

MDMAP

An innovative application utilized in the management of clinical trial metadata

Gregory Ridge and Neeru Bhardwaj
The sanofi-aventis Group, Malvern, PA

Abstract

Since 1999, Sanofi-Synthelabo has delivered all of its Item 11 components for a new drug application to the FDA in electronic format. While the product delivered was compliant with the guidance set forth by the agency, the technologies and methodologies used by the company varied from project to project. In 2003, a diverse working group was formed to standardize the method used by the Biostatistics organization when producing an electronic submission. This group's primary goals were the standardization and automation of the processes that Biostatistics used worldwide to capture, manage and maintain the clinical trial metadata.

After extensive development and validation, MDMAP (Meta Data Management and Publishing) was deployed to the users worldwide. This application, developed in SAS/IntrNet® and the J2EE® framework, has streamlined the sometimes-cumbersome activity of collecting and managing clinical trial metadata. MDMAP implements a robust, scalable and secure enterprise solution that gives users complete functionality as they maintain and support their clinical metadata. This paper will focus on the complex user requirements placed on the development team, and the challenges they faced in achieving deployment to users crossing several clinical research sites.

Introduction

For many organizations, the term electronic submission (ESUB) sends shivers down their back. After all, it is another component to the already demanding procedure of submitting a drug application to the FDA. Initially, Sanofi-Synthelabo too was daunted by the prospect of submitting a dossier in electronic format, but with a few successful submissions under our belt, the requirements for an ESUB soon became routine. What was not routine or standard was the method used in the production of the Item 11 deliverables. Sanofi-Synthelabo used tools such as MS Word®, Excel®, Access®, ISI Toolbox® etc., in the management and production of an ESUB. Over the course of several electronic submissions, we utilized the previous technologies, or various combinations of them, in the production of the final deliverables. In all cases the environment deployed for the management and production of the ESUB package was less than optimal. Limited security and no audit trail were assigned to the process. By 2003, it was decided by department management that a single robust application needed to be developed to manage the process in which the Biostatistics organization could produce its electronic submissions. Thus began the saga of the MDMAP project.

What is an Electronic Submission??

For those unfamiliar with deliverables for the Item 11 section of an electronic submission, it is perhaps best now to identify the required components of an ESUB package. In the eyes of Sanofi-Synthelabo, there are three key deliverables; DEFINE.PDF, a blank study CRF annotated to the submission package and the actual submission database created in transport format. Additional deliverables such as patient profiles are outside the scope of MDMAP. Of the three components, DEFINE.PDF is the most challenging and complex to manage. This document contains the metadata that supports the actual submission data. Some people often ask; what is metadata? Meta data is simply data describing data. Think of it as simply a SAS Proc Contents with additional key information added. Additional details on this key deliverable will be discussed later in the paper. The annotated CRF is key because this document allows FDA reviewers the ability to identify data collected during the clinical trial with data stored in the submission database. The production of this component has typically been an exhaustive manual effort by either statistical programming or clinical data management. The final key deliverable, creation of the transport files, is typically the most routine of all deliverables, however issues such as file size and audit trail can make this effort somewhat challenging.

