

Clinical Research Content Management: A Single Version of the Collaborative Truth

Dave Handelsman, SAS, Cary, United States of America

ABSTRACT

The clinical trials process generates mountains of information across users, trials, departments, divisions and companies. While various point solutions are deployed across research companies, there is growing recognition that a living clinical research repository that spans organizational boundaries provides significant benefit in bringing new therapies to market.

Companies of all sizes can derive substantial business value across their organization, whether they rely on the growing repository as a collaborative workspace for their internal and external business units, or whether they build an accessible institutional memory to intelligently drive their business process forward.

This paper will describe the issues and opportunities associated with deploying a clinical research information management system that not only tames, but leverages, the clinical information mountain.

INTRODUCTION

The process of bringing new therapies to market is long (10-15 years), expensive (> \$900M) and complex. In many cases, research tasks are conducted across multiple organizations, multiple geographies and multiple systems. Although this research data must ultimately be aggregated and integrated as part of the submission process, the research content typically remains dispersed until the submission process begins.

By bringing this disparate information together earlier in the process, standardizing it around CDISC and providing the tools to leverage this rich information repository, the life sciences research organization can work collaboratively and constructively within a common workspace, and bring new therapies to market more effectively. Systems that are web-based, in particular, provide tremendous accessibility to not only users within a research organization, but experts external to the organization, further leveraging the content repository as well as the available technical resources.

Examples where a web-based collaborative repository is recognized as providing value to research companies include the following common situations:

- *Established research companies with staff that span multiple offices or geographies.*
The repository provides the means for the teams to work on a single, recognized, set of research information.
- *Research companies seeking to leverage external technical experts.*
The repository provides the means for the content, along with the company's computing network, to remain safe and secured, while external experts such as SAS programmers, statisticians or PK analysts perform their assigned tasks.
- *Strategically analyzing data across trials.*
By having all relevant data organized centrally, the pool of data is readily available for trial design planning, modeling and safety review.
- *In preparation for submission.*
Submission components can be generated and organized while research is ongoing, and be submission-ready at the conclusion of the research program.

PhUSE 2007

- *Smaller research organizations looking to partner with larger organizations.*
A web-based repository provides the means for the larger company to easily and securely access and review the relevant intellectual property, and as the relationship progresses, provides a neutral location for ongoing collaboration.
- *Research companies looking to efficiently deploy technical solutions.*
Pharmaceutical software solutions typically require significant information technology (IT) efforts to deploy and validate, and a hosted solution frees IT organizations from much of this time-consuming and expensive effort.

CONTROLS AND COMPLIANCE

Any system being developed to support collaborative efforts regarding clinical trials activities must address key elements of controls and compliance. These goals, as with all similar systems, are to fully document the integrity of the research and development content. At a basic level, this includes secure authentication and access measures that allow users to not only authenticate to the appropriate systems, but to have the correct level of access based upon their identity, their business organization, and the permissions granted to the objects themselves. While such controls are always important, they become indispensable when working within a collaborative environment where not all users are affiliated within a single organization, and the effect of sharing data inappropriately can have severe ramifications.

Authentication controls include requiring each user to maintain a unique set of identifying credentials, with password expirations set to an appropriate period of time. Once authenticated, access to objects or groups of objects can be managed via access-control lists or similar constructs. As shown in Figure 1, access may be assigned by individual or by group. In this case, user *dahand* has full access, user *davidsmith* has only read access and members of the *CRO Group* have read and write access.

Name	Type	Read	Write	Delete	Manage
davidsmith	User	✓			
Owner (dahand)	User	✓	✓	✓	✓
CRO Group	Group	✓	✓		
Users with Accounts	Group				

Figure 1. Access controls

It should additionally be possible to define an adjustable inheritance scheme, such that access granted to a folder may be cascaded to other objects residing beneath that folder, or so the cascade may be interrupted as necessary. Such an interruption could occur, for example, when it makes logical sense to group objects in a single hierarchy, but the access to a single data set (such as unblinded treatment information) needs to be restricted.

Access control schemes can quickly grow complex and often become over-engineered, which ultimately will present great difficulty in terms of both long-term maintenance, and true object security. Rules regarding inheritance and multiple affiliations should be simple to understand and implement, and err on the side of always granting more limited access. For example, if a user belongs to two groups, one with read/write access and the other with only read access, the actual access the user should have to an object associated with both groups is the lowest common denominator (in this case, read access). When access controls become overly complex, they often become ineffective due to the difficulties in fully understanding how permissions cascade down a hierarchy, especially when encountering one-off changes to the inherited permissions.

