Paper 208-2008

Standard Management Approved Report Templates
The SMART System

Gary Stevens, Biogen Idec Ltd, Maidenhead, UK

ABSTRACT

In the modern pharmaceutical business arena it is of paramount importance that any new drug or compound is brought to the market as soon as possible. Physicians and patients clamor for additional, more efficacious and safe treatments to meet the needs across a wide range of therapeutic areas. In response, pharmaceutical and biotechnology companies invest vast sums of money into research and development of new therapeutics. To ensure the success of a drug or compound this investment must be recouped and profit margins met in an extremely competitive market before therapeutic patents expire. Therefore it is essential, for the success of a pharmaceutical product to reach patients, that the clinical trial process is as quick and efficient as possible.

From a SAS© programming prospective, the realization that a sizeable portion of our output, namely safety analyses, is repeatable across studies/programs led Biogen Idec to develop, initially the Standard Management Approved Report Templates (SMART) tables, and subsequently the SMART system, which automates the production of these common outputs. The automation of this process brings with it fast reliable system for producing accurate output and hence a quicker and more efficient process helping further meet our aim of ensuring the success of Biogen Idec products. This paper aims to provide a description of the system and how it was implemented, developed and the thought process behind it.

INTRODUCTION

SMART TABLES

Approximately 6 years ago the recognition of the need for standardization of certain tables, common across all Biogen studies was discussed. It was identified that this project would concentrate on the safety reporting as this seemed to be nearly, if not exactly, identical across all studies. Initially, table and listings templates were constructed for Adverse Events, Demography and Laboratory Data. These were soon followed by Concomitant Medications, Subject Exposure and Subject Disposition Tables. This project was referred to as the SMART project and the outputs generated were referred to as the SMART tables. Much thought was put into how to automate the production of these SMART standard tables but none proved a total success. Programmers were having to make too many modifications to generate these tables to fit their input data and sometimes unnecessary modifications because they thought it was "better that way" and eventually the project seemed to lose its direction.

CDISC IMPLEMENTATION

Along with the SMART table project another concurrent initiative in biostatistics was the implementation of CDISC Submission Data Tabulation Model (SDTM) as our new standard for raw data. It was immediately obvious that this could enable us to develop a system using standard input in the form of CDISC v3.1 data. This standard "input" data across drug programs was one of the driving forces behind the SMART system. We could actually build a system if we could ensure standard "input" and standard "output".

SMART SYSTEM

In the summer of 2006 the programming team, with standardized CDISC SDTM raw data now available, determined that automation was possible. The original finalized SMART table shells were re-reviewed and it was decided to explore the idea of having an external vendor automate the process for safety reporting. An external vendor was identified from a number of proposals based on detailed requirements. Biogen Idec assigned a project lead and work began on a system that would suit our needs. The whole process from specifying requirements to the SMART system deliverable going live took around 6 months. The system was up and running in March 2007.

DEFINING THE REQUIREMENTS

INITIAL THOUGHTS

The first step was to define the general requirements for the system. The deliverable from the project was to be a system to generate SAS® code to produce safety tables known as SMART tables. These standard tables will be used across all protocols/programs. The deliverable was therefore to be a system, not tables or outputs based on specific data. The input data will be in CDISC SDTM v3.1 format. The system will generate analysis datasets as needed, as well as executable code to produce final output.

In developing the specifications it soon became apparent that many "Unspoken Assumptions" had to be documented. We take for granted that many specifics of tabulated output are obvious to those involved in the day-to-day production of such output but the external client, although proficient in SAS, had limited experience in our specific clinical requirements and output. Decisions such as handling of missing data, definitions of treatment emergent, other algorithms and imputation rules through to centering of columns in a table were all documented. This documentation had to be as detailed as possible to remove any ambiguity or uncertainty as to what was required. A certain amount of flexibility would also need to be added to the system.

STUDY DESIGN ASSUMPTIONS

The Reporting system was designed with the following study design assumptions.

- Parallel studies with up to 5 treatments and a total column.
- Crossover studies will not be included in the initial development. Reporting modules to handle crossover trials will be added as a future development.
- The reporting system will be designed to handle a single phase of a study, e.g. Treatment Phase. Studies with multiple phases would require programmers to subset out the phase of interest prior to running the SMART system.