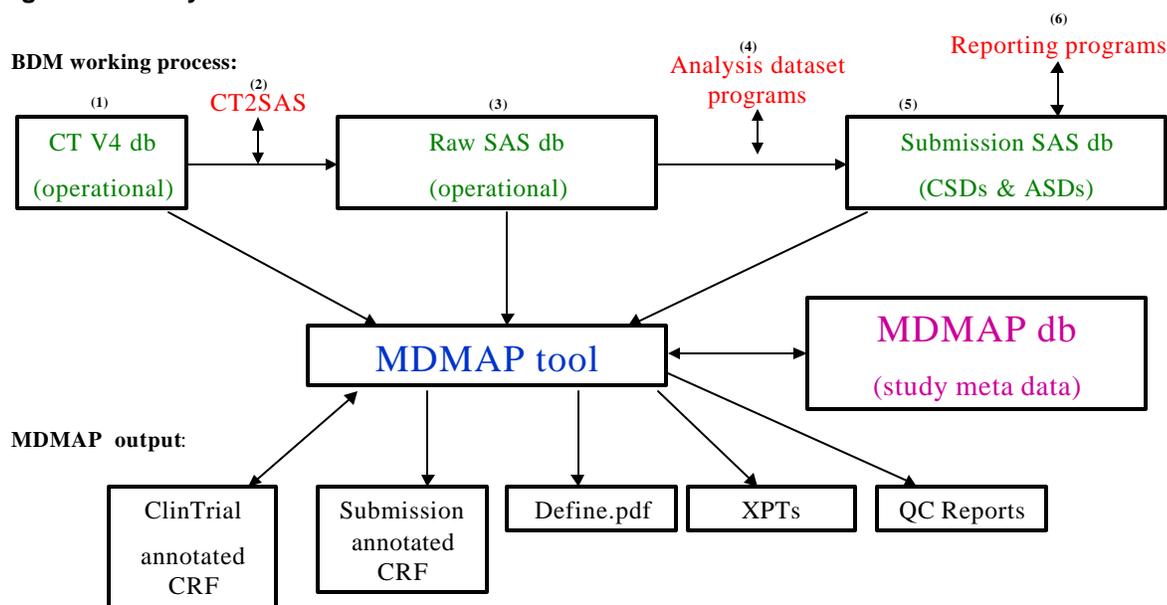
Application Requirements

From the beginning of the project, the working group assigned to MDMAP (**Meta Data Management and Publishing**) had an excellent understanding of the requirements of the project. This can be attributed to the fact that we had already successfully submitted several dossiers electronically. While those submissions were done manually, the requirements for each were identical and this attribute gave us a solid platform to base the MDMAP project. The primary objectives the working group had for the application were:

- Standardize and automate the process that Biostat and CMPK use worldwide to capture, manage and maintain the Clinical Trial Meta Data.
- Final Publishing capability (in define.pdf format), including the capability of identifying selected datasets/variables in the output.
- Improve and streamline submission CRF annotation.
- Standardize and audit trail the creation of XPT submission transport files.
- Develop an application compliant with Sanofi-Synthelabo's interpretation of Regulatory Agency Guidance. (21CFR Part 11)
- Create components of item 11 of electronic submission in a secure environment ready for delivery to Regulatory Operations.

With the objectives clearly in mind, the working group began development of a detailed user requirements document identifying all aspects of the application. The focus of this document revolves around a simple diagram that represents the workflow of the organization. See Figure 1:

Figure 1 – Analysis Workflow



Existing SAS code/applications

Existing SAS and Oracle databases

Proposed application managing processing of meta data

Proposed meta database

With the workflow process well understood, the development of the requirements for the application began to fall into place. We knew that the application would need to have access to both Oracle and SAS data. Oracle is the database format utilized by the CDM department for data collection and cleaning and SAS, of course, is the database utilized by Biostats in the analysis and reporting of study results. SAS data utilized in the development of a clinical trial falls under two categories, raw and analysis. Raw data is simply the Clinical Trial data extracted from Oracle into SAS. Analysis data consists of a tremendous number of derived variables in addition to the original source or raw data. The application would need access to Oracle, as the tool could be used by CDM prior to the execution of a SAS process, CT2SAS, which extracts clinical operational data into SAS. CDM is responsible for producing an annotated blank CRF. This document is used by both Data Management internally and Biostatistic's as they prepare the creation of submission analysis datasets. We knew the application would have to have security and be able to limit access to study metadata to appropriate users. We knew the application would have to have full editing functionality as the users developed and managed the study metadata. Since we had already prepared and submitted previous electronic submissions, users would expect the same level of editing capability they had in the manual development days. This meant access to MS Word type functionality including the ability to write and manage text, incorporate graphics and bookmark to external information when appropriate. We knew the application had to have the ability to generate the final deliverables including the creation of DEFINE.PDF, XPT files and the submission annotated CRF. And finally, we knew the application had

to give the users access to a multitude of audit trail reports. This included information ranging from discrepancies between the metadata actually stored in SAS versus the metadata stored in MDMAP to an exact audit trail recording the actions and status of all domains and variables being managed in the application.

