SDTM in Business Intelligence

Mike Collinson, Oracle Heath Sciences Consulting (HSC), Reading, UK

ABSTRACT

The SDTM standard is widely used for data storage and pooling. This paper looks at the use of SDTM data for cross study analysis using Business Intelligence tools, starting with an introduction to warehousing principles. It covers the steps required to use SDTM domains as dimensions in a star schema, allowing the data to be used for user-driven creation of powerful ad-hoc reports and provision of interactive dashboards.

INTRODUCTION

The SDTM standard is widely used for data storage and pooling. If configured correctly SDTM data may be used for cross study analysis using Business Intelligence tools. Availability of this data in a central reporting environment increases data transparency and provides near real time actionable insight across large volumes of data.

DATA WAREHOUSE PRINCIPLES

Growing volumes of data, global operations and increasing regulatory scrutiny are encouraging pharmaceutical companies and healthcare providers to develop Clinical Data Warehouses. Data warehouses can be a mine of information in a data rich business environment, and can greatly enhance data transparency and visibility. The interoperability of systems is increasing along with interchange standards, and real world data is being collected more widely than ever before.

Data warehouses are often used to aggregate data from multiple transactional systems. Such systems may have data structures designed for collection, and not be aligned with the reporting standard. Typically this data is transformed and then loaded into a central data model that has been optimized for analysis, for example market research or data mining. It is possible to design a Clinical Data Warehouse that follows the model of a traditional data warehouse with a single well-defined data model into which all clinical data are loaded. This can create a powerful tool allowing cross study analysis at many levels. Data is never deleted or removed from the warehouse, and all changes to data over time are recorded.

The main features of a reporting standard must be ease of use and quick retrieval. SDTM is a mature, extensible and widely understood reporting standard with clearly specified table relationships and keys. The key relationships can be used to allow users to select data from different reporting domains without an understanding of the relationships between domains.

Creation of a star schema can allow such ease of use, with the SDTM based relationships between the tables and all the relevant keys stored in a single table, allowing and streamlining any user selection.

AUTOMATED LOADING AND RECODING

Data may be loaded from the source transactional systems in a number of ways. With EDC, new studies are continually brought online, and may be uploaded repeatedly. Ongoing case processing in Safety systems can present new data to the warehouse every day.

Most warehouse systems include a number of interfaces to load data. Many also supply APIs to allow external programs to control the warehouse in the same way as an interactive user. A combination of robust metadata, consistent data standards and naming conventions can allow automated creation of template driven warehouse structures, and automate data loading.

The SDTM table keys enable incremental loading, where only records changed in the source system are updated in the warehouse, saving disk space. We can also use the SDTM keys in our audit processing, and use them to identify deleted records in incrementally loaded data pools.

Once the data is loaded it is easy to manipulate the data at different levels, for example up-versioning the coding information for a selection of compounds within a Therapeutic Area, but leaving the original coding for all others.

USING SDTM DOMAINS FOR DATA POOLING

SDTM conversion, data pooling at Therapeutic Area and Compound level, and Medical Dictionary re-coding can be handled automatically in the warehouse in the reporting standard. Use of SDTM facilitates the pooling of studies to

the maximum version available, accommodating all of the studies in previous versions without destructive changes which would affect the warehouse audit trail.

Once studies are conformed in SDTM, they can be automatically pooled to serve reporting needs. Ad-hoc data pools can be created simply, using metadata driven programs without the need for transformation.

DOMAINS AS DIMENSIONS

The SDTM domains are suitable as Dimensions in our star schema, familiar to users, extensible with Supplementary domains, and aligned with reporting needs. The SDTM domains do evolve slowly over time, presenting challenges in the data integration layer.

As required SDTM domains can be used as dimensions either associated with Supplemental domains, with a logical join specified, or as normalized tables containing the Supplemental variables themselves.

The fact table contains SDTM keys, and outer joins on all satellite tables. User can start at any table, and drag and drop variables to collect data from any combination.

BIRTH OF A STAR SCHEMA

Creation of a star schema can allow such ease of use, with the SDTM based relationships between the tables and all the relevant keys stored in a single table, allowing an streamlining any user selection. At the center of the star is our Fact table. The Fact table needs to contain all of the information required to seamlessly join the tables together, and will typically include common selection criteria and derived variables.

A simple SDTM based star schema will contain a DM-driven fact table, surrounded by DM and all of the other domains as dimensions. The key relationships for each table are specified in the star itself. Supplemental domains can be associated with the parent domain, creating a more complex star, although the two will resolve as a single table using the relationships in the star schema.

