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.

Computer Encoding Set

- ASCII: With coding points 0-127.
- Extended ASCII: Adds coding points 128-255.
- EBCDIC: Wider range coding points 0-255.
- Unicode: Offers wide range of characters.

NPSC and Non-Printable Characters

- NPSC includes non-printable characters (e.g., @, #, $, %, ^, &, *, (, ), {, }, [, ], |, ;, :).
- Non-printable characters are special characters that cannot be displayed properly when they appear in output data.
- NPSC can affect data processing and analysis in SAS.

NPSC in Outputs

Adding directly to the data
Example: Adding Greek letter µ to replace the u in the laboratory unit umol/L.
```sas
data adlb;
  set final;
  if ord=3 then do;
    cmverb=COMPBL(TRANWRD(cmverb, BYTE(id), " "));
    cmverb="\{Unicode \(mu\)\}mol/L";
    ssi_unit=BYTE(181)||SUBSTR(param,14,5);
    if paramcd="CREAT_S" then ssi_unit=si_unit^"\{Unicode \(mu\)\}mol/L";
  end;
  run;
```

Adding directly to the output code
```sas
ods escapechar="^";
line 1 "Total Bilirubin (^{Unicode \(mu\) mol/L});
```
endcomp;
```

Replacing NPSC

BYTE and TRANWRD function
Example:
```sas
data co;
  set raw.co;
  byte=BYTE(id);
  trans=TRANWRD(coval, byte, ",\'\');
  if trans<" Appalachian"
```