



eScholar[®]

Bringing Data Together

Guide to Extracting Data

eScholar CDW[™] for PK-12

Version 2022

Last Updated: 2022-09-06

Table of Contents

INTRODUCTION	1
PURPOSE	1
DEFINITIONS, ACRONYMS, AND ABBREVIATIONS	1
CDW-PK12 DATA INTEGRATION TEMPLATES	2
OVERVIEW.....	2
TEMPLATE CHARACTERISTICS.....	2
THE INDEX WORKSHEET.....	2
THE INFORMATION WORKSHEET	2
INDIVIDUAL TEMPLATE STRUCTURE	3
TEMPLATE FILE FORMATS SUPPORTED	4
TEMPLATE FILE NAMING CONVENTION.....	4
TEMPLATE FILE DEPENDENCIES	5
CREATING EXTRACT FILES	6
RESOURCES AVAILABLE	6
DISTRICT CODE	6
SCHOOL YEAR.....	7
DATE FORMAT	7
FIELD LENGTHS.....	7
TEXT DATA.....	7
LOOKUPS.....	8
CREATING AN EXTRACT RECORD.....	ERROR! BOOKMARK NOT DEFINED.
HOW THE CDW-PK12 LOAD PLANS WORK.....	10
DETERMINING THE CORRECT PLAN.....	10
BASIC FORMAT CHECKING.....	10
DATA INTEGRATION CHECKING	10
DUPLICATE CHECKING	10
VALID VALUES CHECKING	11
OTHER DATA VALIDATIONS	11
DATA TRANSFORMATIONS.....	12
INSERT VS. UPDATE PROCESSING.....	12
ERROR, WARNING, AND ALERT LOGS.....	ERROR! BOOKMARK NOT DEFINED.

INTRODUCTION

PURPOSE

This document is intended to provide guidance on extracting data from source systems into eScholar CDW™ for PK-12 (CDW-PK12) Data Integration Template format. This information can be used to create individual extraction routines or a system of file extraction.

DEFINITIONS, ACRONYMS, AND ABBREVIATIONS

CEDS: Common Education Data Standards. CEDS is a project within the National Center for Education Statistics (NCES) to work with all relevant stakeholders to develop voluntary data standards for use by education institutions across Early Learning, PK12, Postsecondary, and Workforce.

FAQ: Frequently Asked Question

ISO Date Format: A date format defined by the International Standards Organization; the ISO Date Format is **YYYY-MM-DD** and is the required date format within much of CDW-PK12

NCES: National Center for Education Statistics. NCES is the primary federal entity for collecting and analyzing data related to education. NCES previously published online handbooks that offered guidance on consistency in data definitions and in maintaining data so that they can be accurately aggregated and analyzed; the Data Handbooks have been replaced by the CEDS project.

CDW-PK12 DATA INTEGRATION TEMPLATES

OVERVIEW

The CDW-PK12 Data Integration Templates provide the specifications that must be followed to create extract files that can be successfully processed by CDW-PK12 load plans; these load plans are executed via eScholar Data Manager™ (eDM).

TEMPLATE CHARACTERISTICS

- The templates are represented as worksheets in an Excel workbook
- There are two primary types of templates within one workbook:
 - *Master Lookup Definition* worksheet and *Master Lookup Code* Template
 - Data Warehouse Templates
- Each template maps directly to one or more tables, although most templates map to a single table

THE INDEX WORKSHEET

- The Index worksheet contains the list of templates included in the workbook and the order they are presented in the workbook
- The template order is alphabetical with one exception: the *Master Lookup Code* template and accompanying *Master Lookup Definition* worksheet are the first worksheets after the *Information* worksheet and before the alphabetical set of Warehouse templates

THE INFORMATION WORKSHEET

- The Information worksheet contains general information about the templates and a description of the contents of each template worksheet. The following are described:
 - CEDS Information linkage
 - Minimum Compatible Version defined
 - **Code** column values explained
 - Overview of file formats supported
 - Explanation of Load Sequence information
 - File naming convention defined (see [Template File Naming Convention](#) section below)

