

Clinical Development Perspective Report

Clinical and Regulatory Informatics

**Managing the technical infrastructure for
clinical drug development**

Alan Hopkins, Ph.D.
August, 2003



Creation of this document was partially supported by Acumen Sciences LLC while the author was on staff at that organization. Thanks to former Acumen Sciences colleagues Barry Sherman, MD and Sandor Schoichet who provided valuable input in early drafts. Eric Greenberg, CEO of Acumen Sciences, allowed distribution of this report.

The following people graciously served as external reviewers:

Wayne Kubick, Vice President
Lincoln Technologies, Inc.
Wellesley Hills, MA 02481

Jeff Schumack, Associate Director for Information Technology
Human Genome Sciences, Inc.
Rockville, MD 20850

Ira Stoler, Vice President and Chief Information Officer
Tularik, Inc.
South San Francisco, CA 94070

Contents

Acknowledgements	2
Executive Summary	4
Themes	5
Perspective	6
The Development Process Model	7
Clinical Systems Architecture for Drug Development	9
Clinical Information Technology Governance	21
Summary of Action Recommendations	24
Conclusion	25
References	26

List of Figures

Figure 1. PharmaStat Drug Development Process Model	7
Figure 2. Data and Metrics	8
Figure 3. High Level Technical Architecture	12
Figure 4. Conceptual Components of a statistical computing environment	15
Figure 5. Productivity depends on alignment of business process, organizational readiness and technology	22
Figure 6. Governance of Technology Projects	23

Executive Summary

PharmaStat believes that effective implementation of the drug development process requires coordination of science, information technology and clinical operations to acquire the target scientific data and useful process-related metrics. This document describes the PharmaStat perspective on the information technology aspects of drug development, which we call Clinical and Regulatory Informatics (CRI).

The scope of CRI comprises the processes of data collection, aggregation, integration, analysis, review, and regulatory publishing of clinical trial data and results. These activities are critical to business success. PharmaStat believes that companies invest large amounts of capital in technologies but do not receive a commensurate return on investment. This document serves as a perspective for companies just entering clinical development for the first time and as training material for those new to pharmaceutical operations.

In this report, we present a high-level process model for clinical development. In support of these processes, an information technology architecture is proposed. The elements of each process and the corresponding technical architectural element are discussed. Suggestions are provided for corporate governance of key CRI investments and their implementation to achieve maximum return on investment.

The most important product from the clinical drug development process is the resulting body of data which support licensure of a new pharmaceutical in the format of an electronic common technical document (eCTD). In the course of acquiring the data and preparing them for publishing in a regulatory submission, a variety of computer applications systems, usually disjointed, are used to manage several types of data. System integrators make a living helping pharmaceutical and biotechnology companies link these systems to share data and to provide useful information and metrics on internal processes. To satisfy regulatory requirements, the whole clinical study process needs to be transparent, so that there is a clear validated and auditable flow of information from its source in the clinic to its destination in regulatory documents.

Other major areas addressed include:

1. **Data access and integration.** One key objective for CRI systems is integration of the various applications used to support the clinical development process. This results in easy access to study information via web services. This information can be stored in a data warehouse where key management information from all the clinical applications can be integrated and regulatory submissions can be prepared. Data standards are an important part of this discussion.
2. **Regulatory compliance.** All systems used for clinical development need to be validated. Hardware, software, and operating procedures must comply with Title 21 CFR Part 11 for electronic records and signatures. Clarification and implementation of these requirements is still in process. For example, statistical processing of clinical trial information into final reports is an area where practice has largely been ignorant of the electronic records and signature rule.
3. **IT Systems Expense.** The cost of acquisition, installation, and maintenance of all these systems represents a significant expenditure. PharmaStat estimates that the cost of implementing clinical development systems according to the technical architecture is \$25M to \$30M, and continues to grow. This represents a high bar for smaller companies to acquire and maintain the infrastructure and provides a definite niche for contract research organizations (CROs) and other third parties (e.g. SMOs, ASPs) to provide these services.

An effective well-running clinical data platform is a competitive advantage because it can improve data quality and speed of execution.

Themes

- The data are the most important product in the clinical sections of the electronic common technical document (eCTD). It's the content, integrity, structure and documentation of the data that matter.
- Data and technology silos have evolved in clinical research, largely along disciplinary lines, making reporting and operational decision-making difficult. No single integrated solution is available.
- Systems should be aligned to provide transparency of the process of collecting and packaging the data.
- A major responsibility when outsourcing clinical processes is the active management and validation of the data transfers with the sponsor.
- Integrity of electronic data bases are a key FDA concern. A prospective data quality plan is recommended. Creating submission components electronically through the entire development process can build quality into the e-CTD.
- Technology integration cannot truly take place in the absence of data standards.
- Appropriate on-going validation and compliance with 21 CFR Part 11 are major and costly hurdles to successful implementation and maintenance of clinical trials software.
- Statistical processing of clinical trial information into final reports is an area where practice has largely been ignorant of the electronic records and signature rule.
- The pharmaceutical marketplace is so small there is not a lot of upside for good competition and comprehensive solutions. Almost no application software vendors have managed to sustain adequate growth and profitability in this narrow and conservative market.
- Depending on organizational size, PharmaStat estimates that it costs \$5MM - \$30MM to put all the systems in place necessary for effective clinical trial execution and reporting. This will drive increased use of ASP and CRO models for some companies.
- Pharmas are typically very poor at clinical information technology governance, and lack a principle-based approach to CRI systems architecture. New models are needed to effectively manage IT evolution and outsourcing against one-off long-term projects.
- Productivity depends on alignment of work processes, technology, and organizational structure. Balance these variables through a strategic business plan – a data vision and complementary technology plan. Actively review these plans to effectively manage your clinical and regulatory IT investments.
- Since it's all about the data, then systems integration should be the primary responsibility of the clinical IT group.
- The clinical software ensemble is increasingly complicated and expensive while the regulatory requirements for validation and system management are growing.
- An effective well-running clinical data platform is a competitive advantage because it can improve data quality and speed of execution.

Perspective

Dreaded phrases in the biopharmaceutical industry:

- Refusal to file
- Clinical hold
- FDA for-cause audit
- Second FDA review cycle
- Major deficiency letter
- NDA non-approvable letter
- Stock price dive

Any of these results could be due to a failure to provide adequate data to the FDA. This report discusses how information technologies can be optimally utilized to provide the highest quality data in regulatory submissions.

The clinical development process is long, complicated and expensive. As long as testing in human subjects is required, there is an irreducible minimum length of work (at least six months) for almost any clinical study. One place where there is an opportunity to gain efficiency is through the effective use of informatics, but industry has yet to fully realize this opportunity.

Information technology capabilities are a critical success factor in clinical drug development. Lots of clinical data are collected during a series of clinical studies. Technology provides a foundation for potential breakthroughs in operating processes. Multiple software applications are required to support the clinical trials process.

The goals of a clinical information technology strategy should be to develop a sufficiently integrated systems architecture that allows:

- Data to flow smoothly through the process; and
- Transparent, auditable pathway from the investigators to the regulatory submission, despite ongoing evolution of the underlying technology platform;
- Metrics on operational processes readily available from computerized systems. Metrics provide the capability for benchmarking of processes and cycle times, improved planning and decision making.

Companies in the pharmaceutical and biotechnology industries have often viewed their information systems as proprietary. Historically many companies built their own applications because commercial software was not available. As vendors have developed commercial products for various functions, piecemeal technology acquisition has led to a series of data silos corresponding to different functions of the drug development process. To maximize the value of the information collected, expensive system integration efforts seem never-ending. The resulting systems are “brittle” and can break if even small changes are made.

PharmaStat recognizes that effective information systems deployment is only one component of successful drug development. Obviously, the science associated with the drug and the clinical development plan are key to ultimate success. Given the plans, implementation of business processes of study execution is equally important to the science and technology. This report is based on PharmaStat experience and represents our opinions based on that experience.

