<div align="center">

Paper 234-26

Standards for Electronic Submissions to FDA

Kay Obenshain, SAS® Institute Inc.

</div>

ABSTRACT

The FDA is moving towards its goal of paperless submissions by 2002. To meet its goal, the agency is defining standards for regulated industry to use when making regulatory submissions in electronic format. Standards for data are being defined on different levels, including content, structure, and format. The SAS® XPORT Transport format currently serves as an FDA standard format for data sets in electronic submissions. Other SAS technologies supporting electronic submission and review of data sets include the SAS System Viewer, the UODBC driver, and JMP®. This paper presents SAS Institute's role in supporting standards for submission data, and it is intended for users of all skill levels.

INTRODUCTION

There are three main areas covered in this paper. First, some driving forces, challenges, and expectations for defining standards are presented. This is followed by a description of the current role of SAS in submission standards. The third topic is a discussion about future directions for submission standards. Finally, there are some concluding remarks.

DEFINING STANDARDS
Driving forces

Do you remember the Mars Climate Orbiter spacecraft that was lost as it was inserted into orbit around Mars? The loss was attributed to a failure to recognize and correct for an error in transfer of information between the spacecraft team in Colorado and the mission navigation team in California. One team was using English units, while the other team was using Metric for information that was critical to maneuvers for placing the spacecraft in the proper Mars orbit.

Standards, such as standard units in the above example, can help prevent errors in transfer of information. Standards can also create abundance. The fax machine is an example of that.

Challenges

Consider some of the challenges in defining standards. There are legitimate differences that standards must accommodate. As an example, different countries have different requirements for regulatory submissions, and it would be desirable for a standard to accommodate or be consistent with requirements for more than one country. Another example would be in clinical data: "Stacked" or "Strung" describes different ways of arranging the same data. (Some people say "long & thin" or "short & wide." A medical reviewer or clinician might prefer the "strung" arrangement to view data across time, while a statistician may prefer the "stacked" arrangement to calculate descriptive and inferential statistics.

Uneven levels of technology present another challenge. Is your computer at work as nice as your kids' computer at home? Different levels of technology exist within companies, within the US, and especially between different countries.

In some cases, requirements for one FDA center differ from that for another center. The agency is trying to minimize differences whenever possible. Both the Center for Drug Evaluation and Review (CDER) and the Center for Biologics for Biologics Evaluation and Review (CBER) accept and archive data sets in XPORT format.

A company needs to reach consensus on a configuration to accommodate standards. For an international company, reaching consensus and implementing the configuration may be easier said than done.

There are costs associated with developing standards, such costs of meetings. Then there are costs associated with writing and implementing standards. When a new standard is defined, companies have to rewrite their SOPs and retrain personnel.

Communication has to occur between stakeholders in order to define and establish standards. Forums for communication between FDA and industry include PhRMA, PDA, DIA, and CDISC. While standards can accommodate some differences, at the same time it is important for individuals to remain flexible, adaptable, and open to change. Some individuals may have to change their way of doing things.

In defining standards it is important to use language that everyone can understand.

Standards need not always be leading edge technology. Hopefully, they won't be the least common denominator, but somewhere in-between.

Expectations

Standards need to be open so that anyone who wants to can read/write to them.

Regulatory agencies need industry buy-in, and technology providers also want to see this.

According to US law, the FDA must remain vendor neutral. FDA cannot adopt a standard that would force people to use a particular vendor.

Standards should be extensible so that vendors can build on them, and so that the standards can evolve as technology advances.

For a standard data set format, it is important to be able to process the data to/from common database systems.

Vendor support is necessary, and the support need not be only from the vendor that originally developed the technology.

In the instance of a standard format for a data set, it was important to have a free viewer to examine the data directly without invoking SAS or any specific product.

ROLE OF SAS IN SUBMISSION STANDARDS

FDA continues to define and update standards for various types of submissions to the agency, and to issue Guidance documents to assist companies making submissions in electronic format. SAS Institute's role in supporting submission standards as of January 2001 is described below.

SAS output

The output delivery and reporting system in SAS helps users get SAS output into the proper format for electronic submissions. Support for RTF is production in SAS Version 8.1 and beyond. PDF is production in Version 8.2.

UODBC driver

The SAS Universal ODBC driver helps users process SAS data sets and XPORT files to other file formats. The driver is an implementation of the open database connectivity (ODBC) standard that enables you to obtain read-only access to SAS data sources.

The SAS UODBC driver does not require the SAS System in order to access SAS data. It is useful when read-only access to SAS data is required, but the power of the data manipulation capabilities of the SAS ODBC driver is not needed.

JMP
JMP is a product from SAS that Medical Reviewers at the FDA use to explore safety data. Many people in industry use JMP also, especially the scientists. They like it because they can visualize their data and perform some analyses without doing any programming.

