Get Smart! Eliminate Kaos and Stay in Control – Creating a Complex Directory Structure with the DLCREATEDIR Statement, SAS® Macro Language, and Control Tables

Louise S. Hadden, Abt Associates Inc., Cambridge, MA

ABSTRACT

An organized directory structure is an essential cornerstone of data analytic development. Those programmers who are involved in repetitive processing of any sort control their software and data quality with directory structures that can be easily replicated for different time periods, different drug trials, etc. Practitioners (including the author) often use folder and subfolder templates or shells to create identical complex folder structures for new date spans of data or projects, or use manual processing or external code submitted from within a SAS® process to run a series of MKDIR and CHDIR commands from a command prompt to create logical folders. Desired changes have to be made manually, offering opportunities for human error. Since the advent of the DLCREATEDIR system option in SAS version 9.3, practitioners can create single folders if they do not exist from within a SAS process. Troy Hughes[i] describes a process using SAS macro language, the DLCREATEDIR option, and control tables to facilitate and document the logical folder creation process. This paper describes a technique wrapping another layer of macro processing which isolates and expands the recursive logical folder assignment process to create a complex, hierarchical folder structure used by the author for a project requiring monthly data intake, processing, quality control and delivery of thousands of files. Analysis of the prior month's folder structure to inform development of control tables is discussed.

INTRODUCTION

Since 1998, the U.S. Centers for Medicare and Medicaid Services (CMS) has maintained a website, Nursing Home Compare, which provides detailed quality information about every certified nursing home in the country. In December 2008, CMS greatly enhanced the usability of the website by adding an easy-to-understand 5-star rating. Each nursing home receives one to five stars based on performance in each of three key quality domains (health inspections, reported staffing levels, and quality measures derived from mandated assessments of resident health and well-being) plus an overall quality rating. Calculation of ratings requires integration of information from both facility and resident-level data sources. SAS® was used extensively in analysis to support the development of the rating system, and it is currently used to process data to refresh the ratings each month, based on newly collected data in each domain.

Data from two organizations is received monthly, data from two additional organizations is received quarterly, and data from an additional organization is received biannually. Processing is classified into four separate categories: reading in original data, calculating ratings, "other" processing (reading in data not used in ratings, redacting text fields, preparing for various types of output), and production of various types of output. A standard set of programs run in a standard, complex folder structure each month, with additional programs for quarterly and biannual processing.

This project is representative of the adage attributed to Heraclitus "The only constant is change". The incoming data changes every month. Sometimes the underlying structure of the data changes, and there needs to be flexibility in our system. Accordingly, we employ control tables in a number of instances, including date macro include files, hard code include files, self-writing batch files to run sequential programs, etc. This has worked to an extent - our processes have worked well for a long time, over 9 years, perhaps with and/or because of extraordinary effort by the programmers and technicians. As we look at our system with an eye toward software quality, though, it is clear we have a long way to go.

One time-consuming aspect of this project is recreating the complex directory structure every month for the next month's processing. We had been using an empty "skeleton" file structure, copying it over each month, then copying the most recent programs used in the prior month's processing. The files in the "skeleton" and even the structure itself had a habit of changing on a regular basis, and we started to have a clutter of file skeletons, something inevitably would not get copied over, and chaos reigned. The DLCREATEDIR option as described by Troy Hughes offered an alternative to this clumsy and error prone manual process.

DESIRED DIRECTORY STRUCTURE

Lorem ipsum. Lorem ipsum.

CONCLUSION

Lorem ipsum. Lorem ipsum.

REFERENCES

Hughes, Troy Martin. 2016. *SAS® Data Analytic Development: Dimensions of Software Quality*. Hoboken, New Jersey. John Wiley & Sons, Inc.

[i]Hughes, Troy Martin, 2016. "SAS® Spontaneous Combustion: Securing Software Portability through Self – Extracting Code", presented at the SAS Global Forum 2016 Conference in Las Vegas, NV (Paper 11768-2016). http://support.sas.com/resources/papers/proceedings16/11768-2016.pdf

Rosenbloom, Mary F. O. and Carpenter, Arthur L., 2015. "Are You a Control Freak? Control Your Programs – Don't Let Them Control You!", presented at the SAS Global Forum 2015 Conference in Dallas, Texas (paper 2220-2015). http://support.sas.com/resources/papers/proceedings15/2220-2015.pdf

Rosenbloom, Mary F. O. and Carpenter, Arthur L., 2016. "I've Got to Hand It to You; Portable Programming Techniques", presented at the MidWest SAS Users Group Conference in Cincinnati, Ohio (paper BB03). http://www.lexjansen.com/mwsug/2016/BB/MWSUG-2016-BB03.pdf

Williams, Christianna, Hadden, Louise, Mortimore, Edward, Nagy, Frank, Plotzke, Michael, and White, Alan. "Measuring Nursing Home Quality – The Five-Star Rating System", presented at the SAS Global Forum 2010 Conference in Seattle, Washington (Paper 161-2010). http://support.sas.com/resources/papers/proceedings10/161-2010.pdf

Usage Note 56710: The DLCREATEDIR system option creates a single directory

ACKNOWLEDGMENTS

The author gratefully acknowledges the helpful work of Troy Martin Hughes, Art Carpenter, Mary Rosenbloom, Richann Watson and Kirk Paul Lafler, among others.

RECOMMENDED READING

- Base SAS® Procedures Guide

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Louise Hadden: Louise_Hadden@abtassoc.com

Scan me for sample code and additional papers!