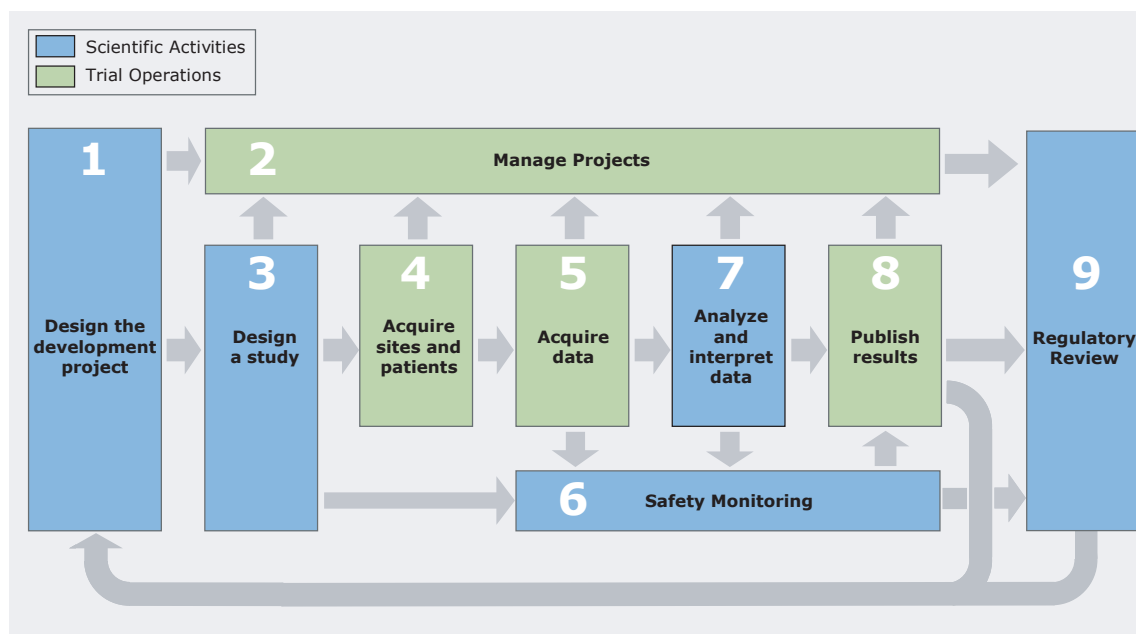
The Development Process Model

Since the most important role of CRI systems is to enable and track a productive clinical development process, a high-level clinical development process model (see Figure 1) is the right place to start. Very simply, the goal of the process is to develop an electronic regulatory submission that supports the vision for the product in the clinical development plan. The key to this effort is the data. Each aspect of the process should be optimized to collect the appropriate data to support the claims of the proposed package insert. Conduct each trial with appropriate endpoints, acquire the data and document its status and integrity - its ability to support study conclusions. After analysis, the database should be “published” for submission in an eCTD including appropriate documentation. In what follows, we assume that this process model is generally applicable.

The PharmaStat clinical development process model is shown in Figure 1. Nine high-level processes are identified. For an individual study, there is a high degree of linearity of the overall process even though there can be substantial overlaps in the duration of one process with another. Investigators (sites) cannot be contracted before the protocol is completed. Subjects cannot be enrolled without the proper site startup. Data cannot be processed before subjects are treated. The study results cannot be determined until data from the last patient become available. Only after analysis can the final study report be completed. Among these processes, there are various deliverables and handoffs among staff and each process is typically supported by one or more information systems.

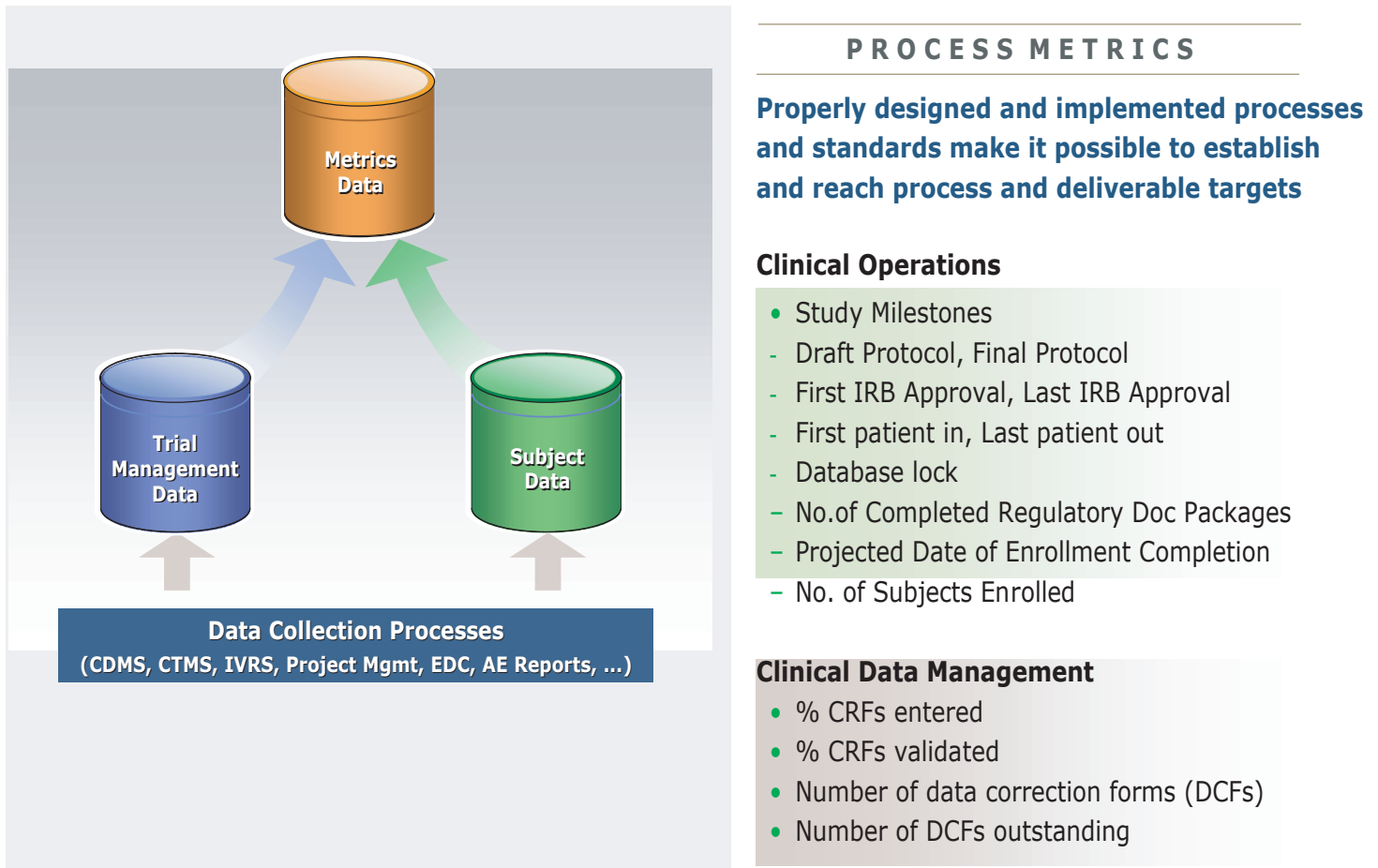
An overview of the process deliverables and supporting computer applications are summarized in Table 1.

Figure 1. PharmaStat Drug Development Process Model



Drug development generates two types of data (Figure 2): the subject data essential to the scientific analysis and FDA submission (e.g. response rates, adverse events) and the process data essential to effective trial operation (e.g. number of patients enrolled, discrepancies, and reports processed as well as history of site visits and dates of audits).

Figure 2. Data and Metrics



The data are the most important product in the clinical sections of the electronic common technical document (eCTD)^{1,2}. It's the content, integrity, structure and documentation of data that matter.

The content of the database will be dictated by the study protocol. Keep the study objectives clear and simple. Collect only the data necessary to support the study objectives which allow streamlining the data monitoring and data management work. Including scientific sub-studies or collecting too much auxiliary information on patients runs the risk of adding additional complexity at the expense of focus on data quality. Without the right study design elegance and efficiencies of technologies and operational processes are worthless.

