Paper 205-2008

Evolution of SAS® in Life Sciences Research and Development

Dave Handelsman, SAS, Cary, NC

ABSTRACT

The widespread use of SAS® in clinical trials data transformations, analysis, and FDA submissions is well-recognized throughout life sciences research companies. Industry meetings such as SAS Global Forum, PharmaSUG, and PhUSE provide prime examples of how loyal SAS users continue to push the envelope in terms of using SAS for these historic activities.

SAS, however, can provide considerably more value to health and life sciences companies. To that end, SAS has implemented a new internal organization—the SAS Health and Life Sciences Global Practice—to lead the way regarding how SAS can provide value not based only on our historical successes, but across health and life sciences enterprises. The practice will serve the growing need for technology and solutions that will fuel the convergence of health care providers, health plans, and biopharmaceutical companies.

The evolution of SAS' capabilities in the health and life sciences markets will improve the quality and affordability of patient care around the world through deeper scientific and business insights driven by focused industry expertise and sales support, the introduction of a broader portfolio of solutions encompassing the entire health and life sciences enterprise, and tighter alignment and coordination across SAS functions.

This session will discuss the SAS vision for the life sciences research industries and the progress that has been made to date.

INTRODUCTION

SAS' long history of success in the pharmaceutical industry is closely associated with its role as the *de facto* standard for producing safety and efficacy data analyses, transforming and transporting data within the life sciences industry, and for submitting data to the FDA. That success, however, is not cause for complacence, and in recent years SAS has taken renewed interest in developing new solutions and services targeted at the pharmaceutical, biotechnology, and health care industries.

These capabilities are seen as industry-specific solutions (such as SAS Drug Development), and targeted services that extend off-the-shelf SAS products to be industry-focused. An example of these packaged services offerings includes SAS Clinical Data Integration, which applies CDISC standards within the SAS Data Integration solution. In the past year, these efforts have been further extended as SAS builds a comprehensive plan to address a variety of business problems in the health and life sciences industries.

This paper will discuss the current and planned solutions specifically targeted at the pharmaceutical, biotechnology, and contract research (CRO) industries. While healthcare and life sciences continue to converge, SAS' plans to address this convergence continue to evolve.

SAS DRUG DEVELOPMENT

SAS is used almost universally as part of the clinical trials process, with a widespread and consistent approach being used to extract and analyze data throughout the clinical research industries. In this approach, data is extracted from a clinical data management system and converted into SAS data sets or views, which are typically stored on a central server.

The SAS data sets that are available, either through the extract or the view, are then converted via SAS programs into analysis-friendly data sets. (Invariably, the structure in which the data has been captured is not directly suitable for statistical analysis). Additional SAS programs are then applied to generate statistical results from the analysis data sets. These statistical results are then included, either as an attachment or directly, within a clinical study report. This process is depicted in Figure 1.