It is wise to limit the size of the Fact table, as it is involved in every request to the warehouse. As this table is linked to all of the others, it should be as stable as possible. Changing the Fact tables will mean changes in the warehousing application as well as visualization application. It can take some tuning to create a powerful query across all of the source tables in the underlying warehouse application to populate the fact table efficiently.
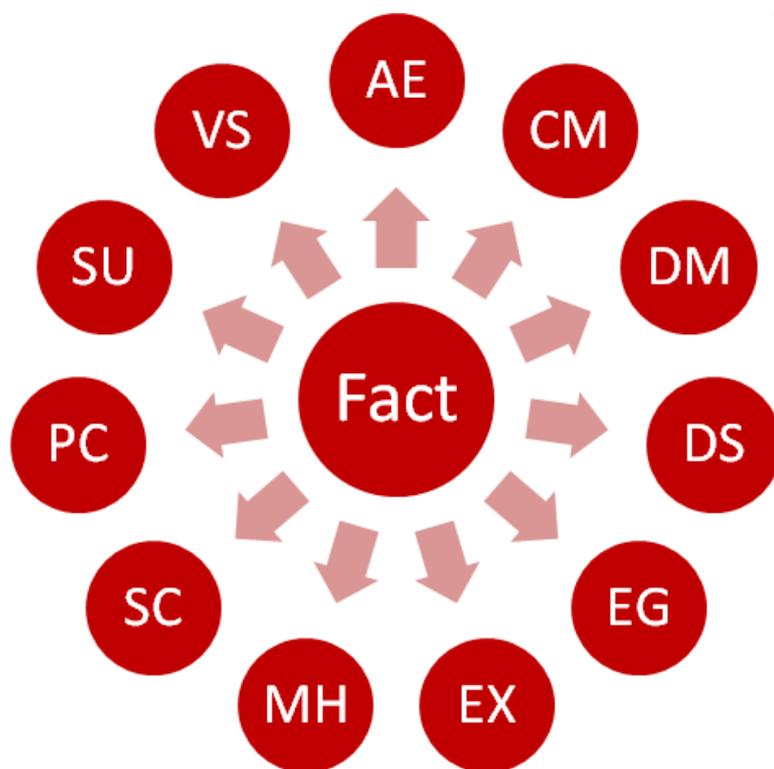

Fig 1. Example of a simplified star schema

DASHBOARDS AND DELIVERY

Uses of a Clinical Data Warehouse include:
- Data reconciliation
- Ongoing medical review
- Streamlined statistical analysis for submission
- Modeling of protocol design and trial simulation
- Responding to regulatory queries
- Safety monitoring and signal detection
- Cross-study analysis

Each of these can deliver value to a customer, but each requires recent consistent data.

Users fall into two groups:

Definers use the reporting interface to design, develop and share reports. Definers create analysis variables, analyses, reports, compile dashboards and add writeback sections. Definers can select any variable from any table and use an expression builder to create measures and derived variables Measures can be defined as part of the analysis, without being added to the underlying data structures Writeback can be enabled to allow comment tracking to improve collaboration. The on demand analyses are easy to export to excel, csv, ppt, pdf etc.

Consumers view, report on, comment on, and export data. They can access dashboards containing multiple analyses in a web browser. Simple selection prompts allow them retrieve an analysis set without programming. External users' access to dashboards and the data that is displayed is controlled by the security set up in the warehouse system itself, so users only have one account across the warehouse and reporting applications. Dashboard results can also be exported to excel, csv, ppt, pdf etc.

RECONCILIATION

Our Clinical Data Warehouse may also be connected to a transactional Safety system. This, coupled with the SDTM data warehouse can allow reconciliation of the two datasources, a crucial task as the Clinical studies are locked and reported, and when periodic reports are due on the Safety data. A set of automated transformations can account for the different vocabularies in the two systems, and the records can be paired together in a dashboard.

The dashboards themselves can be configured to highlight non-matching records, and also to allow data entry to track comments, and acceptance of insignificant differences.

Reconciliation involves both the Clinical and Safety groups, but could also be carried out by CRO users responsible for the studies. As such the application operates as a private cloud, available at any time in the user's browser. This enhances collaboration between the sponsor and CRO, and provides an audited central secure location to capture comments, and record the acceptable of insignificant differences. Security is paramount in an open system, and the warehouses security model is designed to allow CRO users to only see the studies they have been assigned to, hiding other studies from the dashboards and selection prompts.

As a serious adverse event must be reported within 24 hours, it is possible that that event could be reconciled against the clinical data the following day.

CONCLUSION

SDTM can be of huge benefit to the users of a Clinical Data Warehouse system, allowing data pooling for storage, audit and reporting. Critical to this is the creation of a star schema and fact table along with security and accessibility. Through examples of the types of reporting and the roles involved, it is hoped that this paper will inspire readers to maximize the potential of their SDTM data.

REFERENCES

A Practical Guide to Clinical Data Warehousing, Barnden, Palmer. ACDM #80.

RECOMMENDED READING

www.oracle.com

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
Mike Collinson
Oracle Health Sciences Consulting
Oracle Parkway
Thames Valley Park
Reading
RG6 1RA
Mike.Collinson@oracle.com
Brand and product names are trademarks of their respective companies.