

Survey Data Management Using Metadata

A Two-Fold Approach using
Blaise and SAS

Flat File Output – Still?

- ASCII, ANSI
- Universally translatable
- Compress very nicely (up to 90% usually)
- Simple to work with
- Other methods – ODBC, APIs, & native drivers are much more complex

What is Metadata?

- Meta = about, so it is “data about data”

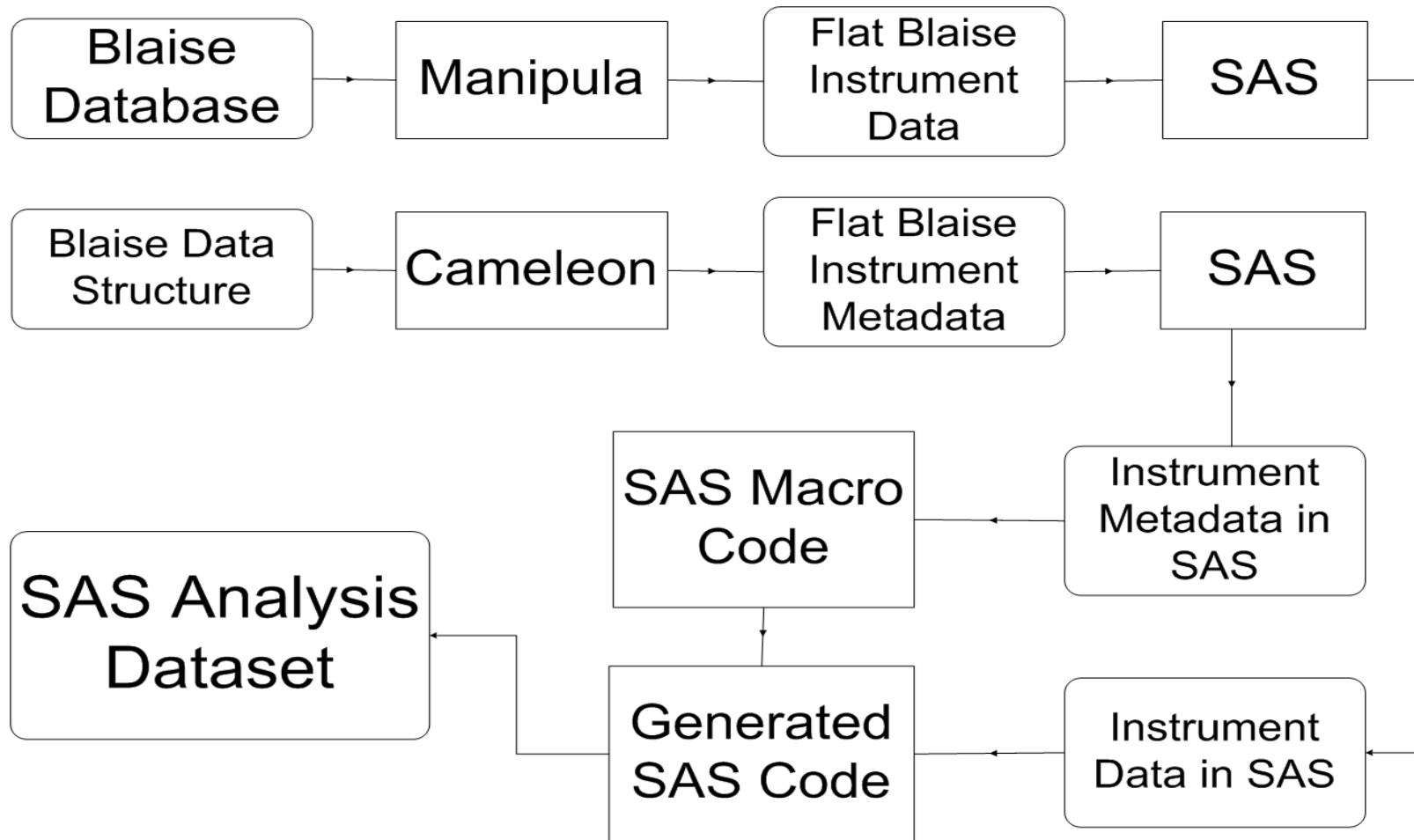
What exactly does it look like?

- Standard dataset table, where each record represents a field in the target data set
- Typically would contain fields such as variable name, type, length, position number, upper and lower limits (for numerics), etc.

