Overview of the CDISC Operational Data Model
for Clinical Data Acquisition and Archive
(based on CDISC ODM DTD 1.1 DRAFT)

October 10, 2001

History and Background

Historically there has been no established industry standard for the context, representation and interchange of the data collected during the course of clinical trials. The lack of standards in these areas contributes significantly to the cost and time associated with drug development by creating barriers to important information-handling processes and activities, including:

- Internal training, preparation and setup for new trials
- Flexible adoption of new and more effective electronic technologies for data capture as they become available
- Migration to new generations of clinical data management software
- Exchange of clinical data with partners, licensees and Contract Research Organizations (CRO's)
- Long-term archiving of clinical data and clinical data management software.

From its inception in 1997, the Clinical Data Interchange Standards Consortium (CDISC) recognized the need for the establishment of standard data models to improve the process of electronic acquisition and exchange of clinical trials information for the benefit of all medical and pharmaceutical product development stakeholders. Objectives were defined that included creation of an open, vendor-neutral standard that covered both data and "metadata" (data about data) and that was capable of supporting all relevant regulatory requirements.

The eXtensible Markup Language (XML) was identified as a key enabling software technology that was gaining wide acceptance as a data interchange framework in other industries, and that was beginning to be utilized by several vendors of clinical trials software products. XML supports the definition and representation of sophisticated data models in a consistent text-based (ASCII) format -- one that can be processed conveniently by a growing set of third-party tools. However, since XML is itself really a "meta-language" (a language for defining industry- or domain-specific sub-languages), to make the most effective use of XML for clinical data interchange, it is necessary for pharmaceutical industry participants to agree on a consistent specific modeling approach to representing the clinical trials data and metadata in XML.

The CDISC Data Acquisition and Interchange Standards (DAS) working group was formed in September, 1999, to further the development of a vendor-neutral clinical data

interchange standard. Within the DAS subgroup, a technical analysis team was created with representatives from Amgen (K. Harter), Domain Pharma (T. Bloom), IBM (A. Diaz), Oracle (J. Rees), Phase Forward (J. Klofft), PHT (G. Gordon), and PPD (S. Cassells). The team was later expanded to include R.Lyons from Nextrials and D. Fram from Lincoln Technologies. A separate requirements team conducted a survey of requirements within the industry; for details on these requirements and further information about ODM 1.0 please refer to the Technical Overview of Version 1.0 of the CDISC ODM Model (available at http://www.cdisc.org/Operational/Overview.pdf).

The initial task of the technical analysis was to examine two different XML-based data interchange models that had been separately put forward by Phase Forward and by PHT and to assess the feasibility of developing an integrated, single XML standard. The results of this assessment were positive, and a high-level overview of the model was presented by J. Klofft and G. Gordon at the DIA Workshop on *Achieving Data Standards for Clinical Development* in November 1999.

The technical analysis group continued to meet in December 1999 and January 2000 to develop an initial version of the XML "Document Type Definition" (DTD) that defined the specific data interchange model. That initial version, called DTD 0.8, was made publicly available for review and comment through the CDISC web site in March 2000. At the DIA National meeting in San Diego in June 2000, a decision was made to develop extensions to the DTD, particularly in including support for an audit trail, and to release Version 1.0 of the DTD by the end of the summer. Version 1.0 of the CDISC Operational Data Model, with increased audit trail support, was released to industry in October 2000.

In January, 2001, a new Operational Data Modeling (ODM) team consisting of a combination of members from the original DAS technical analysis team plus new members was formed to begin work on the next version of the model, version 1.1. The major contributing team members to ODM Version 1.1 included Amgen (L. Baca), Aventis (M. Schmeiszer), CB Technologies (S. Hume), Domain Pharma/Clinsoft (C. Schaffert), GSK (M. Miles), Lincoln Technologies (W. Kubick), Nextrials (R. Lyons) Oracle Clinical (D. Kacher), Phase Forward (J. Klofft), PHT Corporation (R. Ferris, G. Gordon, S. Cassells), Pfizer (J. Streeter), PPDI Informatics (B. Drummond) and Zurich Biostatistics (M. Palmer). This team developed the revised 1.1 model from February through September 2001. Model development was driven by four primary factors:

1. Correction of known problems and limitations of version 1.0
2. Incorporation of expanded functionality to extend the applicability of the models
3. Response to feedback received as a result of a live proof-of-concept demonstration of the model at the 2001 DIA Annual Meeting that involved over twenty industry companies
4. Increasing support for the transfer of electronic lab data from labs to sponsors.