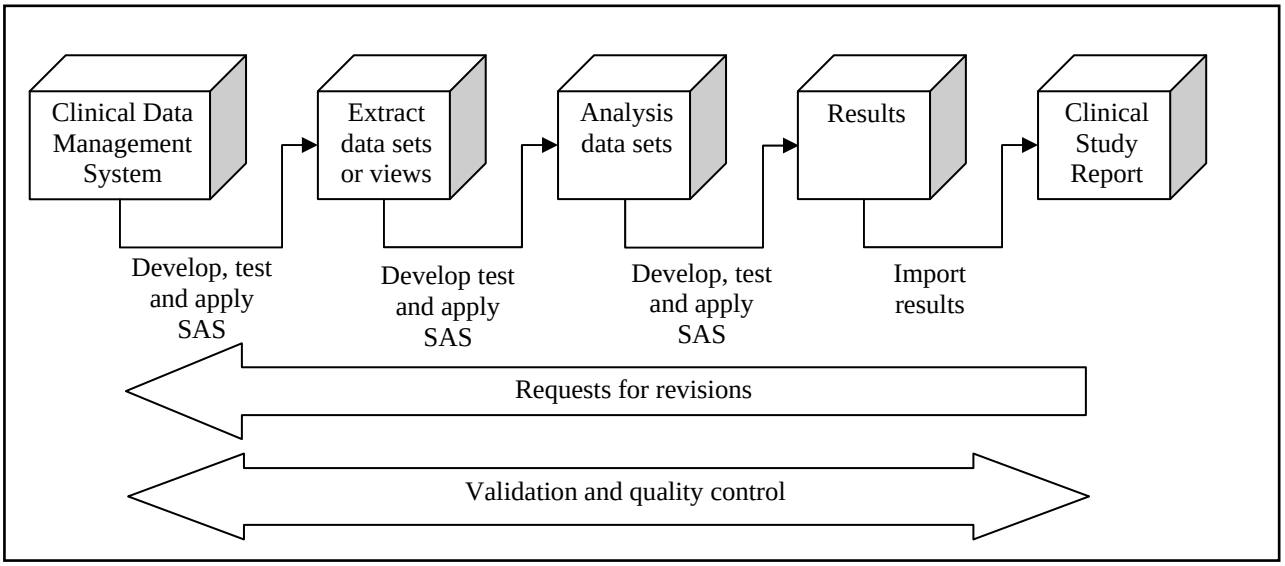


Figure 1: Workflow Associated with Clinical Trial Analysis

The traditional SAS programming process can be replicated within SAS Drug Development, while working in a client-side SAS development environment (using Display Manager, a desktop text editor, or SAS Enterprise Guide) or via the embedded Process Editor application. SAS Drug Development provides the central repository for data, SAS programs, logs, and associated program output, while the relevant program development applications provide the basic tools to write, test and execute SAS programs. With an integrated repository and application environment, the system then provides significant value above and beyond that which is found in the historic SAS programming process.

This additional value includes directly addressing key government regulations such as 21 CFR Part 11 and good industry (clinical, manufacturing, and laboratory) practices (GxPs), collaborating and resourcing across organizational and geographic distances, streamlining the processes from data capture to analysis, and supporting data exploration by scientific and medical consumers for individual and multiple trials.

CONTROLS AND COMPLIANCE

SAS Drug Development automatically provides the necessary controls and compliance to document the integrity of the data transformation and analysis process. For data-related activities conducted strictly within the SAS Drug Development environment, an accurate and thorough history of each programming run (along with relevant inputs and outputs) is automatically captured and preserved. In other words, every time a SAS program is run, an XML-based package is generated within SAS Drug Development that documents the SAS program version being executed, the SAS log file created, and all of the associated program inputs and outputs. This package, referred to as a manifest, is saved for every program run, and can be automatically versioned. In this way, users have the ability to examine the specific version of a SAS program used to create a specific version of a SAS output, along with the accompanying log file. As shown in Figure 2, version 8 of the *get_freq.sas* program used version 2 of the *demo.sas7bdat* data set to create version 8 of the *freq_output.pdf* results file. The details of the program run are documented in version 10 of the *get_freq.log* log file. Each of these manifest entries is hyperlinked to the associated file so that the manifest content can be easily viewed.