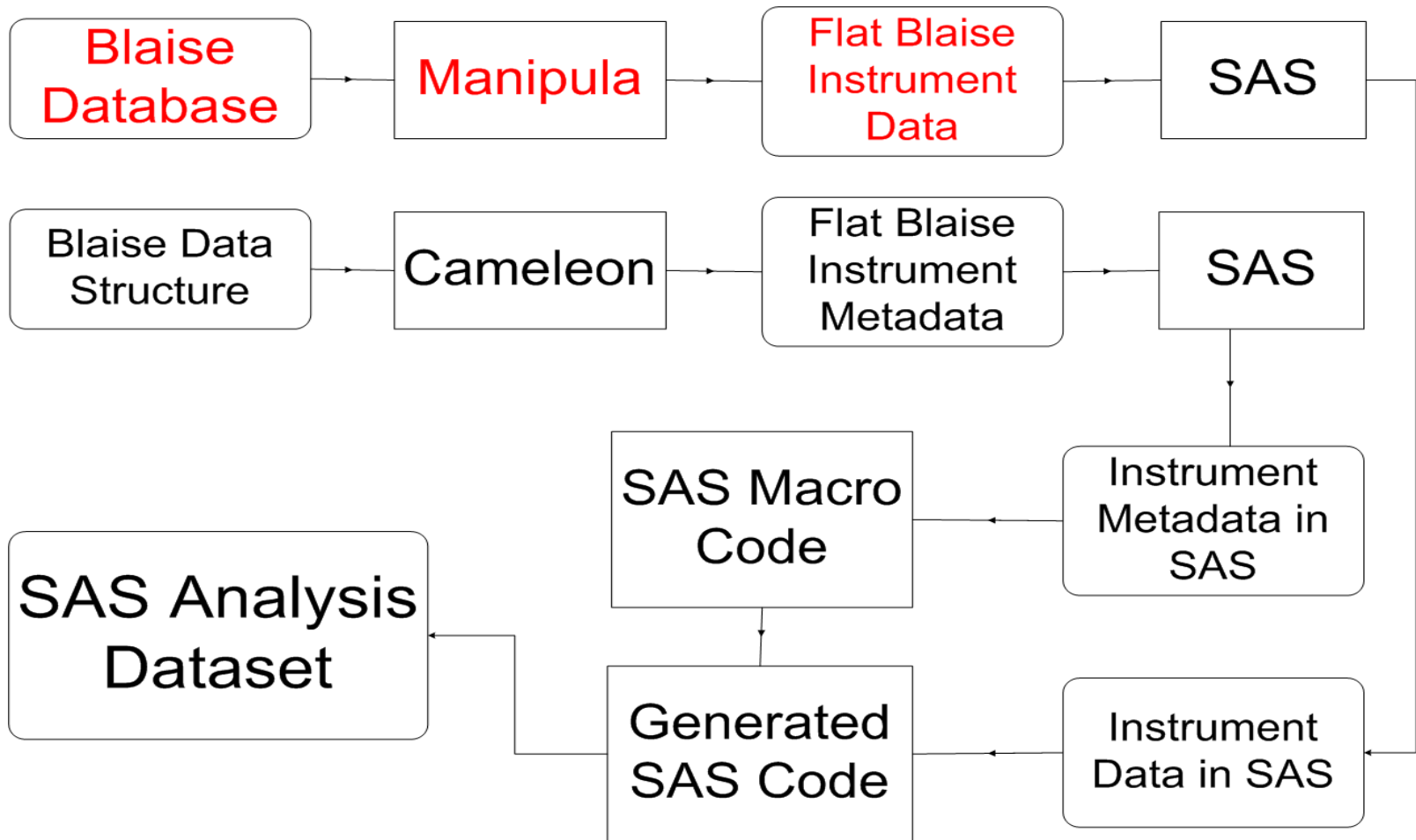
So....What can I do with it?

- There are many uses of metadata in general, but in this context the primary usage is Code Generation for purposes of Data Processing
- This is accomplished by writing small amounts of code (macro code, in the case of SAS) that uses the metadata dataset to write more code about every field in the target dataset.

