Testing CDISC's Operational Data Model in SAS

Michael Palmer, Zurich Biostatistics, Inc., Morristown, New Jersey
Julie Evans, Manager of Technical Services, CDISC, Alexandria, Virginia

ABSTRACT

CDISC recently released version 1-1-0 of its Operational Data Model (ODM v1-1-0) for the pharmaceutical and biotech industries. ODM uses the clinical trial's case report form (CRF) paradigm to represent clinical data and metadata in XML. Pre-release, pharma and vendor companies tested ODM's functionality with clinical data systems. This paper reports on the methods and outcomes for the SAS experience with ODM. The test consisted of four common use cases: (1) CRO to sponsor, (2) lab data, (3) single case report form, and (4) multiple data vendors to sponsor. The test involved exporting data from SAS to ODM XML and importing back into SAS. Because XML consists of named content arranged in a hierarchical structure with content and the content names formatted as text, working with it in SAS is not straightforward. Also, an XML file is really one record that streams into SAS for processing. The test demonstrated that with the appropriate tools and techniques, ODM can be processed in SAS. CDISC is a consortium of 85 pharma companies and vendors with a liaison to the FDA.

INTRODUCTION

CDISC recently released version 1-1-0 of its Operational Data Model (ODM v1-1-0) to the pharmaceutical, biotech, and medical device industries and for clinical research done by non-profit or governmental organizations around the world. ODM uses the clinical trial's case report form (CRF) paradigm to represent both clinical data and the associated metadata in XML.

It is becoming increasingly appreciated in the SAS community that XML has several characteristics that make it's import, export, and processing in SAS an often challenging experience.

XML, and ODM as an XML vocabulary, is

- named content
- organized in hierarchical data structures, and
- formatted completely as text.

The hierarchical structures are often heterogeneous, as they are in ODM, combining metadata and data in a single file and with many optional data and metadata elements.

ODM uses these XML characteristics to provide what is becoming recognized as a clever and efficient way to transmit and archive clinical data. Unfortunately, these clever XML characteristics create a headache in SAS. For instance, in the common SAS file structure, every record in a dataset has the same fields and one or more of those fields are keys that identify the record. In XML, including ODM, there are no records. Data are identified by looking at the values of ancestors and sibs in the stream of XML statements instead of looking at key fields on individual records.

Moving data correctly between the hierarchical, named content in an ODM text stream and flat datasets in SAS was one of the challenges for the CDISC Testing Task Force. Others involved assessing the capability of ODM as a vehicle for clinical data.

CDISC AND THE TESTING TASK FORCE

This paper will discuss the experience of using SAS to test ODM. The testing was carried out by CDISC's Testing Task Force from February until June 2002. CDISC is a consortium of 85 pharma industry companies and vendors with a liaison to the FDA. The organization is an open, multidisciplinary, non-profit body committed to the development of pharma industry standards to support the electronic acquisition, exchange, submission, and archiving of clinical trials data and metadata for medical and biopharmaceutical product development. The mission of CDISC is to lead the development of global, vendor-neutral, platform-independent standards to improve data quality and accelerate product development. CDISC has published standard models for clinical research (also known as operational data), clinical laboratory, regulatory submissions, and statistical analysis data interchange. For more information on CDISC objectives and principles, please see the CDISC website at www.cdisc.org.
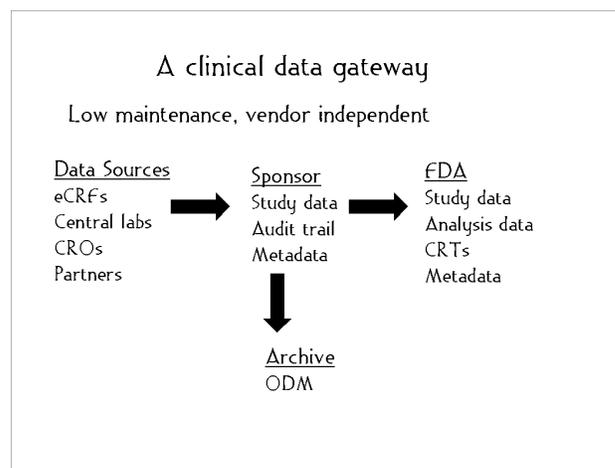
Figure 1. CDISC standards can make clinical data management low maintenance and vendor independent.

The task force consisted of 7 industry representatives from major pharmaceutical companies, CROs, technology vendors, and consultants. Both North American and European testers participated. The CDISC author of this paper led the task force.

