

# Integrating SAS and Adobe Acrobat into Electronic FDA Submissions

Shylendra Kumar, Statsphere, Inc.  
Andy Siegel, Astra USA, Inc.  
Steve Light, Dataceutics, Inc.  
Paul Gilbert, DataCeutics, Inc.

## Abstract

Over the last two years the FDA Center for Drug Evaluation and Research (CDER) has released a series of guidelines for submission of regulatory documents in electronic format. On March 20, 1997, the Agency published the Electronic Records; Electronic Signatures regulation [21 CFR Part 11], which provides for the voluntary submission of parts or all of an application in electronic format without an accompanying paper copy. In September 1997 CDER released the *GUIDANCE FOR INDUSTRY for Archiving Submissions in Electronic format*. This guidance document provides details on submitting records and other documents in electronic format. According to this guidance, the electronic archival document submission should (1) display a clear, legible, easily viewed replica of the information that was originally on paper; (2) provide the ability to print an exact replica of each page as it would have been printed in a paper submission, including retaining fonts, special orientations, table formats and page numbering; (3) include a well-structured index and the ability to easily navigate through the submission; (4) offer the ability to electronically copy text and images; and (5) serve as a substitute for paper copies. These goals can be accomplished by using Adobe Acrobat Portable Document Format (PDF).

In this paper we describe the preparation of a PDF electronic document composed of reports, CRF images, patient data listings, figures and summary tables using customized SAS programs and Acrobat Exchange. The major goals of our electronic submission were: 1) Provide an exact replica of each page as it would have been printed in a paper submission. (2) Allow the reviewer to navigate through the electronic document using at least the standard table of contents currently provided with paper copy submissions. Navigational bookmarks are created within the document according to guidelines and a final CD image is produced for submission. The resulting electronic document can be viewed on any platform that has access to a CD-ROM and supports Adobe Acrobat Reader.

## Goal

The goal of the electronic submission was to deliver an exact replica of the paper submission, and that the table of contents could be used for navigating the submission. It was also essential that the submission required no special hardware for FDA reviewers and that little or no training was required to use the electronic document.

## Introduction

FDA regulations have traditionally specified that a variety of drug regulatory submissions be filed as paper documents. The regulations have required three paper copies of an application for marketing approval: a complete archival copy, a review copy, and a field copy. Since the mid 1980s Pharmaceutical Companies have been providing electronic information in some form (CANDAs and CAPLAs) along with the required paper copy of a New Drug Application (NDA). More recently, Biotech companies have begun submitting Product License Applications (PLA) with some electronic components.

In October 1996 a draft Guidance was proposed by the Food and Drug Administration (FDA) to develop a format for preparing electronic submissions.. This represents a significant milestone in the incorporation of electronic submissions, specifically allowing electronic-only submissions including archival copies. According to the draft guidance, for electronic submissions, Adobe Portable Document Format (PDF)<sup>1</sup> will be an acceptable document format for text and images. In September 1997 CDER released the *GUIDANCE FOR INDUSTRY for archiving submissions in electronic format*.

*(FDA use of specific products does not constitute an endorsement of that product.)*

This paper discusses the preparation of electronic documents based on the guidance. Adobe Acrobat Exchange and SAS programs were used to produce documents composed of text reports, CRF images, patient Data Listings, CTR summary tables and figures. Navigational tools were created within the document to allow the reviewer to navigate through the document using the standard table of contents.

## Why PDF?

Adobe Acrobat Portable Document Format, introduced in 1992, is a platform-independent file format which allows creation of documents from word processors, scanned document images, text or graphic stream files. PDF files are independent of the software used to create the documents and are entirely self contained - they retain original fonts, colors, page layout, etc. PDF files are viewable on many platforms using a freely distributable reader program.

PDF also allows for navigational bookmarks and hypertext links to enable easy movement through (and between) large documents. The files can also be searched and indexed.

The ability to create complex documents with components from various sources (word processors or desktop publishers, scanned images, etc.) which are readable on many popular platforms, without loss of original document formatting, has helped to make PDF an attractive format for electronic publishing.

### The eSUB Preparation Strategy

Since implementation of electronic submissions (eSUB) is an evolving practice, no single software tool or product is available to create entire submissions to our specifications. This was the challenge! We employed a combination of tools, using existing software applications including commercially available Acrobat Plug-ins, and developed SAS and MS Word macro programming to automate the whole process. The overall process is graphically presented in Figure. 1.