Version 1.1 of the ODM was released for comment in October 2001. The ODM team hopes to finalize 1.1 for implementation by the beginning of 2002.

Technical Objectives for ODM Versions 1.0 and 1.1

The overall goal of ODM 1.0 was to make available a first release of the definition of the CDISC XML model, in order to support sponsors, vendors, and CRO's with respect to the design of systems and processes around a standard interchange format.

The technical focus in the development of ODM 1.0 was the definition of structures to represent the three major information components relating to a clinical trial:

- clinical study metadata (item definitions and protocol)
- clinical study administrative data (users and access privileges)
- clinical study data (complete record of patient data and audit trail)

This included representation of metadata capable of supporting either direct electronic, or paper-based, data collection, and capture of clinical data from one system to another.

ODM 1.1 adds support that will greatly increase the likelihood of adoption by vendors of clinical data management systems and other sponsors, including:

- Ability to address changes to key data values
- Expanded transaction support for partial or incremental transfers
- Expanded metadata descriptions of more complex event structures
- Support for including multiple studies and reusable metadata in one file
- Support for depicting non-patient reference data
- Support for vendor extensibility
- Increased compatibility with the CDISC Submissions Data Model
- Increased support for archiving of clinical data and metadata
- Elimination of support for the flat representation of clinical data included in version 1.0
- Corrections to numerous bugs including those related to locales, timezones and signatures.

A major feature of the 1.1 model is a greatly expanded documentation set, which was largely created through the efforts of ODM team member Craig Schaffert.

Several longer-term objectives of the CDISC ODM are not yet fully supported in the ODM 1.1 version. These additional capabilities, which are to be developed in later versions of the standard, include:

- Support for more complex "use case" transaction-level or real-time interoperability between systems
- Representation of queries and query resolution linked to clinical data
- Support for transmission of data in binary format
- Support for XML Schemas.

Overview of the Model

Like ODM 1.0, ODM 1.1 provides a format for representing study metadata, study data and administrative data associated with a clinical trial. It represents only the data that would be transferred among different software systems during a trial, or archived after a trial. It need not represent any information internal to a single product, for example, information about how that data would be stored in a particular database, although ODM 1.1 allows such information to be provided in the form of vendor extensions. The model is intended to be system-neutral and vendor-independent; terminology is described in detail in the ODM 1.1 Specification.

Major Sections of the Model

A top-level view of ODM 1.1 is shown in Figure 1 below. The diagrams depict the model as a single XML structure, rooted in the ODM element which ensures that referential integrity with different DTD versions can be preserved while also allowing for representation of multiple studies in a single file. The diagrams follow XML conventions in describing repeated and optional elements using the "*" (zero or more), "+" (one or more) and "?" (optional) prefixes. The diagrams do not display all elements and attributes of the model – these are discussed in detail in the ODM 1.1 Specification.

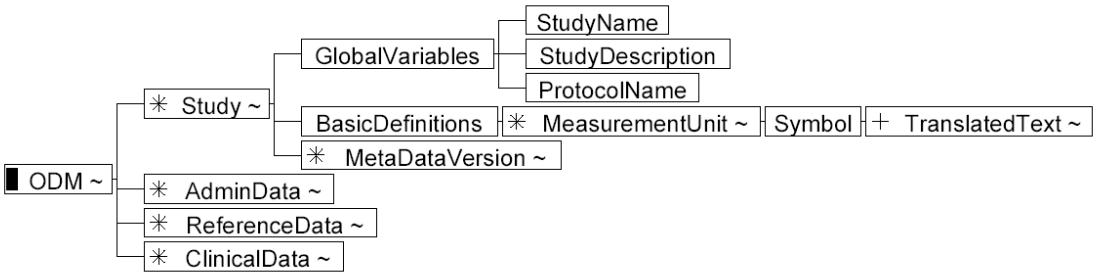

Figure 1 Major Sections of the Model

ODM V1.1 consists of four principal sections:

- Study allows representation of more than one study in a single file
- AdminData includes information about the users of the system, the clinical sites involved in the study, and associated security information
- ReferenceData provides information relevant to the interpretation of data that is not necessarily study specific, such as lab normal ranges
- ClinicalData contains the actual data item values associated with each study.

