

Claim-Segment LDS Record Structure for SAF Files

This document explains the Claim-Segment record structure of the LDS Standard Analytical Files (SAFs); this data structure is the one used to organize Medicare's Research Identifiable Files (RIFs). (Previous LDS data came in DESY Output View structure; for a description of this alternative LDS data structure see the accompanying document entitled "Relating DESY Output View LDS and Claim-Segment LDS Files".) A record in a Claim-Segment record structure comprises one segment of a claim. The vast majority of claims consist of only one segment or record; some of the institutional claims with complicated payment histories contain multiple segments/records.

Three sections make up this document. The first, "Reading Flat Files," describes the flat file structure of Claim-Segment records. The second, "Understanding Segmented Claims," explains why, in some cases, claims are divided into multiple segments or records, and illustrates how and when claim information is stored across multiple segments. The third section, "Managing Segmented Claims," explains how to exploit and integrate information from multi-segmented claims, which in some instances can be done by reducing claims to only one segment and in others can be accommodated by combining segments into a single record, or assigning Claim Numbers for future use.

Reading Flat Files

Claim-Segment LDS files are *flat files*, meaning the number of variables per record is fixed. That is, for a particular year and file type (such as Inpatient or Carrier), each record contains a common set and number of variables, but some variables may be blank.

Figure 1 illustrates the fixed variable record structure. Our example uses two claims. For the sake of simplicity, we show only one group of variables (in this case diagnosis codes) for each claim. The first claim, on the first line of Figure 1, has only two diagnosis codes (as noted in the "Diagnosis Count" variable); these codes are found in variables "DGNSCD01" and "DGNSCD02." The next eight variables DGNSCD03-DGNSCD10 contain blank values. In the second claim, on the second line of the table, there are ten diagnosis codes, so every diagnosis code variable is filled.

Figure 1
Example of Claim-Segment Record Structure

Segment Count	Segment Number	Diagnosis Count	Fixed Variables	DGNSCD01	DGNSCD02	...	DGNSCD10
1	1	2	...	1582	1546		
1	1	10	...	3031	3034	...	8625

Research Data Distribution Center

Acumen, LLC
500 Airport Blvd., Suite 365
Burlingame, CA 94010
650-558-8310
RDDC@acumenllc.com



Understanding Segmented Claims

In some cases a claim is segmented into two or more records. This happens when there are more values in a particular variable group than available spaces. Sometimes a claim will be segmented while showing empty variables in the first record; this happens because Claim-Segment LDS files are built to accommodate different versions of CMS data with inconsistent numbers of variables stored into a single record. We explain this possibility further in the example below.

A segmented claim is different from a non-segmented claim in two main ways. First, the variables “Segment Count” and “Segment Number” distinguish segmented claims. “Segment Count” indicates the total number of records that make up the claim, and “Segment Number” identifies the particular record. Second, all segments of a single claim have the same values in the “fixed” variables, but each segment has unique values for all other variable groups. The first segment holds as many values for a particular variable group as possible, while subsequent segments hold only new values for that group.

To illustrate segmented claims, consider two claims with “Revenue Center” variable groups; one is segmented, one is not. Assume that the first claim has 46 Revenue Center codes while the second claim has 30 Revenue Center codes. Because the maximum number of revenue codes that CMS ever stored in a single claim is 58, the Claim-Segment LDS files contains 58 variables to store revenue codes. As noted above, the presence of 58 variables does not guarantee that 58 revenue code values can be stored in a given record; the actual number of values per record depends on the version of CMS data being used. In this example, only 45 Revenue Center codes are allowed in one record. Thus the first claim, with 46 Revenue Center codes, is segmented into two records: the first contains 45 of the revenue codes, and the second contains one revenue code.

Figure 2 shows these two claims. The first variable in Figure 2, “Segment Count,” indicates that the first claim has two segments, which are identified by “Segment Number.” The “Revenue Center Count” variable gives the number of non-blank variables in this group; for the first record this has the value 45, but for the second record (segment 2 of the segmented claim) it is only 1. All other “fixed” variables, including demographic and identifier variables, are identical in the first two records. The second claim (the third line of the table) is not segmented; “Segment Count” and “Segment Number” both equal 1.

Research Data Distribution Center

Acumen, LLC
500 Airport Blvd., Suite 365
Burlingame, CA 94010
650-558-8310
RDDC@acumenllc.com



Figure 2
Example of Segmented Claims

Segment Count	Segment Number	Revenue Center Code	Fixed Variables	Rvctr01	Rvctr02	...	Rvctr45	...	Rvctr58
2	1	45	...	1010	2040	...	9010		
2	2	1	...	9080					
1	1	30	...	0121	0890	...			

Managing Segmented Claims

This section describes three ways to manage the segmented claims in your data. The first subsection below identifies circumstances under which a researcher can ignore information contained on all claim segments above the first segment, implying that 2nd and higher order segments can be eliminated from research files. The second subsection explains how to combine all claim segments into a single record for immediate use. The third subsection shows how to assign Claim Numbers to records, which would make it possible to combine claim segments in the future.

Before you begin processing your data, it is important that you keep your data in the same sort order in which it was received. The Claim-Segment LDS SAFs were shipped to you with all records of each claim contiguous to each other and in order. As explained above, the Segment Count variable (SGMT_CNT) gives the number of segments (records) in the claim and the Segment Number variable (SGMT_NUM) numbers each record of the claim consecutively. Taken together with the order of the records in the file, these variables allow you to identify which records constitute a single claim. Once you have used one of the methods described below to manage your segmented claims, you will be able to re-sort your data.

