

Paper ###-2019

People Keep ASCIIing Me About These Characters

Dan Konkler, Rebecca Bowermaster, and Stephanie Ann Sanchez, Roche Molecular Solutions

ABSTRACT

Every programmer has run into an issue at least once that was caused by the presence of a control character or extended ASCII character in their dataset. You get extra records in your dataset because there's a line feed character within a CSV field. Your Excel output has "unreadable content" because your data contain "<, >, or &". Something went wrong in your RTF file when your trusty escape character wound up in your dataset. Your fancy validation program that reads in the production output for programmatic comparison fails because it can't process the extended ASCII or control characters. These issues can cause hours of lost time to searching for the problematic character. In this paper, we present two macros that can quickly identify potentially problematic characters in your dataset library and point you to the specific records and fields that need your attention.

INTRODUCTION

Here we will go into more detail about all the things that can go wrong. What are control characters in each destination, FDA/xpt, what is the role of the escape character, and include an ASCII character map.

1. Unexpected results or missing characters may occur in your report if your specified ODS ESCAPECHAR appears as data in your report content
2. An XML special character (, &,'") in ODS EXCEL destination will cause unreadable content errors if option PROTECTSPECIALCHARS=OFF.
3. Extended ASCII characters can translate incorrectly on output and/or read in from XML and HTML files and well as between systems using different transcodings or character sets.
4. Control characters (e.g. line feed) within source data can interfere with data processing (e.g. signal SAS to prematurely end or split the record).

PRESENT THE MACRO(S)

The first main topic will be the presentation of the macro(s).

- 1) The first macro counts all occurrences of specified characters or character sets present within your dataset(s). The macro allows you to search an entire library and print a summary of how many times the specified character(s) appear in each dataset. This is useful for e.g. determining if your favorite ESCAPECHAR is safe to use or if you might have any dataset portability issues when submitting to FDA.
- 2) The second macro, similar to compress, can be used to keep/remove all specified characters from a given dataset and is useful in identifying data fields that contain problematic characters for the data owner or data manager to resolve or replace.

REAL LIFE EXAMPLE

The second main topic will use the macro to identify datasets in SASHELP library that contain problematic characters and demonstrate how the second macro can be used for handling them.

CONCLUSION

The conclusion will summarize our paper, compare/contrast these macros vs. existing tools, and tie together any loose ends.

REFERENCES

TBD

ACKNOWLEDGMENTS

TBD

RECOMMENDED READING

<https://www.pharmasug.org/proceedings/2016/BB/PharmaSUG-2016-BB16.pdf>.

<http://support.sas.com/kb/52/832.html>

<https://www.pharmasug.org/proceedings/2018/QT/PharmaSUG-2018-QT02.pdf>

Commented [SSA{1}]: The following will be included as either recommended reading or references.