

The Changing Nature of SAS Programming in the Pharmaceuticals Industry

Steven Light, DataCeutics, Inc.
Andy Siegel, Genzyme Corporation

ABSTRACT

In the last two years, regulatory initiatives and technology changes have coincided to significantly impact SAS programmers in the pharmaceuticals industry. The Food and Drug Administration (FDA) Guidance for Industry and Electronic Submissions standards (1,2,3) have established new standards and conventions to which we need to adhere. We are also confronted with CDISC (4) data standardization efforts and the maturing of technologies such as XML, desktop publishing and data warehousing. These changes should compel an examination of the strategies, methods and even the mission of SAS programming organizations.

Current trends and issues in the industry will be reviewed, with suggestions on how changes can be addressed.

INTRODUCTION

Over the course of the last several years a series of separate trends and events have transpired that have had or will have a significant impact on the planning, execution, documentation and delivery of statistical programming work that supports drug-related submissions to regulatory agencies, such as the FDA. This presentation will discuss these events and issues that are relevant to statistical programming in a clinical research environment.

For many years now, the nature of the work done by statistical programming organizations, primarily the preparation of summary tables, figures and listings to be included in regulatory submissions, has not changed dramatically. We received data for clinical studies, possibly write SAS programs to prepare summary datasets from the original study data, write SAS programs to build statistical tables, figures and listings (TFLs) reporting the study data. The TFLs are subsequently included in the body of a submission document. While the specific methods we use to reach this endpoint have evolved, for most organizations the fundamental process of preparing data and generating reports has not changed dramatically.

The tradition has been that programming organization 'own' the data sets they use for building TFLs, and the individual TFLs are the final 'product' produced. As a result of changes we will discuss, both of these traditions are no longer viable. We feel that this is a fundamental change that should be recognized by programming organizations, and that there is a need to review the strategies and methods that are currently being used.

CHANGING REGULATORY ENVIRONMENT

With the publication of the FDA Guidance for Industry (5), the expectations of regulators for how we maintain a 'controlled environment' have been made clear. Specific requirements for validation, configuration management, change control and security have been made clear. From our experience, some programming organizations have addressed these issues, many other programming organizations have a long way to go to become fully compliant.

ELECTRONIC SUBMISSIONS

The advent of electronic submission standards and the corresponding data standards have had a significant impact on statistical programming operations. We now have new deliverables: Case Report Tabulation (CRT) datasets, annotated Case Report Forms (CRFs), CRT dataset documentation (define.pdf and variable lists) and related Table of Contents (TOC) files. We also have different deliverable media. The CRTs are delivered as SAS transport data. The CRT dataset documentation is delivered in bookmarked and hyperlinked PDF files.

CRT data sets are intended to replace patient listings. We suggest that it makes sense that the TFL programming and analysis programming use the CRT data sets as the input data. This makes sense for the sake of operational efficiency and as a means to ensure that the CRT data sets are complete and fully useable. CRT data sets need to conform to a set of standards described in the Guidance documents. This means the 'ownership' of data sets has changed. Data sets no longer are created solely for programmer convenience, they are now designed for FDA Reviewer.

CRT data sets be designed as much with FDA Reviewer convenience in mind as for programming convenience. Patient identifiers need to be consistently formatted and unique across a submission. Coding of data values is de-emphasized in favor of more meaningful text values. Overall, datasets should be organized for ease of use, readability and to provide a reviewer 'friendly' format. This may conflict with programmer preferences and prevailing conventions for analysis data.

The regulators have also indicated a growing interest in the delivery of proposed specifications for CRT data sets, prior to an actual submission. This implies that

programming organizations need to be able to project the structure and content of their CRT datasets perhaps significantly in advance of actually creating the data sets themselves. This also implies that CRT data standards across projects can be of value.

Maturing Technology

At the same time that regulator-driven change as taken place, the maturation of document management systems and desktop publishing capability is starting to make inroads into the statistical programming arena.