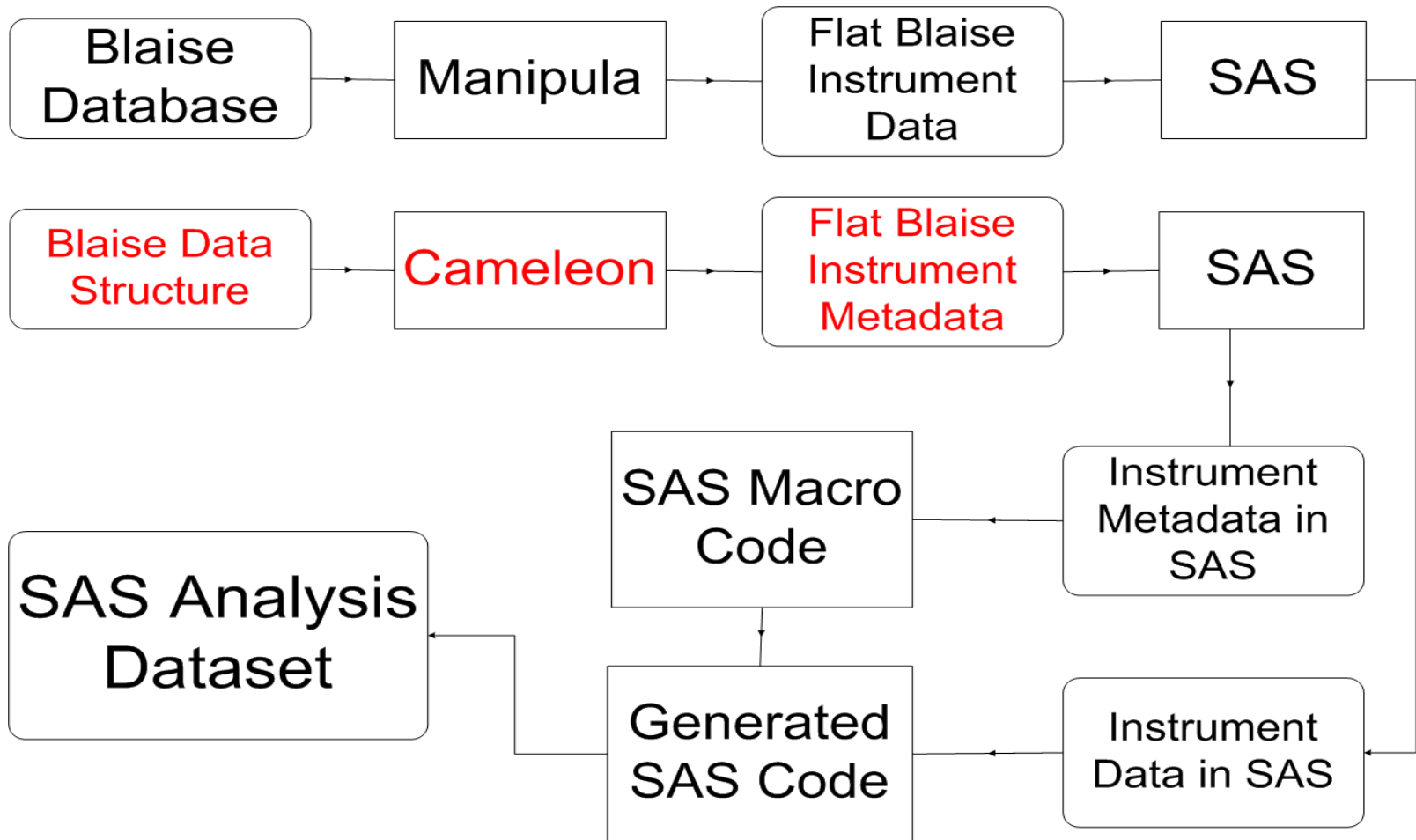
Blaise to SAS Data Flow Diagram



Blaise to SAS Data Flow Diagram



Blaise to SAS Data Flow Diagram



- + AddTab: blkAddTab
 - F BEGINDATE: DATETYPE
 - F ENTRYDATE: DATETYPE
 - F STOPDATE: DATETYPE
 - F BEGINTIME: TIMETYPE
 - F STOPTIME: TIMETYPE
 - F TIMELENGTH_M: 0.00..9999.99
 - F TIMELENGTH_MS: 0.00..9999.99
 - F INSTTIME: 0.00..9999.99
- + Roster: ARRAY[1..20] OF BRoster
- + Housing: BHousing
- Person: ARRAY[1..20] OF BPerson
 - + Intro_Age: BIntro_Age
 - Migration: BMigration
 - F POBSTATE: STRING[3]
 - F POBSTATENAME: STRING[24]
 - F POBFOREIGN: STRING[30]
 - F CITIZENSHIP: TYesNo
 - F HOWCITIZEN: THowCitizen
 - F CITNOPOB: TCitNoPOB
 - F YEARNATURAL: 1890..2006
 - F FIRSTARRIVAL: 1890..2006
 - F NUMARRIVALS: TOnceOrMore
 - F LASTARRIVAL: 1890..2006
 - F POBCODE: 0..999
 - E STATE: EXTERNAL (..\externals\StateCode)
 - E COUNTRY: EXTERNAL (..\externals\CountryCode)
 - A statelookup: STRING[24]
 - A countrylookup: STRING[24]
 - A TheYear: INTEGER
 - A YearsAgoNaturalized: INTEGER
 - A YearsAgoArrived: INTEGER
- + Education: BEducation
- + HealthInsurance: BHealthInsurance

