

Paper 221-2013

Recovering SAS® User Group Proceedings for the SAS® Community

Richard W. La Valley, Strategic Technology Solutions

Lex Jansen, lexjansen.com¹

Kirk Paul Lafler, Software Intelligence Corporation

Abstract

For many years, SAS® User Groups held conferences which had printed only proceedings that were available only to those who attended or those who knew that copies existed in the SAS® Library in Cary, NC. Over the past couple of years, there has been a project to digitize the printed proceedings of SAS® User Groups International, SAS® European Users Group International, North East SAS® Users Group, SouthEast SAS® Users Group, Western Users of SAS® Software, SouthCentral SAS® User Group, MidWest SAS® Users Group, and the Pacific NorthWest SAS® Users Group. This paper will provide an overview of the project and the progress that has been made on this effort.

Introduction

SAS® Users Groups had their start with the first gathering of SAS® Users at SAS.ONE 1976 in Kissimmee, Florida on January 26-28, 1976. Later that year, Julian and Evey R. Horwich edited the first proceedings which were published for those that attended the first meeting. This first SAS® Users Group International (SUGI) was attended by over 200 users of the SAS® Software and they received copies of the proceedings. In addition, copies were later archived in the SAS® Library in Cary, North Carolina. The content of those papers was limited to those who attended and those who found the content in the library. The content of subsequent SUGI conferences was published in print form for the next 22 years and access to the content of those conferences was limited those who attended and those who knew about the content in the SAS® Library.

With the advent of digital content of proceedings for publications, SUGI content began to appear on the Internet at the support.sas.com² around SUGI 21 and on Lex Jansen's website³ around 2003. The copies of the proceedings from SUGI 14 through SUGI 20 appeared to be made available but only in image form only in the late 1990's.

A similar process was followed for each of the Regional SAS® User's Group and SAS® European Users Group International where the proceedings were available in print only format to those that attended and some but not all of those proceedings were retained for the SAS® Library in Cary.

In 2009, approximately 40% of the content of this knowledge base was available digitally and retrievable by the SAS community. A project was sponsored by the SAS® Global Users Group Executive Committee

¹ Lex Jansen is a SAS employee. The views expressed in this paper are the authors' own views and do not represent the views of SAS.

² support.sas.com

³ <http://www.lexjansen.com/sugi/>

⁴ <http://sasCommunity.org>

in 2010 to capture and digitize all of the previously print version proceedings so that content of those proceedings could be made available to the SAS® Community. That project was completed in 2011.

Subsequently, the same effort was undertaken and for all of the Regional SAS® Users Groups and funded by the Regional Groups where proceedings could be located and made available for digitizing. A similar effort was completed for the SAS® European Users Group International and SAS® Forum International proceedings.

In table1 below, the table shows the status of this project and the relative completeness of the content of the various Users Groups Proceedings.

User Group	Years Available	Completeness
SUGI	1976 – 2006	All Available online
SAS® Global Forum	2007 – 2012	All Available online
NESUG	1988 - 2012	All Available online
WUSS	1993 - 2012	All Available online
MWSUG	1990 - 2012	All Available online
SESUG	1993 - 2012	All Available online
SCSUG	1991 - 2012	All Available online
PNWSUG	1996-1998, 2004 - 2009	Partially Available online, Missing proceedings for years going back to 1982.
PharmaSUG	1998, 2000- 2012	Partially Available online, Missing 1999 and all proceedings prior to 1998.
PharmaChina	2012	All Available online
PhUSE	2005 - 2012	All Available online
SEUGI	1983 - 2004	All Available online
SAS® Forum Int'l	2005-2007	All Available online

Table 1: SAS® Users Groups Proceedings Available Online

The amount of content that has been made available through this effort to the SAS® community is reflected in Table 2. This effort has added access to 14,413 papers so far which represents over 60% increase from previous online accessible content.