INDIVIDUAL TEMPLATE STRUCTURE

- Each template includes the following header information
 - Template Group and Version
 - Minimum Compatible Version
 - Best Practices Data Domain(s)
 - Target Table(s)
 - If more than one table is a target of the template, the primary target table is identified with an appendage of **(Primary Target)**
 - Primary Target Table Type – generally Fact, Dimension, or Lookup
 - eDM File ID/JSON Template Name
 - the root value of the file name for eDM (see Template File Naming Convention section, below)
 - the JSON Template Name for clients using web services via the eDM JSON API

- Each template may include the following columnar information
 - Field Number
 - Delim Length, Scale
 - Scale and Data Type information:
 - Scale of any decimal columns
 - Scale = 0 for integer columns
 - If there is no information in this column, the field could be character or date
 - Field Name
 - Note - All date fields have a field name ending in **DATE**
 - Code:
 - **K** - field is a component of the logical key
 - **U** - fields is updateable through the load plan based on matching logical key
 - **M** – field is mandatory and must be supplied
 - **D** – a value is required but a default is supplied
 - **F** – value is inserted once into the table; no updates allowed at all
 - **N** – once a non-null value is supplied, no updated allowed; update is allowed from null to non-null
 - **O** – field is optional
 - **OBS** – field is obsolete and should not be supplied
 - **N/A** – field is not applicable
 - **X** – field is only used for identification purposes and is not stored
 - **Z** – field is associated with special processing, defined in the Rule section
 - How the Code values affect the building of an extract file is discussed further in the *Creating a Record* section below
 - For a complete list of Codes and definitions, see the Information worksheet
 - CEDS Element – the relevant CEDS Element ID(s), if applicable
 - Lookup Name
 - Sample Value

- Each template includes a Rules section

- Any relevant extract rules that must be followed in creating the extract data file are defined
- If any element has a matching CEDS Element ID, a rule references the appropriate version of CEDS
- All Code column values used in any template are defined in one or more rules
- Specific file formats supported by the template are identified in the last extract rule in each section (see [Template File Formats Supported](#) section below)
- Each template includes a Load Sequence section
 - This section defines any pre-requisite files that must be loaded prior to loading the current template-formatted file (see [Template File Dependencies](#) section below)

TEMPLATE FILE FORMATS SUPPORTED

- Each template contains an extract rule on the specific file formats that are supported for that template
 - This extract rule is generally the last rule within the Rules section
 - Starting with v2021, only delimited formats are supported; support for fixed width files has been withdrawn
 - The following delimiters are supported:
 - Comma
 - Tab
 - Additional delimiters may also work but have not been formally tested
- A general recommendation is to create extract files in tab-delimited or comma-delimited format. These formats are supported by every template.
- In a state-level implementation, a standard file format may have been selected

TEMPLATE FILE NAMING CONVENTION

- The Naming convention for template-formatted files is as follows:
 - DistrictCode_EDMFileID_YYYYMMDDHHMM.xxx
 - District Code is the state-supplied or eScholar-supplied value
 - EDMFileID is identified on line 5 of each template
 - One value per template assigned by eScholar
 - The Date and Time information is required to handle processing more than one file for the same template per day
 - xxx represents the file extension.
 - csv – comma-delimited
 - tab – tab-delimited
 - del – other delimiter
- The naming convention is described in the [Information worksheet](#) within each template workbook.
- The extracted file *must* conform to the naming convention in order to be processed by eDM.