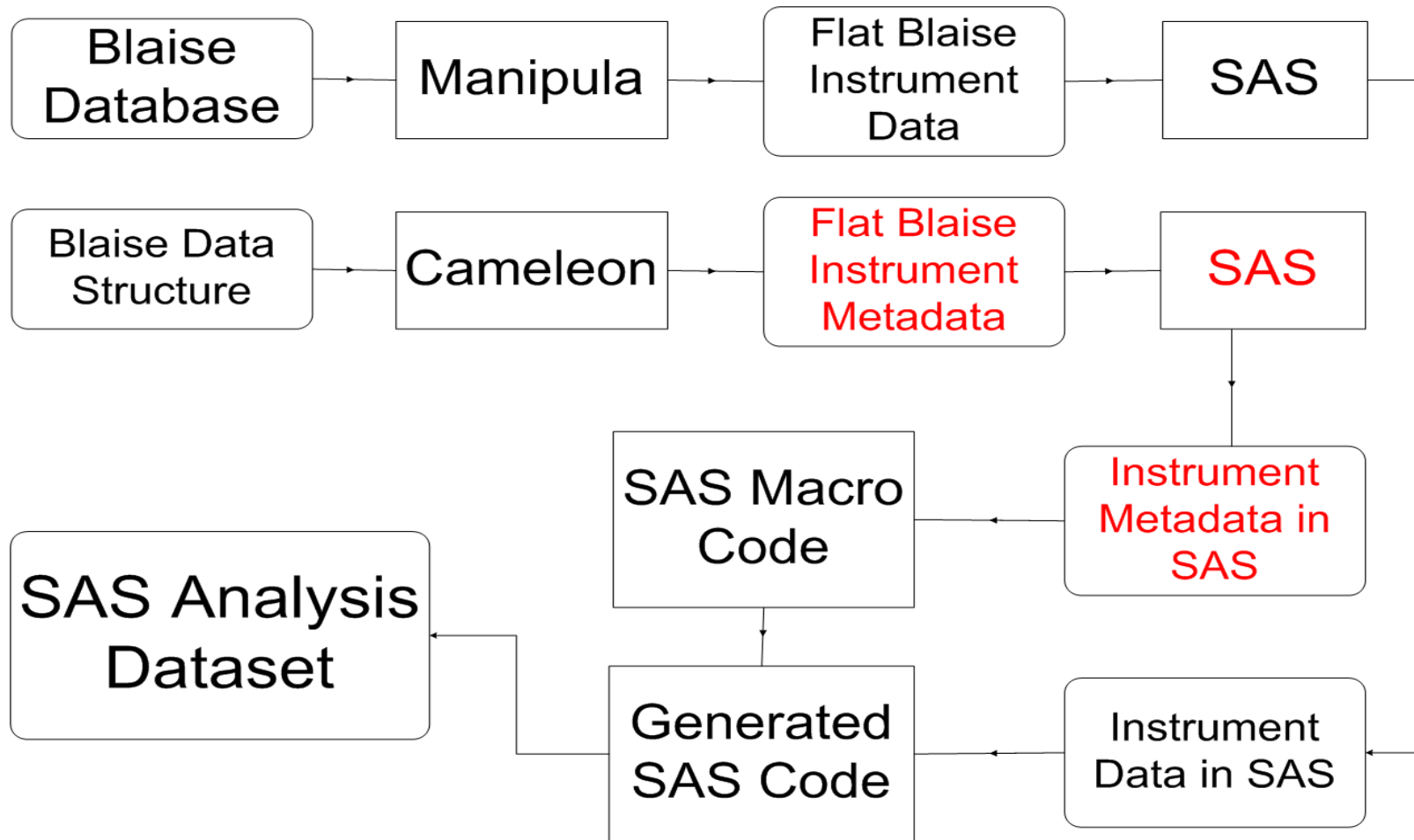
Sample Cameleon Script Segment

```
• [PROCEDURE WriteAllFields]
• [VAR htxt: LONGSTRING
•   ftxt: STRING
•   dtxt : STRING
•   rftxt : STRING
•   minval : STRING
•   maxval : STRING
•   typtxt : STRING
•   lentxt : STRING]
• [BLOCKPROC]
• [FIELDSLOOP]
• [ARRAYLOOP]
• [SETLOOP]
•   [IF TYPE = BLOCK THEN] [BLOCKCALL]
•   [ELSE]
•     [IF TYPE <> OPEN THEN]
•       [htxt:= FIELDPATH]
•       [ftxt:= UNIQUENAME]
•       [if TYPE = STRING THEN][typtxt:= 'STRING']][ENDIF]
•       [if TYPE = ENUMERATED THEN][typtxt:= 'ENUMERATED']][ENDIF]
•       [if TYPE = DATE THEN][typtxt:= 'DATE']][ENDIF]
•       [if TYPE = INTEGER THEN][typtxt:= 'INTEGER']][ENDIF]
•       [if TYPE = REAL THEN][typtxt:= 'REAL']][ENDIF]
•       [if TYPE = SET THEN][typtxt:= 'SET']][ENDIF]
•       [if TYPE = TIME THEN][typtxt:= 'TIME']][ENDIF]
•       [if DONTKNOW THEN][dtxt:= '1']][ELSE][dtxt:= '0']][ENDIF]
•       [if REFUSAL THEN][rftxt:= '1']][ELSE][rftxt:= '0']][ENDIF]
•       [minval:= STR(LOWVALUE)]
•       [maxval:= STR(HIGHVALUE)]
•       [lentxt:= STR(FIELDLLENGTH)]
•       [htxt 80:][ftxt 32:][dtxt 2:][rftxt 2:][minval 10:][maxval 10:][typtxt 20:][lentxt 3:]
•     [ENDIF]
•   [ENDIF]
• [ENDSETLOOP]
• [ENDARRAYLOOP]
• [ENDFIELDSLOOP]
• [ENDBLOCKPROC]
• [ENDPROCEDURE][*WriteAllFields]
```