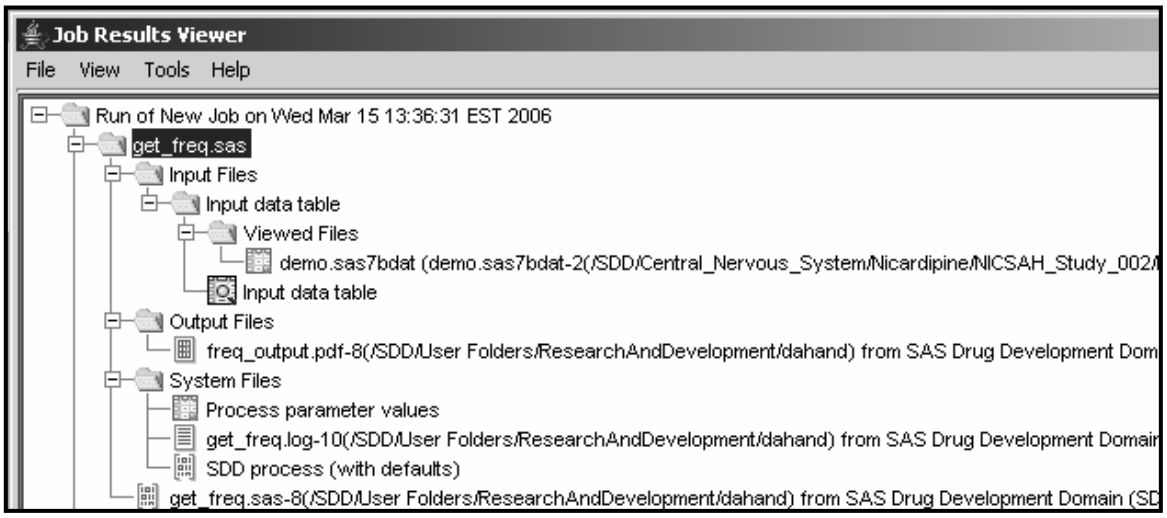


Figure 2: Manifest Contents Showing Relationships of Program Inputs and Outputs

These traceability and documentation capabilities are valuable from the perspective of ongoing quality control and quality assurance activities, and, perhaps most importantly, when addressing questions from regulatory agencies. By using the embedded capabilities within SAS Drug Development, it is a straightforward exercise to electronically search for the crucial bit of information that might be required as part of the review process. Importantly, the metadata necessary to perform this search is automatically embedded in the SAS program execution process, both minimizing any additional work that must be performed by the programmer, and also guaranteeing that the requisite metadata is being captured.

CENTRALIZED INFORMATION MANAGEMENT AND PROCESSING

Information within SAS Drug Development is organized in a familiar file hierarchy mechanism, as shown in Figure 3, allowing authorized users access to the parts of the research content relevant to their work. In this example, *Nicardipine* represents the compound level within the repository, and *Nicsah1*, *Nicsah2,* and *PK* represent protocol levels. The remaining folder icons correspond to other storage levels within the hierarchy. The hierarchy can be customized at all levels based on the individual project requirements. Because the research effort can be consolidated in this centralized repository, the logistics regarding the storage and retrieval across multiple systems at the time of submission is eliminated.

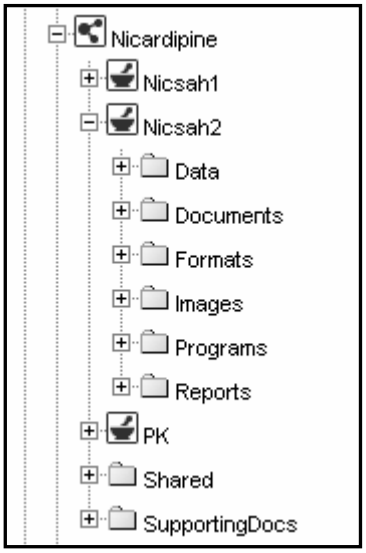


Figure 3: System Hierarchy

This Web-based system is accessible through a standard Web browser, enabling users to access the research content regardless of where they are geographically located. In essence, if they can connect to the server environment—either through the public Internet or a private intranet, depending upon how the system is deployed—they can access the research content. This capability provides research customers with the ability to resource data analysis and review activities wherever capacity exists within their organization—easily transferring work to other departments, other geographies, or other service providers as necessary—and ultimately bringing the right resources to bear on the project at the right time.

HOSTING

SAS Drug Development is commonly, though not exclusively, deployed in a hosted environment. With this approach, the responsibility for hardware procurement, software installation, validation, backups, and product updates falls to SAS, offloading these responsibilities and burdens from stretched clinical research IT departments. Validation, in particular, can be an expensive and time-consuming proposition, and clinical research companies have recognized that the hosted solution approach provides an innovative and effective method to deploy new solutions to their users.

When coupled with the Web-based nature of SAS Drug Development, which requires no client installation, this approach becomes even more appealing.

Although the initial market reaction to the hosted implementation of SAS Drug Development was somewhat guarded, with primary concerns regarding security and stability, successfully hosted implementations have increased steadily in number, and are expected to continue growing. In the United States, virtually all SAS Drug Development implementations are deployed as hosted solutions, and in Europe more than 75% of the implementations are hosted.

ELECTRONIC DATA CAPTURE (EDC) TO ANALYSIS

Clinical trials have traditionally relied on paper-based processes to collect and manage patient data. The emergence of electronic data capture (EDC) eliminates paper from the process, with the investigator site transcribing patient charts directly into the EDC system. The historic practice for this EDC data has been to either periodically, during the course of the trial, or solely at the end of the trial, export the EDC data into SAS for analysis. In some cases, the EDC data has been first back-ported into a legacy clinical data management system, which then produced the extracted SAS data sets for transformation into analysis data sets and ultimately statistical results. While the adoption of EDC technology provides a critical alternative to paper-based data capture, the reliance of the legacy business process to move that data toward analysis is suspect.

Instead, customers are beginning to adopt an information workflow where data flows directly from the EDC system to SAS Drug Development for analysis and exploration. In this emerging model, the EDC system exports, on a daily basis, the incremental changes from the previous day's database, and this incremental data (transported as CDISC ODM data) is automatically loaded into SAS Drug Development for aggregation with other clinical trial data components (such as labs), and subsequent further processing. By synchronizing the metadata between the EDC and SAS systems, the data transfer process is both simpler and faster.