Now that we knew what the application had to do, all we had to do was develop it!!!

Application Features

MDMAP is an internally based, intranet product that has a web application feel to it, unlike other department tools developed in SAS with AF/Frame. Once launched and access has been approved and authorized, the user is prompted with a screen identifying five options. These options are the backbone of MDMAP. They include:

- Administration.
- Study Setup
- Output
- Reports
- CRF Annotation

The blueprint for MDMAP metadata storage philosophies is identical to that of the SAS environment for which it describes. The Biostatistic's organization at Sanofi-Synthelabo utilizes a directory structure keying on three primary variables. These are:

- Compound (Internal drug number)
- Study (Unique internal study number)
- Analysis (Internal analysis type)

The concept of three levels and ability to have multiple analysis levels within a Compound and Study means the user has the freedom to version the metadata. One restricting factor however is that the structure for the metadata within MDMAP must exist on the SAS data server. Therefore, there must be a direct correlation between the data and metadata in terms of storage methodology.

Administration

The administration option gives authorized users access to control admission to study metadata. The controls are set to the Compound, Study and Analysis level, which means MDMAP is able to restrict access at any or all levels of study data. Access to the MDMAP application is profile based. Infrastructure support personnel are given the highest profile. They are responsible for giving access to Compounds to all users. Therapeutic area lead programmers are given access to administer to the study level. Project statisticians and programmers are given access to the analysis levels. Study programmers and statisticians have no administration authority. Users that have administration privileges can give other users access to the study metadata at their limit of the profile. By giving the authority to control access to the metadata to the users, (beyond access to compounds), the departments infrastructure team does not have to get involved with extensive administration activities. They can focus on issues such as dictionary management and database reconciliation.

Speaking of dictionary management, another feature of the Administration option is the supervision of department dictionaries. Sanofi-Synthelabo adopted the CDISC philosophy of data standards and we keep and manage the metadata information for these dictionaries within MDMAP. Currently CDISC V2.3, which focuses on safety data alone, is being maintained in the application but the organization plans to adopt the V3.1.1 standard as early as 4th quarter 2005. This database will be managed within the application giving users access to both CDISC V2 and V3 dictionaries. Access to dictionaries is limited to the department functions group whose responsibilities include standardization of Biostatistic's technologies and methodologies. The primary usage of the dictionaries within MDMAP is the pre-population of study metadata prior database development.

Users also have the ability to lock and unlock an analysis within the application. This feature allows project teams the ability to shut down an analysis from further update. Once an analysis is locked, only team members with appropriate privileges and profiles can unlock it for further modification. All status information, of course, is tracked in the audit trail.

Study Setup

The primary feature of MDMAP is the ability to setup and manipulate study metadata. This is accomplished via the Study Setup option. This selection has five options available:

- Import from Dictionary
- Import form Template

- Import from SAS db
- Import SAS codelist
- Edit Study Metadata

Users have the ability to pre-populate the metadata for a particular study three ways. They are via department dictionaries, existing studies in MDMAP or SAS metadata acquired from the SAS servers. The most common method of pre-population is via SAS db. This is simply the metadata obtained via a SAS Proc Contents imported into MDMAP for a particular analysis. In order to utilize this option, a SAS database with valid metadata must exist prior to the import. We have had cases where project programmers have developed dummy SAS analysis dataset containing no data but simply the mandatory and expected analysis variables for a domain in order to populate the meta database. This allowed the user to utilize the from SAS db option to pre-populate most of the metadata fields in MDMAP. The user would need to enter the comments section of the database of course, but most of the required metadata would be obtained via SAS.