Throughout the data capture, data cleaning and data management stages, there are a significant number of process and electronic controls that ensure that research information is kept safe and unambiguous. Data entry systems

PhUSE 2007

record updates in various audit trails, and care is maintained to ensure that the data at the investigator site remains synchronized with the electronic systems at the sponsor. In effect, there is full traceability between what happens at the site and what resides in the final operational clinical database.

While these rigorous controls have historically been in place throughout the data management process, the controls around data as it leaves the data management system and becomes more widely available for collaboration – such as when it is transformed into analysis data sets and ultimately into statistical results – have only been process-based. Ironically, it is during these stages that wholesale data changes can be made – either erroneously or maliciously – with no automated documentation to capture the details regarding these changes. This is in stark contrast to the strict controls in place for other clinical research systems, which largely protect against individual data point updates.

A successful collaborative system must provide the necessary controls and compliance needed to document the integrity of the data throughout the transformation and analysis life cycle. Such a system would include references to all inputs and outputs associated with programmatic changes to the data as shown in Figure 2. In this example, 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 the referenced content can be easily viewed. Figure 3 shows the manifest history, with version 4 of the manifest corresponding to the content in Figure 2.

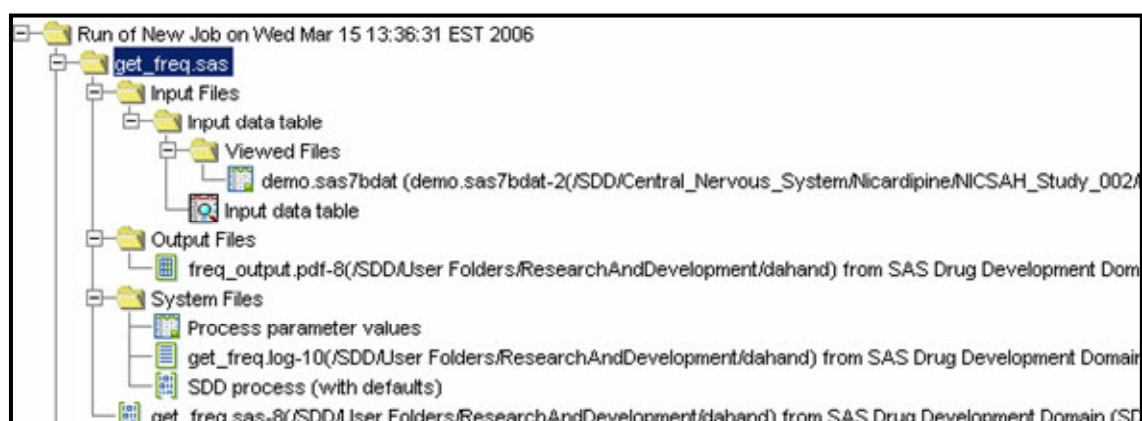


Figure 2. Traceable inputs and outputs

Actions					
	Version ▾	Size	Signature	Created	Created By
☐	6	18881		March 15, 2006 6:51:09 PM GMT	dahand
☐	5	18881		March 15, 2006 6:49:50 PM GMT	dahand
☐	4	22477		March 15, 2006 6:36:51 PM GMT	dahand
☐	3	22324		March 15, 2006 6:17:07 PM GMT	dahand
☐	2	21591		March 15, 2006 6:08:28 PM GMT	dahand
☐	1	22380		March 15, 2006 6:04:22 PM GMT	dahand

Figure 3. Content history

AVAILABILITY AND ACCESSIBILITY

Global companies, in/out licensing arrangements, outsourcing relationships and external analytical expertise require that interested and approved scientists have access to research content. By consolidating research activities within a centralized content repository, the logistics regarding the storage, retrieval and management of this research content can be greatly simplified. Users will no longer need to search among multiple systems (each typically requiring its own set of unique credentials) in order to find the information needed to answer their business questions.