TEMPLATE FILE DEPENDENCIES

- Files may have dependencies
 - Dependencies are noted at the bottom of each template in the **Load Sequence** Section.
 - Some dependent tables are optional – if this is the case, the table will be identified as optional in the **Optional** column within the Load Sequence Section
 - By reviewing the dependencies, extract developers can come up with a prioritized list of when to work on each template in an initial implementation
 - Note that eDM *automatically* handles the load sequence when multiple template-formatted files are submitted together
 - Also note that not all files within the sequence need to be supplied together; as long as required pre-requisite files have been loaded previously, they need not be supplied in conjunction with the template file of interest

CREATING EXTRACT FILES

RESOURCES AVAILABLE

- Data Dictionary
 - If available, a Data Dictionary provides *client-specific* information on the elements to include in the extract files
 - The Dictionary may contain a definition of each element to be collected, including any option sets
 - The definitions may refer to existing data collections
 - The Dictionary may also include the mapping of the elements to the CDW-PK12 template name and field name
- Template Field Names
 - CDW-PK12 Template Field Names can be used for an initial level of understanding of the data represented by the field, especially in the absence of a Data Dictionary
- CEDS Element IDs
 - Each template contains references to CEDS Element IDs, where applicable, in the **CEDS Element** column. If there is a match between a CEDS definition and the definition of the CDW-PK12 template field, the CEDS Element ID is provided. More than one CEDS Element may be a match.
 - Note that the CEDS alignment may be *Identical* or *Similar*
 - The appropriate version of CEDS to use as reference for the Element ID is provided in the Rules section of each template, usually in the third or fourth to last rule.
- CDW-PK12 Best Practices
 - eScholar's CDW-PK12 **Best Practices** documents contain *general* field by field descriptions for all elements in all templates. Usage recommendations are also included for many fields.
 - The documents are divided up into Data Domains
 - The Overview section of each document contains a list of the templates included in that document
 - Some templates are associated with more than one Domain. For example, the *Staff Attendance* template is included in both the *Staff* Domain and the *Attendance* Domain. The information is identical in each section.
 - FAQs are provided, if applicable.
 - When combined with CEDS linkage information, the client can understand the subtleties of an *Identical* CEDS alignment vs. a *Similar* CEDS alignment

DISTRICT CODE

- This value is a required field in every record of virtually every template-formatted file
- This value is supplied directly by the State Education Agency or may be supplied by eScholar using state-specific guidance for district-level implementations

- Any extraction tool must have this value available to insert into each record, if the value is not already stored in the source system
- The value may represent the Local Education Agency or, in some cases, the State

SCHOOL YEAR

- Many templates contain a **School Year Date** field.
- A School Year in CDW-PK12 begins on July 1st and ends on the following June 30th
- The School Year is represented as the last day of the year, i.e., June 30th.
- For example, the School Year that begins on July 1, 2021, and ends on June 30, 2022, would be represented by the value **2022-06-30** and would be referred to as the **2022 School Year**.
- Since School Year Date is a date field, the date format must be ISO: YYYY-MM-DD – See [Date Format](#) Section below

DATE FORMAT

- Any date field include in any template-formatted record must be in ISO date format: **YYYY-MM-DD**. This includes the full ten characters, including the dashes between the components.
- All date fields are identified on the templates with a Field Name ending in **DATE**
- Each template with a Date field also includes the following Rule in the Rules Section:
 - Date fields (field name ends in DATE) must be in ISO format: YYYY-MM-DD

FIELD LENGTHS

- Field lengths are identified in each template in the **Delim Length,Scale** column
- eDM File Manager checks each field for length violations. File Manager rejects any record which contains data in a field that is greater than the maximum allowed length. Text qualifiers are not included in the calculation of the input data field length (see below for a discussion of [Text Qualifiers](#)).
- When creating a delimited file, there is no need to use all of the positions within a single field – it is *not* necessary to pad with blanks, for example
- Both leading spaces and trailing spaces are *always* trimmed by eDM and are not factored into the determination of the field length

TEXT DATA

- Text Data

- For ease of reading reports generated from the data warehouse, eScholar recommends that certain text fields, such as City Address fields, be formatted with Initial Capitals where appropriate
- Text Qualifiers
 - In particular for clients using the comma-delimited format, it is important to use an appropriate text qualifier to handle situations where the delimiter is part of the field value. For example, some name fields are designated to be in the form *Last Name, First Name*, where a comma is part of the field value. Enclosing the value within a qualifier such as a double quote (") enables the value to successfully load.
 - eScholar highly recommends using the double quote ("), rather than the single quote (') as the text qualifier.