This methodology allows the biostatistics and medical teams to have near-realtime access to the EDC data. (The delay of one day is largely a convention that is easily adjusted). From a safety perspective, this approach provides the medical team with access to the most current data possible, and when combined with the self-service reporting and data exploration capabilities of SAS Drug Development offers a unique capability to more fully understand ongoing patient treatment and outcomes. Beyond safety, the immediate availability of clinical trial data provides unparalleled capabilities for operational reporting regarding the trial. These operational reports might include simple status reports such as numbers of patients completing specific enrollment milestones or more advanced reports that bring SAS' predictive analytics to bear. These advanced reports could be used to predict last-patient/last-visit or enrollment-met dates, and would allow resources to be accurately scheduled far in advance of the necessary date. The reports, regardless of their complexity, would be available within the centralized repository to all authorized users.

REDUCING THE BARRIERS TO DATA EXPLORATION

Within clinical research organizations, physicians, medical writers, and other scientific researchers typically have limited direct access to the clinical research data. The data, received and cleaned by the data management team, and analyzed by the biostatistics division, is typically shared with medical staff only through an indirect process whereby the medical team makes specific requests of data management or biostatistics, and then works through several iterations of these requests until the correct information is produced.

Through SAS Drug Development, it is possible to create self-service reports to answer common medical questions (for example, which patients experienced specific adverse events?), as well as dynamic data exploration sessions. These capabilities free the data management and biostatistics staff to leverage their skills to solve the business problems for which they are responsible, while empowering medical researchers to answer their own questions directly. Such questions, whether asked during the execution of the clinical trial to better understand safety concerns, asked at the end to accurately discuss findings of the clinical trial, or asked across multiple clinical trials (a capability unique to SAS Drug Development), are vital to the science of clinical trials, and best addressed by the team that understands the medical nature of the question. By empowering the medical team with the ability to answer their own questions, SAS Drug Development provides research tools throughout the clinical trial process that can be used to not only understand the current trial, but plan for future trials.

SAS CLINICAL DATA INTEGRATION

For the past several years, SAS has been actively supporting CDISC standards in Base SAS software through the CDISC procedure (PROC CDISC). PROC CDISC enables users to read and write XML files formatted using the CDISC Operational Data Model (ODM) and verify that data conforms to the Study Data Tabulation Model (SDTM).

Beyond these basic capabilities, however, SAS provides robust support for transforming data to the SDTM model using the centralized, streamlined, metadata-driven approach of SAS Clinical Data Integration.

At the heart of SAS Clinical Data Integration is the SAS Open Metadata Architecture, a general-purpose metadata management facility that provides common metadata services to SAS and other applications. The metadata architecture reduces development effort because all applications use the same centralized metadata, which makes it the ideal solution for sharing and managing CDISC models and metadata.

SDTM defines the physical structure, metadata, and business rules for creating standard domains of data. Within the SAS Open Metadata Architecture, the SDTM metadata is applied to the SAS metadata model using specific extensions that leverage the controlled terminology and the base model definition. As shown in Figure 4, a complete domain definition is ultimately stored in the metadata.