Systems that operate with a zero-client footprint (such as exclusively through a web browser) provide the greatest ease of deployment. Because, typically, life sciences research companies lock down individual user workstations in order to ensure a validated work environment, applications that require a desktop installation add an undue burden to system accessibility. In many cases, it can take weeks (or months) to get a new software solution deployed within a research company, and this time and expense must factor into the business decisions regarding which software solutions to deploy. Additionally, where the goal is to collaborate on an as-needed basis, a zero-client solution is the only practical option – the time to secure, distribute and install additional licenses often becomes prohibitive in terms of addressing the business problem at hand.

Web-based systems provide the ability for 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 companies 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.

In many circumstances, systems deployed via a centralized hosted model provide even greater flexibility in terms of deployment and use. With this approach, the responsibility for hardware procurement, software installation, validation, backups and product updates falls to the hosting provider, 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 a zero-client application which requires no client installation, this approach becomes even more appealing.

Importantly, a hosted deployment often addresses the serious question of data ownership. In today's business environment, researchers are often looking for a 'neutral environment' in which to work, especially in cases of in/out licensing arrangements. By organizing research content in this neutral territory, and with full measures of controls and compliance, collaborating organizations can freely work together without risk of either group surrendering ownership of their intellectual property.

DATA, METADATA AND MASTER DATA MANAGEMENT

One of the greatest challenges in collaborative efforts is ensure that all users are working with the right version of the research content, and that everyone involved is confident that all participants are working with the correct version. Whether the issue being addressed is reviewing a final study report, or verifying an interim analysis, working from research content that is not fully qualified presents significant difficulties and wasted effort. The problem is compounded when external partners / users are relying on content copied from online systems.

The importance of master data management is only just gaining full understanding in the marketplace. In master data management, there is the recognition that key data exists in more than location, and that this key data should be managed consistently. Otherwise, when key data is brought together to answer business questions, manual intervention is required in order to determine the inconsistencies that arise. Master data management involves uniquely identifying each instance of a business element (investigator name, adverse event, etc.) and representing these instances using a standardized data model, providing a single, consistent view. This entails extracting key data from diverse operational environments to create a system of record and then establishing links to keep operational system files in synchronized. As systems surrender ownership of the master data, the master data management framework must provide fast access across all operational systems to master data without degrading operational performance.

With master data management, a request to associate investigator name with a report becomes a trivial exercise. A request is made within the system for investigator name. Even though that name might be stored in multiple locations, in different formats, the request for the information is processed regardless of the complexities according

PhUSE 2007

to the established business rules. These rules, for example, could state that the investigator name is always extracted from the investigator database, even if it appears in multiple places.

In the pharmaceutical industry, one example of master data management that has been historically addressed indirectly is reconciliation between adverse event databases and clinical trial databases. This reconciliation is typically labor-intensive, and erratic in terms of timing. The adverse event database may be much more current than the clinical database, and requests to – for example – list all adverse events must encompass not only interacting with multiple systems, but then reconciling differences. With master data management, such a request would include a reference for ‘reported adverse events’, and the business logic would automatically determine how to fulfill the business request.

As shown in Figure 4, various adverse event information is captured in both the clinical and serious adverse event databases. While this information is typically reconciled at the conclusion of the trial, queries regarding adverse events frequently happen during the conduct of the trial, and significant effort is typically involved in determining how to bring this data together for review. While programs can be written as necessary to perform these tasks, it makes more sense to employ a master data management approach with regard to this task such that a request for adverse events contains the business logic regarding how to reconcile this information efficiently and effectively, as well as to indicate issues with reconciliation (e.g., different dictionaries being used).

<i>Metadata / Data</i>	<i>Clinical Database</i>	<i>Serious Adverse Events Database</i>
<i>Patient ID variable</i>	<i>ptid</i>	<i>pt_id</i>
<i>Patient ID value</i>	<i>02-001</i>	<i>02001</i>
<i>Adverse event variable</i>	<i>ae_term</i>	<i>adverse_event</i>
<i>Adverse event verbatim</i>	<i>multiple headaches</i>	<i>headaches</i>
<i>Adverse event coded term</i>	<i>headache</i>	<i>headache</i>
<i>Adverse event dictionary</i>	<i>MEDDRA 10.0</i>	<i>MedDRA 9.0</i>
<i>Adverse event severity</i>	<i>severe</i>	<i>n/a</i>

Figure 4. Master data management

NOTIFICATION AND SEARCHING

With the volume of data doubling at a phenomenal pace, there can be great difficulty in staying current on relevant topics of interest. As such, a collaborative system should support the idea of user subscription and notifications. At a minimum, this system will provide the ability for a user to ‘subscribe’ to an object so they can be alerted when the object is changed. An example of such a system can be seen in Figure 5, where a user has subscribed to be notified when the object is deleted or electronically signed. In this system, notifications can be sent via either an internal messaging system or via a traditional e-mail system. Figure 6 indicates a typical e-mail notification showing the full description of the updated content.