LOOKUPS

- CDW-PK12 load plans use Lookups to validate data and to translate codes into a uniform set of descriptions (e.g., the Race or Ethnicity Lookup translates Race Code values such as **1** or **A** into uniform descriptions, such as **African American** or **Asian**).
- A single Master_Lookup Table contains the code sets for all Lookup Names
- The *Master Lookup Code* Template is included in the Data Integration Templates workbook
- An Extract process may need to create Master Lookup files, in addition to warehouse files
 - Note - once data is *initially* loaded into Master_Lookup the eScholar Utility **Extract Data for New School Year** may be used to create a *Master Lookup Code* template-formatted file for a new School Year. See the **Reference Guide – eScholar Utility Plans** document for more information
- Template fields that are associated with Lookups include the name of the appropriate Lookup in the **Lookup Name** column in the template
- Some Lookups are only used within eDM validation processes. These Lookups are identified as **(eDM Only)** in the Lookup Name column in the template
- District-level extract files sent to a state-level data warehouse may need to translate local code sets to the state-defined code sets. In this case, the Master Lookup table is loaded with the state-defined code sets so the extract routine must perform the translation from local to state codes and then the CDW-PK12 load plans translate the state codes to state descriptions.
- See [Valid Values Checking](#) Section below for additional information on how the load plans handle Lookups

CREATING A RECORD IN AN EXTRACT FILE

- When creating a record in an extract file, each field that is coded with a **K** is treated as part of the logical key. It is important to understand the logical key when constructing a record as this affects both [Duplicate Checking](#) and [Insert vs. Update](#)

[processing](#) (see sections below within **How the CDW-PK12 Load Plans Work** for explanations).

- Some fields may be *optionally* treated as part of the logical key. These fields are generally identified with a Code of **KU**. A parameter controls whether the field should be treated as part of the logical key and cannot be update (**K** processing) or whether the field should *not* be treated as part of the logical key and *can* be update (**U** processing)
- See the **Load Plan Parameter Guide** for a description of these parameters and see the relevant **Best Practices** document for a more detailed description of how this works, where appropriate
- When creating a record, each field coded with an **M** (denoting a mandatory field) is required by eScholar and *must* be supplied. The CDW-PK12 load plans reject records where **M** fields are not supplied. Note that clients can specify their own Mandatory fields by using the eDM File Manager *Required* validation, which can vary by field from one collection to another. Also note that, for eScholar-required fields, the *Required* validation can *never* be turned off.
- When creating a record based on a template, it is *not* necessary to populate every non-mandatory field. Since CDW-PK12 is based on a single model there will always be fields that do not apply to a given client.
- However, each field from the template must be *accounted for* in the record. So, when building a delimited file, there must be a delimiter for each field in the template, even if there is no value in between delimiters.

HOW THE CDW-PK12 LOAD PLANS WORK

DETERMINING THE CORRECT PLAN

- The file is matched up with the appropriate load plan based on the standard file name
 - See the [Template File Naming Convention](#) section above for a description of the file naming convention. The same description is included in [the Information worksheet](#) of the Data Integration Template workbook.

BASIC FORMAT CHECKING

- Are the field lengths correct? – if any field would be truncated then the entire record is rejected
- Are the data types correct? – for example, if there is any text data in a numeric field, then the entire record is rejected
- Are dates valid? – if any date field supplied is an invalid date (e.g., February 30th), the entire record is rejected
- Are the date formats correct? – if any date field supplied in a template-formatted file is *not* in ISO format, the entire record is rejected

DATA INTEGRATION CHECKING

- If any key fields or other fields need to be present in a parent table, the load plan performs the necessary integration checks.
 - For tightly coupled tables, the integration check is strict. Any integration check that fails causes the record to be rejected
 - For loosely coupled tables, the integration check is not as strict. In these cases, the client may designate how tightly to integrate the tables. In some load plans, there is no configurable option - the record is processed but a warning message is generated if the integration check fails.
- The integration checks that apply to each field are described in detail in the **Best Practices** documents.