For each study, the following sub-elements are included:

- GlobalVariables contains descriptive information about the study as a whole, such as StudyName, StudyDescription, and ProtocolName

- BasicDefinitions contain definitions of information define other elements within the XML DTD, such as measurement units

- MetaDataVersion contains the metadata definition for the study, not the data itself. It includes definitions of the visits to be scheduled within the protocol; the forms associated with each visit; and the information to be collected on each form.

Figure 2 displays the major elements for these sections of the model (all of which are described in detail in the ODM 1.1 Specification).

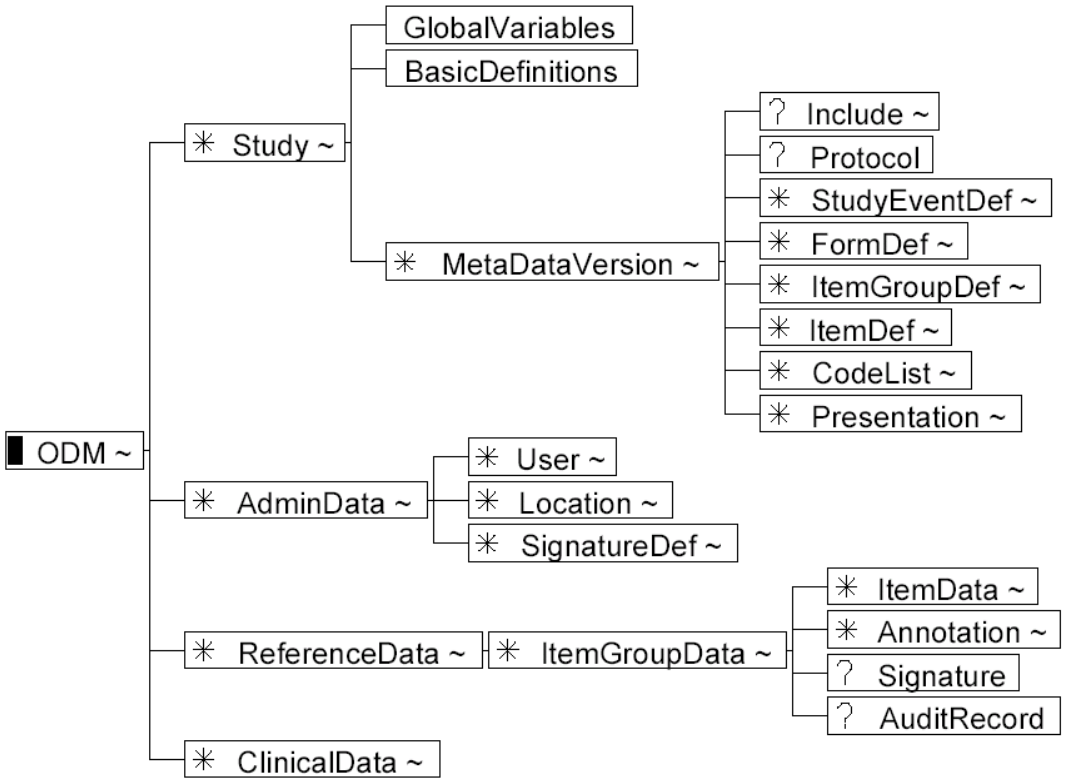

Figure 2 - Structure of MetadataVersion, AdminData and ReferenceData

As shown in Figure 3, Metadata is arranged in a hierarchical structure, and can be "versioned" to support revisions in the study definition. ODM 1.1 includes increased transaction support for mid-study changes and associated partial data transfers.

