

SHIONOGI Global SAS System Renewal Project – How to improve Statistical Programming Platform

Yura Suzuki, Shionogi & Co., Ltd.;
Yoshimi Ishida, Shionogi Digital Science;
Malla Reddy Boda, Shionogi Inc.;
Yoshitake Kitanishi, Shionogi & Co., Ltd.

ABSTRACT

SHIONOGI Global SAS® (G-SAS) System is statistical programming platform to support Shionogi & Co., Ltd. (Japan) and Shionogi Inc. (US) to develop and validate statistical programming deliverables such as SDTM, ADaM and TLFs. G-SAS is the core system to collaborate programming activities between Japan and US SHIONOGI locations, and it operates around the clock. As increasing Clinical Database, Datasets size, Multiple languages and various statistical analyses are causing insufficient memory and hard disk drive. And also to upgrade current SAS 9.2 version to SAS 9.4 version to a new physical server with higher specifications with 64-bit architecture, to handle both SAS versions, and larger datasets, SHIONOGI G-SAS Server Renewal Project was established.

In Addition, method of access control for specific folders post data base lock for double-blind study was reconsidered to control users in accessing analysis results including actual treatment code and potential un-blinded data to avoid the risk of insider trading. To avoid the potential risk, G-SAS Project Team has developed a SAS macro to remove/grant user groups from/to specific folders. This macro has reduced task for System Operation Team.

As a result, SHIONOGI's Global Statistical Programming platform has been improved to be able to handle large datasets, multiple versions of SAS and to control access rights to specific folders and users efficiently. G-SAS System continues to contribute to collaboration of statistical programming activities between multiple locations in SHIONOGI, and contribute to preparation of data submission to regulatory agencies of various countries including FDA and PMDA.

INTRODUCTION

SHIONOGI Global SAS (G-SAS) System is statistical programming platform to support Shionogi & Co., Ltd. (Japan) and Shionogi Inc. (US) to develop and validate statistical programming deliverables of clinical studies such as Study Data Tabulation Model (SDTM), Analysis Data Model (ADaM), Tables, Listings and Figures (TLFs) and so on. G-SAS is an on-premises system and is on the intranet of SHIONOGI, and users can connect to other systems (e.g., data repository system, document management system, electronic trial master file (eTMF) system, etc.) via G-SAS System. Therefore, the users can upload statistical deliverables from G-SAS System to other systems easily.

The first G-SAS System was released in 2012 as a global system to be used by all Statisticians, Data Scientists, Statistical Programmers and Clinical Pharmacologists in both Japan and US SHIONOGI locations. Since SHIONOGI develops multiple compounds globally, operating G-SAS System around the clock gives wings to collaborate programming activities globally as shown Figure 1. For example, Programmers in Japan create deliverables (e.g., SDTM for Study A and ADaM for Study B) in Japanese business hours. Programmers in US proceed subsequent works (e.g., ADaM creation in Study A and ADaM validation for Study B) in US Business hours. In this way SHIONOGI shares statistical programming resources efficiently and preserves quality of each deliverable which will be submit them to Regulatory Agencies including FDA and PMDA by using G-SAS System.

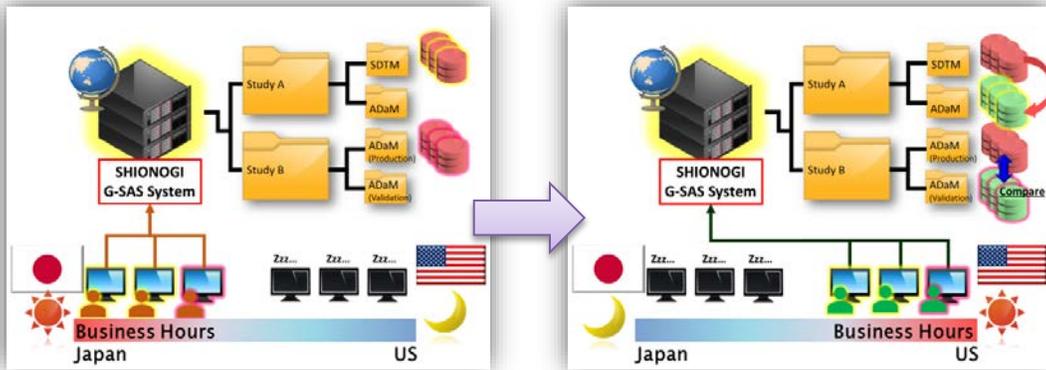


Figure 1. Example of global collaboration for statistical programming activities of each clinical study

Six years have been completed since the first G-SAS System was released, so there are some challenges to overcome. First, users need to handle huge data and data of multiple languages, and to perform various statistical analyses. It causes insufficient memory and insufficient hard disk drive of the server. Second, we have been using SAS 9.2 as SHIONOGI default version to create statistical deliverables, however SAS 9.4 was released and SAS 9.2 is getting to be a previous version. Third, Operating System (OS) of G-SAS System is getting to be a previous version and it has a risk of end of support.