CDISC SDTM

As previously stated the driving force behind the project of automation of the production of the standard outputs was the availability of standard input data. As Biogen Idec moved over to an Oracle Clinical® database system, minimal remapping of the data from the database facilitated CDISC Study Data Tabulation Model (SDTM) availability. This served two purposes, firstly as the preferred structure and format of data for submission to the FDA and secondly a standard structure that facilitates the automated system to work across all of our studies. At present the Derived Analysis Datasets created are not CDISC Analysis Data Model (ADaM) compliant.

IDENTIFICATION OF OUTPUT

The original SMART tables had been defined some time ago. Finalization of these standard outputs resulted in 39 standard Safety Reports. The breakdown was as follows:
- 18 Adverse Event Reports.
- 1 Demography Report.
- 2 Concomitant Medication Reports.
- 4 Laboratory Shift Reports.
- 7 Subject Disposition Reports.
- 7 Subject Exposure Reports.

THE SMART SYSTEM ARCHITECTURAL DESIGN

SMART SYSTEM DESIGN COMPONENTS

The SMART system consists of 6 distinct components.

SMART-START

This is a set-up module where the macro and program libraries are defined along with the global macro variables.

DERIVED DATASET CREATION

Set of SAS v8.2 programs/macros to create derived datasets from CDISC v3.1 input datasets based on the specification for each domain.

REPORT MODULE

Set of SAS v8.2 programs/macros to create tables/listings based on the requirements provided by Biogen Idec Ltd.

SMART EXECUTION MODULE

This module is the central component of the SMART system that will interact with the other components to ensure logical interface and execution.

PARAMETER SETUP MODULE

This module is an xml file that creates user entered parameter metadata that will provide report specific program id, program name, titles and footnotes.

KICK OFF MODULE

This module is a SAS macro that serves as the user interface to the system and allows the user to select and run specific reports with user supplied options.

DESIGN COMPONENTS DETAIL

CREATING THE ANALYSIS DATASETS

DCM Driver Setup

In order to execute the creation of derived datasets for the SMART reporting modules and to properly define libraries and study specific locations for the DCM input/output, the DCM driver will be created. This file is used to assign the location of raw study data (CDISC v3.1) libraries, derived dataset libraries and the global macro libraries.

The DCM Driver includes the following:

- General options
- SAS Autos
- Data directories
- Global macros
- Macro calls for DCMs

Derived Dataset Creation Module

DCMs are SAS Programs that generate derived analysis datasets for each domain based on a specific dataset specification document. The derived datasets provide the data used to generate standard reports based on the SMART requirements. The DCM is designed to provide flexibility for the customization of the study specific needs. A DCM is created for each derived dataset or domain that is specified.

Each DCM contains the following sections:

- Pre-processing Macro
- Core SAS code
- Post processing Macro
- Permanent dataset creation
- Diagnostic & debugging

The Pre-processing macro reads the raw study data from the input file and merges the domain with DPOPS.

This provides the input dataset for the Core SAS code where derivations and manipulation of data take place. In order to maintain integrity of the system, no modifications are to be made to the Core SAS code. If modifications are required then these should be added in the pre- or post-processing macros.

Diagnostic checks are built into the SMART system to allow the user to identify specific data issues that are pre-specified by domain in the derived dataset specifications. For the purpose of review and debugging, a data listing is provided for every data step of the SAS code. The data listings create a PROC PRINT output for each auxiliary work dataset created during data manipulation. These can be turned off during production runs.

The post-processing macro is a placeholder for any user defined post-processing. This would be used if the user needs to deviate from the standard algorithms or create additional variables that are study specific. This macro does not contain any code and will serve as an area where the user can interact with the system as necessary to make study specific modifications.

DERIVED DATASETS CREATED

DPOPS   This needs to be created first as it contains all the "Key" variables for each dataset. These key variables need to be applied to each derived dataset created. These key variables contain, amongst other variables, Treatment Labels and Population variables. This dataset also contains all demographic variables as well as calculated variables such as BMI.

DAE   This dataset contains all adverse event details along with all the common variables from DPOPS.

DCM   This dataset contains all concomitant medication details along with all the common variables from DPOPS.

DLBS   Contains results of Laboratory Tests used in creating standard laboratory shift tables along with the common variables from DPOPS.

DEX   Contains Treatment Exposure variables along with all the common variables from DPOPS.

DLBUTIL   Used to support Laboratory reports and provide flexibility; inclusion/exclusion of labs, reporting order, control of shift Table for Urinalysis.

DEXUTIL   Used for allowing flexibility when calculating compliance; assumes a fixed dose is given per treatment arm (e.g. 50mg bid); user provides the expected number of doses from protocol.

