

The U.S. Census Bureau's utilization of Qualex's Cataloger Pro software tool

OVERVIEW

Qualex Consulting Services, Inc. granted utilization of its Cataloger Pro SAS software tool to the Testing Service & Coordination Branch at the Economic Statistical Methods & Programming Division of the U.S. Census to improve their regression testing process.

Cataloger allowed the Census team to perform their regression testing process in four to five days, a fifty percent improvement on typical regression time frames.

COMPANY

The U.S. Census Bureau is the premier source of information about the American people and the economy. More than just numbers, this information shapes important policy decisions that help improve the nation’s social and economic conditions. The Testing Service & Coordination Branch (ESMPD) is responsible for testing and validating software development changes on the “StEPS I” system*, which is used to process more than a hundred economic surveys.

CATALOGER

Written entirely in SAS®, Qualex’s Cataloger Pro application allows SAS developers to easily view, compare, search and manipulate catalog entries, as well as create publication-quality documents. Qualex Cataloger Pro tool utilizes an easy point-to-click interface. Developers can compare the properties of two SAS data sets as well as review and set many aspects of SAS data sets, aspects such as variable order, indexes and passwords.

One primary validation on their Regression Testing is ensuring that program changes on “StEPS I” do not affect the numeric results generated on the respective system modules like Estimation, Edits or Imputation. If so, the differences can be reviewed for confirmation.

The Qualex Cataloger Pro tool possesses a utility to compare

Contents:	
Overview.....	1
Company.....	1
Cataloger.....	1
Screenshots.....	2
Conclusion.....	5

Cataloger allowed the Census team to perform their regression testing process in four to five days, a fifty percent improvement on typical regression time frames.

“What took days, now takes minutes.”
Kelvin Hackett
Census Branch Chief

values of data files in a way that users can rapidly identify the differences between data files on a easy to understand point-and-click graphical user interface.

Testers who are using the Qualex Cataloger Pro tool have reduced the time they spend on reviewing the accuracy of their data by seventy percent and have a standard mechanism of clearly identify differences between data files.

Extremely useful in seeing trends/outliers whereas sometimes very small differences on the data can be ignored and focus of review is placed on the outliers.

One advantage of the Cataloger Pro Tool when comparing SAS data sets is you can select the variable you want to compare, which is really helpful specially when comparing large data sets that may contain hundred variables or more.

The "StEPS I" data resides on a remote Linux server, users are able to define the location (directory) of this data by entering the information on a predefined/formatted personal Microsoft Excel sheet that gets read in by the Qualex Cataloger Pro tool. This customization/functionality provides great flexibility on the universe of data testers can get access.

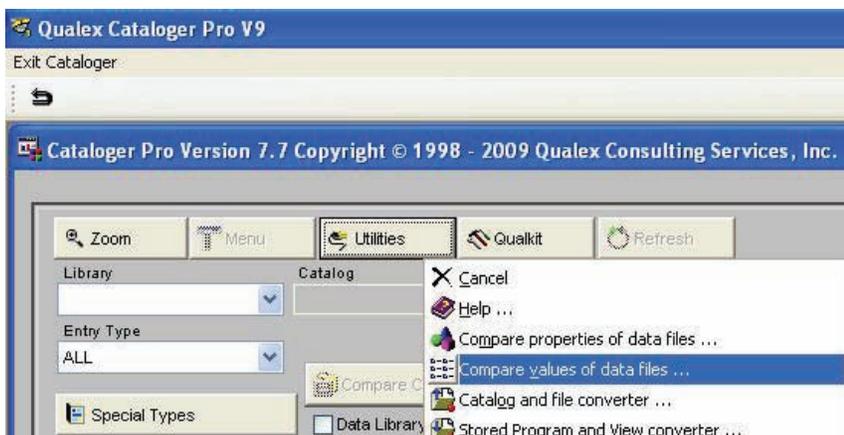


"I have been using the Qualex Cataloger tool for two software releases. It has been an excellent tool and was able to solve pretty much any issue I encountered. I was very impressed with how results were generated."