Notify me when this is:

	Internally		By e-mail	
Updated	☐		☐	
Copied	☐	Select All	☐	Select All
Moved	☐		☐	
Deleted	☒	Clear All	☒	Clear All
Signed	☒		☒	

Figure 5. Subscription

From: David.handelsman@sas.com Sent: Mon 8/20/2007 3:05 PM

To: David Handelsman

Cc:

Subject: [SAS Drug Development Notification] The content of the entity updated

User dahand updated the contents of the entity "/SDD/Central Nervous_System/Nicardipine/SAS Global Forum 2007/Planned Safety and Efficacy Reports/Results/AE_Listing_Report.pdf".

Figure 6. Notification

PhUSE 2007

Systems should additionally allow users to subscribe to folders, such that when information is published to a folder, they are immediately alerted. If, for example, a folder is identified to store the production safety analyses, the trial's medical writer can subscribe to the folder in order to learn when new production analyses are available. Similarly, if the production analyses are regenerated, the medical writer would be alerted to the changes as well. In this way, not only is the team collaborating on a single version of the truth, but they are actively being notified of changes to that truth.

In addition to subscription and notification, searching plays a key role in successfully leveraging the information repository. Searches should be provided at multiple levels – basic metadata (e.g., name, date, etc.), advanced metadata (e.g., column names, research compound, trial phase, etc.), and by embedded content. Importantly, secure object access must be managed through all search capabilities in order to ensure that secured objects are not made available through content searches.

ELECTRONIC SIGNATURES

One goal of a collaborative content management system is the ability to review and approve content as necessary. There are various ways to accomplish this goal via electronic signatures and several mainstream products (such as Adobe) provide integrated electronic signature capability. These tools, however, are typically limited in their ability to sign objects stored in a proprietary structure, such as PDF. Ideally, it should be possible to sign any type of object within the collaborative environment, where objects can be documents, data sets, programs, and others.

Electronic signatures should behave similarly to ink-based signatures, in that they are applied to address a certain condition, where the reason for signature is clearly visible in the same context as the signature itself. As new versions of objects become available, the applied signatures are still relevant to the versions to which they are associated – as with a paper document. Signatures should not lose their meaning or association over time, and they should be attributable to only a single user. The implementation of the signature is less important than the requirements the signature meets.

Figure 7 indicates an object with multiple versions, of which the sixth version has been signed. The signature itself is displayed in Figure 8, which indicates that a specific user has signed a specific object for a reason of *Final quality control*. Electronic objects, just like paper objects, can often require more than one signature, and systems should support the ability to not only track these signatures but indicate a measure of completion.

Actions						
	Version ▾	Size	Signature	Version Creation Date	Created By	Check-in Comment
	11	499712		July 3, 2007 5:48:04 PM GMT	dahand	
	10	499712		June 12, 2007 4:24:59 PM GMT	dahand	
	9	499712		June 4, 2007 11:52:13 PM GMT	dahand	
	8	499712		June 4, 2007 7:53:24 PM GMT	dahand	
	7	499712		June 4, 2007 3:15:11 PM GMT	dahand	
	6	499712	signed	April 17, 2007 9:37:25 PM GMT	cahill	
	5	499712		April 13, 2007 7:35:57 PM GMT	dahand	

Figure 7. Signatures and versioning

Version Signed	User ID	Signed By*	Date Signed	Reason
6	dahand	Dave Handelsman	May 14, 2007 9:13:01 PM GMT	Final quality control.

Figure 8. Electronic signature

In Figure 9, two key capabilities are depicted. For both objects shown, electronic signatures have been applied to multiple versions as indicated by the paged checkmarks. As depicted, the current version of each object is signed, as is a previous version. If the latest version of the object was not signed, and a previous version had been signed, a different representation would be provided in order to show the need for a pending signature.