Another method to pre-populate the metadata for an analysis is from a template. This is simply copying an existing study being managed in MDMAP into a new analysis level. This option is used commonly when you go from one study to another within the same compound. Once the copy is accomplished, all that needs to be done is the addition or subtraction of required variables and the modification of any domain specific metadata.

As mentioned previously, MDMAP manages department dictionaries containing CDISC compliant database standards. This allows users the ability to pre-populate the metadata for a study via these dictionaries. Users have the ability to specify specific desired domains or datasets to import. They can import all variables, just mandatory variables, or some subset of their choosing. When importing from dictionaries, users will also be given a default data definition or comment for the variables imported. This definition corresponds to department or industry standards for that element.

One of the fields displayed on the DEFINE.PDF document is the Codes information. It is very common in the pharmaceutical industry to receive data from CDM in codelisted format. That is, data which is in either numeric or character codes with a SAS format attached to give the user the ability to print the decoded or textual value. As specified in the 1999 FDA guideline, sponsors must submit the decoded value at a minimum, with the actual coded value being optional. Sanofi-Synthelabo adopted the philosophy that dictated we would submit both the code and decode for all formatted variables in a submission database. The coded value would drop the SAS format. To print the allowable/desired values for the coded variables, MDMAP allows users to specify whether the format catalog values or actual data values found in the domain will be printed. Users specify which type of data is to be printed in the DEFINE.PDF via the EDIT STUDY option. In order to acquire the values that will be printed on the document, MDMAP must reconcile with the SAS server, getting needed SAS format and/or data values respectfully. Once these values are acquired, MDMAP creates a hyperlink to be displayed in the DEFINE.PDF under the CODES column. If a user clicks on the hyperlink, the decoded values will appear, complaint to guidance.

The most advanced and complex feature of MDMAP is the ability to create, edit and modify study metadata. Accessing the metadata in MDMAP is accomplished via the EDIT STUDY option. Users have the ability to create or modify the following elements:

- Dataset Name
- Dataset Label
- Dataset splitting (FDA 100MB size limit)
- Dataset CRF page number
- Dataset variable Sort order
- Dataset comments
- General comments

- Variable Name
- Variable Label
- Variable Type (C / N)
- Variable Length
- Variable codelist information (data driven / format driven)
- Variable note information
- Variable submission CRF page number
- Variable dataset link

- Variable comments

Most of the information listed above is fairly straightforward and can be found in a typical SAS Proc Contents. There are a few elements however, that are not routine and must be created and managed via innovative methods. These are the dataset comments, general comments, variable comments and notes. Historically, Sanofi-Synthelabo has provided the FDA with highly detailed comments and observations regarding the datasets and variables in the submission. Since the format of the DEFINE.PDF does not lend itself to extensive text writing due to it's design and format, we have incorporated the use of comments and notes linked to the main document via hyperlinks. Users have the ability to develop extensive and detailed notes and comments regarding the datasets and variables in the submission and do not have to be concerned with the size constraints imposed by the columnar format of the DEFINE.PDF. When a user develops a note or comment and links it to a variable or dataset, the DEFINE.PDF indicates a link or bookmark. When a reviewer or statistician clicks the link, they are automatically transported to the appropriate section of the submission dossier that contains the text for that element. This feature allows the staff to be very descriptive when detailing datasets, generic comments regarding the study or specific variables within a domain.

One of the most unique and innovative features of MDMAP is its ability to incorporate advanced editing capabilities when producing comments and notes. As a primary goal of the project, giving users the ability to develop and maintain detailed and advanced comments containing highly descriptive text was paramount. As a result, the developer needed to incorporate MS Word type editing capabilities within the application. In it's final design, MDMAP allows users the ability to enter text with the following options:

- Create defined paragraphs
- Underline
- Bold
- Number Lists
- Bullet Lists
- Insert text from external files developed in MS Word (notes/general /dataset comments)
- Insert graphics from external files developed in MS PowerPoint®, etc. (notes)