Data integrity as used here refers to a state of the database which allows valid conclusions to be drawn about the study objectives. The data must “hang-together” at a basic level. Data integrity allows for different standards of data quality for different data components, depending on their business and scientific importance. Documentation of the quality of the data can be provided in a data integrity report. A data integrity report can be used to enumerate anomalies (e.g. missing data, discrepancy query status, visits outside the protocol-prescribed window etc.) in the database and particularly for critical analysis-related variable status. These reports can also be useful in monitoring the performance of outsourced data management activities.

Data integrity plays a key role in the FDA review process. The more confident the reviewers are in the database (i.e, the data reflect what actually happened to patients) the less time they will spend trying to validate data during the review process. We believe that regulators would like more reporting of data integrity in clinical study reports. This is requested in the ICH and FDA guidelines on format and content of study reports: “The quality assurance and quality control systems implemented to assure the quality of the data should be described.”

Clinical Systems Architecture for Drug Development

Data and technology silos have evolved in clinical research, largely along disciplinary lines, which are not integrated consequently making reporting and operational decision-making difficult. No integrated solution is available.

Systems should be aligned to provide transparency of the process of collecting and packaging the data.

Data integrity relies on process transparency, a critical architectural issue. Transparency (or data traceability) is best achieved through a simple, integrated, well-documented workflow. Informatics can help by providing better systems integration and process simplification.

It follows that a critical architectural requirement is to provide a framework for integrating multiple components from a variety of vendors and third-party services, all of which can be expected to evolve over time, independently from each other and on a much faster timescale than that of the clinical projects they support.

An idealized high-level systems architecture is depicted in Figure 3. We anticipate access to applications will be through a web interface. In addition, we'd expect access to the products of the analysis process to be available for viewing to project

Table 1. The Drug Development Process: Deliverables and Computer Applications

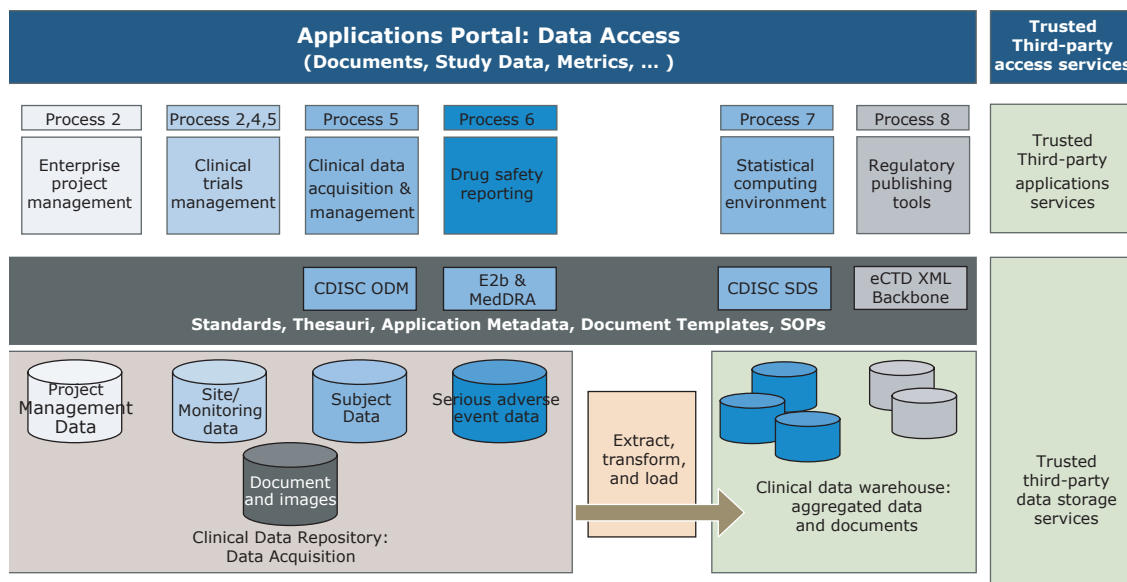
Process	Description	Deliverables	Supporting Applications
1. Design a Development Project	<p>The clinical development plan is the roadmap of clinical studies leading to submissions of marketing applications in all target geographies. The entire implementation team uses this plan to determine the number and timing of the studies, the resources required to execute the plan and the cost of the development program. The clinical development plan includes the strategy for each phase of clinical development. It specifies optimal endpoints, the necessary clinical studies, takes into account regulatory guidance in the therapeutic area, and safety monitoring concerns all oriented toward a specific label for the pharmaceutical product.</p>	<ul style="list-style-type: none"> • Clinical Assessment • Target Product Profile • Drug label • Clinical Development plan • Integrated Development Plan • Regulatory plan • Global submission plan 	<ul style="list-style-type: none"> • Collaborative planning environment • Program budgeting & scheduling • Document management
2. Manage Projects	<p>The project needs a project manager to help teams coordinate tasks and milestones. The manager may also coordinate outsourcing of services and usually closely monitors the budget. The whole process is facilitated if the team has readily available information on the study status – the sites, the data collected, etc. This is one reason for integrating the study management databases with the subject databases.</p>	<ul style="list-style-type: none"> • Project management plan • Clinical supplies management • Site management • Vendor management • Financial management • Study metrics definition 	<ul style="list-style-type: none"> • Project & financial management • Investigator & site management • Regulatory interaction tracking • Document management • IVRS • Clinical supplies management • Clinical portal
3. Design a Study	<p>Study design includes writing the protocol as well as various other plans critical to trial execution. Computer simulations are sometimes conducted based on existing pharmacologic data to help define dosing and sample sizes. Otherwise, technologies usually play a small part in the study design process.</p>	<ul style="list-style-type: none"> • Protocol • Data Monitoring plan • Statistical analysis plan • Safety monitoring plan • Create randomization • Design CRF and clinical DB • Clinical supplies plan 	<ul style="list-style-type: none"> • Clinical simulation tools • Protocol design tools • CRF design tools • Document management
4. Acquire sites and patients	<p>Active knowledge of medical groups active in the therapeutic area being studied helps recruit investigators who in turn recruit subjects. Data on the number of subjects eligible for study help to estimate the number of investigators needed and the time the study will take to complete. Negotiations about contracts and grants at each individual site takes time as do institutional review board approval and training of study personnel. Efforts to make trial information more accessible to patients use advertising, data bases, and web-based recruiting tools.</p>	<ul style="list-style-type: none"> • Investigator Qualification • Site Contracting • 1572s, etc. • Investigators Meeting • IRB process • Informed consent process • Patient enrollment • Monitor & Site training 	<ul style="list-style-type: none"> • Site & investigator DBs • Clinical trial management tools • Patient recruitment services • IVRS & randomization • Site & monitor training services • Document management
5. Acquire data	<p>Data are usually collected on paper case report forms (CRFs) which are subsequently entered into a data management system. These forms may be faxed from the site to the sponsor for automatic scanning and entry into a database. Many vendors provide web-based systems for electronic data capture at investigator sites. Data may also be acquired electronically from service providers such as contract laboratories or IVRS (interactive voice response system) for subject randomization and site management. As data are entered into the database, checks are performed according to the site monitoring process specified in the data management plan. Data clarifications are usually required from the site for data that fails the entry checks.</p>	<ul style="list-style-type: none"> • Data management plan • Database design • Data capture system implementation • Data capture process • Site monitoring process • Validation/Query resolution 	<ul style="list-style-type: none"> • Electronic data capture • Clinical data management • Document & image management • Clinical portal • Coding tools