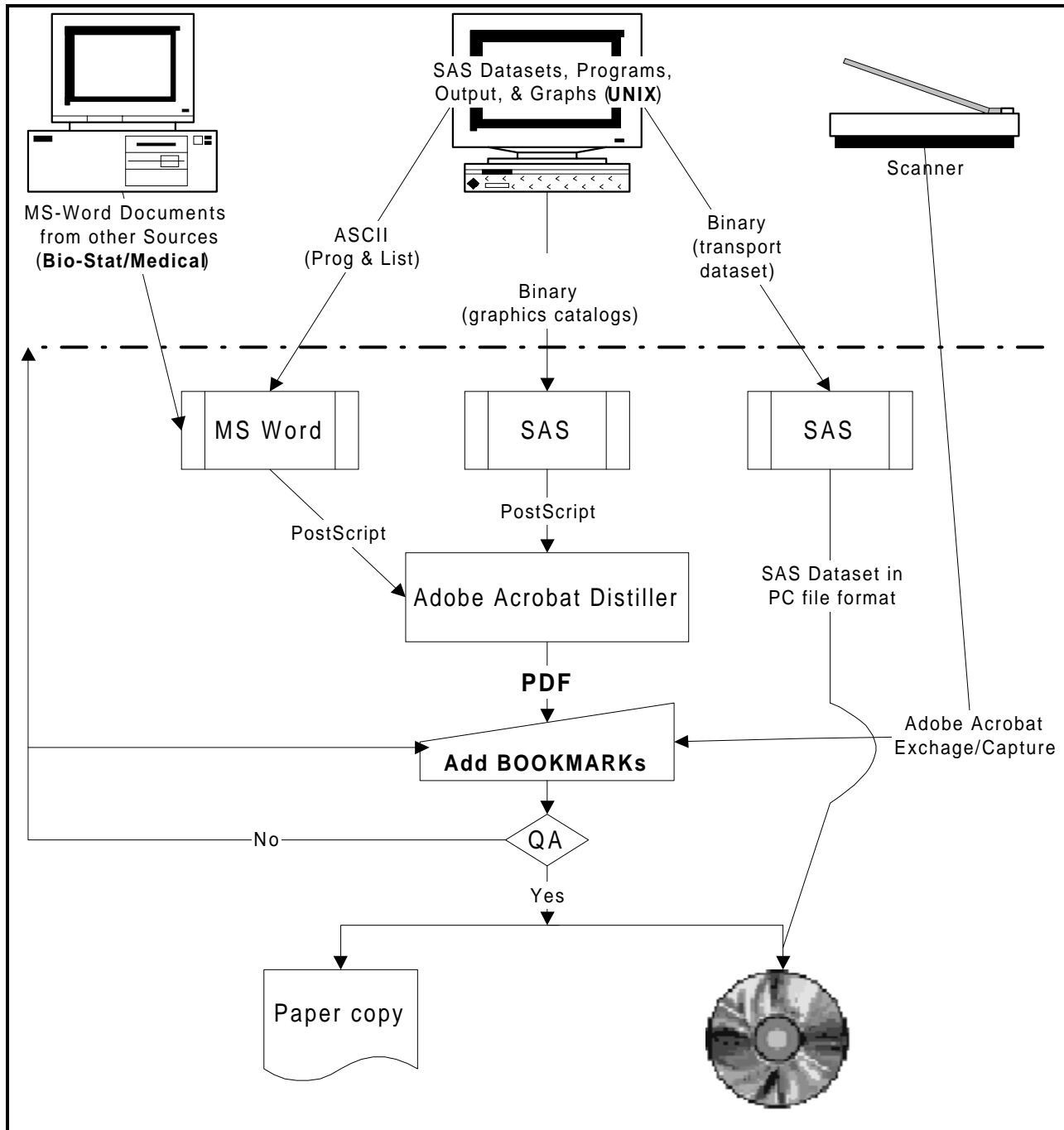


Figure 1

## Working Environment

The Astra USA SAS programming environment is configured in a distributed UNIX environment using IBM AIX workstations and an IBM RS/6000 mainframe. SAS data sets, programs and SAS output are stored in project-specific directories.

The electronic submission was prepared on a network of Pentium PCs, under Windows 95 and Windows NT 4.0. To build the electronic submission it was necessary to transfer SAS programs, output and data from the UNIX environment to the PC network.

## Document Sources and Types

The source document(s) are created in various departments and are stored in the following areas: 1. The SAS dataset, program, output, graphs are stored on IBM RS/6000. 2. The study protocol and other study report related documents come from medical department, these documents are on Windows NT network. 3. The CRFs (Case Report Forms) which are not available in electronic format are provided on paper by the Clinical Data Management department. There are a few other documents, such as Investigator CVs and signed report cover sheets, which are not available electronically.