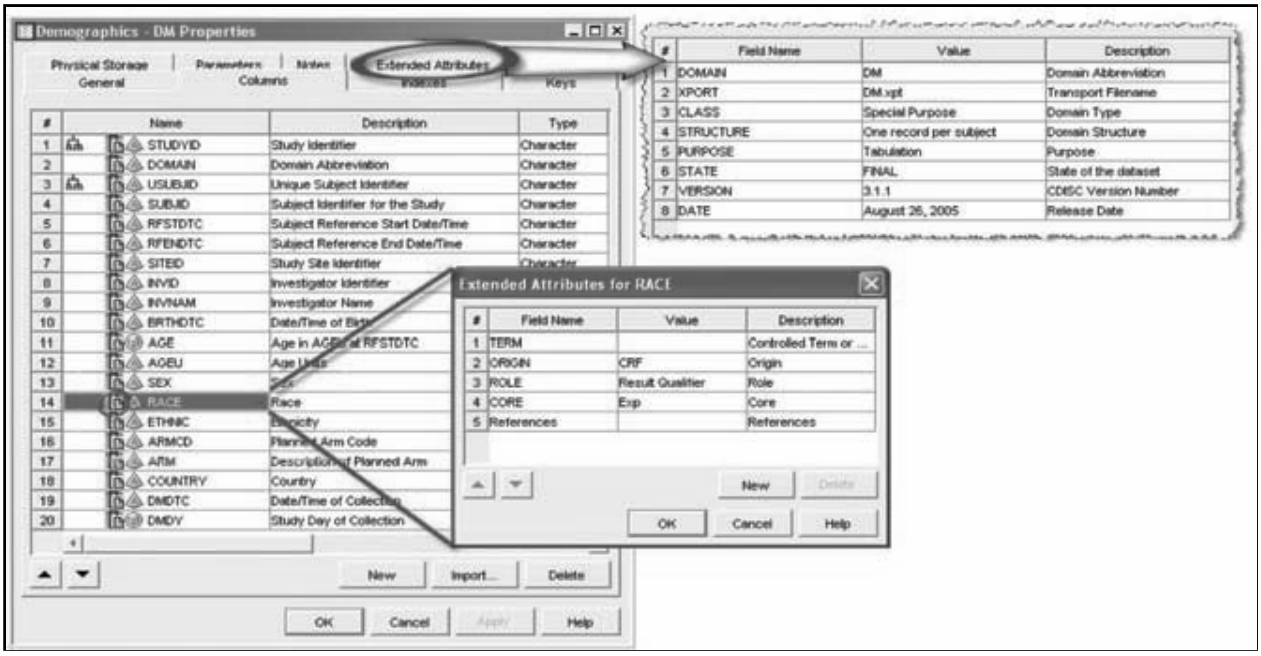


Figure 4: Metadata Created for an SDTM Domain

When a new study is ready to be processed, a new study repository is created that depends on information from the previously created standards repository. This leverages the CDISC metadata from the common standards repository, ensuring that all users are using the same core model definitions. At the study repository level, however, it is then possible to extend the standard model to accommodate study-specific information not accounted for in that core model. By using this configuration, business needs such as archiving and concurrent study development can be more easily managed.

PUTTING THE METADATA TO WORK: SAS® DATA INTEGRATION STUDIO

SAS Data Integration Studio is a visual design tool for building, implementing, and managing data integration processes regardless of data sources, applications or platforms. The drag-and-drop workflow design interface reduces programming time and training needs, and automatically captures and manages standardized metadata from any source. Because of all this metadata-driven capability, SAS Data Integration Studio brings intelligence into the standardization process by automatically identifying column inconsistencies, maintaining traceability, enforcing change management, and evaluating impact analysis throughout the data workflow.

SAS Data Integration Studio combined with the SDTM metadata becomes a streamlined, metadata-driven approach for creating SDTM data. By using SAS Data Integration Studio, driven by SDTM metadata, CDISC metadata is proactively managed while final data is generated. In particular, SAS Data Integration Studio provides the following key capabilities:

direct access to CDISC metadata. When metadata changes are required, SAS Data Integration Studio provides the ability to edit the metadata directly rather than having to work through other tools (like Excel).

visual transformation and mapping. Each step in the process enables users to specify how the data flows through the defined process.

leveraging trial design and controlled terminology data sets. Because these data exist as tangible entities, they can be accessed within the process itself. As shown in Figure 5, controlled terms for VSTESTCD are being verified, and additional information found in the controlled terminology table is added to the data stream. Likewise, Trial Visits is being used to expand visit codes found in the source data.

integrated term management. As terms are being used in the data stream, issues encountered, such as unknown terminology, can be copied to exception tables. Because these tables are also registered in the metadata, users can be immediately alerted to the details of the problem in order to resolve discrepancies.

