OLMS Electronic Reporting and Disclosure System

Data Specifications Document

September 28, 2006

The U.S. Department of Labor
Employment Standards Administration
Office of Labor-Management Standards

Table of Contents

List of Tables

1. Overview

1.1 Audience

This document is intended for technical users responsible for creating data access or data transformation methods supporting Form LM-2, Form LM-3, and Form T-1 filing requirements. Primary technical users include software analysts, software developers, information engineers, and data managers. This document helps guide technical users through the process of creating data access or data preparation tools. It is assumed that readers have prior experience using data definition mechanisms, including XML schemas and/or comma separated value file specifications.

1.2 Purpose

This document and the accompanying files provide specifications and guidance for developing formatted import files that are compliant with DOL's Electronic Reporting and Disclosure System (ERDS). ERDS enables labor union annual report filers to enter data into electronic LM-2, LM-3, and/or T-1 forms and allows them to submit the forms to DOL for processing and publication. If filers submit large amounts of data or wish to further automate the filing process, ERDS can import data from properly formatted files directly into the forms.

This data import process provides a mechanism for filers to load data extracted from financial accounting systems into ERDS. Filers can export or compile their financial data into Extensible Markup Language (XML) or Comma Separated Value (CSV) data files that adhere to the corresponding formats prescribed in this document and import the data files into ERDS forms using the ERDS data import facilities.

This document addresses only the content and structure of data files prepared for import into ERDS forms. Data validation services are provided by both the import process and the electronic forms preparation package itself. For detailed instructions on completing and submitting the electronic LM-2, LM-3, and T-1 forms and on the import process, refer to each form's corresponding User Guide.

The Data Specifications Document (DSD) will not be used by all filers. Some unions, particularly smaller unions that do not have an in-house or contracted technology staff, may instead elect to manually enter information directly into the forms. The DSD will be used by those unions that choose to create import files to populate the forms. The U.S. Department of Labor (DOL) is prepared to offer compliance assistance to help filers create the file formats described in this document. If you have general questions about the Form LM-2, Form LM-3, and Form T-1 filing requirements or specific questions about this document, you may call DOL's toll-free number at (866) 4-USA-DOL (487-2365), or send an e-mail to olms-public@dol.gov. Further information may also be obtained from the "Frequently Asked Questions" section of the Office of Labor-Management Standards Web site at http://www.olms.dol.gov.

1.3 Documents Overview

The ERDS Data Specification Document is intended to be used in conjunction with several other accompanying files. Those files include the XML schema files, sample LM-2, LM-3, and T-1 XML data files, the CSV format file, and the LM-2, LM-3, and T-1 XML schema reference guides.

The XML schema files that accompany this document are all suffixed with the extension ".xsd" and are generally referred to as "xsd" or "schema" files. These files include lm2.xsd, lm3.xsd, t1.xsd and

common.xsd. The common.xsd provides the XML schema for elements that are common to the LM-2, LM-3, and T-1 forms. The lm2.xsd file provides the XML schema for the LM-2 form. The lm3.xsd file provides the XML schema for the LM-3 form. The t1.xsd provides the XML schema for the T-1 form.

The CSV format file is a Microsoft Excel spreadsheet file. The file provides format specifications for CSV data files.

Also accompanying this document are three HTML files, one for each form (LM-2, LM-3, and T-1). These documents comprise the XML schema reference guides.

The remainder of this document is divided into two main sections and provides descriptions of the specific file formats that are accepted by the ERDS data import tools. The first section describes the XML format. The second section describes the file format for comma separated value files.

Please do not hesitate to contact DOL if you have any questions about the purpose or content of this document or the accompanying files.

2. XML Import Format

This section provides an overview of the XML data import format. It assumes the technical user is familiar with XML basics, including core XML syntax, XML namespaces, and the XML Schema Definition language. Technical users not familiar with these concepts can find numerous references on the World Wide Web, such as the World Wide Web Consortium (W3C) at http://www.w3.org/xml.