Sample Flat Metadata Output

Housing.TOILET	TOILET	1 1 1	2	ENUMERATED	1
Housing.SHOWER	SHOWER	1 1 1	2	ENUMERATED	1
Housing.SINK	SINK	1 1 1	2	ENUMERATED	1
Housing.STOVE	STOVE	1 1 1	2	ENUMERATED	1
Housing.REFRIGERATOR	REFRIGER	1 1 1	2	ENUMERATED	1
Housing.TENURE	TENURE	1 1 1	4	ENUMERATED	1
Housing.VALUE1	VALUE1	1 1 1	19	ENUMERATED	2
Housing.VALUE2	VALUE2	1 1 0	9999999	INTEGER	8
Person[01].Intro_Age.PERSONINTRO	PERSONIN	0 0 1	1	ENUMERATED	1
Person[01].Intro_Age.AGEFIX	AGEFIX	1 1 0	116	INTEGER	3
Person[01].Intro_Age.AGERANGEFIX	AGERAN41	0 0 1	4	ENUMERATED	1
Person[01].Migration.POBSTATE	POBSTATE	1 1 0	0	STRING	3
Person[01].Migration.POBSTATENAME	POBSTAT2	1 1 0	0	STRING	24
Person[01].Migration.POBFOREIGN	POBFOREI	1 1 0	0	STRING	30
Person[01].Migration.CITIZENSHIP	CITIZENS	1 1 1	2	ENUMERATED	1
Person[01].Migration.HOWCITIZEN	HOWCITIZ	1 1 1	2	ENUMERATED	1
Person[01].Migration.CITNOPOB	CITNOPOB	1 1 1	4	ENUMERATED	1
Person[01].Migration.YEARNATURAL	YEARNATU	1 1 1890	2006	INTEGER	4
Person[01].Migration.FIRSTARRIVAL	FIRSTARR	1 1 1890	2006	INTEGER	4
Person[01].Migration.NUMARRIVALS	NUMARRIV	1 1 1	2	ENUMERATED	1
Person[01].Migration.LASTARRIVAL	LASTARRI	1 1 1890	2006	INTEGER	4
Person[01].Migration.POBCODE	POBCODE	1 1 0	999	INTEGER	4
Person[01].Education.ATTENDSCHOOL	ATTENDSC	1 1 1	2	ENUMERATED	1
Person[01].Education.TYPESCHOOL	TYPESCHO	1 1 1	3	ENUMERATED	1
Person[01].Education.LEVELSCHOOL	LEVELSCH	1 1 31	46	ENUMERATED	2
Person[01].Education.EDATHOME	EDATHOME	1 1 1	2	ENUMERATED	1
Person[01].Education.HOMEGRADE	HOMEGRAD	1 1 1	4	ENUMERATED	1
Person[01].Education.OTHERED1	OTHERED1	1 1 0	0	STRING	30
Person[01].Education.TOWDEGREE	TOWDEGRE	1 1 1	9	ENUMERATED	2
Person[01].Education.VOCOURSES	VOCOURSE	1 1 1	2	ENUMERATED	1
Person[01].Education.FORVOCDEG	FORVOCDE	1 1 1	2	ENUMERATED	1
Person[01].Education.TOWVOCDEG	TOWVOCDE	1 1 1	9	ENUMERATED	2
Person[01].Education.EDUCATTAIN	EDUCATTA	1 1 51	74	ENUMERATED	2
Person[01].Education.EDUCWOVOC	EDUCWOVO	1 1 51	73	ENUMERATED	2
Person[01].Education.TYPEHSORED	TYPEHSOR	1 1 1	4	ENUMERATED	1
Person[01].Education.TYPEALT	TYPEALT	1 1 0	0	STRING	30
Person[01].Education.ANYCOLLCRED	ANYCOLLC	1 1 1	2	ENUMERATED	1
Person[01].Education.COLLEGE1PLUS	COLLEGE1	1 1 1	2	ENUMERATED	1
Person[01].Education.TYPEUDEGREE	TYPEUDEG	1 1 1	5	ENUMERATED	1
Person[01].Education.OTHERED2	OTHERED2	1 1 0	0	STRING	30
Person[01].Education.TYPEPDEGREE	TYPEPDEG	1 1 1	10	ENUMERATED	2
Person[01].Education.OTHERED3	OTHERED3	1 1 0	0	STRING	30
Person[01].HealthInsurance.HICEMPLOYER	HICEMPLO	1 1 1	2	ENUMERATED	1