Companies are approaching these technologies in very different ways; the point in the statistical programming process at which these technologies is integrated varies from one organization to the next. Programming organizations need to consider where they fit in and how they may alter their working processes.

Publishing technology can be integrated into statistical programming operations to the degree that the fundamental product of programming organizations is altered from a set of TFL text files or printouts, to an integrated set of study report appendices in PDF with complete sets of bookmarks, hyperlinks and related TOC files.

XML is an emerging standard for data interchange that is gaining wide acceptance. Future FDA standards for delivery of CRT data or other submission components may well specify XML as the format of choice. We anticipate that we will also begin to encounter delivery of external data in XML format. Programming organizations will need to build internal expertise in XML programming.

In a similar context, we see a trend developing toward use of data warehousing tools to automate and control processes of creation of CRT datasets and other programming outputs. The nature of trends already discussed, including increased emphasis on data standardization and requirements for documenting a controlled development environment, make these kinds of tools a logical extension of current working processes.

Conclusions

The trends described above lead us to the conclusion that need to reconsider our working processes and what our true deliverables really are. All signs point to a need to establish and maintain tight control of our programming environment, our data and the tools we use. Programming organizations need to manage the structure and contents of CRT data proactively, by adopting a set of data standards that are adhered to by programming staff. The FDA guidance and the CDISC submission data standard efforts provide some guidance in this area.

Our feeling is that the programming organizations will continue to be prodded for more and more information

earlier in the process. This means programming organizations need to have tightly controlled processes and need to be able to provide detailed information about analysis-level data well in advance of creating the datasets themselves.

Considering the increased requirements from the regulators and the changing technological environment we feel that this is a good time for programming organizations to evaluate their working processes, and to consider the nature of the final products created.

Suggestions

Programming organizations can take this opportunity to look at their position relative to regulatory issues, how they are prepared to meet regulatory requirements, and how the technology issues described above can be integrated into existing working processes. A few suggestions are listed below.

- Start with reviewing your position in regulatory compliance. If gaps exist, efforts to address your issues can incorporate additional change you may want to consider concurrently.
- If you have not done so already, establish standards for CRT data. Keep the emerging CDISC standards documents in mind when determining data standards.
- Consider the audience or ‘customer’ for CRT data. Data needs to meet the needs of programmers and of FDA reviewers. Recognize that many reviewers will not use SAS to work with data.
- Recognize the rising importance of metadata about your CRTs. Managing and using this information efficiently can impact on many aspects of your work.
- Use CRT data sets as the primary data for both analysis and TFL programming.
- Consider what the final deliverables should be. It is now reasonable to consider production of integrated, bookmarked and hyperlinked documents to be the deliverable rather than the traditional listing files or paper.
- Incorporate publishing steps and integration with a document management system into ongoing working processes. Treating them as ‘add-ons’ to existing processes will be inefficient.
- Begin to build internal knowledge of XML programming and data warehousing to stay current with industry trends.

REFERENCES

1. Providing Regulatory Submissions in Electronic Format--General Considerations, Food and Drug Administration, January 1999.
2. Providing Regulatory Submissions in Electronic

Format – NDAs, Food and Drug Administration,
January 1999.

3. Guidance for Industry Providing Regulatory Submissions to the Center for Biologics Evaluation and Research (CBER) in Electronic Format – Biologics Marketing Applications, Food and Drug Administration, November 1999
4. CDISC Submission Metadata Model, Version 1.1 April 30, 2000, Clinical Data Interchange Standards Consortium, Inc.
5. Guidance for Industry, Computerized Systems Used In Clinical Trials, Food and Drug Administration, April 1999

CONTACT INFORMATION

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

Steven Light
DataCeutics, Inc.
1610 Medcial Drive, Suite 208
Pottstown, PA 19464
Email: lights@dataceutics.com
Web: www.dataceutics.com

