

Cost Report Database Format

The Research Data Assistance Center (ResDAC) is a CMS contractor that provides free assistance to academic and non-profit researchers interested in using Medicare and/or Medicaid data for their research. ResDAC is staffed by a consortium of epidemiologists, public health specialists, health services researchers, biostatisticians, and health informatics specialists from the University of Minnesota.

*TN-005
October 2003
Updated March 2007
Author: Faith M. Asper*

Prior to April 30, 2002, CMS offered researchers all cost report files (the complete cost report master file in addition to subset files) in a flat file format. After April 30, 2002, CMS offers the complete provider cost report master file in a relational database format.

The purpose of this technical publication is to explain the cost report relational database format. The change from a flat file to a relational database has allowed CMS the ability to provide this information as a downloadable public use file.

The cost report data files are now available for hospitals, skilled nursing facilities, renal dialysis facilities, hospices, and home health agencies. Information about the downloadable cost reports is found on the CMS cost report web page at http://www.cms.hhs.gov/CostReports/01_Overview.asp#TopOfPage.

Cost Reports in a Relational Database Format after April 30, 2002

A relational database is a database that is organized into tables. Each table contains like or related data elements. A commonly known software application that uses a relational database model is Microsoft Access. Advantages of a relational database are that researchers can easily exclude or perform cross-sectional analysis¹.

CMS has categorized each variable into one of three groups, a report variable, an alpha-numeric, or a numeric variable. The report variables are contained within the Report table and it contains descriptive information, such as the record id, provider id, and the date that the cost report was submitted (See Figure 1 for a complete list of variables in the Report table for the Hospital cost reports). The Alpha table contains character and numeric values, such as provider address, urban/rural designation, and the date that the hospital began operation. The Numeric table contains only numbers and would include variables such as number of discharges, total number of beds, and total cost by cost center.

The unique identifier that links records across the cost report tables is the Report Record Number. A unique report record number is assigned to each cost report that is recorded in the Report table. The Report table contains one record for every cost report filed by a provider per year whereas the Alpha and Numeric tables have

¹ Readme.doc (version June 2008) from the CMS Hospital Cost Report website at <http://www.cms.hhs.gov/CostReports/Downloads/HospReadme0608.pdf> retrieved on 8/27/2008.

multiple entries with the same report record number, which indicates that these variables all belong to the same cost report.

Figure 1: Cost Report Relational Data Tables

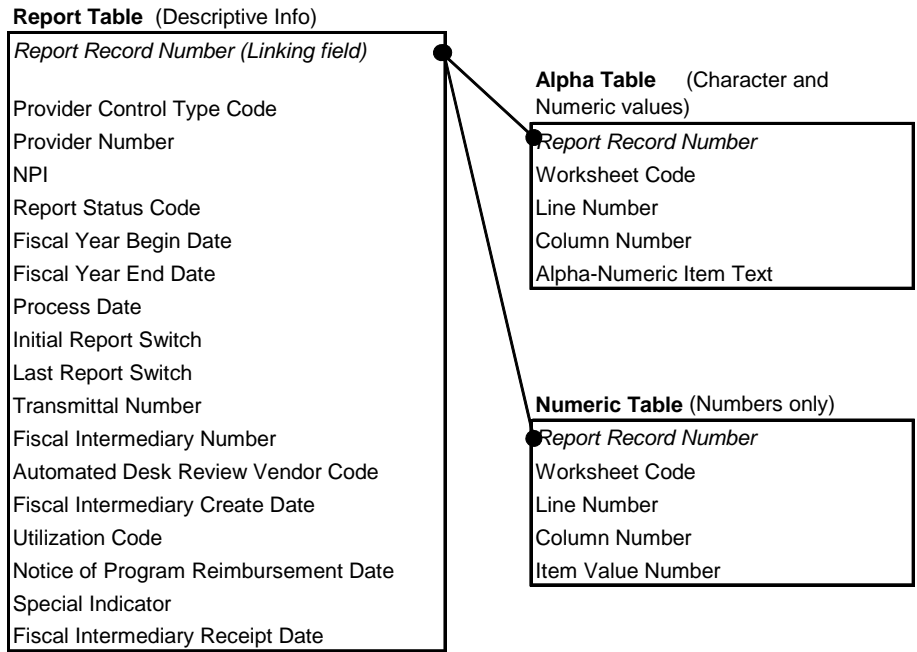


Figure 2 is an example of how the data appear in the Alpha file.

Figure 2: Illustrative Example of Alpha Table contents

Report Record Number	Worksheet Code	Column Number	Line Number	Alpha Item Text
9876	S200000	0100	02200	Y
9876	S200000	0400	00101	LOS ANGELES
9876	S300001	0000	02600	00600INTENSIVE CARE UNIT
9876	S300001	0000	02700	00700CORONARY CARE UNIT
9876	S300001	0000	02800	00800BURN INTENSIVE CARE UNIT

Creating Subsets of the New Relational Database and Finding Worksheet Variables

Because the new relational database contains the highest status5 cost report submitted electronically for all providers, some of the data files are quite large, and often researchers are interested in creating subsets of the cost report data by provider or fiscal year before working with the variables. If, for example, a researcher wanted to create a subset of the cost reports for a group of providers, the researcher would use the Report table and extract the providers of interest based on their Medicare provider number. Once the providers were selected from the Report

table, this subset could be linked to the Alpha-Numeric and/or Numeric tables using the report record number. This would yield a subset of the cost report data containing all worksheet variables for all fiscal years for the particular providers of interest.

