



Form 5500 Datasets Guide

GENERAL INFORMATION

Generally, certain types of private- sector employee benefit plans file a Form 5500 or 5500-SF, along with required schedules and attachments, annually to satisfy the Employee Retirement Income Security Act (ERISA) and the Internal Revenue Code.

The Data Dictionaries, provided for each form year's dataset, map the fields in the datasets to line items on the Form 5500, 5500-SF, and schedules.

The Form Year Changes document, provided for each year's dataset, summarizes changes to the specific year's form version compared with the previous form version.

Information regarding the content of a filing and who must file is described in the [Form 5500/5500-SF instructions](#).

SCOPE

This document describes the data structures and relationships of 2009 and later datasets. If you need assistance with data structures and relationships of 2008 and earlier datasets, please contact the EBSA Public Disclosure Room at (202) 693-8673.

ALL VERSUS LATEST DATASETS

While applicable plans are required to file a Form 5500 or Form 5500-SF annually, multiple filings may exist for a plan's single form year if a plan amends their original filing or if a duplicate filing was erroneously submitted. To help address this, two sets of data are provided for each form year:

- The dataset listed as "All" includes all filings received by the Department for that form year and may contain multiple filings for a single plan due to amended or duplicate filings. Valid filing statuses in the "All" datasets include "FILING_RECEIVED", "PROCESSING_STOPPED", and "FILING_ERROR".
- The dataset listed as "Latest" includes only the latest filing received for a plan with a filing status of "FILING_RECEIVED".
- For more information on filing status, see [EFAST2 FAQ 38](#).

The datasets are provided as zipped CSV files – comma delimited Excel files with "(double quote)" as the text qualifier. The first row of each spreadsheet contains the field names. In addition to filer-provided data, each dataset includes three system-generated fields: filing status, date received, and an indication of whether there is a valid signature.

A “File Layout” document is provided for each CSV file, and includes data type detail for each field. Fields may be designated as Text or Numeric. Dates are included as Text fields and are formatted as mm/dd/yyyy. Numeric fields with names ending in _CNT are typically whole numbers, while Numeric fields with names ending in _AMT typically contain 2 decimal places.

FORM YEAR CHANGES – REMOVED FIELDS

Fields may be removed from the face of the forms and schedules as part of the annual form development cycle. Fields that are removed will continue to be maintained in the structure of the dataset file layout, but will contain blanks for all future form years. Retaining the removed fields in the file layout ensures consistency across years to minimize changes needed to scripts and other programmatic processing developed by consumers of the datasets.

RELATIONSHIP OF FORM 5500/5500-SF TO ASSOCIATED SCHEDULES

Filing data from the main Form 5500 and 5500-SF are contained in tables “F_5500_[year]” and “F_5500_sf_[year]”, respectively, and a filing can have one or more schedules associated to the main Form 5500 or 5500-SF.

Each Form 5500/Form 5500-SF submission, and its schedules, is uniquely identified by an acknowledgement ID (ACK_ID field). The ACK_ID field is the first field noted in each dataset table, and is the best way to connect the data from a single filing submission to the various form and schedule tables.

Some schedules may require joining multiple tables to capture their complete data. For example, a given ACK_ID in table F_5500_2015 may have associated records in tables F_SCH_C_2015, F_SCH_C_PART1_ITEM1_2015, F_SCH_C_PART1_ITEM2_2015, and F_SCH_C_PART1_ITEM3_2015. The data noted in the tables may be related to the actual forms and schedules via the Data Dictionaries listed at the top of each data set.

UNIQUE IDENTIFIERS

ACK_ID = Unique identifier assigned to each filing.

FORM_ID = Unique identifier for a schedule when there are multiple schedules per filing.

ROW_ORDER = Unique identifier for repeating rows/elements in a schedule.

CODE_ORDER = Unique identifier for each service code listed in a single row/element of Schedule C. Service codes are linked to service providers in Schedule C by the Row Order field.

Schedule A (Insurance information) A single filing may have multiple Schedule As attached.	If SCH_A_ATTACHED_IND=1 in the Form 5500/5500-SF file, the filing has Schedule A(s) attached. Similarly, the field NUM_SCH_A_ATTACHED_CNT can be used to determine the number of Schedule As attached to the filing.
Schedule C (Service Provider Information)**	If SCH_C_ATTACHED_IND=1 in the Form 5500/5500-SF file, the filing has a Schedule C attached.
Schedule D (DFE/Participating Plan Information)	If SCH_D_ATTACHED_IND=1 in the Form 5500/5500-SF file, the filing has a Schedule D attached.
Schedule G (Financial Schedules)	If SCH_G_ATTACHED_IND=1 in the Form 5500/5500-SF file, the filing has a Schedule G attached.
Schedule H (Financial Information for Large Pension Plans)	If SCH_H_ATTACHED_IND=1 in the Form 5500/5500-SF file, the filing has a Schedule H attached.
Schedule I (Financial Information for Small Pension Plans)	If SCH_I_ATTACHED_IND=1 in the Form 5500/5500-SF file, the filing has a Schedule I attached.
Schedule MB (Actuarial Information for multiemployer defined benefit pension plans)	If SCH_MB_ATTACHED_IND=1 in the Form 5500/5500-SF file, the filing has a Schedule MB attached.
Schedule R (Pension Plan Information)	If SCH_R_ATTACHED_IND=1 in the Form 5500/5500-SF file, the filing has a Schedule R attached.
Schedule SB (Actuarial Information for single employer or multiple employer defined benefit pension plans)	If SCH_SB_ATTACHED_IND=1 in the Form 5500/5500-SF file, the filing has a Schedule SB attached.

*All schedule data can be linked to the Form 5500 by matching the ACK_ID across files.

Schedule C Tables

Note that the following Schedule C tables contain a field for the service provider code:

F_SCH_C_PART1_ITEM2_[year] (PROVIDER_OTHER_SRVC_CODES)
F_SCH_C_PART1_ITEM3_[year] (PROVIDER_INDIRECT_SRVC_CODES)
F_SCH_C_PART2_[year] (PROVIDER_FAIL_SRVC_CODE)

However, the service provider code field in these tables is empty. Because multiple service provider codes are reported in this one field, three tables are used to provide the codes:

F_SCH_C_P1_ITEM2_CODES_[year]
F_SCH_C_P1_ITEM3_CODES_[year]
F_SCH_C_P2_CODES_[year]

Each of these three “codes” tables contain the identifier fields:

ACK_ID
ROW_ORDER
CODE_ORDER
SERVICE_CODE

To establish the relationship between the service provider information and the service codes, a one-to-many join will need to be done from the “parent” table to the codes table. For example, the 2009 Schedule C dataset joins would be:

(parent table)F_SCH_C_PART1_ITEM2_2009 (ack_id,row_order) -< (child)
F_SCH_C_P1_ITEM2_CODES_2009 (ack_id,row_order) sorted by ack_id,row_order, code_order

(parent table)F_SCH_C_PART1_ITEM3_2009 (ack_id,row_order) -< (child)
F_SCH_C_P1_ITEM3_CODES_2009 (ack_id,row_order) sorted by ack_id,row_order, code_order

(parent table)F_SCH_C_PART2_2009 (ack_id,row_order) -< (child) F_SCH_C_P2_CODES_2009
(ack_id,row_order) sorted by ack_id,row_order, code_order