To resolve these issues, SHIONOGI has upgraded the server to a new one which has sufficient server capacity and sufficient processing speed of SAS with 64-bit architecture. SHIONOGI G-SAS System Upgrade Project was established and it consisted of Business, IT and Global CSV-QA (Computerized System Validation Quality Assurance) teams from both Shionogi & Co., Ltd (Japan) and Shionogi Inc. (US).

FEATURES OF THE UPGRADED G-SAS SYSTEM

SPECIFICATION OF SERVER

To handle huge data and to perform various statistical analyses, server specification of G-SAS System has been upgraded with 64-bit architecture, higher CPU cores, higher memory size, and larger hard disk drive than the original server of G-SAS System. Also, to handle data of multiple languages, SAS Unicode Support has newly been installed as well as DBCS (double-byte character set) version and SBCS (single-byte character set).

After the upgrade and G-SAS System is released, SAS 9.4 is a default version instead of SAS 9.2. To facilitate regulatory authorities' requests to perform additional analyses with the same original submission version of SAS (i.e., SAS 9.2), the validated environment for SAS 9.2 is still necessary. G-SAS Project Team discussed how to retain execution environment of SAS 9.2, and we decided to build one physical server and have multiple environments - SAS 9.4 production and validation environments, and SAS 9.2 production and validation environments. The environments for SAS 9.4 were newly developed. On the other hand, the environments for SAS 9.2 were migrated from the previous server, as shown Figure 2. This allows flexible allocation of memory between environments depending on our usage rate of SAS 9.2 and SAS 9.4. Also, it reduced task of computerized system validation, just prepared one document of design specification for one physical server, and two documents of Installation and Operational Qualification (IOQ) for validation environment and for production environment.

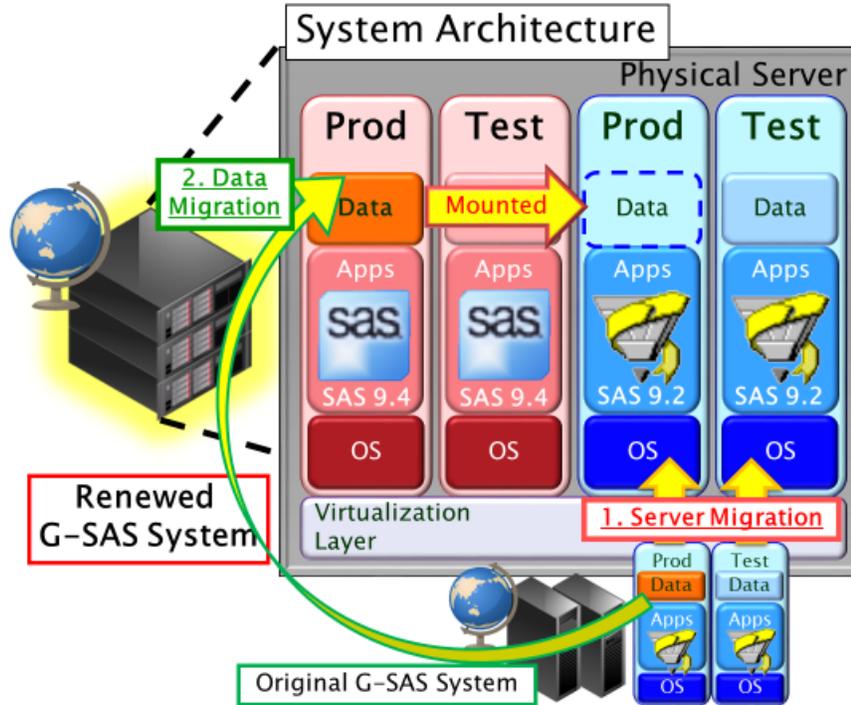


Figure 2. System architecture and server/data migration

Clinical study data were migrated from the original server to the new server, as shown Figure 2. Storage of the data is located at the production environment of SAS 9.4, and the drive is mounted as a folder in both the production environment of SAS 9.4 and SAS 9.2. This makes it an open possibility to manage clinical study data under the two environments.