Process Undertaken

The process of scanning or digitizing these paper proceedings created a series of images for each page in the paper proceedings. These images were converted into text using optical character recognition (OCR) and saved as pdf formatted files. The OCR process is error prone and some of the problems encountered were documented in the paper by Wooding and La Valley at NESUG 2011 and SESUG 2011. An effort was undertaken to partially clean the files of SUGI but it was abandoned for many of the earlier years because of changing fonts, condition of the scanned proceedings, and time constraints. Images of each page are available to SAS® Users.

Each paper was data mined for the name of the author(s), title of the paper, organization of the author(s), year paper was published, conference, section or stream the paper was part of, and whether

the paper received an award. This information was tabulated and is part of a master index of papers presented at SAS® User Conferences.

Group	Total	Added	% Added
SUGI/SAS®Global Forum	9134	6540	71.6%
PharmaSUG	1749	37	2.1%
PharmaChina	28	0	0.0%
WUSS	2108	1236	58.6%
PNWSUG	286	117	40.9%
MWSUG	1381	869	62.9%
SCSUG	981	644	65.6%
NESUG	3013	1884	62.5%
SESUG	2049	651	31.8%
PhUSE	799	0	0.0%
SEUGI	2435	2435	100.0%
Total	23963	14413	60.1%

Table 2: Number of Papers which were made available with this project

Why Do It?

The need to undertake this effort is not obvious to everyone since the content in the papers presented is not usually relevant to current users. However, the effort was undertaken to preserve the history of the various user group's proceedings and make that content available to the entire SAS® User Community for research. There are many early SUGI papers which have presented key statistical concepts that are available in no other publication. A good example of this is the paper by James Goodnight in 1976 which documents a SAS® procedure for GLM but also provides some foundational GLM theory for statistics.

And the use of such a historical archive online, allows researchers to track the historical development of the SAS® software from beginning to the present. An example of this might be the ability of a researcher to track the papers which discussed software interfaces to access to various databases to the development of the SAS®/ACCESS software and subsequent papers describing the product from SAS® and the various enhancements since it was introduced.

Another important by-product of this effort is a single data store of all authors and co-authors of papers that have been presented at SAS® Users Group Conferences. This allows researchers (and conference attendees) to investigate the previous papers and content of the various presenters. This index includes the name of the conference, year of the conference, author first name, author last name, author name, company of author if given, title of paper, award received, stream or track the paper was in, and the link to the paper on line. This index provides an excellent resource for research. A simple example of the potential use is illustrated by the following Word Cloud in Figure 1 which can be easily generated of all the authors who have the word MACRO in their title of their paper. The names of those who presented the most papers dealing with Macros (Carpenter, Whitlock, King, Fehd and Kromrey) clearly are visible in

this nice tool. If a new user wanted to see who had generated the most papers on that topic, he could use such a tool.

