



Non-Printable and Special Characters? ... BYTE me!

Louise Sims

PHASTAR, Unit 2A, Bollo Lane, London, W4 5LE

Louise.Sims@phastar.co.uk

What are NPSC?

Computer Encoding Sets

- American Standard Code for Information Interchange (ASCII):** a 7-bit SBCS which contains 128 coding points (0 to 127).
- Extended ASCII set:** an 8-bit extension of the ASCII coding set which contains coding points 128 to 255. BEWARE: Different versions of the Extended ASCII set, e.g. ISO 8859-1 version, CP437 version.
- Extended Binary Coded Decimal Interchange Code (EBCDIC) set:** an 8-bit SBCS with coding points 0 to 255.
- Unicode set:** Much larger than ASCII and EBCDIC with each character between 8-bits and 32-bits in size, contains characters which cover most of the world's languages.

ASCII control characters		ASCII printable characters		Extended ASCII characters	
00	NULL (Null character)	32	space	64	@
01	SOH (Start of Header)	33	!	65	A
02	STX (Start of Text)	34	"	66	B
03	ETX (End of Text)	35	#	67	C
04	EOT (End of Trans.)	36	\$	68	D
05	ENQ (Enquiry)	37	%	69	E
06	ACK (Acknowledgement)	38	&	70	F
07	BEL (Bell)	39	'	71	G
08	BS (Backspace)	40	(	72	H
09	HT (Horizontal Tab)	41	)	73	I
10	LF (Line feed)	42	*	74	J
11	VT (Vertical Tab)	43	+	75	K
12	FF (Form feed)	44	,	76	L
13	CR (Carriage return)	45	-	77	M
14	SO (Shift Out)	46	.	78	N
15	SI (Shift In)	47	/	79	O
16	DLE (Data link escape)	48	0	80	P
17	DC1 (Device control 1)	49	1	81	Q
18	DC2 (Device control 2)	50	2	82	R
19	DC3 (Device control 3)	51	3	83	S
20	DC4 (Device control 4)	52	4	84	T
21	NAK (Negative acknowl.)	53	5	85	U
22	SYN (Synchronous idle)	54	6	86	V
23	ETB (End of trans. block)	55	7	87	W
24	CAN (Cancel)	56	8	88	X
25	EM (End of medium)	57	9	89	Y
26	SUB (Substitute)	58	:	90	Z
27	ESC (Escape)	59	;	91	[
28	FS (File separator)	60	<	92	\
29	GS (Group separator)	61	=	93	]
30	RS (Record separator)	62	>	94	^
31	US (Unit separator)	63	?	95	_
127	DEL (Delete)				

Can cause NPSC

Since the control characters are non-printable, these can appear as blank spaces or as odd symbols. Note this does not apply to 127 (Delete) which is printable.

Do not cause NPSC

Since these are printable, they usually do not cause any problems within data.

Can cause NPSC

Since there are different versions of the extended ASCII set, if the input encoding set is different to the encoding set used to view the data, different coding points will correspond to different characters. These cannot be displayed properly so can cause NPSC.

Check your SAS session encoding version!

```
proc options option=encoding;
run;
```

Shows the current encoding version is **WLATIN1**. This is the default encoding version in Western European countries.

```
2911 proc options option=encoding;
2912 run;

SAS (r) Proprietary Software Release 9.4 TS1M3

ENCODING=WLATIN1 Specifies the default character-set encoding for the SAS session.
NOTE: PROCEDURE OPTIONS used (Total process time):
   real time          0.00 seconds
   user cpu time      0.00 seconds
   system cpu time    0.00 seconds
   memory             20.46k
   OS Memory          3552.00k
   Timestamp          08/01/2016 04:06:31 PM
   Step Count         43   Switch Count   0
```

Identifying NPSC

NOTPRINT Function

Syntax: NOTPRINT ("character string" <,start>)

where *character string* is the text in which to search for NPSC, *start* (optional) is the starting position within the character string to start searching from.

Example:

```
data ie;
  set sdtm.ie;
  ntpnt=NOTPRINT(ietest);
run;
```

IENTEST	IECAT	NTPRNT
Subject is pregnant or breast feeding, or planning to become pregnant while enrolled in the study, up to the subject last study visit including the 30-day follow-up period.	EXCLUSION	76

The NOTPRINT function shows the first occurrence of NPSC is at position 76

RANK function

Syntax: RANK (expression)

where *expression* is a character string.

Example:

```
data ie;
  set sdtm.ie;
  ntpnt=NOTPRINT(ietest);
  id=RANK(SUBSTR(ietest,ntpnt,1));
run;
```

IENTEST	IECAT	NTPRNT	ID
Subject is pregnant or breast feeding, or planning to become pregnant while enrolled in the study, up to the subject last study visit including the 30-day follow-up period.	EXCLUSION	76	10

RANK function shows the NPSC is ASCII coding point 10; the line feed control character.

Removing NPSC from Data

COMPRESS function

Syntax: COMPRESS (source <, characters> <,> modifiers>)

where *source* is the character string, *characters* (optional) are specified characters which should be removed from the string, *modifiers* (optional) are constants which modify the function.

Example:

```
data co_raw;
  set raw.comments_all;
  cmnt=COMPRESS(comment, , "kw");
run;
```

COMMENT	CMNT
PATIENT DID NOT COME TO HIS W32 VISIT IN THE SCHEDULED TIME WINDOW BETWEEN 10 JUL - 24 JUL	PATIENT DID NOT COME TO HIS W32 VISIT IN THE SCHEDULED TIME WINDOW BETWEEN 10 JUL 24 JUL

NPSC removed from string using the k and w modifiers in the COMPRESS function.