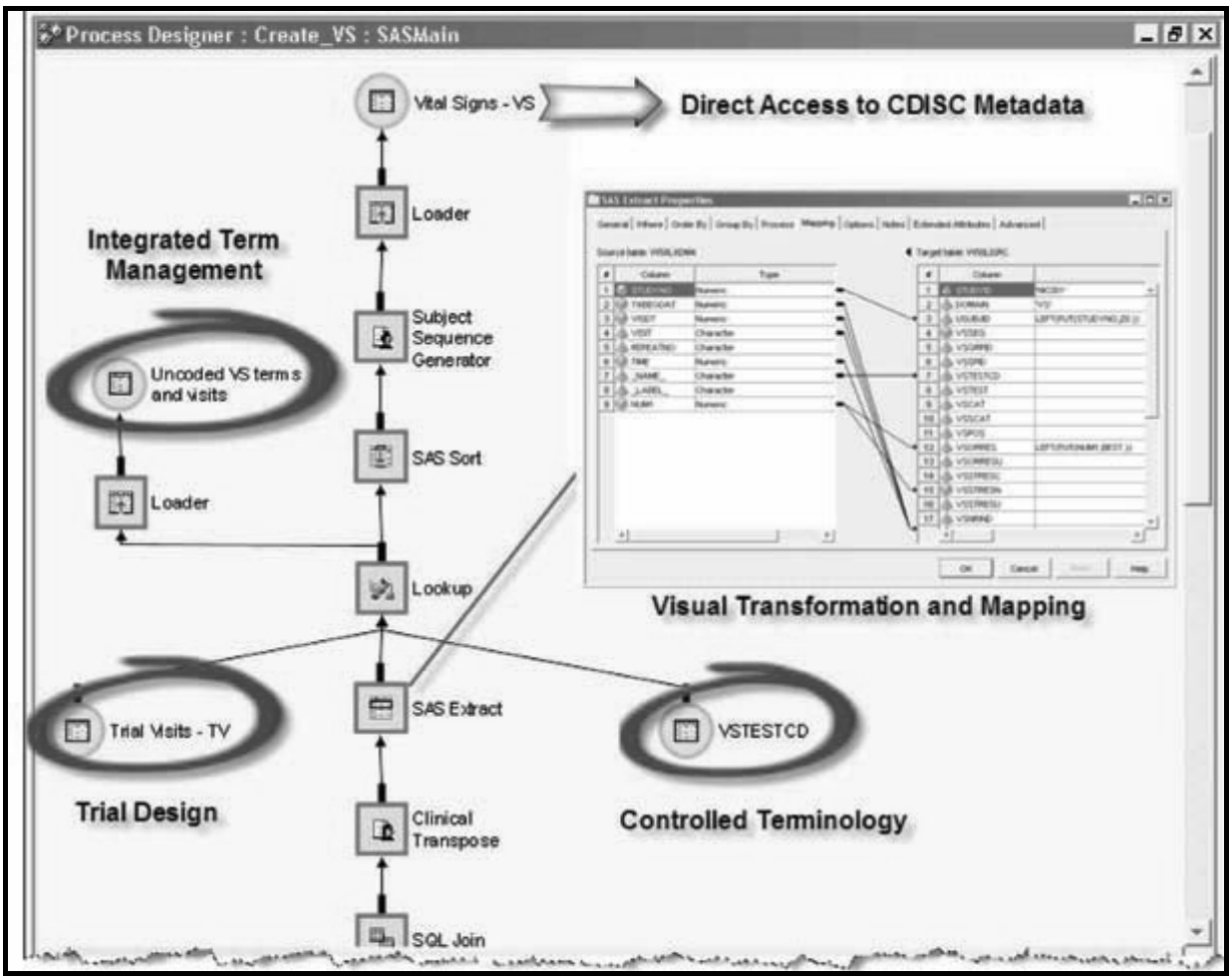


Figure 5: Key Benefits of Working with SDTM in SAS Data Integration Studio

DOING MORE WITH THE METADATA

SAS Data Integration Studio clearly adds value to proactively leverage and manage the CDISC standards while standardizing data. However, standardized data is not the only output of this process. In addition to the CDISC models, a significant amount of valuable metadata is available that can be further leveraged.