PhUSE 2007

Similarly, objects are at a different status regarding their signatures. The second object is shown having a complete set of signatures, while the first object is depicted with an incomplete set of signatures. This object requires a certain set of signatures to be considered fully signed, and only a partial set of signatures is available as shown in Figure 10.

☐	Name	Signed	E Sign Status
☐	ae_demog.sas7bdat		
☐	hema_demog.sas7bdat		

Figure 9. Signature status

User Name	Date Signed	Required Signer
David Handelsman (dahand)	August 20, 2007 7:19:42 PM GMT	✓
David Smith (davidsmith)		✓

Figure 10. Partially completed signatures

Additionally, electronic signatures should be implemented such when the object is transferred from the system, the associated signature is transferred as well. This can be fairly complex to implement, since the signatures should be agnostic with regard to object type, and not necessarily embedded in the object itself as with PDF. Such transfers are possible, however, and serve to mimic ink-based signatures, which are always associated with the signed object regardless of the location of the object.

AGGREGATION AND APPLICATION

A collaborative workspace provides tremendous value in terms of working more efficiently and effectively within a research organization. It additionally provides tremendous opportunity in altering business processes such that the business of conducting clinical trials and submitting results to regulatory agencies is fundamentally changed. These business process changes can be seen especially clearly in the areas of trial design and planning, early safety detection and submission preparation.

TRIAL DESIGN

By building an institutional memory of clinical research data and metadata, the process of assessing what types of trials to run, and which designs will be the most successful becomes more straightforward. Rather than having to dredge through data stored under different structures on dispersed systems, all of the available data is available in a single centralized location. At a basic level, inclusion / exclusion criteria can be reconciled with screening failures, and newly proposed criteria can be simulated based upon the body of information that has been aggregated in order to determine detrimental patient recruitment effects.

The aggregated information additionally provides the ability to better design trials in terms of collecting information that will be most relevant to analysis, data that will be the least error-prone and trials that run for an optimized duration. Given the extraordinary cost associated with bringing new therapies to market, the ability to better design a single trial can reduce costs considerably. The ability to avoid running trials that have flawed designs can reduce costs even more, with the added benefit of reducing time to market.

EARLY SAFETY DETECTION

For all therapies under development and already approved, safety is a key consideration. Being able to review safety data in aggregate, across all studies, provides a much better opportunity to identify safety concerns than investigating safety issues within a single study. Again, by having all of the data appropriately structured and available, more sophisticated safety reviews can be conducted, and these can be conducted more efficiently. Preparation of annual safety reports becomes a simple and straightforward task.

It is important to recognize that safety evaluations can also include analyses that span not only research compounds, but classes or therapeutic areas. By having the data readily available, the ability to look for hidden trends becomes significantly more practical.

PhUSE 2007

SUBMISSION PREPARATION

At most research organizations, the process of building the electronic submission does not begin until the end of the research process. At that point, considerable effort is spent locating, organizing and packaging the research content for submission to the regulatory agencies. This effort often spans offices, companies and geographies, and almost always contains unexpected surprises with regard to missing information, or protocols that were started, discontinued and forgotten. By aggregating the research content on an ongoing basis, it is also possible to pre-build the submission as the research program progresses. At the conclusion of that research program, the process of packaging the submission becomes simple and straightforward, with the vast majority of the assembly and organization process already completed.

CONCLUSION

Organizational silos, disparate data and distributed outsourcing networks create a variety of clinical research issues that can be directly addressed with industry-tuned content management and collaboration systems. These systems allow research and development efforts to be synchronized with less effort and more consistency, while at the same time maximizing resource utilization and domain expertise. When coupled with embedded analytics, such as those that can be applied to trial design, safety analyses and submission preparation, a collaborative environment to manage research content can enable true business process improvements in bringing new therapies to market.

REFERENCES

Master Data and Master Data Management: An Introduction, David Loshin, Knowledge Integrity, Inc
(<http://www.sas.com/events/cm/98829/index.html>)

CONTACT INFORMATION

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

Author:	David Handelsman
Company:	SAS
Address:	SAS Campus Drive Cary, North Carolina, United States
Phone:	919 531 0303
Email:	david.handelsman@sas.com
Web:	http://www.sas.com/industry/pharma/

Brand and product names are trademarks of their respective companies.