

16.1 CDISC ODM XML

The Clinical Data Interchange Standards Consortium (CDISC) is a clinical research standards body formed to encourage maximum sharing of information and minimum duplication of efforts. One of the standards CDISC has created and endorsed is the Operational Data Model (ODM), which facilitates the archive and interchange of the metadata and data for clinical research. ODM is represented in XML format and is designed to collect data from many different sources into one document.

Purpose of This Document

OpenClinica provides and/or consumes CDISC ODM XML representations in its Extract Data and Import Data modules and other parts of the software. This document describes how OpenClinica represents study metadata and data that is stored in its internal database as CDISC ODM XML documents. It assumes a working knowledge of CDISC ODM 1.3 and of OpenClinica, and attempts to describe how each OpenClinica field or element is represented in ODM, and under what conditions. The document is best read when accompanied by the CDISC ODM standard. It is geared towards developers, but is also intended for data managers who want to know more about the capabilities of ODM XML export in OpenClinica. Additionally, parts of this document will find its way into the online documentation, for all end users. OpenClinica's ODM representation has changed iteratively from version to version of OpenClinica, and the appendix to this document charts these changes since version 2.5 and the addition of the custom extension to the ODM, introduced with OpenClinica 3.0.

Scope of This Document

This document provides a detailed specification of the OpenClinica [CDISC ODM XML version 1.3](#) with OpenClinica Extensions as implemented in the OpenClinica 3.1 and later releases.

Definitions and Acronyms

- CDISC [Clinical Data Interchange Standards Consortium](#)
- DOM Document Object Model
- ODM CDISC [Operational Data Model](#)
- XML Extensible Markup Language
- XSL Extensible Stylesheet Language, a language for displaying an XML file as a given type.

General Issues

CDISC defines its Operational Data Model, version 1.3, as a vendor neutral, platform independent format for interchange and archive of clinical trials data. The model includes the clinical data along with its associated metadata, administrative data, reference data and audit information. All of the information that needs to be shared among different software systems during the setup, operation, analysis, submission or for long term retention as part of an archive is included in the model. An XML document must meet certain basic criteria to be considered conformant to the ODM standard. These are briefly discussed below:

Syntactic Constraints

The syntactic constraints defined by the ODM standard are:

1. The ODM file must be a well-formed XML file. See the [XML standard](#) for details.
2. The ODM file must conform to the XML Namespace standard. See the [XML Namespace standard](#) for details.
3. The ODM file must contain only elements and attributes defined in the ODM standard schema or in a valid vendor extension schema, and must satisfy the rules about element nesting and the formats of attribute values and element bodies.
4. The ODM file must contain a prolog and a single (top-level) ODM element.
5. The namespace for version 1.3 of the ODM is <http://www.cdisc.org/ns/odm/v1.3>.

Currently, the ODM study definition file (available from the View Study page at the URL /DownloadStudyMetadata?studyId=#) does not meet these requirements for the following reasons:

1. The file generated from the View Study page is only a fragment of XML, and does not contain the initial tag which defines the character set and version, i.e. `<?xml version="1.0" encoding="UTF-8"?>`.
2. The file generated does not contain references to any XML Namespaces, including the namespace for version 1.3 of the ODM itself.
3. The file generated does contain elements defined in the ODM standard schema, but lacks the single top-level ODM element.
4. The files suffix is txt instead of xml.

OpenClinica ODM Data Import meets the above constraints, but note that OpenClinica parses everything within the ClinicalData element only, and it does not read anything in the Study element, and, as such, cannot import Study metadata at this time.

Entities and Elements

Entities and elements in OpenClinica use the same names as their counterparts in ODM. For example, the ODM definitions for study event and Study Event Definition are valid for the entities of the same name in OpenClinica (see Section 2.6, Entities and Elements, of the ODM specification):

- A *study event* is a reusable package of forms usually corresponding to a study data-collection event.
- A *Study Event Definition* describes a particular type of study event (mostly by listing the types of forms it can contain).
- The clinical data of a study will typically have many actual study events corresponding to each StudyEventDef,

Where the usage of these entity names in OpenClinica diverges from the ODM definition, it will be noted in this document.

Clinical Data Keys

The ODM standard uses the concept of Internal Clinical Data Keys to uniquely address clinical data entities within the model. The following table details the key, or combination of entity identifiers, that you would need to uniquely and specifically address a clinical data entity.

[table id=63 /]

For example, an XPath query to retrieve a specific item data value in an OpenClinica ODM Extract

would be of the form:

```
/odm:ODM/odm:ClinicalData[@StudyOID='S_P12345_2818']/odm:SubjectData[@SubjectKey='SS_101']/odm:StudyEventData[@StudyEventOID='SE_INITIALT' and @StudyEventRepeatKey='1']/odm:FormData[@FormOID='F_AGEN_V10']/odm:ItemGroupData[@ItemGroupOID='IG_AGEN_DOSETABLE-F_AGEN_V10' and @ItemGroupRepeatKey='1']/odm:ItemData[@ItemOID='I_AGEN_AGENT_NAME']/@Value
```

In the image below you can see that the latter half of the XML file (the part contained in the <ClinicalData> tags) links to specific tables in the OpenClinica database. We then link back to the Study metadata through those OIDs. Internally we don't use OIDs in those tables, but instead the conventional methods of primary keys and foreign keys in the database is good enough. For more on how ODM, OIDs, and Clinical Data Keys are implemented in OpenClinica see the blog post, ["Thoughts on Code: OpenClinica and Open Standards with CDISC"](#).

Data Representations in ODM XML (Extract)

When OpenClinica outputs ODM XML, the five basic XML entities (gt, lt, quot, amp, apos) are escaped using XML Entity notation (For example: "bread" & "butter" => "bread" & "butter"). Whitespace is represented literally linebreaks and tabs in ItemData values and other fields will be preserved. Note that, while tabs and carriage returns are limited in the data entry side of the application, (tabs will automatically shift focus from one Item to the next, for example) all spaces and linebreaks are saved to the database, and will export into ODM XML. Items saved in the database with non-ASCII characters will be extracted to XML entities using their ASCII decimal value equivalents; please see the next section, ["OpenClinica Data Representations in ODM XML \(Extract\)"](#) for an example of this.

Item Data Types

OpenClinica supports a subset of the Item Data Types defined in ODM. The data type mapping is shown below, along with the allowed string pattern used to validate item values for a given data type. Note that a listing of no definition in the table below means that the data type is not supported in OpenClinica.

Item Data Types

[table id=64 /]

Mapping of OpenClinica Elements to ODM

[table id=65 /]

Approved for publication by Kerry Tamm. Signed on 2020-12-15 1:46PM

Not valid unless obtained from the OpenClinica document management system on the day of use.