In order to find a specific worksheet variable, the researcher needs to have an understanding of where this variable is found within the worksheets that providers submit to CMS. For example, if the researcher wanted to know whether or not a provider is a referral center (a Yes/No field), the researcher would go to the worksheets that providers complete and find the location of the variable. Figure 5 shows that the value for Referral Center is located on Worksheet S-2, column 1, line 22. Next, the researcher needs to determine that the Referral Center value is an alpha-numeric variable and therefore, it would be found in the Alpha-Numeric table.²

Figure 3: Worksheet S-2 from the Hospitals and Health Care Complexes Form 2552-96

3690 (Cont.) FORM CMS-2552-96 0

HOSPITAL AND HOSPITAL HEALTH CARE COMPLEX IDENTIFICATION DATA		PROVIDER NO.:	PERIOD: FROM _____ TO _____	WORKSHEET S-2 (CONT.)		
---	--	---------------	-----------------------------------	------------------------------	--	--

Hospital and Hospital Health Care Complex Address:

1	Street:	x	P.O Box:	x		
1.01	City:	x	State:	x	Zip Code:	x
			County:	x		

Hospital and Hospital-Based Component Identification:

	Component	Component Name	Provider Number	Date Certified	Payment System (P, T, O, or N)		
					V	XVIII	XIX
	0	1	2	3	4	5	6
2	Hospital	x	x	x		x	x
3	Subprovider	x	x	x		x	x
4	Swing Beds-SNF	x	x	x		x	x
5	Swing Beds-NF	x	x	x			x
6	Hospital-Based SNF	x	x	x		x	x
7	Hospital-Based NF	x	x	x			x
8	Hospital-Based OLTC	x					
9	Hospital-Based HHA	x	x	x		x	x
11	Separately Certified ASC	x	x	x		x	x
12	Hospital-Based Hospice	x	x	x			
14	Hospital-Based Health Clinic (specify)	x	x	x		x	x
15	Outpatient Rehab. Clinic (specify)	x	x	x		x	x
16	Renal Dialysis	x	x	x			

17 Cost Reporting Period (mm/dd/yyyy) From: _____ To: _____

18 Type of Control (see instructions) **1** 2

Type of hospital/subprovider (see instructions)

19	Hospital	x	
20	Subprovider	x	

Other Information

21	Indicate if your hospital is either (1) urban or (2) rural	x	
21.01	Does your facility qualify and is currently receiving payment for disproportionate share in accordance with 42 CFR 412.106?	x	
21.02	Has your facility received geographic reclassification? Enter "Y" for yes and "N" for no. If yes, report in column 2 the effective date (mm/dd/yyyy). (See instructions.)	x	x
22	Are you classified as a referral center?	x	
23	Does this facility operate a transplant center? If yes, enter certification date(s) (mm/dd/yyyy) below.		
23.01	If this is a Medicare certified kidney transplant center, enter the certification date.		x
23.02	If this is a Medicare certified heart transplant center, enter the certification date.		x
23.03	If this is a Medicare certified liver transplant center, enter the certification date.		x

² Providers may submit more than one cost report, an initial report and then subsequent reports that would contain adjustments and revisions. Each cost report is assigned a status and the order is from 1 to 4 with 1 being the lowest. 1- as submitted, 2-settled without audit, 3-settled with audit, 4-reopened.

In the Alpha-Numeric table, this variable is denoted as Worksheet code S200000, column code 0100, and line code 02200. (Please see the Appendix for an explanation of the worksheet, column, and line coding schemes.)

Figure 6 gives an example of how the data would appear in the data tables for Provider Number 12345 for the Fiscal Year 1996 cost report. According to the Alpha-Numeric table, Provider Number 12345 submitted a value of "Y", denoting that it is a referral center.

Figure 6: How "Referral Center" would be identified in the cost report database

Report Table (Descriptive Info)

Variable Name	Illustrative Value
<i>Report Record Number</i>	9876
Provider Number	12345
Fiscal Year Begin Date	1/1/1996
Fiscal Year End Date	12/31/1996

Alpha Table (Character and Numeric Values)

Variable Name	Illustrative Value
<i>Report Record Number</i>	9876
Worksheet Code	S200000
Line Number	02200
Column Number	0100
Alpha-Numeric Item Text	Y

Summary

In the relational database format, the data are organized into tables and one row in the data files (alpha or numeric) represents one variable within the cost report. With the relational database format, the researcher needs to understand the cost report worksheet contents and variable types in order to extract the items of interest.

If you have any questions or comments, ResDAC staff can be contacted at 1-888-ResDAC or resdac@umn.edu

Internet Citation:

Cost Report Database Format. Technical Brief, ResDAC Publication Number TN-003, January 2003, Updated September 2008. Research Data Assistance Center, University of Minnesota, Minneapolis, MN. <http://www.resdac.umn.edu>.

Technical Document authored by Faith M. Asper, ResDAC staff.

Copyright © 2008 Regents of the University of Minnesota. All Rights Reserved.

The analyses upon which this publication is based were performed under Contract Number 500-01-0043, entitled, "Research Data Assistance Center (ResDAC)", sponsored by CMS, Department of Health & Human Services.