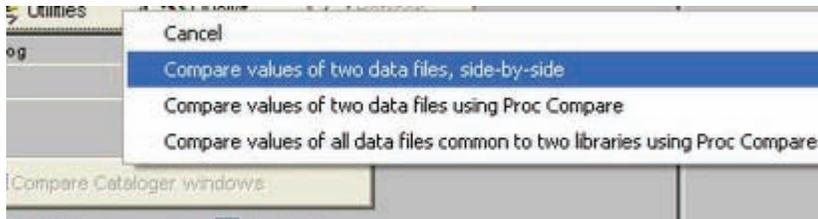
Ivy Nguyen, Software Tester
Testing Services & Coordination
Branch

Cataloger Steps

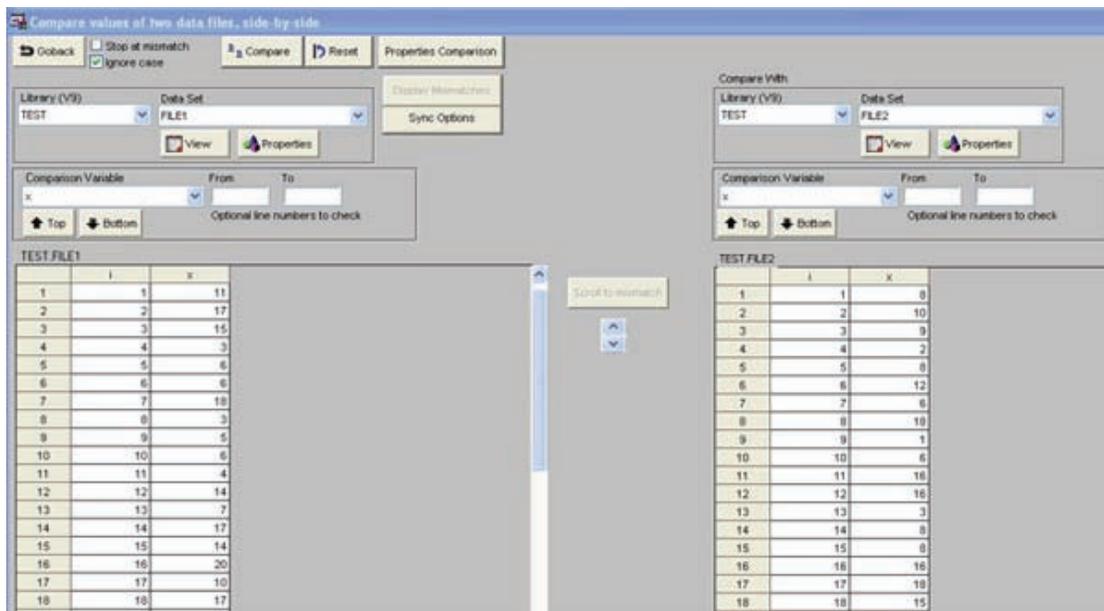
After entering Cataloger Pro Version 7.7, users press the "Utilities" button, to select the menu option "Compare values of data files ..."



Users can then choose either the “Compare values of two data files, side-by-side” or the “Compare values of two data files using Proc Compare” menu options.



When selecting “Compare values of two data files, side-by-side”, they get the following screen:



In the screen above, users select two datasets to compare and indicate the variable that needs to be compared. After pressing the “Compare” button a screen showing the mismatches gets displayed. See how row 10 is not a mismatch:

The screenshot shows a 'Mismatches' window with a table comparing 'TEST.FILE1' and 'TEST.FILE2'. The table has columns for 'ROW' and the file names. Row 10 is highlighted, indicating it is not a mismatch.

ROW	TEST.FILE1	ROW	TEST.FILE2
1	11	1	8
2	17	2	10
3	15	3	9
4	3	4	2
5	6	5	8
6	6	6	12
7	18	7	6
8	3	8	18
9	5	9	1
10	6	10	6
11	4	11	16
12	14	12	16
13	7	13	3
14	17	14	8
15	14	15	8
16	20	16	16
17	10	17	18
18	17	18	15

When users select the menu option “Compare values of two data files using Proc Compare” they get the following screen after pressing the “Compare” button:

Compare values of two data files using Proc Compare

Library (V9) Data Set: TEST FILE1

Compare With Library (V9) Data Set: TEST FILE2

Buttons: Compare, Print, Export, View observations that are different

Comparison details		Frequency
Number of Observations with Some Compared Variables Unequal		95
Number of Observations with All Compared Variables Equal		5
Number of Variables Compared with All Observations Equal		1
Number of Variables Compared with Some Observations Unequal		1
Number of Variables with Missing Value Differences		0
Total Number of Observations Read from TEST.FILE1		100
Total Number of Observations Read from TEST.FILE2		100
Number of Observations in TEST.FILE1 but not in TEST.FILE2		0
Number of Observations in TEST.FILE2 but not in TEST.FILE1		0
Number of Variables with Differing Attributes		0
Number of Variables in TEST.FILE1		2
Number of Variables in TEST.FILE2		2
Number of Variables in Common		2
Number of Variables in TEST.FILE1 but not in TEST.FILE2		0
Number of Variables in TEST.FILE2 but not in TEST.FILE1		0

Variable	Number of Observations where different
x	95

Users then select “View observations that are different”, which brings the following screen:

View observations that are different

Display only variables having some differences (1 with differences)

Observations with: 95

Observation start range: 1

Buttons: Print, Export

Variable	Number of Observations where different
x	95

TEST.FILE1		
Observation Range Sequence Number	Observation number where different	x
1	1	11
2	2	17
3	3	15
4	4	3
5	5	6
6	6	6
7	7	18
8	8	3
9	9	5
10	11	4
11	12	14
12	13	7
13	14	17
14	15	14
15	16	20
16	17	10
17	18	17
18	19	10
19	20	10
20	21	5

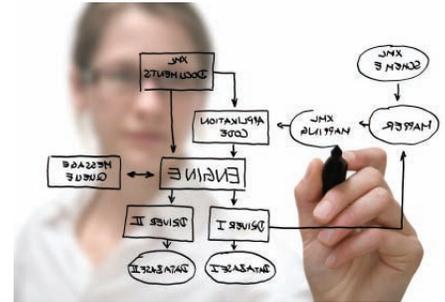
TEST.FILE2		
Observation Range Sequence Number	Observation number where different	x
1	1	8
2	2	10
3	3	9
4	4	2
5	5	8
6	6	12
7	7	6
8	8	18
9	9	1
10	11	16
11	12	16
12	13	3
13	14	8
14	15	8
15	16	16
16	17	18
17	18	15
18	19	8
19	20	15
20	21	7

Conclusion

By utilizing the “Qualex Cataloger Pro” tool to compare data values on their Regression Testing, Census’ employees were able to increase efficiency in their regression testing by up to fifty percent.

The changes were noticed immediately by Kelvin Vonzell Hackett, Branch Chief, Testing Services & Coordination Branch, who stated, “What took days, now takes minutes.” It was a notion that Mrs. Ivy Nguyen, Software Tester, Testing Services & Coordination Branch confirmed, adding, “I have been using the Qualex Cataloger tool for two software releases. It has been an excellent tool and was able to solve pretty much any issue I encountered. I was very impressed with the way how results were generated.”

At the start of the project, Qualex had several objectives for their work at the Census, including fast implementation, process standardization, as well as helping the Census save time in their Regression Testing and, even though only a limited amount of the Cataloger Pro tool’s capabilities were utilized, each and every objective was not only met but also surpassed.



“There is no more scrolling through trying to decipher the difference in Steps. With Qualex Cataloger Pro tool it is easy to read highlighted color and side by side comparison makes my job much more efficient.”

*Dane-Ann Massey,
Software Tester
Testing Service & Coordination Branch*

* The Standard Economic Processing System (StEPS) is a generalized system being developed in the Economic Directorate to process over 100 current economic surveys. It is written entirely in SAS® and operates in a UNIX environment. StEPS is designed to handle different surveys with different needs by using generalized programs and data structures. A development team comprised of programmers, mathematical statisticians and survey statisticians are identifying requirements; standardizing concepts, terminology, and data structures; and developing integrated SAS code to perform various survey processing activities. Such activities include edit/imputation, data review and correction, estimation and variance estimation, and select data collection activities.

For more information:

Qualex Consulting Services, Inc.
4300 Biscayne Blvd.
Miami, FL 33137
Telephone: + 1 877-887-4SAS (4727)
www.qlx.com
E-mail: andrew@qlx.com