1. **SAS datasets, programs, output and graphs:** The statistical analysis documentation, programs, CTR Tables and Figures, and datasets were downloaded from the UNIX network. Due to the variety of types of data downloaded, there was no single tool that would support the downloading and conversion of all these documents at once. In order to overcome this challenge we wrote a few simple SAS programs:

a. **SAS Programs and Table/PDL output files(ASCII):**The SAS program executed from Windows completed the following processing:

- Using FTP options the file containing the overall table of contents (SAS output) is read from UNIX.
- A temporary dataset with all program names and output file names used to create the Patient Data Listing and CTR Tables is created from the TOC file.
- All programs and output files referenced are then downloaded to the PC using SAS FTP functionality. Since Windows95/NT4.0 supports long file names, during the download process all the output files are renamed according their respective section number and Table/PDL numbers. With Section and Tables numbers (according to ICH guidelines) as the output file names it was then easy for us to organize these documents into their respective volumes at the end of the publishing process.
- A DDE call is then made to MS-Word and the downloaded output file is inserted into a new MS-Word document using a specially designed template. The MS-Word document is

then printed to a PostScript file. The PostScript file was selected as the better source for PDF generation due to font embedding problems encountered while using the PDF Writer. After the PostScript files are generated Adobe Acrobat Distiller is used to convert the files to PDF.

b. **CTR Figures(Binary):** The next program developed processes graphical output as follows:

- All graphs and figures are stored in UNIX catalogs, one catalog per protocol.
- The catalog is converted to transport format in UNIX and downloaded to the PC.
- The catalog is imported and all graphs replayed using a Postscript graphic driver supplied by SAS into a postscript file. Again, during this process all Figures were renamed according to their respective Section and Figure numbers. Once again Adobe Acrobat Distiller was used to create PDF files from these postscript files.

c. **SAS Datasets (Binary):** All datasets were put in a SAS transport file, downloaded and imported to PC format. Since various FDA reviewers may use other versions of the SAS software tools during the review process we provide the transport dataset and the v6.12 SAS datasets in SAS windows compatible format. Analysis dataset program documentation, SAS Programs, proc contents and listing of 10 observations are part of the statistical program documentation. All these items were compiled into one SAS output file using a standard macro within SAS work environment on UNIX. This file was downloaded and converted to PDF using the steps described in section (a) above.

d. **Macros and Formats:** Customized SAS Macros and the SAS program which creates the format catalog were includes both in ASCII and PDF format.

2. **MS-Word Documents from other Sources:** Most of these documents come from medical department and some from Biostatistics and clinical data management. Most of the medical writing, which includes clinical study reports, are stored electronically. These documents are converted into PDF by printing them through Acrobat Distiller from their original source such as MS Word and Excel.

3. **Scanning:** As mentioned earlier that, some of the documents are not available in electronic format. These documents include signed consent forms, Clinical Study Report, Investigator CVs, and Case Report Form (CRF). By FDA regulation a blank copy of the CRF each protocol should be included in the submission. Also, the completed CRF for

discontinued and deceased patients should be included in the submission.

All these documents are converted to PDF using scanning and Adobe Acrobat Exchange/Capture software

## Bookmarking

The PDF format provides several methods to add navigational tools to a file. Electronic submissions make extensive use of these features, and making the PDFs convenient for reviewers to use has proven to be possibly the greatest challenge in creating our eSUBs.

Bookmarks, represented by text in the overview area, act as hypertext links to specific areas in a PDF document or to other documents. While several Acrobat Plug-ins provide tools for creating bookmarks we ended up producing most bookmarks manually.

## Assembling Volumes

All documents were compiled into volumes according to regulatory guidelines and page numbers were added. Assembling all the PDF files into standard volumes was a major, labor-intensive task. The table of contents for all tables, figures and PDLs became handy in this process. Using Document composer(Adobe Plug-in from AMBIA) and WINBATCH we automated the assembling process.

## QA

Quality Assurance and reliability checks were performed at several stages of building the eSUB.

1. Before converting electronic files to PDF the electronic files were checked for accuracy, special characters, symbols, fonts, style and format to ensure compliance with Astra USA SOPs.
2. Scanned documents were checked for content, number of pages, order of pages, orientation, brightness, darkness, contrast, alignment etc.,
3. After the PDFs were created and relevant bookmarks were added, QA was done to ensure the bookmarks were pointed to the proper destination.
4. At the end a paper copy was printed from PDF and was checked against PDF to ensure the presence, location, page count, page numbers and the content.