Output

MDMAP has the ability to publish study metadata that it stores for an analysis. The publishing options available are:

- Create DEFINE.PDF
- Create XPT files
- Final Publishing Capabilities

Creation of the DEFINE.PDF is important for several reasons. First, it is used by the programming and statistical team throughout the development of a study. All programmers and statisticians utilize Proc Contents extensively in the development of required databases and reports. Since the DEFINE.PDF is a more advanced version of the common SAS contents, we utilize this file in its place. The primary and ultimate goal of the DEFINE.PDF is with the reviewers and statisticians at the FDA. This output gives the staff the opportunity to evaluate and QC the final deliverable prior to the final publish option. To review a sample of a Sanofi-Synthelabo DEFINE.PDF please see Figure 2.

Figure 2 – Sample DEFINE.PDF

Datasets For Study EFC4505			
Dataset	No. of Obs.	Description of dataset	Location
CODE4505	1	CODE dataset * One record per patient	crtdatases/EFC4505 CODE4505.XPT

Study: EFC4505		CODE dataset CODE4505.xpt		
Variable	Label	Type	Codes	Comments
AGE	Age in years at screening	N		
CECITY	City	C		
CENAME	Centre name	C		
COMPLETE	Completers population	C		
COUNTRCD	Country code	N	COUNTRY	derived country decodes
COUNTRY	Country	C		

Creation of the XPT files is a fairly routine event when producing an ESUB package. The only limiting factor is the 100mb size limit restriction mandated by the FDA. Creating smaller subset SAS files for those domains that exceed the 100mb limit must be done on the SAS server first. The advantages of MDMAP creating the XPT files are the audit trail that is created and the ability to lock the analysis to prevent accidental deletion and/or overwrite. We also feel more confident that the production of the FDA deliverable is completed in a 21 CFR Part 11 compliant environment.

The ability to execute a complete final publish is a key element of the application. Final publishing includes the creation of the DEFINE.PDF, annotation of the submission CRF, creation of the XPT files for delivery to a client and the automatic transfer of all components to a secure network file share. Once again, all actions are tracked in a detailed audit trail per guidance.

Reports

MDMAP has the ability to create a large number of detailed QC and status reports. These reports are extremely useful for users of the tool to understand the status of the metadata and to reconcile any differences that may be found. They are also very helpful for project programmers in understanding the status of the project in terms of quality and completeness. By providing an audit trail for the application, we can track all operations and manipulations of metadata. Basically it allow us to understand who did what, when it was done, and why the action was needed to be performed. Sample reports available include:

- Audit trails for domains, variables, comments and notes including reason for action.
- Dictionary reports for domains, variables and versions
- QC reports for meta data discrepancies, missing attributes, XPT creation and import errors
- Status reports for all Compounds/Studies/Analyses

CRF Annotation

One of the truly unique features of MDMAP is the ability of the tool to automate the process of CRF annotation. Annotation is done by both CDM and Biostatistic's in the course of study development and analysis. CDM annotates the CRF to the operational data, that is the data stored in the Oracle database, and Biostats annotates to the reporting or submission database. The reason that the annotation of the CRF could be different between departments is due to the fact that the Biostatistics organization adopted CDISC data standards for its analysis database. Variables may have different names in the operational versus analysis database. Users need to have the blank CRF available in PDF format. Once the final page numbering of this document is understood and the page numbers have been entered for all domains and variables in the EDIT STUDY option, users can automatically have the CRF updated with the appropriate information. Both domain and variable names are annotated on the CRF. See Figure 3.