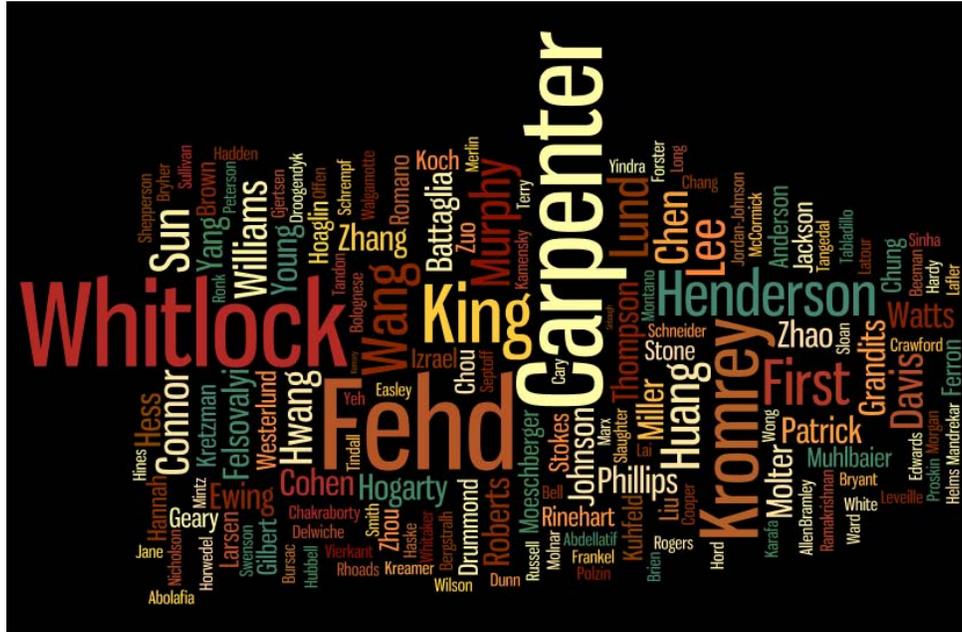


Figure 1: Word Cloud of authors who used the term Macro in SUGI/SAS® Global Forum Paper Titles

If someone wanted to explore the different topics of a particular author of SUGI/SAS® Global Forum papers, Figure 2 shows the kind of display that could be generated with a Word Cloud. As can easily be seen, the author's (Kirk Paul Lafler) key areas of papers are PROC SQL, Output Delivery, JMP, and others.

The content can be used to provide vital information for planning of future conferences for SAS® Users.

It is important for the SAS® Community to begin to explore this wonderful knowledge base and understand how it can be fully utilized to improve the user experience. This collection represents one of the key user contributed sources of information for the SAS® Community. The others include but are not limited to SAS®-L, SAS®Community.org, VIEWS and the evolving SAS® Communities.

Making the Proceedings Available to Everyone

All of the papers which have been captured and indexed by this effort are currently available on various websites on SAS®.support.com, SAS®Community.org, the regional's websites and Lex Jansen's website. For many years, Lex Jansen has provided a consolidated view of these proceedings to the SAS® User Community that provides access all of the papers on his website - <http://www.lexjansen.com/>. It allows the user access to the content provided at user conferences by conference, year, and author as well as

SUGI Conference Papers from 1976-1996

SUGI and SAS Global Forum Proceedings are copyrighted to SAS Institute Inc. All rights are reserved. Any unauthorized reproductions are prohibited.

Read the story of how these historical SUGI proceedings were made available online. [View SEUGI papers here](#)

Click proceedings cover image to [view all](#) papers from that year:



Figure 3: Book Shelf Approach for SUGI Proceedings 1976 to 1996 on sasCommunity.org

SEUGI Conference Papers from 1983-2007

SEUGI and SAS Forum International Proceedings are copyrighted to SAS Institute Inc. All rights are reserved. Any unauthorized reproductions are prohibited.

Read the story of how historical SEUGI proceedings were made available online. [View SUGI papers here](#)

Click proceedings cover image to [view all](#) papers from that year:

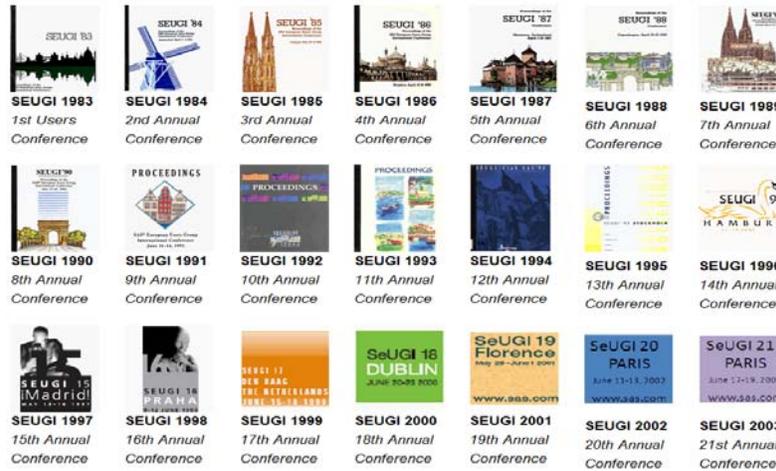


Figure 4: Book Shelf Approach for SEUGI Proceedings

engines adhere to rules and have built-in algorithms to interpret search requests while delivering the “best” possible results; in the end, the derived results are only as good as the search terms entered. To minimize the challenges of finding the right combination of keywords or phrases along with the