Blaise to SAS Data Flow Diagram

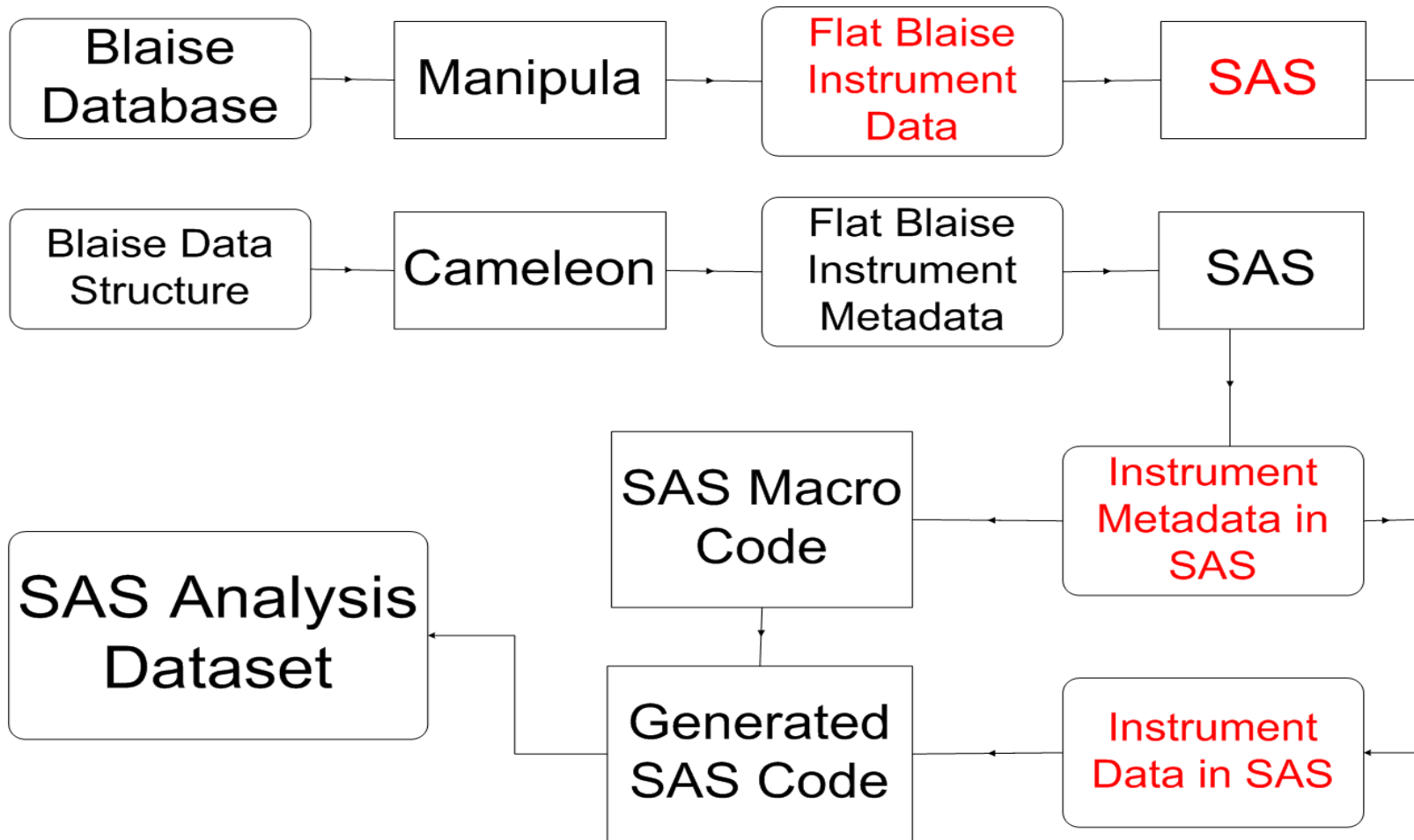


```

• libname capitest 'J:\blaisedata\ACS_CAPI\data\test';
• DATA capitest.cxvt_meta ;
• INFILE 'J:\blaisedata\ACS_CAPI\data\test\instcxvt.meta' lrecl=170;
• length blsvar $ 80
•     acsvar $ 20
•     root $ 20
•     lngadjst $ 1
•     rghtzero $ 1
•     arrvar $ 20
•     order 8
•     file $ 1
•     blocklvl1 $ 30
•     blocklvl2 $ 30
•     blocklvl3 $ 30
•     personnum 3
•     dk 3
•     rf 3