THE ODM TESTING

ODM uses the clinical trial's case report form (CRF) paradigm to represent both clinical data and the associated machine-readable metadata in XML. ODM's clinical data model will hold whatever a traditional paper CRF or electronic data capture (EDC) form would collect. The metadata is very much like what an annotated CRF has: field names and attributes such as datatype and length, code lists, and file names and record layouts. One very significant difference between an annotated CRF and ODM metadata is that ODM metadata, as XML, is machine-readable. In addition, ODM has a reference data section for information that is not associated with specific study subjects. An example of such data is clinical lab normal values. The fourth and final category of data in ODM is administrative data such as site personnel names and addresses. Complete documentation on ODM is available at the CDISC web site.
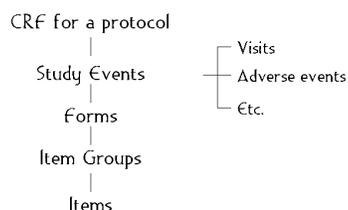
Operation Data Model (ODM) v 1.1

Top-level structure

```
          ┌ Study
          ├ AdminData
   ODM ───┤
          ├ ReferenceData
          └ ClinicalData
```
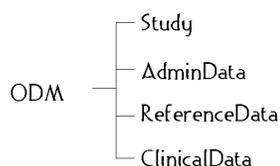
Figure 2. The ODM is a single data source for all clinical study information.

Operational Data Model (ODM) v1.1

Conceptually, ODM data is in the image of a CRF

```
   CRF for a protocol
          |                    ┌ Visits
   Study Events         ──────┤ Adverse events
          |                    └ Etc.
       Forms
          |
    Item Groups
          |
        Items
```

Figure 3 The ODM provides an XML-formatted case report form.

USE CASES

The Task Force used a use case-based approach to testing the model and defined four of the most commonly applicable use cases:

Use Case 1: CRO to Sponsor Transfer. The scope of the test involved creating an ODM XML file from SAS datasets, and then creating SAS datasets from the ODM XML file. For this use case, the testers used a small set of data for a phase I single-dose crossover study.

Use Case 2: Lab Data Transfer. This use case is now covered by the CDISC Lab Model.

Use Case 3: Single CRF Browse. This use case is now covered by the ODM Viewer.

Use Case 4: Multiple Vendors Sending Segments of Data. This situation occurs when several smaller vendors are providing only part of the study data, e.g., the enrollment and/or diary data. This use case is intended to simulate data a CRO (vendor) would collect and transfer for a specific subset of a study's data that is not collected as part of the standard CRF, e.g., patient-reported outcomes data. This case is intended to demonstrate that a subset of a subject's data collected by a third party can be merged with the core CRF data for a subject.

For each use case, the team's goal was to create an ODM XML file from the source data and then create a SAS file from the ODM XML file. The team used several methods in creating and reading ODM files. Three team members wrote custom programs (perl, XSL) to create ODM XML files. One team member used a SAS-based toolkit, the Tekoa Toolkit from Zurich Biostatistics, Inc., to read the ODM XML files and create SAS files from them as well as to write XML from SAS files.

Overall, ODM passed the test, and all the methods tested were able to move data correctly between XML and non-XML target environments, such as SAS itself or an intermediate environment on the way to SAS.

TEKOA TECHNOLOGY

Of the several methods used, Tekoa Technology[sm] was the only one that could meet all of the test criteria entirely within SAS, without resorting to perl, XSL, or other non-SAS technology.

Tekoa works by indexing XML in a way that both preserves the hierarchical information inherent in an XML file and is invariant to the hierarchy levels that may or may not be present in a particular XML file. Tekoa imports XML to and exports XML from the base SAS DATA-step. SAS programmers work with XML in the DATA-step just as they work with any other kind of data. The testing proved that the ODM model and Tekoa could successfully be used in a SAS environment to transfer data between two organizations for all of the use cases.

TESTING OUTCOME

In the course of evaluating and approving ODM, the Testing Task Force also identified several issues (e.g., data mapping, documentation, file security) that are being considered by the ODM technical committee at CDISC.

The Testing Task Force reported its findings to the CDISC organization and ODM v1-1-0 was released for general use, including use with SAS.

CONCLUSION

1. ODM works for the transfer of clinical study data between a data source, like a CRO, and a study sponsor.
2. SAS with suitable tools works as the platform for ODM XML, on both the file sending and the file receiving sides of the process.
3. Tools for SAS to work as an ODM platform include XSL, perl, and Tekoa Technology.

REFERENCES

An ODM testing toolkit is available for free from Zurich Biostatistics, Inc. by emailed request to Michael Palmer.

ACKNOWLEDGMENTS

We would like to acknowledge and thank the Testing Task Force members and their companies for their diligence in evaluating the ODM and their generous contribution of time during the testing process.

CONTACT INFORMATION

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

Michael Palmer
Zurich Biostatistics, Inc.
45 Park Place So., PMB 178
Morristown, NJ 07960
973-727-0025
mcpalmer@zbi.net
www.zbi.net

Julie Evans
CDISC
P. O. Box 162033
Austin, TX 78716
julieevans@mindspring.com
www.cdisc.org

Zurich Biostatistics, Inc. is a Registered CDISC Solution Provider.