Process	Description	Deliverables	Supporting Applications
5. Acquire data	Data are usually collected on paper case report forms (CRFs) which are subsequently entered into a data management system. These forms may be faxed from the site to the sponsor for automatic scanning and entry into a database. Many vendors provide web-based systems for electronic data capture at investigator sites. Data may also be acquired electronically from service providers such as contract laboratories or IVRS (interactive voice response system) for subject randomization and site management. As data are entered into the database, checks are performed according to the site monitoring process specified in the data management plan. Data clarifications are usually required from the site for data that fails the entry checks.	<ul style="list-style-type: none"> Data management plan Database design Data capture system implementation Data capture process Site monitoring process Validation/Query resolution 	<ul style="list-style-type: none"> Electronic data capture Clinical data management Document & image management Clinical portal Coding tools
6. Safety monitoring	Trial sponsors track serious adverse events and provide expeditious reports of all unexpected adverse events during a clinical trial and after drug approval. Reporting systems which manage electronically filing and tracking of these reports are necessary.	<ul style="list-style-type: none"> SAE Reporting Data Monitoring Committee Clinical data/SAE reconciliation 	<ul style="list-style-type: none"> Safety data management Risk analysis and monitoring Document and image management Coding tools
7. Analyze and interpret data	Before database lock and unblinding of results, a review of the blinded data gives an opportunity prior to analysis to make any modification to the disposition of subjects or data analysis plan. A data integrity report may be helpful. After database lock, the analysis data sets are prepared including the treatment code for each subject. The prespecified analyses are run according to the statistical analysis plan. The results of the final analysis will need validation prior to finalizing the final study report.	<ul style="list-style-type: none"> Statistical programming & validation Statistical analysis Submission datasets 	<ul style="list-style-type: none"> Data extraction and integration tools Statistical programming environment Team collaborative tool
8. Publish results	The eCTD requires not only final clinical study reports, but also the clinical overview and clinical summary. Equally important to the documents, well-organized and documented electronic data are crucial for regulatory review.	<ul style="list-style-type: none"> Global submission module design Template update/design Section authoring & review process Data Publishing Document publishing process 	<ul style="list-style-type: none"> Submission authoring tools Collaborative review environment Regulatory interaction tracking Regulatory publishing tools (both paper and electronic) Document and image management
9. Regulatory Review	Much of the output of the processes ends up at the FDA for review. Often the feedback received from FDA is used to fine-tune the development plan.	<ul style="list-style-type: none"> Regulatory feedback Approvable letter, etc. 	<ul style="list-style-type: none"> FDA review environment

members through a web browser. Note that the applications are mapped to the processes from Figure 1 that they support. Data is stored in either the clinical data repository (raw data and documents) or in a data warehouse (aggregated data, results, and filings) after the raw data has been transformed into analysis-ready data sets. Note that data standards play an important part in this process – whether raw data or transformed data – it is best to store according to standards. The process of archiving the electronic records stored in the database should also be addressed.

Creating the elements of the data warehouse requires a distinct “Extract, Transformation, and Load” process. The cornerstone to clinical information processing is a database where all information is stored. The database may be a relational database such as Oracle which contains raw data but perhaps also the data warehouse elements which may be SAS datasets, documents, etc. The database contains information from multiple systems: a document management system, software for clinical site management, clinical data management, and safety systems.

Figure 3. High Level Technical Architecture



The need for efficiencies in the clinical trials processes is beginning to be addressed with data standards and internet-based process automation. The internet’s capability for rapid transactions and exchange of data among participants makes it a great research communications medium. In addition to the need for efficiency, the US Food and Drug Administration will now accept new product licensing applications electronically. This requires software for publication of all the components of the submission. The eCTD is the ultimate deliverable of the development process must be kept in mind at all steps.

Individual firms may deviate from the full-fledged architecture in a number of ways. The most frequent deviation is not to have one or more of the key applications available. For example much of the clinical trials management system capabilities could be done manually or with one or more simple desktop databases designed for a specific study. The downside, of course, would be that the information in individual desktop databases would not be easily accessible and designed to work with the other information systems. Experience also demonstrates that an effective enterprise project management system is absent in many implementations. Another way to deviate from the architecture is to outsource one or more of the applications to a CRO or other service provider. For example, maybe clinical data monitoring and data management are outsourced to a CRO. This lead to the following caveat.

A major responsibility when outsourcing clinical processes is the active management and validation of the data transfers with the sponsor.

Because data interchange is so critical, strive to use public data standards, especially when working with external third-parties and negotiate discounts if you are using say, CDISC standards. The cost of a data transfer of a clinical study from a CRO can be \$30,000 or more if each data transfer is a custom job.

When using electronic data capture, it becomes very important to map the migration of data from the EDC system to internal systems. How would this occur? Where would the adverse events be mapped? What about other coding? How would subject laboratory information be handled? The following applies.

Integrity of electronic data bases are a key FDA concern. A prospective data quality plan is recommended. Creating submission components electronically through the entire development process can build quality into the e-CTD.

The data quality plan would lay out system plans, interfaces, validations and handoffs to demonstrate prospectively that these issues were handled properly. This is one part of assuring and documenting data integrity.

We now discuss elements of the technical architecture in more detail.

Applications portal for data access. A web portal provides a secure collaborative framework for clients, associates, partners, CROs, and other parties to exchange information in a regulated environment. Clinical study documents, process information and study data can be available for viewing through a web portal for all clinical studies to project teams, external investigators development partners, CROs, and management. Data provided to external sources (e.g. investigators, collaborators, regulatory agencies) can be presented in a user friendly, organized and well-documented manner.

Project Management and metrics. Project management software can help track the activities across a study and across a project. This facilitates resource management and identification of bottlenecks in the process. Collectively information management systems offer the opportunity to capture metrics on project/study progress during the course of daily work. Transparent access to project/study information will facilitate ease of information flow, decision-making, and management of development and commercial projects.

Clinical Trials Management. A clinical trials management system (CTMS) can be used to track the following types of information:

- Trial administrative information – source for information on study details
- Site details – investigator information, regulatory document status, etc.
- Patient details – ID, visit schedule, CRF pages collected, etc.
- Project planning and management - schedules, resource tracking
- Platform for essential support of site monitoring visits and associated reports
- Cost tracking – manage budget and initiate investigator payments
- Clinical supply tracking forecast supply/resupply allocation and shipping, track expiration dates, etc.

- Flexible reporting from the central database

Availability to store and access the types of information managed by a CTMS system is critical to study management. The CTMS can facilitate a distributed workforce of CRAs or partnership with a contract research organization. Much of the data captured in this system can be used to gather process performance information to effectively manage studies.

Clinical Data Management. The clinical data management system is where the raw subject data is stored for each clinical trial. Typically this software provides data entry capabilities in a secure environment supported by audit trails, error checking, data status, etc. Effective use of the clinical data management system is critical for data integrity. Ultimately, it's the clinical trial data that is the most important part of an application for new drug approval.

As readiness for electronic data capture matures, there is an opportunity to collect data in near real-time. There is tremendous economic benefit to rapid data collection, as well as scientific usefulness, e.g. speedy preparation of reports for study data monitoring committees is compelling. From the information technology perspective, it is important the the clinical data management system be able to collect data electronically.

Global, project specific and industry standards provide the basis for:

- Selection of variables to be collected per study
- Data collection instrument design (CRFs, Electronic Data Capture screens, etc.)
- Operational database structures
- Data quality checking
- Transformation of data into data structures for analysis and reporting
- Programming and analysis
- Integrated safety information

Statistical processing of clinical trial information into final reports is an area where practice has largely been ignorant of the electronic records and signature rule.

The Statistical Computing Environment (SCE). Complementing the data itself are processes to insure the statistical integrity of data analyses and the reporting of results. Just as the design, analysis and interpretation of results depends upon planning, the statistical analysis and reporting requires the following:

- A well-defined, and timely completion of the relevant sections of a Study Analysis Plan (SAP)
- Table, Listing, and Figure specification and creation
- Analysis file specification and creation
- Specified program validation/verification requirements
- A specific analysis plan for any interim analyses
- A clear process for the creation, review, and quality assurance of final study reports
- Creation of integrated safety and efficacy databases

We define the SCE as the electronic implementation of the statistical analysis plan for a clinical trial. A collection of tools and operating procedures comprises a statistical environment for computing in a regulated environment. The output of the environment are controlled documents published in portable document format (PDF). The statistical tools used are primarily SAS Software and S-Plus for tables and graphics.