DUPLICATE CHECKING

- The Data Loading process checks for the presence of duplicate records *within each input file*.

- This check is based on the components of the logical key – the fields identified with a **K** in the Code column of the template
- Note that this could mean that two records that are *not* identical could fail the duplicate logic test if their *logical keys are identical*
- When duplicates are detected, *all* records in a duplicate set are rejected
- All records involved in a duplicate set are written to the Error table for client review.
- Please note that an input file record with a matching record in the database is *not* a duplicate situation – this is an update situation (see [Insert vs. Update Processing](#) Section below).

VALID VALUES CHECKING

- Each field associated with a Lookup (unless the Lookup is labeled *eDM Only* in the Template) may optionally be validated in the load plan using one of the following options:
 - **REJECT** – record is rejected if the input code does not match a lookup value
 - **REJECTNULLOK** - record is rejected if the input code does not match a lookup value but nulls are not rejected
 - **WARN** – record is still processed if the input code does not match a lookup value but a warning log record is generated
 - **WARNNULLOK** - record is still processed if the input code does not match a lookup value but a warning log record is generated; no warning is generated for null values
 - **ACCEPT** - record is still processed if the input code does not match a lookup value and no error or warning log record is generated
- Note - eDM File Manager also provides the ability to set *Valid Values* checking for fields with Lookups – this validation occurs prior to Data Loading but does not allow for all of the options above
- The load plan validation option is set separately for each field with a Lookup and for each execution of the load plan
- See the **eScholar Reference Guide – Valid Values** document on the Support Site for more information

OTHER DATA VALIDATIONS

- Other types of data validations may be executed by the load plan:
 - Data comparisons – e.g., that an end date is greater than or equal to a start date
 - These data validations are shown as Rules in the template and are described in detail in the **Best Practices** documents

DATA TRANSFORMATIONS

- For each record that passes the basic format checking, the load plan performs any necessary transformations
 - Lookups – codes are validated and may be translated to descriptions
 - Derived columns are created – e.g., Student_Name is derived from individual Last Name and First Name components

INSERT VS. UPDATE PROCESSING

- Once a record has passed all the required checks it is ready for database processing. There are two kinds of processing: Insert and Update
- The input record is compared to the target table to see whether the record already exists in the table. This is performed based on logical key – the fields in the template where Code = **K**.
- If *no match is found*, the record is new and *Insert* processing is performed
 - The record is prepared and then inserted into the warehouse. This may be done in batches for enhanced performance
- If *a match is found*, the record exists in the table and *Update* processing is performed
 - Each field in the input file record is compared to the matching table record to determine whether any values have changed. Only fields identified with a **U** in the template are checked. If any updateable field has changed, the table record is updated appropriately.
 - If no updateable fields have changed, no further processing is done
 - The load plan uses a hashing technique to quickly compare all the relevant fields values to target table column values
- Delete processing
 - In general, the load plans do *not* include Delete processing; only Insert and Update processing described above
 - For example, there is no ability to exclude a record from an input file and trigger delete processing in this way.
 - Delete processing is included in the **eScholar Record Delete Utility™** module (RDU™)
 - This exists both as a stand-alone Utility and as a process integrated with Data Loading
 - e.g. You may set up eDM to always Delete from Programs_Fact based on the matching District and Collection just prior to loading a Programs Fact file

ERRORS AND WARNINGS

- The CDW-PK12 load plans produce *Error* information when records are rejected based on any of the checks described above.

- The CDW-PK12 load plans produce *Warning* information when records are processed but violations are detected.
- All *Errors* and *Warnings* are written to the Error Table and displayed in eDM for client review