<p>9134 SAS papers presented at SGF/SUGI 1976-2012.</p> <p>SAS Global Forum 2013: April 28 - May 1, San Francisco, CA</p>	<p>1749 SAS papers presented at PharmaSUG 1998, 2000-2012.</p> <p>PharmaSUG 2013: May 12-15, Chicago, IL</p>	<p>28 SAS papers presented at PharmaSUG China 2012.</p> <p>PharmaSUG China 2013: September 6-7, Shanghai</p>
<p>2108 SAS papers presented at WUSS 1993-2012.</p> <p>WUSS 2013: September 13-15, Las Vegas, NV</p>	<p>286 SAS papers presented at PNWSUG 1996-1998, 2004-2009.</p> <p>PNWSUG has merged with WUSS</p>	<p>1381 SAS papers presented at MWSUG 1990-2012.</p> <p>MWSUG 2013: September 22-24, Columbus, OH</p>
<p>799 SAS papers presented at PhUSE 2005-2012.</p> <p>Share your knowledge interactively on the PhUSE Wiki. Read the PhUSE Blog.</p> <p>PhUSE 2013: October 13-16, Brussels, Belgium</p>	<p>981 SAS papers presented at SCSUG 1991-2012.</p>	<p>3013 SAS papers presented at NESUG 1988-2012.</p> <p>NESUG 2013: September 8-11, Burlington, VT</p>
	<p>2049 SAS papers presented at SESUG 1993-2012.</p> <p>SESUG 2013: October 20-23, St. Pete Beach, FL</p>	<p>2435 SAS papers presented at SEUGI 1983-2003.</p>
		<p>418 SAS papers related to CDISC. The CDISC Forum.</p>

Figure 5: Lex Jansen’s Website front page for accessing SAS conference proceedings.

frustration associated with unsuccessful searches and massive listings, the authors are exploring next generation of search tools. In image 6, you can see a sneak preview of the current research which the authors hope to share in the near future called sasNerd®.

Better Searches = Better Results

- [SAS software](#)
- [hash object](#)
- [arrays](#)
- [best practices](#)
- [sascommunity](#)
- [SAS Excel](#)
- [big data](#)
- [PROC SQL](#)
- [ODS](#)
- [Enterprise Guide](#)
- [BI](#)
- [Dictionary tables](#)
- [SAS user groups](#)
- [DATA step](#)
- [macro](#)
- [performance](#)
- [PROC REPORT](#)
- [debugging](#)
- [SAS functions](#)
- [JMP software](#)
- [SAS-L](#)

Search for:

Search Engines:

Filetype:

Time Period:

Conference:

Copyright © 2012-2013, Kirk Paul Laffler
sasNerd - A New Way to Search
All Rights Reserved.

SAS is the registered trademark of the SAS Institute Inc., Cary, NC, USA.

[About sasNerd](#) [Send Feedback](#) [FAQ](#) [Blog](#) [Contact](#)

Figure 6: Sneak Preview of current research called sasNerd

Conclusion

As the quantity of SAS user content continues to grow, the importance of producing successful searches is not only important, it may be an essential activity in conducting effective research required by an employer. The SAS user community will greatly benefit from having powerful and quality search tools to help perform better searches and achieve better results. The user community will need new tools to take advantage of the content made available through this content.

References

Bedford, Denise and Richard W. La Valley (2011), "Improving Searchable Access to SAS® Global Forum Conference Papers Using Semantic Analysis Methods and Text Analytic Technologies," SAS Global Forum (SGF) 2011 Conference Proceedings, Kent State University and SAIC, Arlington, VA.

Harper, Renee and Lainie Hoverstad (2010), "Discovering the Road Less Traveled to SAS® Information: A Guide for Your Journey," SAS Global Forum (SGF) 2010 Conference Proceedings, SAS Institute Inc., Cary, NC, USA.