Routinely statistical analysis is conducted by invoking software to run pre-specified statistical procedures most usually

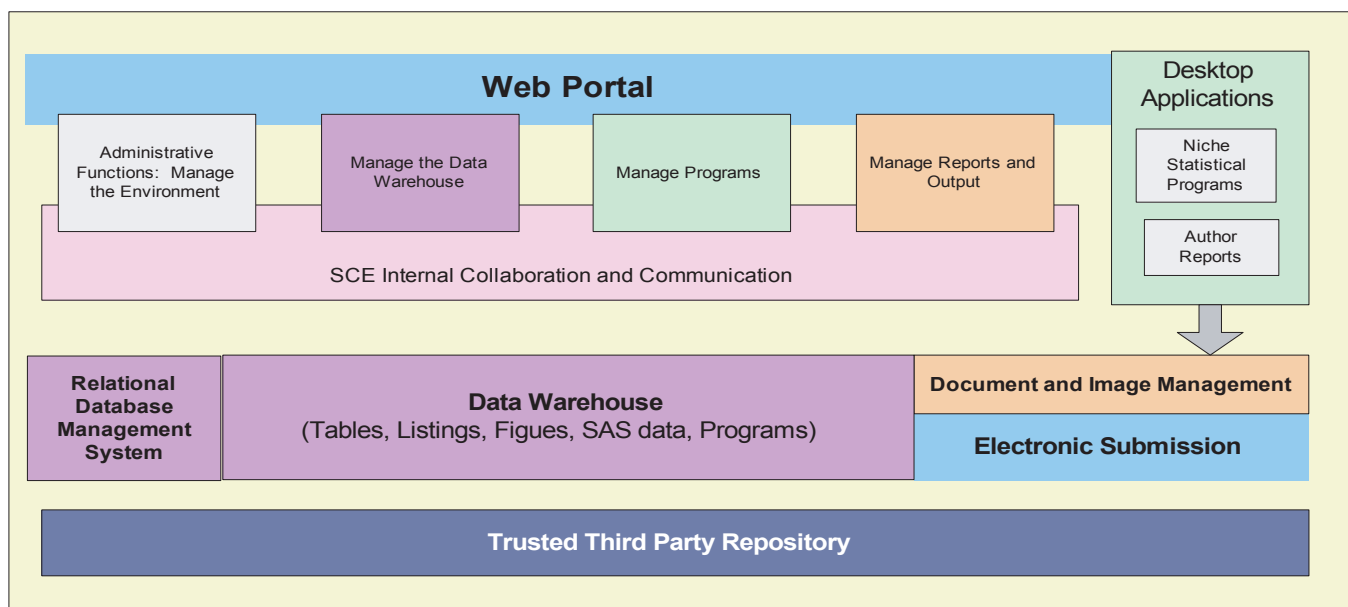
using commercial software. The issue of transparency here becomes traceability of the reported statistical results from the extraction of the data from the clinical database, the transformation of that data into statistical analysis datasets, the invocation of the statistical analysis with the correct parameters and data, and the integration of statistical results into the final clinical study report. Sometimes the statistical results are dependent upon complicated programming which formats the results from multiple statistical functions for a comprehensive table of results.

Several objects need to be managed here: the programs for data manipulation and statistical analysis, the analysis data sets, the results of the analysis, and the integration of the statistical results into a publishable report. How do we manage all of these objects in a way that will guarantee the traceability and transparency we must have from this process? The best way is to work in an environment that tracks all of the objects. Let's take a look at what the components of such an environment might be.

Analysis data sets should be organized according to the recommendations of the Clinical Data Interchange Standards Consortium (CDISC). Data sets are documented according to FDA recommendations. And of course validated processes and systems are used to enter / import, manage, transform and summarize all data.

Tables, listings, figures, SAS data sets, analysis programs are stored in the clinical data warehouse. These functions are diagrammed in Figure 4.

Figure 4. Conceptual components of a statistical computing environment



Document and image management. A secure repository for study documents and standard operating procedures is a must to conform to FDA electronic records and signature rules. This requirement can be satisfied by use of a properly configured electronic document management system (EDMS) and standard operating procedures for its use. These systems come with all the audit trails and data security mechanisms to effectively manage corporate documents. Documents not collected electronically should be scanned into the document system usually using a separate software system.

Final clinical study reports should be prepared for electronic submission in accordance with current FDA guidance on electronic submissions. The electronic submission will contain the entire final report (text, tables, figures, appendices and CRFs) and associated datasets.

Publishing tools are used to create table shells and populate these tables with the appropriate statistical information. Tables are specified using a general table specification using eXtensible Markup Language (XML) and based on standard technology for table description (OASIS). Publishing tools are used to create submission-ready datasets and documents with tables, listings and figures.

Data Warehouse. Data from various operational systems can be integrated and stored in a data warehouse for display on the web. Thus, study metrics, data displays, and documents stored in the data warehouse can all be viewed from the web portal. Objects in the data warehouse can be stored in a relational database. The warehouse will also serve as storage for regulatory submissions.

Regulatory publishing. In addition to documents, documentation of a studies' data structure and the elements in the database are crucial to its use for review. Both FDA guidelines and the CDISC group have recommendations for documentation of the data structure in submissions. Failure to properly organize and document the data could lead to a failure to file at the FDA so the activities discussed here are not to be overlooked.

Archive. The archive stores regulatory data which might need to be retained for 20 years or more. Electronic records must be maintained in a safe compliant environment and be suitable for access and use in the future. Since hardware and software change so frequently, native file formats may not be suitable. Data may need to be converted to standard formats to insure records maintain their integrity – their metadata (or definition/documentation), electronic signature link, audit trail etc. Procedural controls are needed to address systems and access for archiving. Archiving is an important step and should be given adequate thought and attention.

Trusted third-party data storage services. In the future, we see trusted third parties developing repositories for helping industry share information with partners, investigators, regulatory agencies, independent data monitoring boards, etc. We expect that instead of buying software, the industry can buy full applications services from vendors. Application service providers in conjunction with third party trusted repositories can provide many of the services that the industry needs to manage collect and analyze their data without managing large information centers themselves.

Table 2: Clinical & Regulatory Systems Architecture

Architectural Components	Description	Selected Vendors & Products
1. Clinical Portal	Web-based access to applications and data for on-going clinical projects. The portal should also provide a means of collaboration which is 21 CFR Part 11 compliant.	Intrasphere Intralink STATPROBE
2. Enterprise Project Management	A system for tracking tasks and milestones for the entire corporate-wide drug development effort including research, manufacturing, preclinical, clinical, regulatory, and sales and marketing launches	Many, many vendors including: Microsoft Project Primavera ...
3. Clinical Trials Management	Software used to track trial administrative information, site details including monitoring reports, enrollment, budget, clinical supplies, etc.	FW Pharma IMPACT ClinSource TrialXS Siebel eClinical
4. Clinical Data Acquisition & Management	<ul style="list-style-type: none"> Paper and electronic data capture applications link the sponsor, the monitors, clinical laboratories, and sites together. Includes initial data capture, discrepancy resolution, site monitoring, and source data management. Capture, validation and storage of subject data in a compliant environment. 	Data Acquisition: PhaseForward InForm CB Technologies PHT Aris Worldwide Oracle RDC (under development) Marcon Data Management: Oracle Clinical PhaseForward Clintrial eResearch Technologies services MAJARO ClinAccess DZS ClinPlus
5. Safety Data Management	Processing, analysis, and reporting of serious adverse events.	PhaseForward Clintrace Oracle AERS Relsys Argus ARIS G
6. Clinical Data Repository	The repository consists of multiple databases associated with collection of data and documents.	Oracle Documentum OpenText Livelink
7. Extract, Transform & Load	These processes take the data from the operational data base and create new data sets that are amenable to analysis and regulatory submission.	Oracle and/or SAS
8. Clinical Data Warehouse	The warehouse is where the final information is stored for analysis and submission.	Oracle SAS
9. Statistical Computing Environment	A compliant and validated environment for creation of statistical tables, data listings, and graphics for final study reports.	SAS Meta-Xceed Waban Software
11. Regulatory Publishing Tools	Preparation of reports for submission of to regulatory agencies in the format required by electronic guidelines.	Liquent IPEDO CDC Solutions Datafarm Image Solutions

Technology integration cannot truly take place in the absence of data standards.