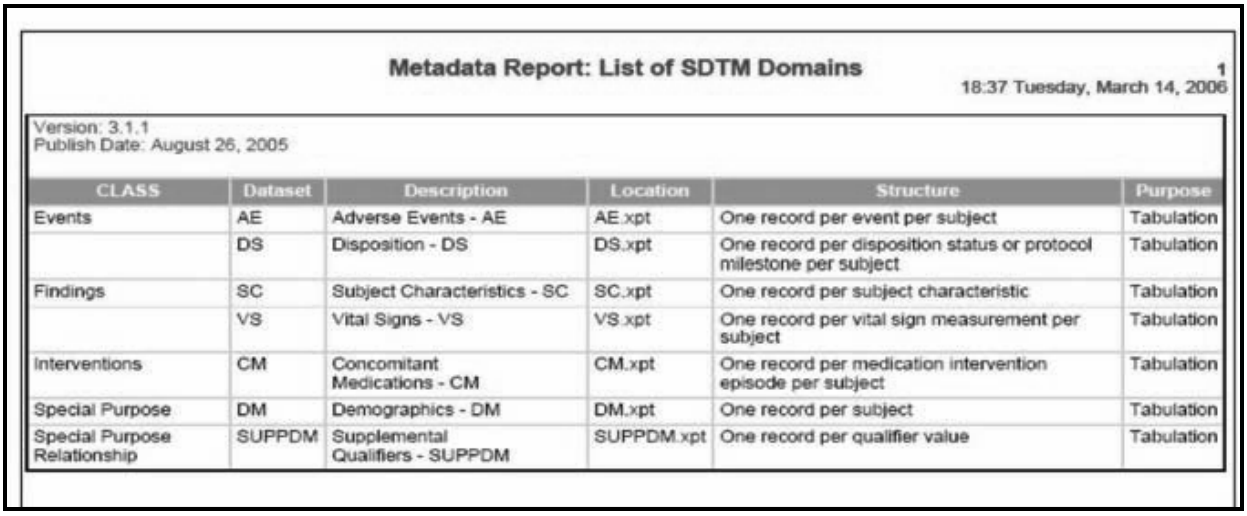
impact analysis. Although great lengths can be taken to thoroughly define standards and processes, these activities are subject to change over time. Impact analysis can be applied to variables to understand how they are created throughout the entire standardization process. When something changes anywhere in that process, it is a simple matter to fully understand the implications, or impact, of that change.

validation. During study data processing, the model metadata can also be used to verify that the data that has been created conforms to the standard model definition. Many of the business rules applied to the models can

be validated by using the metadata. The validation process can be implemented in the metadata, and added to any visual flow so that validation executes automatically as part of the standardization process.

metadata reporting. The SAS Metadata Server supports standard XML exchange through the SAS Metadata Interface, and this feature enables users to extract metadata and reformat it into meaningful reports. For example, for each SDTM domain produced, the metadata is extracted and then used to document the library as well as create a report detailing the change records associated to each table. A sample report is shown in Figure 6.

CRT-DDS /define.xml. By using SAS Data Integration Studio with the CDISC data models to standardize data, much of the information that is required for the Case Report Tabulation Data Definition Specification document (CRT-DDS) is already available. To easily generate this document, it is a straightforward process to apply the metadata to create the CRT-DDS report, an example of which is shown in Figure 7.



Metadata Report: List of SDTM Domains

1
18:37 Tuesday, March 14, 2006

Version: 3.1.1
Publish Date: August 26, 2005

CLASS	Dataset	Description	Location	Structure	Purpose
Events	AE	Adverse Events - AE	AE.xpt	One record per event per subject	Tabulation
	DS	Disposition - DS	DS.xpt	One record per disposition status or protocol milestone per subject	Tabulation
Findings	SC	Subject Characteristics - SC	SC.xpt	One record per subject characteristic	Tabulation
	VS	Vital Signs - VS	VS.xpt	One record per vital sign measurement per subject	Tabulation
Interventions	CM	Concomitant Medications - CM	CM.xpt	One record per medication intervention episode per subject	Tabulation
Special Purpose	DM	Demographics - DM	DM.xpt	One record per subject	Tabulation
Special Purpose Relationship	SUPPDM	Supplemental Qualifiers - SUPPDM	SUPPDM.xpt	One record per qualifier value	Tabulation

Figure 6: Sample Metadata Report

Figure 7: Sample CRT-DDS (define.xml) Report

BEYOND CDISC

Although SAS strongly encourages the use of the CDISC standards, it is widely recognized that individual businesses might choose to use variants of these standards or their own organizational standards. These non-CDISC standards can be constructed, applied, and managed using the same core business processes outlined above. Virtually all of the benefits described are still relevant under these circumstances, but the burden is on the customer to define, implement, and validate these non-CDISC standards.

BUSINESS INTELLIGENCE

The business of bringing new therapies to market has never been more complex or more costly. After decades of increasing revenue, the life sciences research industries face financial threats brought on by limited research pipelines, increasing time to market, expiring patents, competition from generics, and increased safety review. It is more important than ever for research companies to thoroughly understand the scientific and operational nature of their business and gain insight into their completed, ongoing, and planned clinical trials.