•     minval 8
•     maxval 8
•     type $ 20
•     len 3
• ;
• INPUT
•     blsvar $ 1 - 80
•     arrvar $ 81 - 100
•     dk     113 - 113
•     rf     115 - 115
•     minval 117 - 130
•     maxval 132 - 145
•     type $ 147 - 166
•     len    167 - 170;
• ;
• if index(blsvar, '.') > 0 then do;
•     blocklvl1 = trim(substr(blsvar,1,index(blsvar,')-1));
•     temp1 = trim(substr(blsvar,index(blsvar,')+1));
•     .
•     .
•     .
•     .

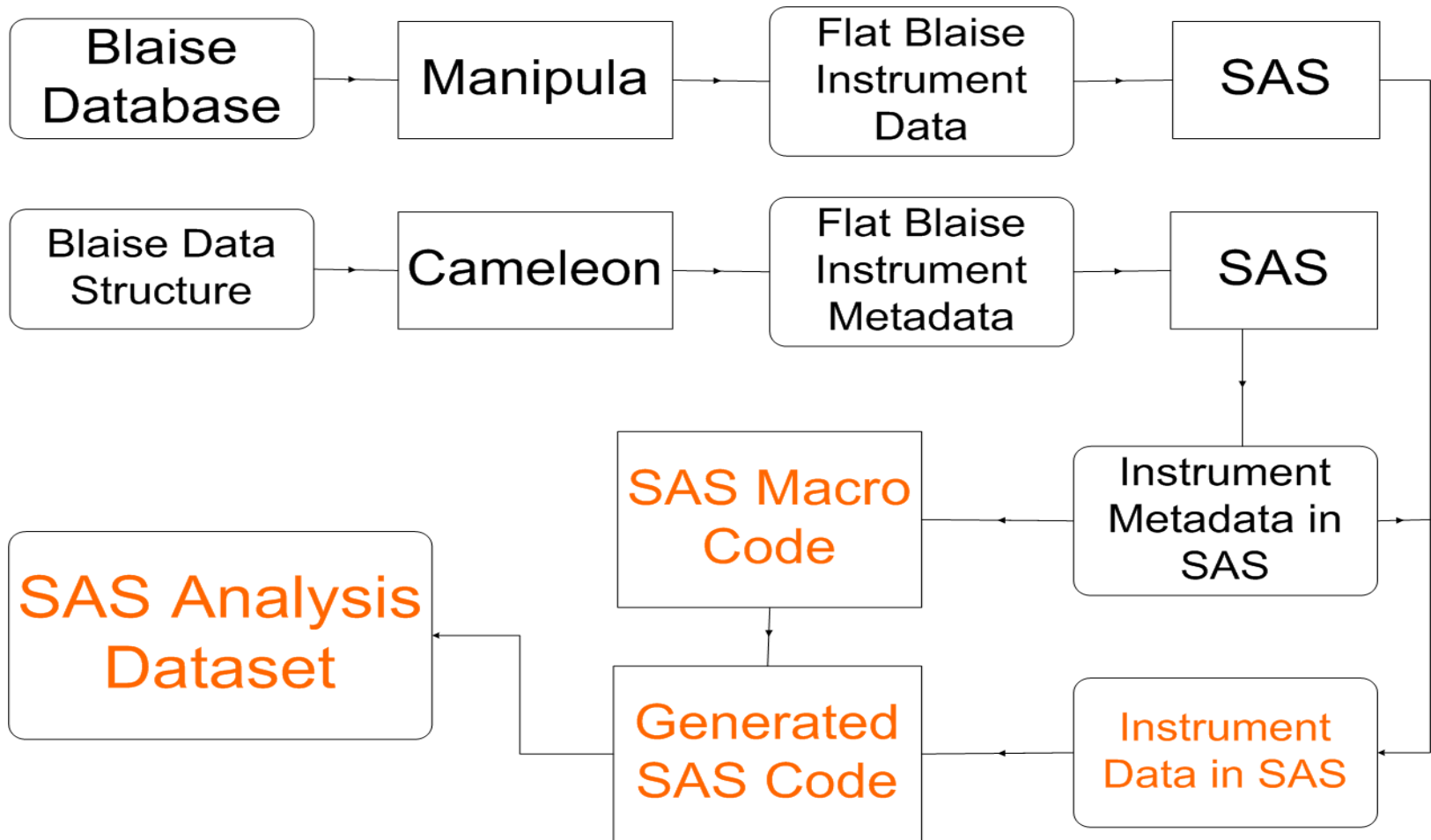
```