Figure 3 – sample annotated CRF

Physical examination		DATASET=DEMOGA	
Date of examination	day <input type="text"/> <input type="text"/> / month <input type="text"/> <input type="text"/> / year <input type="text"/> <input type="text"/> <input type="text"/> <small>U1 <PHYSDT></small>	time (24 hour)	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <small>U2 <PHYS TM></small>
Blood pressure	systolic <input type="text"/> <input type="text"/> / diastolic <input type="text"/> <input type="text"/> mm Hg <small>U3 <BPSYS></small> <small>U4 <BPDIA></small>		
Heart rate	<input type="text"/> <input type="text"/> beats per minute <small>U5 <HR></small>		
Weight	<input type="text"/> <input type="text"/> <input type="text"/> Kg <small>U6 <WGT></small>	Killip class <small>U7 <KILLIP></small>	<input type="checkbox"/> I <input type="checkbox"/> II <input checked="" type="checkbox"/> III <input type="checkbox"/> IV
Height	<input type="text"/> <input type="text"/> cm <small>U7 <HGT></small>	Atrial fibrillation <small>U8 <AF></small>	<input type="checkbox"/> yes <input type="checkbox"/> no
Abnormalities		DATASET=PHYSEXAM	
Skin	<small>U9 <SKINYN></small> yes <input type="checkbox"/> no <input type="checkbox"/> <small>U10</small>	please give details	<input type="text"/> <small><SKINSP></small>
Eyes	<small>U11 <EYESYN></small> <input type="checkbox"/> yes <input type="checkbox"/> no <input type="checkbox"/> <small>U12</small>		<input type="text"/> <small><EYESSP></small>
Ears/nose/throat	<small>U12 <EARSYN></small> <input type="checkbox"/> yes <input type="checkbox"/> no <input type="checkbox"/> <small>U13</small>		<input type="text"/> <small><EARSSP></small>

Technical considerations

The objective of the various technical considerations was to provide a very scalable, stable, standards-based platform for consolidating meta-data management around a single set of tools.

The heterogeneous environment creates a challenge for organizations seeking to streamline operations, control costs, and reduce operational complexity. This section of the paper describes the technical solution adopted by the technical team after the evaluation of other possible technical options.

Technical solutions

The main objective of the MDMAP architecture design was to provide a robust, scalable and secure enterprise solution based on the already available infrastructure within the organization.

The different technology/architecture was evaluated prior to finalizing the particular technical solution. Initially a client/server solution using the ADO technology to access SAS data was explored, but this solution had many disadvantages, which resulted in exploring a web-based solution.

The move towards web-based applications continues in all software arenas. What was before accessed using thick client-server models or monolithic batch systems, are now moving towards a distributed and lightweight web interfaces.

The advantages of a web-based architecture are many: minimal, if any, client side application footprint, quick and global access to multiple services residing on multiple remote systems, automatic software updates that do not require client side changes, multiple and concurrent access to these services, and interoperability when accessing differing services.

There are also disadvantages of web-based systems: network latency can affect the response time, and therefore the user experience when interacting with remote systems, unforeseen load by concurrent users can affect response times for all users, failover mechanisms need to be implemented to guarantee uninterrupted access to remote services, and schemes for data locking and simultaneous access have to be implemented.

MDMAP application Architecture

Architecture is crucial for the software development because it is the blueprint of a software system. It provides an overview of the composition and functionality of the given software system.

MDMAP application integrates various J2EE® technologies/components such as XML, JNDI, JDBC, etc. The MDMAP application uses the Struts framework, an open source framework for building web-based Java® application. The MDMAP application uses open source FOP (Formatting Object Processor), a Java API for converting the XML document to PDF. A Java pdf library Itext is used to generate the annotated case reports form.

