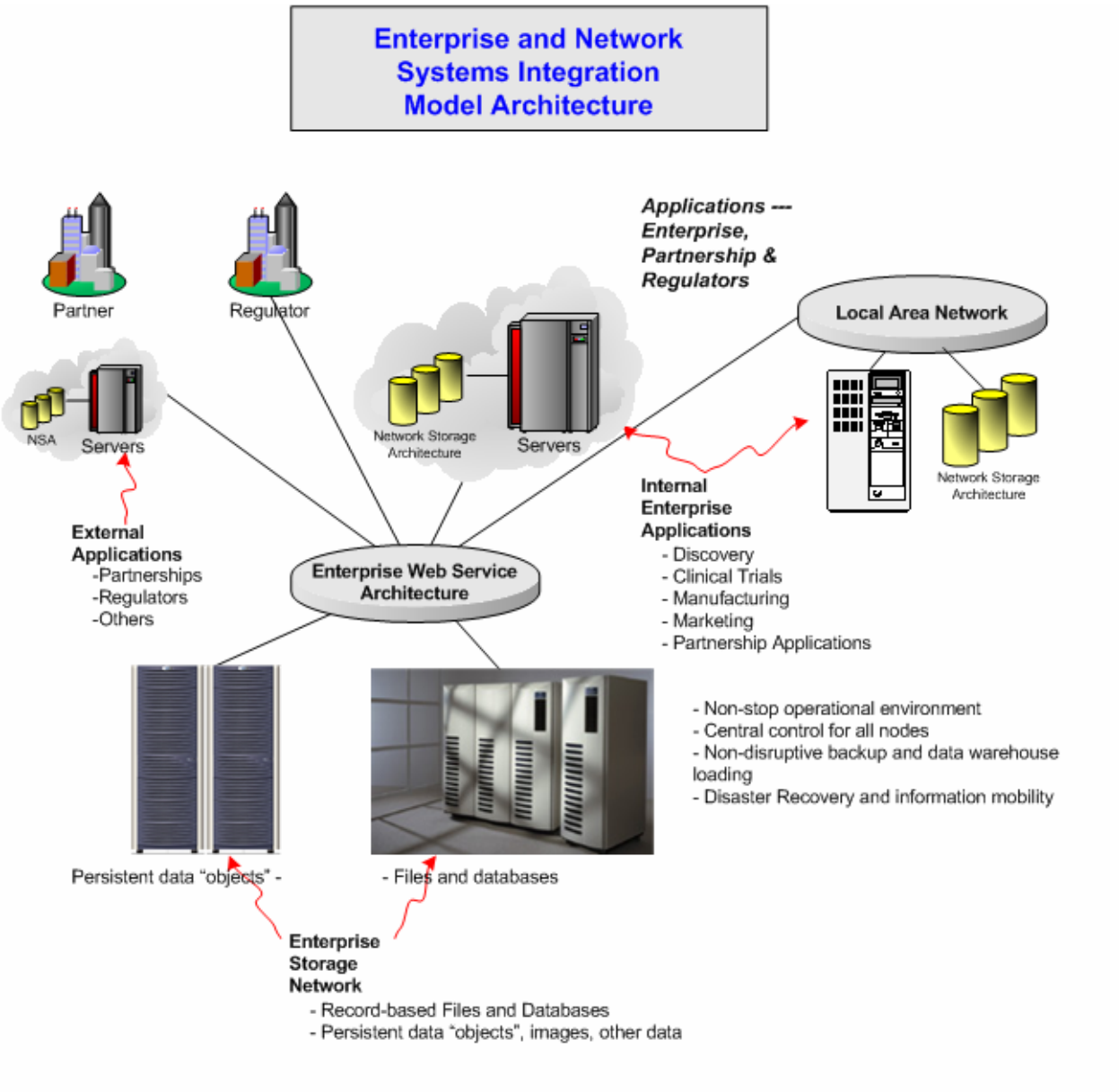

Figure 5 A Model Architecture for the Enterprise and Local Networks

3 Conclusion

The rapid growth in the development of solutions and automated systems in the value chain within healthcare and the pharmaceutical industry has resulted in a lack of systems and data integration within the extended enterprise that includes the modern pharmaceutical or life sciences enterprise and providers and payers. Not only is there a lack a standardization of data gathering and reporting, but also there has been a tendency to accumulate disparate applications, computers, and storage without first developing a systems and computer architecture that encourages regulated data exchange and systems integration.

As a result, there is difficulty in moving data through the life sciences R&D and production value chain and in monitoring business processes, some of which, like clinical trials, that take years. Also, there is a need for security, access and control systems to be implemented across the entire regulatory, contractual and organizational "chain of command" to facilitate sharing the right data with the right party under the right circumstances and at the right time. This is now mandated by the FDA, EMEA in the EU and HIPAA.

Implementing systems to selectively allow access on a full- or part-time basis to integrated data is the solution to many of the problems plaguing systems integration in the life sciences. Using web services to execute, monitor and manage application, operating system and storage technology, we have developed a model systems integration solution that integrates business processes, disparate data, files, databases and operating system platforms into an Enterprise Application Architecture that is secure and reliable.

The model web services-based architecture, when combined with data standards that are emerging in the life sciences value chain such as the eCTD[10], becomes a standard for creating, storing and effectively sharing local and enterprise-scale applications and data within the life sciences industry.

Well formed and easily accessible web service protocols for data and processes access will improve the discovery, development, marketing, production and sales processes in the pharmaceutical and biotechnology industry by dramatically reducing systems integration, documentation, validation and verification costs of software and business processes.

The alternative is business process chaos, and given the costs of modern life sciences discovery and development—and the chaos will become, if it is not already, unacceptable to senior management, regulators and shareholders.

For more than 1,000 years the construction industry has used *architecture* and *standardized components* to facilitate construction. Given the clinical and economic risks associated with life sciences discovery, manufacturing and distribution, no less should be done among life sciences enterprises, partners, regulatory agencies and healthcare care providers.

The tools exist, but the will to use them must also be there.

[10] The electronic common technical dossier mandated by EMEA in the EU for NDA filing as of 2003.

4 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