Until industry standards are created which permit data exchange among different silos, the truly integrated clinical system will continue to be the holy grail of the industry despite the best efforts of systems integrators. Integration is not a one-time achievement. Changing technologies, components and organizations can drive new integration needs. Standards help cut time and effort to maintain integration. PharmaStat believes that the most important aspect of the high-level architecture in Figure 3 is the standards backbone separating the applications from the data repository and the data warehouse. This layer can help align the various systems or make it impossible for them to all work together. This standards layer is by far the most difficult aspect of the technical architecture to manage and leverage.

Probably the oldest data standards in the drug development industry have been those associated with coding of medications/ procedures (ICD-9, etc) and adverse events. The MedDRA dictionary has become the standard for reporting of adverse events.

Clinical data standards are a work in progress. The Clinical Data Interchange Standards Consortium (CDISC) has proposed standards for operational databases, lab results transfer, and submission data standards.^{3,4,5} At this time, there are no competing standards and it is likely adoption will occur over time. Some software vendors are beginning to support CDISC standards and as biopharma adopts them, more efficiency from both sponsors and vendors can be expected.

Subject-data standards such as CDISC only represent part of the information that is generated during a clinical trial. The Regulated Clinical Research Information Management (RCRIM)⁶ a technical committee operating under Health Level 7 has ambitions to develop standards for other aspects of the research process. For example, a protocol representation project is underway. This effort seeks to identify and organize protocol information so that it might be used in other processes. RCRIM seeks to develop a common information management framework across all of the clinical trials processes.

Also, other standards have been developed by the International Conference on Harmonisation (ICH).^{7,8,9} From a technological point of view, ICH guidance on the electronic Common Technical Document (eCTD) provides a path forward for the creation of an electronic licensing application that can be used in the United States, European Union, Canada and Japan. Also, ICH has developed standards for electronic transmission of postmarketing adverse event data to regulatory authorities^{10,11}. The US FDA has a pilot program in place to accept electronic submission of adverse events and has received over one million electronically reported events.

It is notable that none of the standards discussed here were driven by the pharmaceutical industry, though the industry participated in many of the ICH and CDISC working parties. It is worthwhile to speculate how the industry itself can use its collective influence to move informatics development for clinical research in a direction that truly influences vendors and drives best practices.

We believe that the eXtensible Markup Language (XML) is a key technology that facilitates integration among systems components and regulatory agencies as long as the systems have a common vocabulary. For example, the exchange of post-marketing safety reports between sponsors and the FDA is based on XML data elements.¹¹ The eCTD has an XML backbone.² The CDISC operational data sets can be defined in terms of XML elements. Therefore we support the ICH and CDISC and the move towards emerging standards. It just cannot happen fast enough!

Table 3. Pharmaceutical Applications using XML

Application	Description
Adverse event reporting with E2B	The ICH created a regulatory e-submission format for individual case study reports using XML. Currently in pilot status with US FDA.
eCTD (Electronic common technical document)	Another ICH standard which based on an XML backbone for navigation. A DTD has been published by ICH and FDA for the eCTD.
Product Information Management (PIM) Initiative from EMEA	XML schema standards have been developed for the summary of product characteristics, the product information leaflet, and package labeling. An optional component of the eCTD.
Clinical Data	An XML standard has been developed by CDISC for interchange of information into and from an operational data base
Clinical Study Reports	XML authoring provides the opportunity to develop structured documents with statistical output inserted directly into documents avoiding retyping or cut and paste. XML authoring is not widely used at this time but usage can be expected to increase.
Data Interchange	PharmaStat believes that XML will also become the standard for submission of clinical data to the FDA in the future, replacing the SAS transport format.

Appropriate on-going validation and compliance with 21 CFR Part 11 are major costly stumbling blocks to successful implementation and maintenance of clinical trials software.

There is a special guidance for the use of computerized systems in drug development.¹² All the clinical systems need to be validated.¹³ The validation process – Installation Qualification, Operating Qualification, and Production Qualification (IQ, OQ and PQ) – must be completed for every GCP system. Extremely time consuming and costly, validation requirements contribute to slow acceptance of new technologies in clinical drug development. In addition, the systems (both hardware and software) must comply with Title 21 CFR Part 11 for electronic records and signatures.^{14,15} Cost of acquisition, installation, integration and maintenance of all these systems represents a significant initial and on-going expenditure. Thus the entry cost to support clinical development is non-trivial.

Therefore on-going business process and documentation tools required to actively maintain validated and audit-ready state of all clinical and regulatory systems over time. This has spawned a new marketplace – the regulated systems market with demands for security, access and control systems to manage data and other intellectual property in the pharmaceutical industry. There are many small companies and large systems integrators which provide services in this area.

The pharmaceutical marketplace is so small there is not a lot of upside for good competition and comprehensive solutions. Almost no application software vendors have managed to sustain adequate growth and profitability in this narrow and conservative biopharmaceutics market.

Vendors come in several flavors: silo software vendors, CROs, applications providers, and integration consultants, and e-clinical services. Despite the proliferation of biotechnology companies, there are only a few small vendors effectively

servicing the marketplace. For example, the marketplace for clinical data management software is dominated by Oracle (Oracle/Clinical) and Phase Forward (ClinTrial). We refer to “e-clinical services” as internet-based applications that facilitate project team workflow and collaboration, helping to accelerate and improve the productivity of the clinical data collection process. The e-clinical vendors tend to provide innovative solutions that have largely been absent in this marketplace. There are far too many of these vendors for the marketplace to support and so consolidation can be expected.

Depending on the size of an organization, PharmaStat estimates that it costs \$5MM - \$30MM to put all the systems in place necessary for effective clinical trial execution and reporting. This will drive increased use of Application Service Providers (ASP) and CRO models for some companies.

For each of the critical applications in Figure 3 - clinical trials management, clinical data management, the statistical computing environment, adverse event reporting, data repository, document management, and regulatory publishing tools - there are several layers of cost. First is the cost of the application itself, then the hardware servers to run the applications, professional fees for installation, integration and validation, software maintenance fees and the in-house infrastructure to support the computer operations and applications systems. Since no single integrated package exists, there is overhead in dealing with a variety of vendors of applications.

The applications software packages are generally sold by the number of “seats” or users of the application. Prices are usually negotiated, so it is difficult to quote specific prices for an entire suite of applications. Think six figures for most applications. The cost of large systems used by many (e.g. document management) can easily reach seven figures. Some vendors are unwilling to quote the retail price for their software. PharmaStat has talked with both vendors and customers to estimate the total cost of ownership.

The price of the applications themselves is just the tip of the iceberg of the total cost of ownership. The applications need to be installed on servers. Sometimes, an application will require both a database server and an application server for optimal performance. Usually the database software will be purchased separately from the applications software and the cost is non-trivial. The size (and cost) of the servers will be dependent on the number of users of the applications system. Usually separate servers are used for each application. This can facilitate smooth application performance by not competing with other applications and making software upgrades easier.

In addition to acquiring the application software and hardware, expect significant “professional fees” from the vendor or third parties to help integrate each application into the corporate environment. These fees may include validation and training in addition to software customization specific to a firm. Expect significant staff time to develop internal business rules and operating procedures for each application. Sometimes extensive “change management” is required to get employees working in a new way. The internal human cost of instituting these systems cannot be over emphasized. Systems support staff will be required for implementing and maintaining the array of software applications. This usually requires information technology specialists who understand the business needs of the organization as well as a technical understanding of the software itself. A database administrator will be required to maintain all the databases. Don't overlook facilities costs to simply maintain the architecture: networks, firewalls, computer rooms, daily backup and recovery, standard operating procedures, disaster planning, desktop computers and associated productivity tools all in a 21 CFR Part 11 compliant environment.