2.1 XML File Format

XML import files must be well-formed, valid XML 1.0 document instances (as defined in http://www.w3.org/TR/REC-xml). Documents must include an explicit XML declaration indicating the XML version (1.0) and the document encoding. The ERDS import tool supports the following encodings: UTF-8, UTF-16, ISO-8859-1, US-ASCII. Other encodings may be supported, but are not guaranteed.

The following example shows the first few lines of a sample LM-2 import file. Note that all data elements are in the default "lm2" namespace.

```
<?xml version='1.0' encoding='UTF-8'?>
<lm2-import xmlns='http://www.dol.gov/esa/olms/2003/lm2-1_1' >
  <assets>
    <accountsReceivable>
      <itemizedAccounts>
        <account>...</account>
        <account>...</acount>
```

Complete LM-2, LM-3, and T-1 examples are included with this distribution for reference.

2.2 XML Document Schemas

XML files used for data import must conform to the appropriate XML document format as specified by a set of schema files defined using the W3C 2001 XML Schema Recommendation (namespace http://www.w3.org/2001/XMLSchema). Table 1 lists the locations of the schema definitions used by the ERDS import tool.

Table 1: XML Schemas

File name	Description
lm2.xsd	Top-level schema definition for Form LM-2 import files
lm3.xsd	Top-level schema definition for Form LM-3 import files
t1.xsd	Top-level schema definition for Form T-1 import files
common.xsd	Common types shared by multiple schemas

These schemas define an import file's valid data elements, including hierarchy and sequencing of these elements, data types, maximum lengths, and number of occurrences. Where an element is restricted to a fixed set of values, the schema lists the complete set; unless explicitly indicated, values must be provided in the exact form shown (e.g. state abbreviations and status codes must be in uppercase as defined by the schema). Copies of these files are included with this distribution, or you

can download the latest versions directly from the Department of Labor's web site at http://www.dol.gov/esa/regs/compliance/olms/dsd.htm.

The ERDS import schema definitions impose minimal restrictions on required data elements, allowing labor organizations to import as much or as little data as is available. For example, an organization wishing to populate Schedule 16 could do so by creating an XML data file containing only Schedule 16 data, omitting the other top-level data elements. Other schedules could be populated from other import files or directly through the forms preparation tool. This approach provides maximum flexibility for supporting a particular organization's filing preparation process.

For help navigating and understanding the XML import structure, this distribution includes supplemental XML schema reference guides for each of the schema files. The XML schema reference guides present the definitions in a hyperlinked graphical format. Table 2 lists the locations of the xml schema reference guides.

Table 2: XML Schema Reference Guides

File name	Description
lm2/lm2.html	XML schema reference guide for lm2.xsd
lm3/lm3.html	XML schema reference guide for lm3.xsd
t1/t1.html	XML schema reference guide for t1.xsd

Where discrepancies between these XML schema reference guides and the .XSD schema files exist, the schema files are the final authority.

Note that an XML schema cannot capture all the rules concerning data content. Data supplied in import files must also conform to the specific validation rules defined in each form's User Guide (such as filing thresholds, explanatory notes, and other instructions).

3. CSV Import Format

Import data files may also be created using a comma-separated value format. The data type, value, and field-level format restrictions can be found in the accompanying spreadsheet titled csv-formats.xls. Unlike XML import files, CSV files may contain only itemized schedule data, where each record in the input file has identical structure and format. Additional data, such as schedule totals, non-repeated information, or general forms information, must be entered manually into the electronic forms or imported through XML data.

3.1 CSV File Format

An ERDS CSV import file consists of a single header record, followed by a series of data records. Each data record typically represents a single schedule item. A record is usually a single text line, terminated by a line feed (ASCII LF=0x0A), a carriage return (ASCII CR=0x0D), or a CR/LF pair. Because fields might contain embedded line-breaks (see below), a record may span more than one line.

The first record in an imported CSV file **must** be a header record containing field names. Field names are separated by commas. An example of the header record is shown below.

```
name, sex, ssn, date_of_birth
```

The header record, like all other CSV records, is encoded in accordance with the rules stated in this section.