SAS Analytic Tools
Statisticians at FDA and industry use SAS/STAT® and other analytical tools for statistical analyses.

SAS System Viewer
The SAS System Viewer is analogous to Adobe's Reader. The Viewer is available for free at www.sas.com, and it is a quick way to view SAS data sets and XPORT files. The Viewer supports typical spreadsheet functions including selecting which columns to view, ordering, and freezing columns.

XPORT
XPORT is one of 2 transport formats SAS has developed for data sets. The other transport format is CPORT. XPORT was developed in the mid-80s as a way to move data sets from one computer to another, and it has not changed over the years. CPORT, on the other hand, has been updated as technology has advanced. SAS cannot publish specifications for CPORT because CPORT supports password-protected data sets.

You can download a set of macros for processing data sets to/from XPORT format for free. The macros take a subdirectory of data sets and process them to XPORT files, or vice-versa. If there is a format catalog in the subdirectory, it will be processed too. To download the macros, the free viewer, or specifications for XPORT, go to www.sas.com/fda-esub. There is additional information about XPORT there also.

Some reasons why XPORT works as an archival format for data sets are listed below.

- XPORT is easy for industry to use because most regulated companies already use SAS software.

- There are other technology providers whose products process data to/from XPORT.

- XPORT is well tested. Industry and FDA have successfully used it for years.

- The free viewer is important, for reasons described above.

- It is possible to process XPORT files to other common file formats.

- XPORT is just a text file.

- Specifications for XPORT are published at www.sas.com/fda-esub.

- XPORT works on any operating system on which SAS runs.

- Long-term support is important for an archival format. The life cycle of a drug can be 20-30 years or more. XPORT will live as long as anybody thinks there might be a need for it. It is a long-lived policy at SAS that new releases maintain compatibility with older versions of the SAS system.

FUTURE DIRECTIONS
While the XPORT format works as an FDA archival standard, technology has advanced since XPORT was written. Our customers have asked for a format that supports some of the newer technology.

The XPORT format does not allow for extensibility. For example, one cannot alter the size of variable names without altering the layout and requiring revisions to all programs that process XPORT.

It may be reasonable to consider using the evolving XML standards to define a transportable representation of a data set. XML is an industry standard that shows promise for defining a newer transport format. The W3C describes a standard way to encode data sets in XML.

At first glance, XML looks a little wordy and verbose. The size need not be a problem because public domain ZIP software is readily available that can compress XML nicely.

The current transport format can deal with labels assigned to values in data sets. The SAS terminology for these is formats. As a simple example the value M may be assigned a format Male, and F could be assigned Female. The formats can be saved in an "adjunct" file, and our macros support processing the format file along with the data sets. With XML, it is possible to wrap up the formats with the data.

XML can bridge incompatibilities of computer systems and vastly improve web applications, basically because XML tags say what the information is, not what it looks like. This facilitates more precise declarations of content and more meaningful search results across multiple platforms. Once data has been located, it can be handed off to other applications for further processing and viewing.

The Open Information Model, or OIM, is a set of meta data specifications that we believe could work well as a XML schema language for submission data. OIM was initiated by Microsoft, but now it is vendor neutral. It is maintained by the Meta Data Coalition, which will maintain and evolve OIM as a technology-independent and vendor-neutral meta data standard. Being upward compatible means that they cannot kill anything in OIM. So if the model is updated, the old model will still work. OIM is very rich, and it offers more than needed for submission meta data. But we could take the parts we need and use them- that is an advantage XML offers.

CONCLUSION
FDA is accepting and reviewing data sets in XPORT transport format as part of submissions from regulated industry. Electronic submissions provide efficiencies in the review process and eliminate tons of paper. In short, they are working.

Considerations for updating the standard archival format for data sets have been presented. Currently the CDISC submissions group is making good progress defining content for clinical data domains and associated meta data. When it is time to consider updating the format, XML offers promise.

Regulatory agencies, regulated industries, and technology providers are working to establish and evolve open standards for data. Where do we go from here? It is exciting to think of the possibilities. Let us all work together to improve the quality and efficiency of electronic submissions to regulatory agencies.

REFERENCES
FDA, *Guidance for Industry: Providing Regulatory Submissions in Electronic Format - General Considerations*, 1999.

FDA, *Guidance for Industry: Providing Regulatory Submissions in Electronic Format - NDAs*, 1999.

CONTACT INFORMATION

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

> Kay Obenshain
> SAS Institute Inc.
> SAS Campus Drive
> Cary, NC 27513
> Work Phone: 919-531-4594
> Fax: 919-677-4444
> Email: Kay.Obenshain@sas.com
> Web: www.sas.com

_____