As a result, the upgraded G-SAS System ensures reproducibility for legacy programs which were created by using the original G-SAS System. If regulator agency requests new analyses for submitted studies, the environment of SAS Version 9.2 in the upgraded G-SAS System is available for the requested analyses.

OPERATION - HOW TO RESTRICT ACCESS RIGHT FOR UNBLINDING?

Access control under project/protocol folder of the G-SAS System was automated by using the method of Koretaka et al. [1]. Especially, method of access control for specific folders post data base lock for double-blind study was reconsidered in the G-SAS Project Team.

Why do we need to restrict access right?

In preparing deliverables, project team member (Project/Trial Statistician, Data Scientist and Programmers) need to grant read/write permission for each folder/file. When we prepare deliverables for double-blind study, dummy data of treatment code (i.e., information of planned and actual treatment group) is used before unblinding. Figure 3 shows an example of access right of Compound X. Project/Trial Statistician, Data Scientist and four Programmers (Programmer A, B, C and D) belong to Project Team of Compound X, and they are able to access the folders/files of Compound X. Let's assume that Study A is conducted as double-blind, and Project Team of Compound X is working on preparation of each deliverable of Study A.

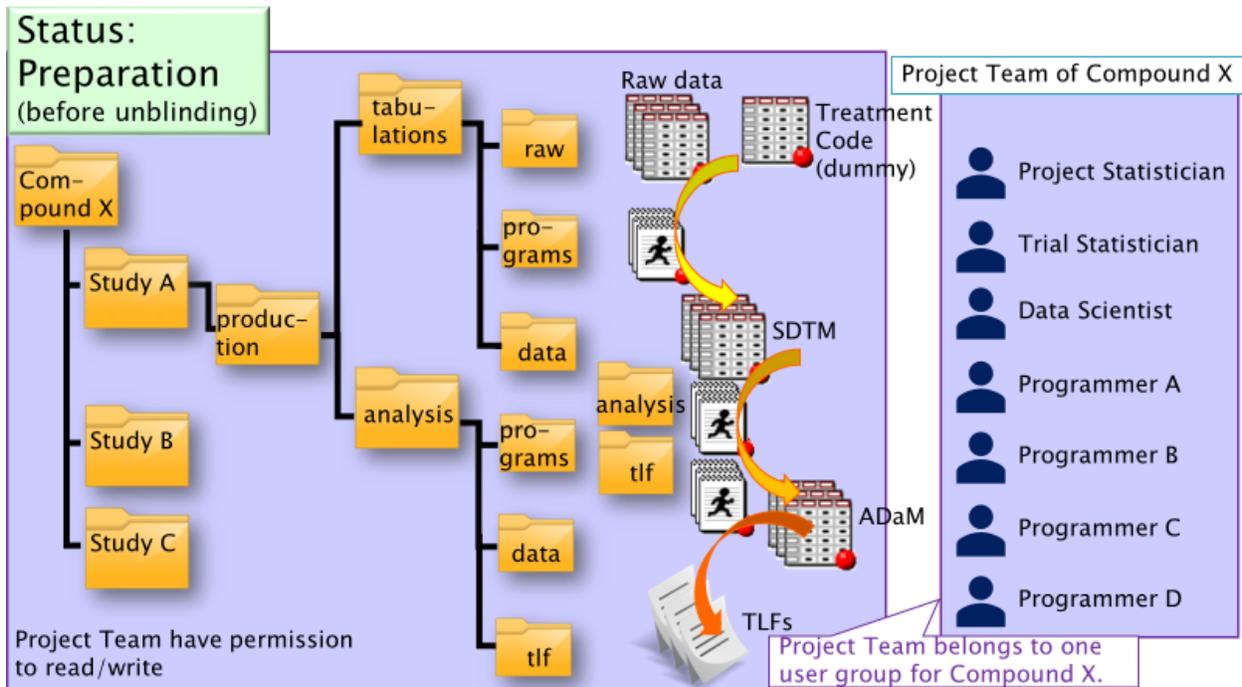


Figure 3. Access right of Compound X - before unblinding

Now preparation is completed before data base lock, and Study A is unblinded. After unblinding, the Project Team receives actual data of treatment code. By using the final raw data including real treatment code, Programmer A and B produce SDTM, ADaM and TLFs, as shown Figure 4. After that, folders/files highlighted in pink contain data of real treatment code.

Let's assume that Programmer C and D were just assigned to support Study A before unblinding. In that case, they should not access folders highlighted in pink before analysis results are disclosed to the Project Team, otherwise they are able to run programs to produce TLFs and they may know the study results. This is a potential risk of insider trading or information leakage.