After the header record, the CSV file is made up of one or more data records. A data record contains values in one or more fields, separated by commas. A sample data record is shown below.

```
Frank Smith, Male, 111-22-3333, 01/01/1930
```

White space characters (spaces or tabs) are allowed. However, leading and trailing space-characters adjacent to comma field separators are ignored—for example,

```
Frank Smith  ,...
```

resolves to "Frank Smith".

To preserve whitespace, data fields that contain leading or trailing spaces must be surrounded by double-quotes. For example, the correct format for a data field that must contain a space at the end of the data would be the following:

```
"Frank Smith ", ...
```

Data fields containing embedded commas must be delimited with double-quote characters. In the example below. “January 1, 1930” is delimited by double quotes because it has an embedded comma.

```
Frank Smith, Male, 111-22-3333, "January 1, 1930"
```

Data fields that contain double-quote characters must be surrounded by double-quotes. For example, if a nickname is part of the name—as in, Frank "Mad Dog" Smith, the correct format is

```
"Frank ""Mad Dog"" Smith", Male, 111-22-3333,...
```

In this example, Frank “Mad Dog” Smith is the value in the `name` field. Note that each embedded double-quote (in this example, the “ before Mad and the “ after Dog) must be represented by a pair of consecutive double quotes.

A field that contains embedded line-breaks must be enclosed by double-quotes.

```
date, message, email_address
01/01/2004, "Please being the following:
    tent
    sleeping bag
    matches
    rope
  Frank
  ",frank@email.com
```

In the above example, the `date` data field contains the value: "01/01/2004". The `message` data field contains the value: "Please being the following:<cr> tent<cr> sleeping bag<cr> matches<cr> rope<cr>Frank". And the `email_address` data field contains the value: "frank@email.com". Though the example record takes up more than one line in the CSV file, it is a single CSV data record. This is valid because the line breaks are embedded inside the double quotes of the field.

Fields may always be delimited with double quotes. The delimiters are always discarded.

If a field value is only white space (tabs and spaces), the value is presumed absent. To indicate an empty string, use double quotes. Trailing missing fields may be omitted.

CSV-formatted import data for schedules with individually itemized payments (for example, Schedule 15 on the LM-2) must be provided using two related data files. The related data files are a master file and a payment (or detail) file. The master file lists the top-level payer/payee details, containing one record for each individual. The detail file contains details of all itemized transactions.

Records in the detail file are linked to individual records in the master file through a payer/payee ID. This ID must be unique within each schedule related master file. The ID is used only to match records between CSV master and detail files; it is not saved in the form, transmitted to DOL or used for any other purpose once the import process completes. Records in the detail file must occur in the same order as the individual records in the corresponding master file. The following shows an example of these structures.

sched25.csv (master file)

```
id,name,title
p1,"John Jones","lobbyist"
p2,"Mary Smith","benefactor"
p3,"Don Johnson","office manager"
```

sched25-payments.csv (detail file)

```
id,purpose,amount,date
p1,"Political representation",50,2002-10-31
p1,"Organizational duties",100,2003-01-25
p1,"Educational assistance",75,2003-02-26
p2,"Grants and gifts",100,2002-06-3
p2,"Petty cash",100,2002-06-18
p3,"Office supplies",50,2002-08-15
```

3.2 CSV Record Formats

Each importable schedule has a corresponding CSV format specification indicating the record fields and their types. Schedules with more than one type of item, such as Schedules 5 and 6, have multiple associated CSV formats, one for each type of schedule item.

CSV field types are based on the standard XML Schema data types, as defined in http://www.w3.org/TR/xmlschema-2/, augmented with additional simple types defined in the OLMS "common.xsd" schema (zip codes, state abbreviations, etc.). The names in the header record must match the name and order of the fields defined in the format specification.

The file named "csv-formats.xls" lists all the supported import schedules and their corresponding format specifications.

As with the XML format, data supplied in CSV import files must also conform to the additional validation rules defined in each form's User Guide (such as filing thresholds, explanatory notes, and other instructions).