Blaise to SAS Data Flow Diagram



```
data tmoin.&surveyid.;
  infile "&tmoinpath.\&surveyid..dat" missover pad lrecl =
    &totreclen;
  input
  %do i = 1 %to %eval(&blaisemetacnt);
    &&blaisevar&i %cmpres($%cmpres(&&blaiselen&i).)
  %end;
  ;
run;
```

Blaise to SAS Data Flow Diagram




```

/* convert DK's/RF's to standard format using LNGADJST */
  %if %quote(&&lngadjst&j) = %str(*) %then %do;
    %if &&varlen&j = 1 %then %do;
      if &&arrvar&j = '9' then &&acsvar&j = 'D';
      else if &&arrvar&j = '8' then &&acsvar&j = 'R';
    %end;
    %else %do; /* length of var > 1 */
      if &&arrvar&j = REPEAT('9',%eval(&&varlen&j) - 1) then &&acsvar&j =
right('D');
        else if &&arrvar&j = REPEAT('9',%eval(&&varlen&j) - 2)||'8' then
&&acsvar&j = right('R');
      %end;
      else if &&arrvar&j ne "" then &&acsvar&j =
put(input(&&arrvar&j,%eval(&&varlen&j).),z%eval(&&varlen&j-1).) %str(;)
        else &&acsvar&j = "";
    %end;

```

File Edit View Insert Format Tools SAS Data Window Help

Type a question for help

100% Arial 10 B I U

Active Data: Active Worksheet My SAS Favorites

	A	B	C	D	E	F	G	H	I	J	K	L	
1	Module	Variable Long Name	VARNAME	VARNUM	LENGTH	INVAR	INVALUE	CODEVAR	FILTER	LL	UL	VARTYPE	
2	Both	CMID	CMID	1	9	CMID			**			CHARACTER	
3	Both	Panel	Panel	2	6	Panel			**			CHARACTER	
4	Both	site	site	3	5	site			**			CHARACTER	
5	Both		SOURCE	4	1	SOURCE						CHARACTER	
6	Housing		TYPREC	5	1	TYPREC						CHARACTER	
7	Housing		NPER	6	2	NPER				0	20	NUMERIC	
8	Housing		ADDC1	7	44	ADDC1			04			CHARACTER	
9	Housing		ADDC2	8	44	ADDC2			04			CHARACTER	
10	Housing	Reported Number of Persons	RPER	9	2	RPER			13		0	29	NUMERIC
11	Housing	Respondent Last Name	RLN	10	20	RLN			02				CHARACTER
12	Housing	Respondent First Name	RFN	11	13	RFN			02				CHARACTER
13	Housing	Respondent Middle Initial	RMI	12	1	RMI			01				CHARACTER
14	Housing		RLNO	13	2	RLNO					1	20	NUMERIC
15	Housing	Respondent Telephone Number	RTEL	14	10	RTEL			13				CHARACTER
16	Housing	Reported Date of Mail Completion	RDATE	15	8	RDATE			16				CHARACTER
17	Housing		PERMIS	16	1	PERMIS			13		1	2	CHARACTER
18	Housing		INTLANG	17	1				13		1	2	CHARACTER
19	Housing		RDCOD1	18	2	RDCOD1							CHARACTER
20	Housing		RDCOD4	19	2	RDCOD4							CHARACTER
21	Housing		RDCOD5	20	2	RDCOD5							CHARACTER
22	Housing		RDCOD6	21	2	RDCOD6							CHARACTER
23	Housing		HHLNO	22	2	HHLNO					1	20	NUMERIC
24	Housing	Current Population of HH	CPOP	23	2	CPOP			13		0	20	NUMERIC
25	Housing		CEDIT	24	2	CEDIT			11				CHARACTER
26	Housing		CTFU	25	2	CTFU			11				CHARACTER
27	Housing	Census Just-In-Case 1	CJIC1	26	1	CJIC1			11		0	9	CHARACTER
28	Housing	Census Just-In-Case 2	CJIC2	27	1	CJIC2			11		0	9	CHARACTER
29	Housing	Census Just-In-Case 3	CJIC3	28	1	CJIC3			11		0	9	CHARACTER
30	Housing	Census Just-In-Case 4	CJIC4	29	1	CJIC4			11		0	9	CHARACTER
31	Housing		TMOYR	30	4	TMOYR							CHARACTER
32	Housing		TMOMO	31	2	TMOMO							CHARACTER
33	Housing	Outcome	OUTCOM	32	1	OUTCOM							CHARACTER
34	Housing	Last Name - Person 6	LN6	33	20	LN6			02				CHARACTER

In Conclusion...

- Methodology can be used for various modes (CATI, CAPI, Mail, NR Followup)
- Can be applied to different treatments (Control and Variant)
- Extremely efficient; the methodology performed six times the work in a very short period of time.
- This approach will work on any other similar survey, which is the ultimate goal in code re-use and programming efficiency.