PRODUCTION OF REPORTS

Report Module

SMART Report Modules are a set of SAS programs designed to generate the tables and listings required. Before creating the reports for any domain the user must first create the DCM for that domain. The Report Module is called through the SMART engine and reads in the appropriate parameters from the Kick-Off Module and parameter metadata. The Report Module reads in the variables from the derived dataset and creates the desired report. The output includes executable SAS programs, SAS logs and report output (.lst and .rtf files).

SMART START Module – smartstart.sas

The SMART Start Module is where the user defines global and study specific locations for the system. This file is used to assign following:

- Report Module location
- Parameter Set-up Module
- Program location
- Output location
- Input data library
- Derived dataset location
- Parameter dataset location
- Page format and set-up (orientation, linesize, pagesize and page numbering)
- Data listing option for edit checking and debugging
- Reporting of Treatment Ns in either the column heading or in the body of the table.

Kick-Off Module

The Kick-Off Module is where the user determines which Report Modules are to be run and what options are designed for this run. Each time the user wants to create a new set of reports they will modify and submit the Kick-Off Module.

Parameter Set-up Module

The Parameter set-up file is a user interface to create the Parameter dataset that will be read by the SMART Execution module. The Parameter set-up module is an xml file (Extensible Mark-up Language). It typically contains headers and footnotes for all the reports to be run.

SMART Execution Module

The SMART Execution Module is a SAS driver program designed to integrate the Report Module, the DCM Driver, parameter metadata and the user interface (Kick-Off Module). Before execution of the SMART reports the following set-up steps need to be performed.

- Smartstart.sas needs to be set up to contain the appropriate path locations for all macro and program libraries and the desired values for all the global macros.
- DCMs have to be generated so that the data is current
- Parameter Set-up module needs to be modified as appropriate for the reports that are requested
- The Kick-Off module has been set up to generate specific reports along with the desired report specific parameters.
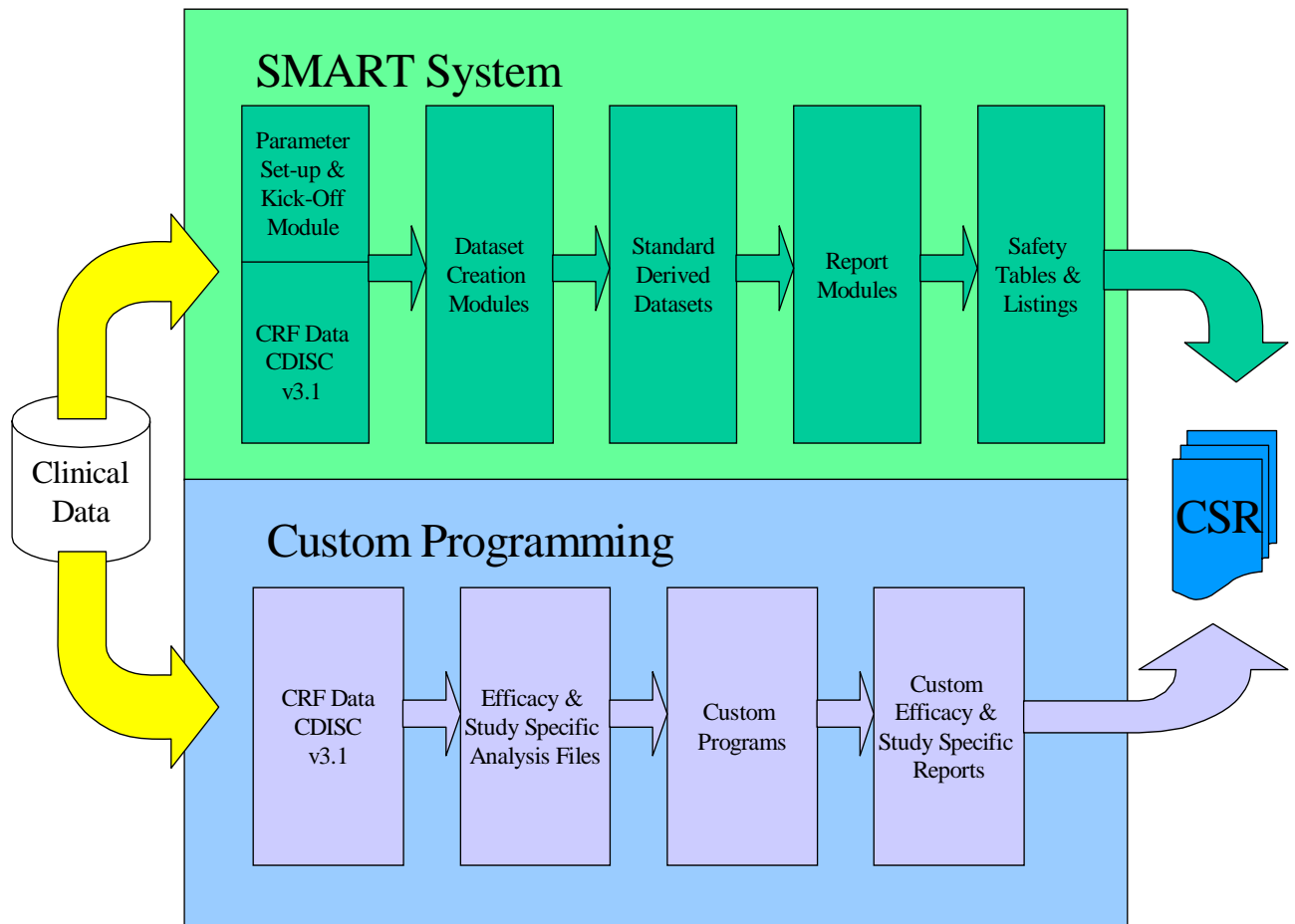
After the set-up has been completed the User will execute the Kick-Off Module. This initiates execution and production of SMART reports. Execution will proceed as follows:

- The Kick-Off module performs the edit checks of the Parameter Set-up Module, which is contained in an XML file.
- The Kick-Off module reads the parameter values specified in the smartstart.sas, Parameter set-up module and kick-Off module and combines them and loads them into a SAS dataset.
- The Kick-Off module starts the SMART Engine.

The SMART Engine integrates the parameter datasets and derived datasets, and connects them with the corresponding Report Module and executes reports. The execution of the SMART Engine proceeds as follows:

- SMART Engine runs the macro loadparm. This macro reads the parameter values and assigns macro variables.
- SMART Engine uses the report identification parameters to merge the Report Module with the corresponding macro variables from the parameter dataset and create a stand-alone SAS program.
- The SAS program is saved to the designated library/file location.
- The stand-alone SAS program runs to generate the report.

PROCESS FLOW DIAGRAM

BENEFITS AND ADVANTAGES

FOCUS ON SAFETY

Get summaries of safety data faster for internal safety review or reporting to Data Safety Monitoring Boards.

CONSISTENCY

New safety standards will provide consistency in the way the data are reported to regulatory authorities and will aid reviewers in cross-study comparisons.

APPROVALS

Faster review and better safety profiles can lead to faster approvals.

SPEED

Results are available faster.

ACCURACY

Validated output produced quickly from CDISC SDTM standard input datasets.

6

INCREASED EFFICIENCY

Less time is required by programmers to produce standard safety tables.

FLEXIBILITY

System has built-in flexibility to allow various user-chosen reporting options.

FOCUS ON EFFICACY

More time to focus on producing analyses of efficacy endpoints or on other study measures of interest.

THE FUTURE OF THE SMART SYSTEM - CONCLUSIONS

Since the SMART Systems inception there have been numerous updates of the system. These have been done on a quarterly basis. These releases have focused on enhancements to the system.

- Conversion and validation of system to SAS Version 9.
- Modifying the system to use our Standard Validated Macros.
- Tweaking the system to comply with our Programming Standards – style of coding, naming conventions etc.
- Having the file create XPT datasets (SAS Transport Files).
- Having the System create free standing programs for analysis files.
- Adding AE, Death and Laboratory Data Listings.
- Updates of the Analysis Files to support new output.
- Making system output compliant with our documentation repository system.

As Biogen Idec moves over to the CDISC standard for all of our projects within biostatistics, one of the important future enhancements is to ensure the Derived Analysis Datasets are CDISC ADaM compliant.

As of present there have been no documented metrics on the efficiency and time savings of the system but on questioning users it seems at most it can take up to a couple of hours to produce the 39 fully validated tables plus any requested listings

CONTACT INFORMATION

Gary Stevens
Biogen Idec Ltd.
Thames House   Tel.   +44 (0) 1628 501000
Foundation Park   Direct   +44 (0) 1628 512583
Maidenhead   Fax   +44 (0) 1628 501010
Berkshire SL6 3UD
UK   gary.stevens@biogenidec.com