Maintenance is a big-ticket item. The annual renewal fees for the applications software is usually a large fraction of the first year fee and is just the beginning. Testing application software and installing updates is a time-consuming and therefore an expensive process. Not to be missed is the need to also maintain special testing environments for most products to facilitate upgrade testing and validation. Ultimately the applications users themselves will have to perform acceptance testing of the software (production qualification). This means the workers who actually process information for the organization must be taken “off-line” for a period of time to do the testing temporarily lowering operational productivity.

An important consideration in preparing the strategy is the realization that expensive technology investments will achieve the highest return on investment in firms where they are used for many studies. So what makes sense for a large pharma company may be quite different from what is needed by a small biotech company just entering clinical research. The cost of the applications software is high, and integration, maintenance and training add to the cost. Smaller firms need to balance the costs with the possibility of outsourcing to a contract research organization or using an applications service provider. In any case, it's the data that's important and we believe that each firm should manage their data processes actively to use consistent standards to facilitate aggregation and regulatory publishing, and achieve data integrity with good documentation. This process should not be left to third parties.

In addition to the major software applications just discussed there are other software components which may be used from time to time which more appropriately fall into the category of services provided by third-parties. This includes electronic data capture services, voice randomization systems, patient recruitment, etc. The web-services model is allowing a proliferation of specialized third-party services. This makes a permanently more complicated services marketplace and room for integrators to help connect all the pieces of the environment.

PharmaStat believes a development organization should be spending 10-12 percent of the clinical and regulatory budget for systems enhancement and maintenance as a rule of thumb. If some applications are outsourced, the accrued costs should be recognized as information technology expenditures.

Clinical Information Technology Governance

The good news is that effective use of technology can help speed development and get new therapies to patients faster. The bad news is it can be an endless money pit from which return on investments are not realized if not properly managed.

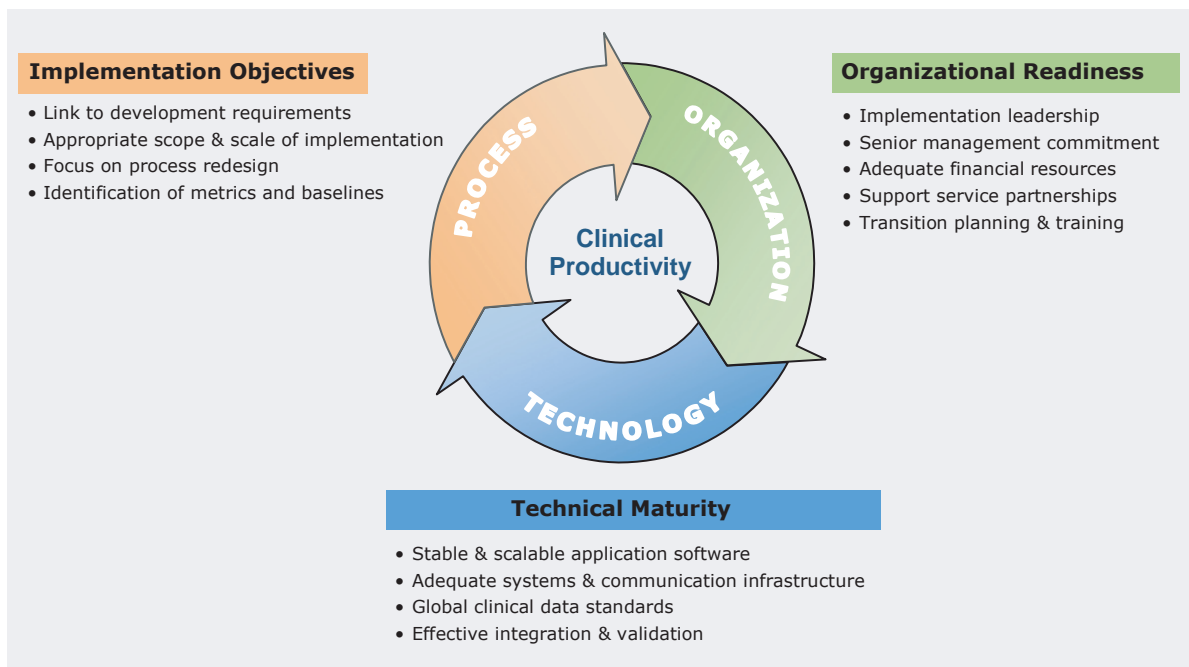
Productivity depends on alignment of work processes, technology, and organizational structure. Balance these variables through a strategic business process plan – a data vision – and a complementary technology plan. Actively review these plans to effectively manage your clinical and regulatory IT investments.

Adopting a new technology often means changes in business processes. For example, electronic data capture changes the role of the data manager, the clinical research associate (CRA) and the site study personnel. From the data management perspective, electronic screens need to be designed to capture the subject information. Data entry will be performed at each study site with edit checks at the time of entry. Fewer discrepancy corrections would be expected to be generated. The CRA will find that site monitoring of study results will be based on electronic reports from the database and not from ordinary paper case report forms. For these process changes to be successful, there needs to be readiness to accept them by all parties concerned. For success, there needs to be project leadership and

commitment of financial resources and training at both the sponsor and at each investigator site. Finally, the technology should be mature and stable. Nothing could frustrate acceptance of a new process faster than bugs in the software that was supposed to streamline the business process. Only by attention to the alignment of technology, process, and organizational readiness will the introduction of new application systems produce productivity gains expected (Figure 3).

To obtain this balance we believe that the organization should have a strategic vision for technology that is driven by business needs. The strategy should be written and developed by the business process owners. In the clinical-regulatory realm we call this strategy a data vision or data blueprint which focuses on gathering, managing, aggregating and reporting information that is critical to the clinical trials process and regulatory reporting.

Figure 5. Productivity depends on alignment of business process, organizational readiness and technology



Pharmas are typically very poor at clinical information technology governance, and lack a principle-based approach to CRI systems architecture. New models are needed to effectively manage IT evolution and outsourcing against one-off long-term projects.

Governance refers to an organization's decision-making process and the assignment of decision-making authority and accountability. Technology investments are very expensive. Inappropriate decisions to deploy systems and implementation mistakes can waste not only money but precious time. While recognizing that information technologies, properly deployed, offer great potential benefit, pharma and biotech firms constantly struggle to find a balance between technical infrastructure spending versus direct costs of product development projects.

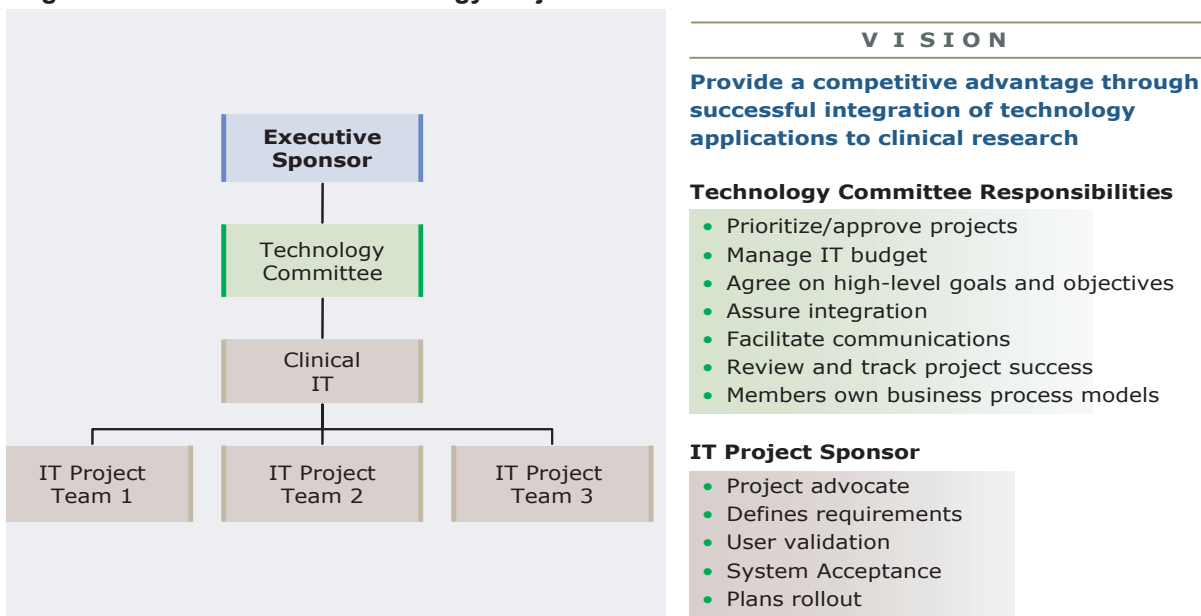
For most organizations we recommend chartering a clinical IT steering committee. Because there are multiple stakeholders in these systems, membership should be the business process leaders of the systems identified in Figure