BYTE function

Syntax: BYTE (n) where *n* is a numeric value between 0 and 255 and represents the coding point in ASCII/EDCDIC.

Example:

```
data cm;
  set raw.conmeds;
  check=NOTPRINT(cmverb);
  if check>0 then id=RANK(SUBSTR(cmverb,check,1));
  cmverb_=COMPBL(TRANWRD(cmverb, BYTE(id), " "));
run;
```

Replacing NPSC

BYTE and TRANWRD function

Example:

```
data co;
  set sdtm.co;
  id=RANK(SUBSTR(coval,83,1));
  coval_=TRANWRD(coval, BYTE(id), "'");
run;
```

COVAL	COVAL_
CONTINUATION OF PREVIOUS COMMENT: AND ALSO BECAUSE OF CONSIDERATION OF THE PATIENT'S RENAL FUNCTION	CONTINUATION OF PREVIOUS COMMENT: AND ALSO BECAUSE OF CONSIDERATION OF THE PATIENT'S RENAL FUNCTION

The BYTE function identifies the ASCII value of the arrow and the TRANWRD function replaces all occurrences of the arrow with an apostrophe.

NPSC in Outputs

Adding directly to the data

Example: Adding Greek letter μ to replace the u in the laboratory unit $\mu\text{mol/L}$.

```
data adlb;
  set adam.adlb;
  where paramcd="CREAT S";
  si_unit=BYTE(181) || SUBSTR(param,14,5);
run;
```

PARAM	PARCAT1	AVAIL	SIUNIT
Creatinine (umol/L)	STANDARD UNIT	79.560	$\mu\text{mol/L}$
Creatinine (umol/L)	STANDARD UNIT	79.560	$\mu\text{mol/L}$
Creatinine (umol/L)	STANDARD UNIT	88.400	$\mu\text{mol/L}$

BYTE function can be used to add NPSC to data to follow-through to the output.

Adding directly to the output code

ods escapechar="^";

```
compute before page / style={just=1};
  line @1 "Total Bilirubin (^{Unicode mu}mol/L)";
endcomp;
```

Baseline Grade	NA n (%)	0 n (%)
Total Bilirubin ($\mu\text{mol/L}$)		

The ODS escape character can be used to add NPSC directly to an output title, heading or label.

Hidden Dragon/Invisible Character

Example:

```
data final1;
  set final;
  if ord=3 then do;
    if LENGTH(txt)>52 then coll=SUBSTR(txt,1,51) || " " ||
      SUBSTR(txt,52);
    else coll=txt;
  end;
```

Blank spaces obtained by pressing Alt+255 on the number keypad 7 times.

run;

Category Type	Primary Cause	Overall Variable Dosing (N = 179) n (%)
Death		1 (0.6)
Cardiovascular		1 (0.6)
Stroke		1 (0.6)
Sudden cardiac death		0 (0.0)
Non-cardiovascular		0 (0.0)
Hemorrhage (excluding hemorrhagic stroke or bleeding in setting of coronary revasc)		0 (0.0)
Suicide		0 (0.0)

BEFORE

Category Type	Primary Cause	Overall Variable Dosing (N = 179) n (%)
Death		1 (0.6)
Cardiovascular		1 (0.6)
Stroke		1 (0.6)
Sudden cardiac death		0 (0.0)
Non-cardiovascular		0 (0.0)
Hemorrhage (excluding hemorrhagic stroke or bleeding in setting of coronary revasc)		0 (0.0)
Suicide		0 (0.0)

AFTER

Importing Data with NPSC into SAS

Step 1: Open the external data within Notepad++.

Step 2: Open the Find and Replace tool in Notepad++, ensuring the option "Extended (\n, \r, \t, \0, \x...)" is selected under Search Mode.

Step 3: Search for NPSC and replace with a blank space.

\n = line feeds, \r = carriage returns, \t = tab, \0 = null character and \xdd = ASCII character where ddd is the ASCII/Unicode coding value.

BEFORE	
Study Code	Country
D5160C00003	Australia
Patient E0304301 CT schedule has been changed from 6 weekly to 12 weekly due to discomfort during the procedure; this is related to the cold contrast required to be contrasted prior to scan. This was approved by Global Study Physician and Safety Team	

External data with NPSC imported into SAS.

```
1 *****
2 Study Code,Country,Site No,Enrol No,Rand No,Issue No,Catg
3 D5160C00003,Australia,304,304301,1014,V16,Major Protocol Devia
4 Patient E0304301 CT schedule has been changed from 6 wee
5
6
7 *,Patient will continue to have 12 weekly CT scans up to
8 D5160C00003,Canada,1002,1002311,1095,V32,Major Protocol I
```

External data with NPSC in Notepad++ where it is easier to see the NPSC affecting the data.

AFTER	
STUDY_CODE	COUNTRY
D5160C00003	Australia

External data with NPSC removed imports into SAS correctly.

```
1 *****
2 Study Code,Country,Site No,Enrol No,Rand No,Issue No,Catgory
3 D5160C00003,Australia,304,304301,1014,V16,Major Protocol Devia
4 D5160C00003,Canada,1002,1002311,1095,V32,Major Protocol Devia
5 D5160C00003,Canada,1003,1003301,1840,V7,Major Protocol Devia
6 D5160C00003,Canada,1005,1005301,1029,V6,Major Protocol Devia
7 D5160C00003,Canada,1005,1005302,1035,V6,Major Protocol Devia
```

External data in Notepad++, with NPSC removed. Each record appears on one line, enabling SAS to process the data correctly when it is imported.