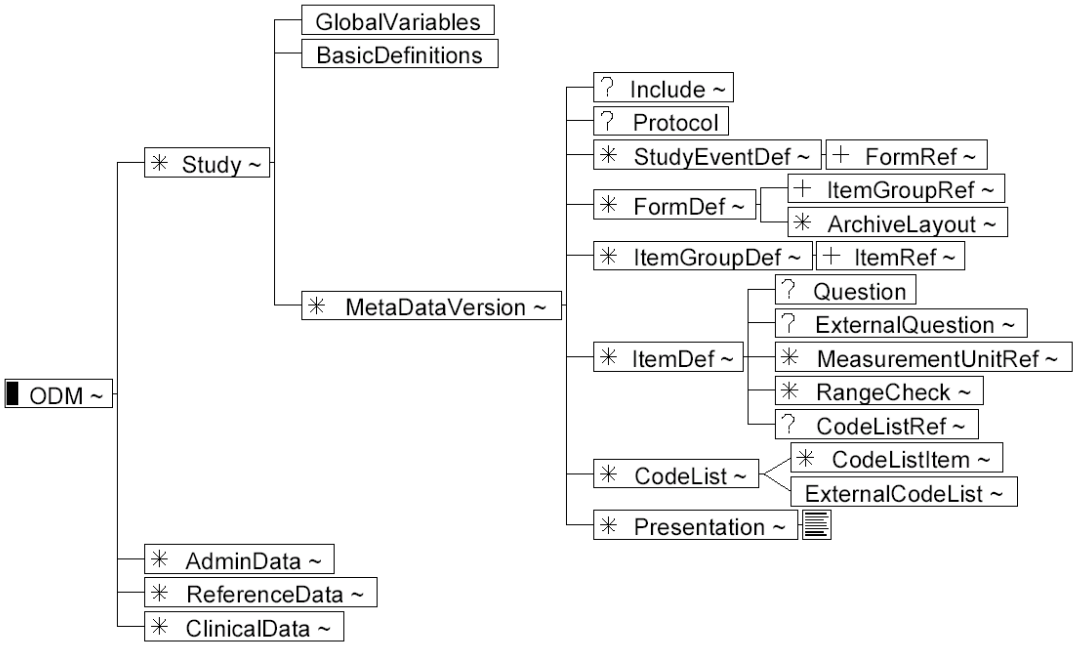

Figure 3 Metadata Elements

Clinical Data Section of the Model

Unlike ODM 1.0, which allowed Clinical data to be represented in either a hierarchical or flat format, Version 1.1 requires all Clinical data to be represented according to a specific logical hierarchy for each subject (StudyEventData, FormData, ItemGroupData, ItemData) that parallels the metadata hierarchy (StudyEvent, Form, ItemGroup, Item).

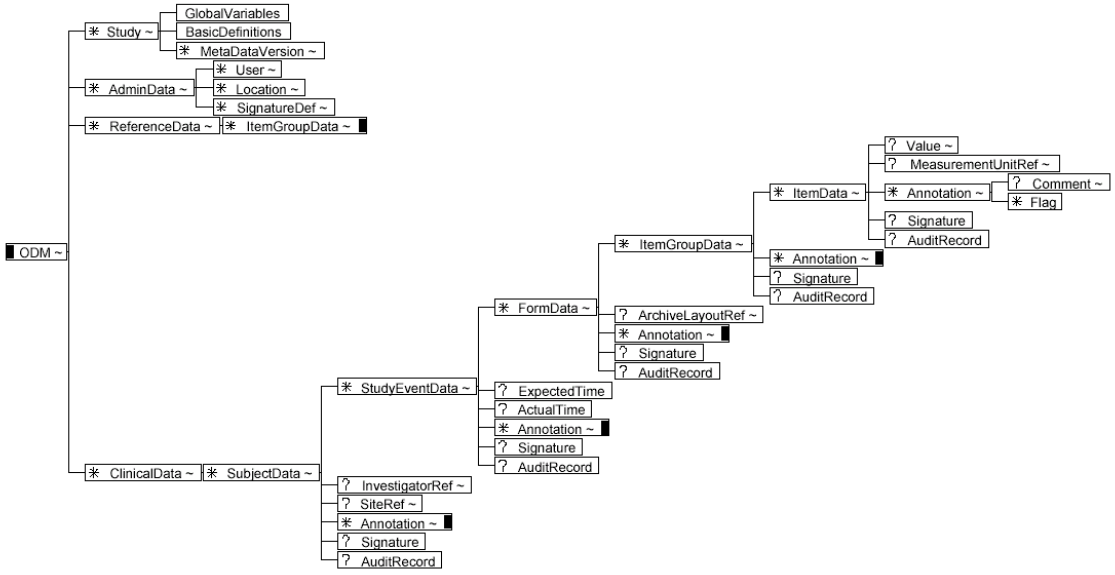

Figure 4 – Clinical Data Elements

An actual clinical data value would be stored in the Value element, shown in the DTD diagram above on the far right. At each level, there can be an associated measurement unit, a signature, an annotation comment and an audit record.

Request for Feedback from Industry

ODM 1.1 Draft is being posted on the CDISC web site for open comment by all industry stakeholders. Feedback should be submitted through the CDISC Discussion Facility at http://www.cdisc.org/forum/default.asp?page=1&b=1 in the folder titled "Operational Data Model Version 1.1 DTD Comments".