Disregarding Unnecessary Claim Segments

Many LDS applications can be performed using only the information from the first segment of a claim. In such instances, 2nd and higher order claim segments can be deleted from created research files to avoid confusion. Fewer than one percent of the Hospice, Inpatient, Outpatient, and Skilled Nursing Facility SAFs claims have two or more segments. For those claims that do have more than one segment, only the Revenue Center variable group ever contains new values in the 2nd and higher segments. Carrier and Durable Medical Equipment (DME) LDS SAFs never contain claims with more than one segment.

Research Data Distribution Center

Acumen, LLC
500 Airport Blvd., Suite 365
Burlingame, CA 94010
650-558-8310
RDDC@acumenllc.com



In short, you can disregard all but the first segment of a claim when:

1. You are working only with Carrier or DME files, or
2. You are not using information about revenue center codes in your analysis.

If (1) or (2) apply, then you may reduce your data sets to one segment per claim by keeping just the first segment of each claim, i.e., keep just the records with Segment Number (SGMT_NUM) equal to 1.

Combining Segments into a Single Record

If your data has claims with multiple segments and you want to use all information stored on the Revenue Center data, then we recommend that you combine multiple segment claims into a single record. As explained above, all records of each claim are currently contiguous to each other and in order. The Segment Count variable (SGMT_CNT) and the Segment Number variable (SGMT_NUM) taken together with the order of the records in the file allows you to identify which records constitute a single claim.

The general strategy is to use the first segment of each claim as the starting point and to add on variables to store the additional Revenue Center values from the second and higher segments. This strategy can be accomplished in X steps. First, start with all the variables in the first record of the existing claim—the record with SGMT_NUM = 1. This segment contains all of the non-Revenue Center variables that will be needed for the combined claim record, as well as the first set of Revenue Center variables¹. Second, create additional variables to hold the Revenue Center data from the second and higher segments. Because CMS stores a maximum of 450 occurrences of each Revenue Center variable (although in practice this limit is rarely if ever reached), you should create variables numbered up to 450 for each type of Revenue Center variable. For example,

¹ Revenue Center Group Variables include: RVCNTR01–RVCNTR58, REV_DT01–REV_DT58, APCPPS01–APCPPS58, HCPSCD01–HCPSCD58, MDFCD101–MDFCD158, MDFCD201–MDFCD258, MDFCD301–MDFCD358, MDFCD401–MDFCD458, MDFCD501–MDFCD558, PMTTHD01–PMTTHD58, DSCTND01–DSCTND58, PCKGND01–PCKGND58, PRICNG01–PRICNG58, OTAF_101–OTAF_158, IDENDC01–IDENDC58, RVUNT01–RVUNT58, RVRT01–RVRT58, RVBLD01–RVBLD58, RVDTBL01–RVDTBL58, WGDJ01–WGDJ58, RDCDCN01–RDCDCN58, RVMSP101–RVMSP158, RVMSP201–RVMSP258, RPRPMT01–RPRPMT58, RBNPMT01–RBNPMT58, PTNRSP01–PTNRSP58, REVPMT01–REVPMT58, RVCHRG01–RVCHRG58, RVNCVR01–RVNCVR58, RVDDCD01–RVDDCD58.

Research Data Distribution Center

Acumen, LLC
500 Airport Blvd., Suite 365
Burlingame, CA 94010
650-558-8310
RDDC@acumenllc.com



add variables RVCNTR59 through RVCNTR450 and REV_DT59 through REV_DT450. You can adjust the number of variables downward from 450 depending on the actual maximum number of segments that occurs in your data. Third, once you have created these additional variables, you will use values from the second and higher segments to fill in their values. If the claim has multiple segments, only the first 45 variables in each segment are used and variables 46 through 58 contain no data. Thus the first segment contains occurrences 1 through 45, the second segment contains occurrences 46 through 90, the third segment contains occurrences 91 through 135, and so on. To fill in values for additional revenue variables, take the Revenue Center values in the second segment of the Claim-Segment file, and use them to fill in variables 46 through 90. For example, assign the value of HCPSCD01 in the second segment to HCPSCD46 in the new combined record, assign the value of HCPSCD02 to HCPSCD47, and so on, all the way up to assigning HCPSCD45 to HCPSCD90. Proceed through each segment of the claim, assigning the Revenue Center values in the third segment to variables 91 through 135, assigning values from the fourth segment to variables 136 through 180, and so on.

Creating Claim Numbers for Later Use

An alternative to combining multiple segmented claims into one record is to create a Claim Number. This method allows users to retain the possibility of integrating information from multiple segments, while leaving the structure of the LDS Claim-Segment file intact. Because LDS Claim-Segment files are already sorted by DESY SORT KEY, you can identify a new claim every time the Segment Number takes the value of one. Simply start a new Claim Number each time Segment Number has the value of one. Once each claim segment is identified by Claim Number, you can feel free to re-sort your data. The example below shows how to create a Claim Number using SAS.

Using SAS, perform the following steps to create a Claim Number variable:

1. Set the file 'by DESY_SORT_KEY'.
2. If `SGMT_NUM=1` then Claim_No+1;

Figure 3 illustrates how Claim Numbers are assigned with this method

Research Data Distribution Center

Acumen, LLC
500 Airport Blvd., Suite 365
Burlingame, CA 94010
650-558-8310
RDDC@acumenllc.com



Figure 3

Claim Number	Claim Total Segment Count	SGMT_NUM
1	2	1
1	2	2
2	1	1
3	1	1
4	3	1
4	3	2
4	3	3