Jansen, Lex, SAS and JMP Paper Repository Website, <http://www.lexjansen.com/>.

La Valley, Rich and Nat Wooding, Nat (2011), "Using SAS® to Ease the Proofing of Messy Text", 2011 NorthEast SAS Users Group (NESUG) Conference and 2011 SouthEast SAS Users Group (SESUG) Conference, Strategic Technology Solutions, Herndon, VA, USA and Independent Consultant, Midlothian, VA.

Lafler, Kirk Paul and Charles Edwin Shipp (2012), "Google® Search Tips and Techniques for SAS® and JMP® Users," 2012 South Central SAS Users Group (SCSUG) Conference, Software Intelligence Corporation, Spring Valley, CA, USA and Consider Consulting Inc., San Pedro, CA, USA.

Lafler, Kirk Paul and Charles Edwin Shipp (2012), "The Incredible Google®: Tips and Techniques for Better Searches and Better Results," Odyssey Press, A Division of Software Intelligence Corporation, Spring Valley, CA, USA.

Shaik, Zubair, Satish Garla and Goutam Chakraborty (2012), "SAS® Since 1976: An Application of Text Mining to Reveal Trends," (2012), SAS Global Forum (SGF) 2012 Conference Proceedings, Oklahoma State University, Stillwater, OK.

Acknowledgements

The authors thank Jennifer Waller, Erik Larsen, Stephanie Townsend and Diana Suhr, Poster and Video Presentations Section Chairs, for accepting our abstract and paper. We also thank Rick Mitchell, SAS Global Forum (SGF) 2013 Conference Chair, the SGF Executive Committee, SAS Institute, and Conference Leaders for organizing a great conference!

The authors would also like to thank the Executive Board of the SAS Users Group for funding the initial work on this effort. And would like to thank the Executive Boards of the Regional Users group who provided funding for the digitalization process for each of their respective proceedings. In addition, this

effort would not have been possible without the support of SAS Publishing and the SAS Library staff. And we would like to thank Lainie Hoverstad for posting the SUGI/SAS Global Forum data on sasCommunity.org.

The authors would especially like to acknowledge several key individuals who donated their personal proceedings for the effort: Ray Pass from NESUG, John Xu and LeRoy Bessler from MWSUG, Tom Winn from SCSUG, MariBeth Johnson from SESUG, Peter Lund from PNWSUG and Diana Suhr and Kim LeBouton from WUSS. In addition, there are many others such as Don Henderson who encouraged this effort.

Trademark Citation

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. [sasNerd](http://sasNerd.com) is the registered trademark of Kirk Paul Lafler and Software Intelligence Corporation. Other brand and product names are trademarks of their respective companies.

About the Authors

Richard W. La Valley is the founder and president of Strategic Technology solutions in Arlington, Virginia. As a SAS user since the mid-1970's, Richard has been involved with many of SAS users groups and was the Conference Chair of the 1989 SAS Users Group International (SUGI) Conference in San Francisco, California. He also is responsible for spearheading the projects which have resulted in the additional 60% of SAS community content from the various proceedings.

Lex Jansen has been using SAS software for more than 20 years and works for SAS Institute as a Principle Software Developer. As the developer of lexjansen.com, the web portal for all published SAS and JMP-related content since the beginning of time, Lex has written and presented numerous papers at international, regional, special-interest and local SAS users group conferences and meetings.

Kirk Paul Lafler is consultant and founder of Software Intelligence Corporation and has been using SAS since 1979. He is a SAS Certified Professional, provider of IT consulting services, trainer to SAS users around the world, and sasCommunity.org emeritus Advisory Board member. As the author of five books including PROC SQL: Beyond the Basics Using SAS, Second Edition (Spring 2013); PROC SQL: Beyond the Basics Using SAS (2004), Kirk has written more than five hundred papers and articles, been an Invited speaker and trainer at three hundred-plus SAS International, regional, special-interest, local, and in-house user group conferences and meetings, and is the recipient of 22 "Best" contributed paper, hands-on workshop (HOW), and poster awards.