

Harvesting Metadata From SAS®

Exporting SAS Datasets to DDI 3 XML files - Joachim Wackerow and Larry Hoyle

In the SAS system a dataset exists in a kind of ecosystem. Metadata relevant to a particular dataset must be gathered from a variety of places, not just from the data table itself. Metadata are contained in user formats (which may exist only in SAS code), integrity constraints, dictionary tables, and SAS built-in formats (together with the documentation for those formats).

	RealName
1	Joe <In the know> Schmo
2	Bill Hill
3	Donna O'Fauna
4	Rob "Bob" Cobb
5	Tom Applebaum
6	Louise Mac&Cheese
7	Jack Black
8	Jill Hill
9	Gno Avocado
10	Poppy LaFleur
11	Pickett Wilson
12	Olive Hoyle

```
proc format;

value sexAll
0 = 'Young Male'
1 = 'Adult Male'
2 = 'Young Female'
3 = 'Adult Female' ;

value sex
0 = 'Male'
1 = 'Male'
2 = 'Female'
3 = 'Female';

value avocadonumber
low-<0 = 'avocados owed'
1 = 'lonley avocado'
1<-6.02214149e23 = 'too few avocados'
6.02214149e23-6.02214209e23 = 'guaca mole'
6.02214209e23<-high = 'a party';
```

Format name	Starting value for format	Ending value for format	Format value label	Minimum length	Maximum length	De
AVOCADONUMBER	LOW	0	avocados owed	1	40	
AVOCADONUMBER	1	1	lonley avocado	1	40	
AVOCADONUMBER	1	6.02214149E23	too few avocados	1	40	
AVOCADONUMBER	6.02214149E23	6.02214209E23	guaca mole	1	40	
AVOCADONUMBER	6.02214209E23	HIGH	a party	1	40	
BMI	LOW	18.5	Underweight	1	40	
BMI	18.5	24.9	Normal weight	1	40	
BMI	25	29.9	Overweight	1	40	
BMI	30	HIGH	Obesity	1	40	
ONE	1	1	One	1	40	
SEX	0	0	Male	1	40	
SEX	1	1	Male	1	40	
SEX	2	2	Female	1	40	
SEX	3	3	Female	1	40	
SEXALL	0	0	Young Male	1	40	
SEXALL	1	1	Adult Male	1	40	
SEXALL	2	2	Young Female	1	40	
SEXALL	3	3	Adult Female	1	40	
STEXT	F	F	Girl	1	40	
STEXT	FEMALE	FEMALE	Girl	1	40	
STEXT	Female	Female	Girl	1	40	
STEXT	M	M	Boy	1	40	
STEXT	MALE	MALE	Boy	1	40	
STEXT	Male	Male	Boy	1	40	
STEXT	f	f	Girl	1	40	
STEXT	m	m	Boy	1	40	

Foreign Key integrity constraints

User formats – in SAS code

SAS Cntlout/Cntl in Dataset based on user formats

VIEWTABLE: Test Data for SAS to DDI 3 program									
	ID	name	avocado	sex	percentTime	DOB	TOB	entryDT	fee
1	1	Joe <In the know> Schmo	1	Male	10.0%	1952-02-29	11:49:00	2008-04-22T13:07:23	€11.000,00
2	2	Bill Hill	6.021E23	Male	20.0%	1972-01-01	14:01:00	2008-04-22T13:07:23	€10.000,00
3	3	Donna O'Fauna	6.0221E23	Female	30.0%	1985-06-08	05:25:00	2008-04-22T13:07:23	€9.000,00
4	4	Rob "Bob" Cobb	6.0221E23	Male	40.0%	1964-12-25	01:23:00	2008-04-22T13:07:23	€8.000,00
5	5	Tom Applebaum	6.0221E23	Male	50.0%	1975-03-15	15:15:00	2008-04-22T13:07:23	€7.000,00
6	6	Louise Mac&Cheese	6.0221E23	Female	60.0%	2007-06-05	08:09:00	2008-04-22T13:07:23	€6.000,00
7	7	Jack Black	6.0221E23	Male	70.0%	2011-11-11	11:11:00	2008-04-22T13:07:23	€5.000,00
8	8	Jill Hill	6.0221E23	Female	80.0%	1972-01-01	14:04:00	2008-04-22T13:07:23	€4.000,00
9	9	Gno Avocado	-3	Male	90.0%	1999-04-01	03:03:00	2008-04-22T13:07:23	€3.000,00
10	10					1960-01-01	01:01:00	2008-04-22T13:07:23	€2.000,00

SAS Dataset

Member	#	Integrity Constraint	Type	Variables	Where Clause	Reference	On De
1	WORK.MYSASDATA	1 DOB_present	Not Null	DOB			
2	WORK.MYSASDATA	2 avocado_unique	Unique	avocado			
3	WORK.MYSASDATA	3 id_GT_0	Check		ID>0		
4	WORK.MYSASDATA	4 name_fkey	Foreign Key	name		WORK.REALPEOPLE	Restric
5	WORK.MYSASDATA	5 prim_key	Primary Key	ID			
6	WORK.MYSASDATA	6 sex_MF	Check		sex in (., 1, 2)		

- format fee EUROX10.2;
- format entryDT IS8601DT.;
- format TOB IS8601TM.;
- format DOB IS8601DA.;
- format percentTime percent8.1;

SAS built-in formats

Integrity Constraints captured from PROC CONTENTS

```
describe table dictionary.columns;

create table DICTIONARY.COLUMNS
(
  libname char(8) label='Library Name',
  memname char(32) label='Member Name',
  memtype char(8) label='Member Type',
  name char(32) label='Column Name',
  type char(4) label='Column Type',
  length num label='Column Length',
  npos num label='Column Position',
  varnum num label='Column Number in Table',
  label char(256) label='Column Label',
  format char(49) label='Column Format',
  informat char(49) label='Column Informat',
  idxusage char(9) label='Column Index Type',
  sortedby num label='Order in Key Sequence',
  xtype char(12) label='Extended Type',
  notnull char(3) label='Not NULL?',
  precision num label='Precision',
  scale num label='Scale',
  transcode char(3) label='Transcoded?';
);
```

Dictionary.columns

```
describe table dictionary.tables;

create table DICTIONARY.TABLES
(
  libname char(8) label='Library Name',
  memname char(32) label='Member Name',
  memtype char(8) label='Member Type',
  dbms_memtype char(32) label='DBMS Member Type',
  memlabel char(256) label='Dataset Label',
  typemem char(8) label='Dataset Type',
  crdate num format=DATETIME informat=DATETIME label='Date Created',
  moddate num format=DATETIME informat=DATETIME label='Date Modified',
  nobobs num label='Number of Physical Observations',
  obslen num label='Observation Length',
  nvar num label='Number of Variables',
  protect char(3) label='Type of Password Protection',
  compress char(8) label='Compression Routine',
  encrypt char(8) label='Encryption',
  npage num label='Number of Pages',
  filesize num label='Size of File',
  pcompress num label='Percent Compression',
  ...
);
```

Dictionary.tables

EUROX Currency-euros
 "Writes numeric values with a leading euro symbol (E), a period that separates every three digits, and a comma that separates the decimal fraction"

From the external documentation for SAS built-in formats

The Dictionary tables expose the metadata stored in the SAS dataset itself.