### Problems encountered during the QA process:

- (1) After the bookmarked PDFs were compiled into volumes and paginated, the content of the PDF volume was checked against the paper copy originally submitted to the regulatory department. This was a manual process and the QA person had to go through the documents page by page. Several times we found missing or incorrect documents in the electronic copy because of various reasons: a last minute change was made to a document and the electronic version was not updated; during compilation some documents were accidentally not included; during downloading process certain pages of the documents lost their

margin settings and the result was either part of the information or certain pages were missing in PDF. The PDF volumes were updated either by downloading the file from the original source and converting it into PDF or by reproducing PDF from the originally downloaded file.

- (2) During the QA process we also found that some of the final PDF volume pages were not in the same order compared to the original paper copy. During the compilation process the user and placed the pieces in the wrong order. Since we cannot remove page numbers within a PDF file we could not move out of order pages to the right order. In order to resolve this problem we maintained two copies of volume in separate directories, one with page numbers and without page numbers. Whenever we found problems like this, the reordering was done in the unpaginated copy and then page numbers were added and saved as a final paginated copy.
- (3) When the paper copy was compared with the final PDF volume. we noticed :some of the special characters we used in our documents and other embedded fonts were not interpreted properly and the result was junk characters in the paper copy. This was due improper setup within Adobe PDF Writer or Distiller.

## CD Preparation

We used combination of software tools to build the CD. The reviewer has two options, a). install the contents of the CD using custom installation tool provided on the CD to their local HD or on to a network, or b). review all documents and datasets from CD without installing any software.

### Tools provided to browse the CD:

A copy of Adobe Acrobat Reader, Plug-ins and SAS Viewer. Also, copy of Acrobat reader and SAS Viewer were installed on the CD which will let the user to view the documents and datasets without installing additional software on his/her PC.

### The contents of the CD are as follows:

- \* CD viewer.
- \* Setup program – to install the contents to network or HD.
- \* PDF files organized by volumes under section numbers (CTR, attachments and appendices in multiple volumes)
- \* A MS-Word copy of CTR.
- \* Navigator file used by Aerial Navigator Plug-In.
- \* SAS Datasets in PC SAS format and XPORT format.
- \* SAS programs and Macros and documentation.
- \* Adobe Acrobat Reader and Navigator plug-in installed on CD.
- \* SAS Viewer Installed on CD.

## Conclusions

The electronic submission using PDF and SAS data was an acceptable and effective means of presenting information to the FDA. The submission we developed was easy to learn and use, since everything was in PDF format and convenient bookmarks and hyperlinks were provided for navigation. The submission included clinical summary reports, tables, data listings, graphs, and SAS programs. Statistical as well as the medical reviewers at the agency have found this system to meet their needs from a functional and user-friendly standpoint.

Our first eSUB was based on the Draft Guideline. We developed and learned many new techniques while creating this first submission. Because it was a new learning experience, the majority of the work done was a manual process. By the time we started our second project, the Final Guideline was published and we had the experience from the first project. Since there were no major changes in the Final Guideline, we were able to incorporate all the processing steps we had previously developed. The second submission we created took much less time than the first because we were able to replace many of the manual processes with automated systems.



## Profiles

**Shylendra Kumar**, is President of StatSphere, Inc., specializing in statistical data analysis, SAS/PH-Clinical implementation, Electronic Submission and SAS system administration. His six years of experience include academic research, data analysis, SAS/PH-Clinical implementation, eSUB, programming, network and hardware support for SAS software.

**Andy Siegel**, Assistant Director, Medical & Regulatory IT at Astra USA, is responsible for implementing SAS/PH-Clinical as the data review module of the Astra USA electronic submission. His fourteen years of experience with clinical information systems include SAS systems development and management, CANDA project development and management, and programming.

**Steve Light**, a Project Manager at DataCeutics, Inc., is responsible for SAS report macro development and Biostatistics/clinical SAS project management. His twelve years of experience with clinical information systems include SAS systems development and validation, NDA-driven project management, SAS/PH-Clinical implementation, programming and clinical data management.

**Paul Gilbert**, is Vice President of DataCeutics, Inc., a consulting group specializing in software solutions, system integration, programming and support in the areas of clinical data management and statistical reporting. His fifteen years' experience includes clinical data management, implementing and maintaining Clintrial, designing SAS based biostatistics reporting systems, managing SAS based NDA programming support, SAS/PH-Clinical implementation and CANDA integration.