Advanced analytics provides the ability to explore existing data stores and plan for the future. By being able to tackle sophisticated tasks—such as identifying hidden safety signals, forecasting and optimizing key business predictors, or more soundly building cost and revenue models—these more difficult, and more critical, business issues can be successfully addressed. It is not sufficient to focus analytics on only the routine data integration and safety/efficacy analytical results; advanced analytics must be brought to bear on the business of bringing new therapies to market.

INDUSTRY ARCHITECTURE FRAMEWORK

Traditional boundaries within and between life sciences organizations are disappearing. Research is evolving from independent silos into collaborative communities and into connected ecosystems of internal and external partners and stakeholders. The solutions previously described, as well as others on the SAS drawing board, will drive new efficiencies in how therapies are developed and how product portfolios are managed. More effective business processes, made possible by broader distribution of deeper knowledge, should drive significant new progress in improving human health. At the present time, however, information does not flow seamlessly across the organization, between business partners, or outside the industry.

For the first time since the introduction of information technology into clinical research, technologies are actually up to the task. Computing and networking standards are maturing. Processing horsepower and storage space are more powerful and affordable than ever. In short, the life sciences industries can exploit information technology in ways that were unimaginable even a few years ago.

That being said, there are still obstacles to be overcome. Interoperability between information systems is often rudimentary and initially starts with data standards. The life sciences industries have made great progress here with such standards as CDISC and Health Level 7 (HL7). However, true integration is much broader in nature.

THE SAS LIFE SCIENCE ARCHITECTURE

The SAS solutions to address these life sciences business issues discussed in this paper will provide this true integration. Although products may have historically been developed in silos, SAS is defining the architecture that will support all industry solutions. By leveraging the SAS Enterprise Intelligence Platform, core capabilities such as data integration, intelligence storage, and advanced analytics will be brought together within a single interoperable infrastructure. Industry components and services, such as SAS Drug Development, SAS Clinical Data Integration, and other offerings will be managed from a single point, reducing the administrative effort for maintenance of applications, users, and security. Data consistency will be assured because metadata is stored in a single metadata repository and shared across all industry solutions. Each element of the life sciences platform—from data integration, to storage, to analytics and reports—will add incremental value for the organization. But the most value will be obtained when these elements are individually optimized and integrated across the enterprise architecture.

Organizations that have adopted service-oriented architecture (SOA) will appreciate that the SAS Life Sciences Architecture can, through Web services and message-oriented middleware, integrate with a service-oriented architecture to add value and performance to an SOA fabric. SAS programs can be deployed as Web services, and the platform can consume third-party Web services as necessary.

ENSURING INTEGRITY

Life sciences data that is used to determine the safety and efficacy of new therapies must be handled in accordance with good industry practices, sound business practices, and Title 21 of the Code of Federal Regulations (21 CFR Part 11).

Industry solutions should automatically document data management activities and users interactions with this critical data. Integrated processes provide versioning, audit trails, and electronic signatures, as well as fully describe the relationships between process inputs, transformations, analyses, and results. This degree of transparency and documentation of clinical information processes ensures ongoing quality control and quality assurance, but moreover, makes it easy to confidently address inquires from regulatory agencies.

CONCLUSION

For nearly 30 years, SAS solutions have provided accurate, consistent, and reliable analysis of large volumes of information across all major industries. SAS is globally recognized as the industry leader in analytics, the de facto standard for clinical data analysis, and an FDA standard for electronic submissions. SAS solutions are used at more than 40,000 sites, including the U.S. Food and Drug Administration (FDA) and 100 percent of the Fortune 500 pharmaceutical companies.

The tools and solutions that SAS has provided to the life sciences industries to date are valued and valuable. SAS, however, is not complacent, extending our life sciences research capabilities to key business areas such as clinical

data integration, robust CDISC capabilities, and beyond. SAS' renewed focus on the life sciences research industries is important to SAS and will provide game-changing capabilities to the life sciences research industries.

REFERENCES

SAS Institute Inc. 2007. SAS Institute white paper. "Implementing CDISC data models in the SAS® Metadata Server." http://www.sas.com/apps/whitepaper/index.jsp?cid=3504.

SAS Institute Inc. 2007. SAS Institute white paper. "Interoperability and architecture for the life sciences industry." http://www.sas.com/apps/whitepaper/index.jsp?cid=3502.

Shipway, John. 2007. SAS Institute news feature. "The Era of Architecture." http://www.sas.com/news/feature/hls/sep07era3.html.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Dave Handelsman
SAS
Cary, NC 27513
919 531 0303
david.handelsman@sas.com
http://www.sas.com/industry/pharma/