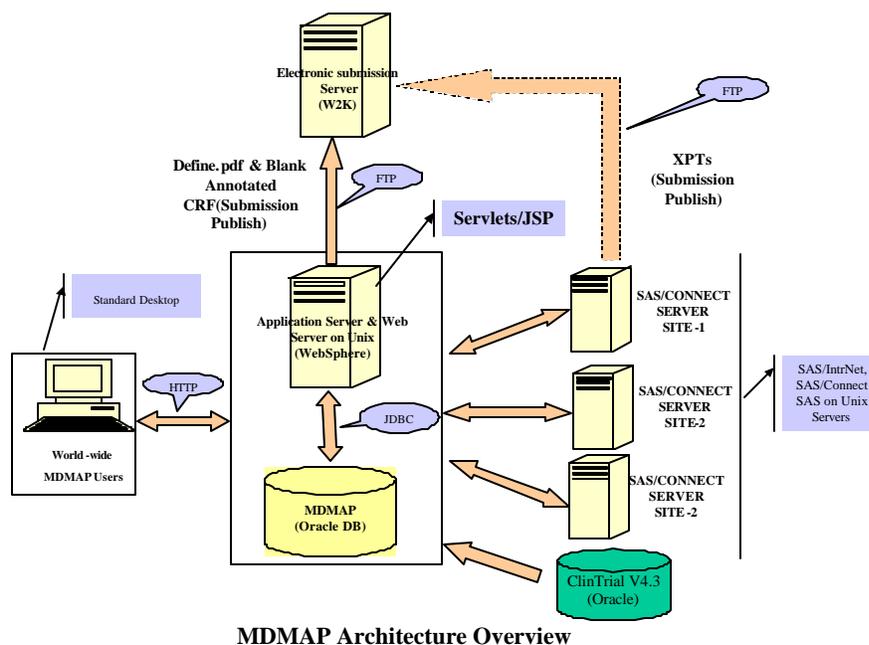
The MDMAP application uses the various component of SAS/IntrNet software to access information and services available from SAS servers

The application was built using a multi-tiered architecture. The business tier was implemented using Java classes, the presentation tier Servlets and JSP, and the client tier web browser. Oracle database is used for data persistence.

Communication with SAS/Server

MDMAP uses the SAS/CONNECT driver for Java, to start a SAS session, connect to that session, access existing SAS data, run procedures to access the SAS data, and retrieve the results. The SAS/CONNECT driver for Java program communicates with the SAS/CONNECT server using a connection daemon, the SAS spawner. The following diagram, Figure 4, shows the complete MDMAP architecture:

Figure 4 – MDMAP Architecture



Conclusions

MDMAP's goal was to be able to produce all electronic submission deliverables in a single package compliant within our comprehension of 21 CFR Part 11. In production since January 2005, we feel as though the application successfully automates the production of the required ESUB components in a secure environment. Users now do not need to have understanding of multiple technologies and processes when producing an electronic submission. Having a single application manage and publish the ESUB delivery process provides users a consistent and validated environment from which to work. The application is both very robust and designed in a method that promotes easy scalability should the user requirements or department population increase.

Acknowledgements

Sanofi-Synthelabo Inc. is a member of the sanofi-aventis Group.

SAS is a registered trademark of the SAS Institute, Inc. of Cary, North Carolina.

Microsoft Word, Access, Excel, PowerPoint are registered trademarks of the Microsoft Corporation

ORACLE is a registered trademark of the ORACLE Corporation

JAVA is a registered trademark of the Sun Microsystems.

J2EE is a registered trademark of the Sun Microsystems.

ISI Toolbox is a registered trademark of Image Solutions

Mr. Ridge would like to thank the IT project lead of this project, Neeru Bhardwaj, for her outstanding technical skills and dedication to the project. He would also like to thank the validation leader, Richard Spearman for his exhaustive efforts on the project.

Ms. Bhardwaj would like to take this opportunity to thank Gregory Ridge for providing the IT project team his SAS expertise and business knowledge. She would also like to thank Mr. Bill Friggle for providing the project team his expertise in ESUB area.

Finally, both of the authors would like to thank the users of the tool and their management for being extremely patient so the development team could do it right the first time!!

Contact Information

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

Author Names	Greg Ridge / Neeru Bhardwaj
Company	The sanofi-aventis Group
Address	9 Great Valley Parkway
City State ZIP	Malvern Pa, 19355
Work Phone:	610 889 6321 / 610 889 6336
Fax:	610 889 6932 / 610 889 8894
Email:	Gregory.ridge@sanofi-aventis.com Neeru.Bhardwaj@sanofi-aventis.com