2 and the head of IT. The committee should have ownership of the data blueprint and the overall technology plan. Decisions would be made in the following areas:

- Budget – How much should we spend and what projects can be funded
- Leadership – project prioritization, resource and change management
- Oversight – Review implementation, manage the data blueprint, technology plan, and coordination with corporate-wide planning

Who should run clinical IT projects? We recommend designating business unit sponsors for each IT system or initiative. They will assign implementation resources and establish success criteria. Appropriate training and management of new skills and job responsibilities is required for the successful rollout of technological systems. This forces the user community to be stakeholders in the implementation process.

Figure 6. Governance of Technology Projects



Since it’s all about the data, then systems integration should be the primary responsibility of the clinical IT group.

Clinical IT should not be focused on systems development but on integration, validation and maintenance. Use well made preexisting components whether you rent them through a CRO or an application service provider. Avoid internal software development whenever possible.

Managing the systems hardware and software environment is non-trivial. 21 CFR Part 11 and validation requirements along with short software lifecycle and a long drug development lifecycle implies an on-going systems maintenance problem. You need to have loose coupling between the different systems. Every application could be running on a different machine. Validated releases of the underlying data base or operating system will probably not be common among the applications supported. One must be able to upgrade one system without having to upgrade the others. The systems team has to be much more involved in the clinical projects. Every time you touch something, it needs re-validation.

Summary of Action Recommendations

A number of recommendations were mentioned in the text. Here we summarize those recommendations that we believe will help make the data pathway validated, integrated among various applications, and data readily accessible.

Table 5. Recommendations for Action

Recommendation	Remarks
1. Plan for electronic regulatory submissions	PharmaStat believes that fully electronic data processing activities provide greater quality in the eCTD.
2. Document data integrity	Leverage the CDMS system to provide metrics on the data.
3. Adopt clinical data standards	Leverage standards for importing data from vendors and exporting data to FDA.
4. Integrate clinical information systems	Integrated systems provide metrics on processes for decision making and provide readily available subject data when needed.
5. Manage the entire data pathway as one process. Develop a data vision and a complementary technology plan. One executive should be accountable for the overall process.	The data vision should describe the preferred way of doing business. SOPs should be available to support the data vision. The technology plan illustrates how the clinical applications in the data vision are managed.
6. A prospective data quality plan is recommended as a guide for the team and as trial documentation.	The data quality plan should describe the data flow for the trial, including transfers from vendors and quality assurance plans.
7. Actively manage the transfer of databases from outsourcing vendors.	It is very important to be able to trace the transition of data into the sponsor's database and to assure validated processes are used at both ends of the transfer.
8. Develop a plan to bring statistical computing activities into compliance with 21 CFR Part 11	Statistical analysis traditionally has been an ad-hoc activity. Rigor in validation and compliance is needed to pass an audit.
9. Set-up a governance structure for clinical information technology decision-making	Use the governance structure to assure appropriate technology priorities, oversight, and successful implementation.
10. Systems integration should be the primary responsibility of clinical information technology.	Buy commercial software or fund its development through third-party integrators or vendors. Insiders should be concerned with integration and validation of systems.

Conclusion

An effective well running clinical data platform is a competitive advantage because it can improve data quality and speed of execution.

It is not the individual components of the system, but the extent of integration that is an advantage for firms. We believe that the pharmaceutical industry has not realized the business value of their clinical technology investments. There is huge value to be gained so change in behavior is indicated in the governance of these activities in the firm and the continuing development of data standards for each clinical process.

References

1. International Conference on Harmonisation (ICH); Guidance on M4 Common Technical Document, October 2001.

M4: Organization of the CTD
M4E: The CTD - Efficacy
M4Q: The CTD - Quality
M4S: The CTD - Safety
M4S: The CTD - Safety Appendices
2. International Conference on Harmonization: Electronic Common Technical Document Specification V 1.0 24 June 2001
3. Clinical Data Interchange Standards Consortium (CDISC) <http://www.cdisc.org/>
4. Introduction to the CDISC Submissions Data Domain Models Version 2.0 http://www.cdisc.org/pdf/Submission_Data_Domain_Models_2.pdf
5. CDISC Submission Metadata Model <http://www.cdisc.org/pdf/SubmissionMetadataModelV2.pdf>
6. Health Level Seven. Regulated Clinical Research Management (CRIM). <http://www.hl7.org/Special/committees/rcrim/rcrim.htm>
7. International Conference on Harmonisation E3: Structure and Content of Clinical Study Reports, July 1996. <http://www.fda.gov/cder/guidance/iche3.pdf>
8. International Conference on Harmonisation E6: Good Clinical Practice: Consolidated Guideline, April 1996. <http://www.fda.gov/cder/guidance/959fnl.pdf>
9. International Conference on Harmonisation. E9: Statistical Principles for Clinical Trials, September 1998. http://www.fda.gov/cder/guidance/ICH_E9-fnl.PDF
10. Draft FDA Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Postmarketing Expedited Safety Reports, May 2001. <http://www.fda.gov/cder/guidance/4153dft.pdf>
11. FDA Guidance for Industry: E2BM Data Elements for Transmission of Individual Case Safety Reports, April 2002. <http://www.fda.gov/cder/guidance/4627fnl.pdf>
12. FDA Guidance for Industry. Computerized Systems Used in Clinical Trials. http://www.fda.gov/ora/compliance_ref/bimo/ffinalcct.htm.
13. Guidance for Industry and FDA Staff: General Principles of Software Validation. Jan 11, 2002. <http://www.fda.gov/cdrh/comp/guidance/938.pdf>
14. 21 CFR Part 11—Electronic Records; Electronic Signatures, Final Rule; Federal Register, 62 FR:13430–13466, March 20, 1997.
15. Draft FDA Guidance for Industry. Part 11, Electronic Records; Electronic Signatures - Scope and Application, August 2003. <http://www.fda.gov/cber/gdlns/prt11elect.pdf>

Other PharmaStat Technical Reports

Optimizing the Use of Data Monitoring Committees in Clinical Trials by Alan Hopkins, December 2002.

Abstract: Data Monitoring Committees (DMCs) are being used more and more frequently in industry-sponsored clinical trials. They are being asked for frequently by institutional review boards (IRBs) and are becoming standard in trials with endpoints of mortality or irreversible morbidity. The added complexity of an external monitoring committee is discussed and some best practices are suggested. Implications of using a DMC by the sponsor project team is considered.

Now FDA has developed a Draft Guidance on the establishment and use of clinical data monitoring committees. We will comment on this guideline through the lens of key parties involved in the clinical trial.

About PharmaStat

PharmaStat provides design, analysis, and reporting services for clinical drug development in the biotechnology, pharmaceutical and medical device industries. Consulting is offered to manage integration of science, information technologies, and experienced implementation for successful drug development.

PharmaStat can organize expert multidisciplinary teams to implement and manage any aspect of the data analysis and document preparation process.

Web Site

www.Pharmastat.com

Contacts

ahopkins@Pharmastat.com

Main: 650-598-9860

Fax: 650-598-9861