Shionogi manages user group by compound (e.g., CompoundX_PROG for Project Team of Compound X), not by study level. Once a member is added to CompoundX_PROG user group, he or she can access all studies of Compound X. This helps to consider development strategy of each compound for Project Statistician, Trial Statistician and Data Scientist. Also, if multiple studies for one compound are ongoing, the same Programmer may work on multiple studies of one compound. Let's go back to the example shown in Figure 4. If Programmer C and D are working on other study (e.g., Study B) for Compound X, it's impossible to delete these users from CompoundX_PROG user group. On the other hand, we need to control access right appropriately to avoid the potential risk. If we grant access right to unnecessary person, it renders him/her at increased risk of information leakage whether consciously or not. To grant access right to essential personnel, we can minimize the potential risk.

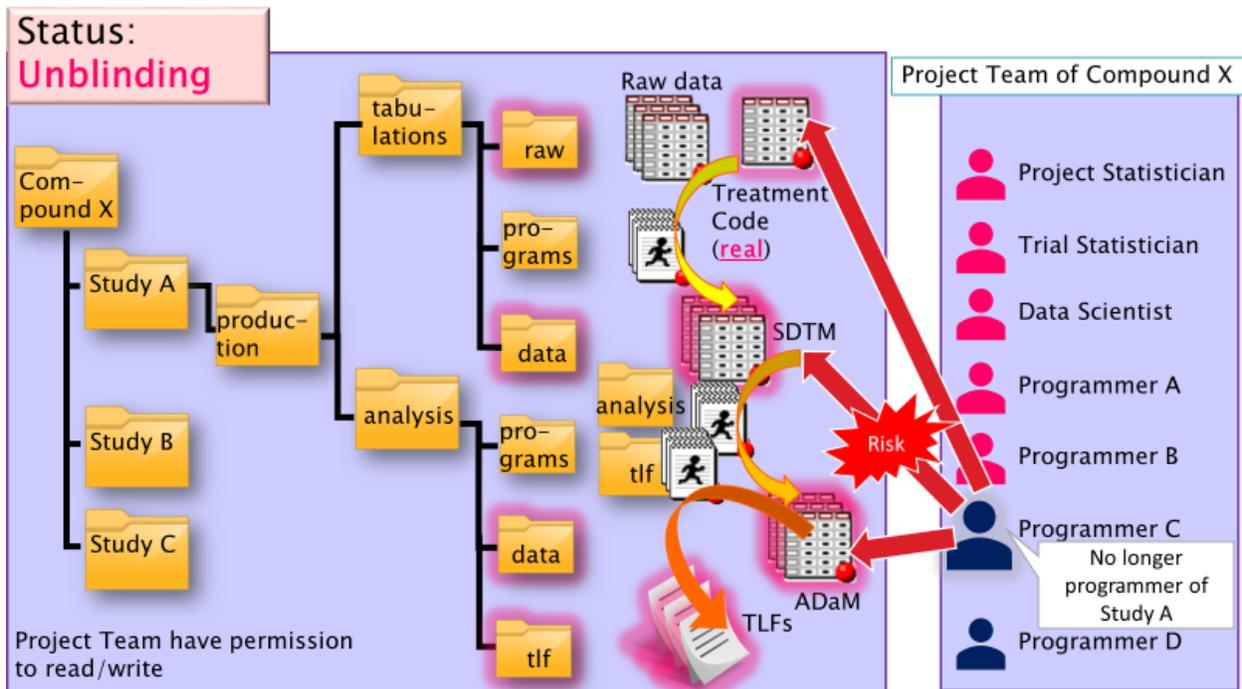


Figure 4. Access right of Compound X - after unblinding

How to restrict access right?

To avoid the risk, G-SAS Project Team has developed a SAS macro to remove user groups from specific folders which contains data of real treatment code (i.e., folders highlighted in pink in Figure 4) just before unblinding, and to grant user groups to the specific folders.

Prerequisite

Project Team Members are assigned to one or two user groups of G-SAS System based on their roles;

- If a user should access analysis results including real treatment code, he or she needs to be assigned to “(Compound name)_PROG” (user group for Project Members by compound) and “(Compound name)_(Study name)_RPROG” (user group for restricted Project Members by study). For example, Programmer A is assigned to CompoundX_PROG and CompoundX_StudyA_RPROG.
- If a user is a programmer and is not allowed to access analysis results including real treatment code, he or she needs to be assigned to “(Compound name)_PROG” only. For example, Programmer C is assigned to CompoundX_PROG only.

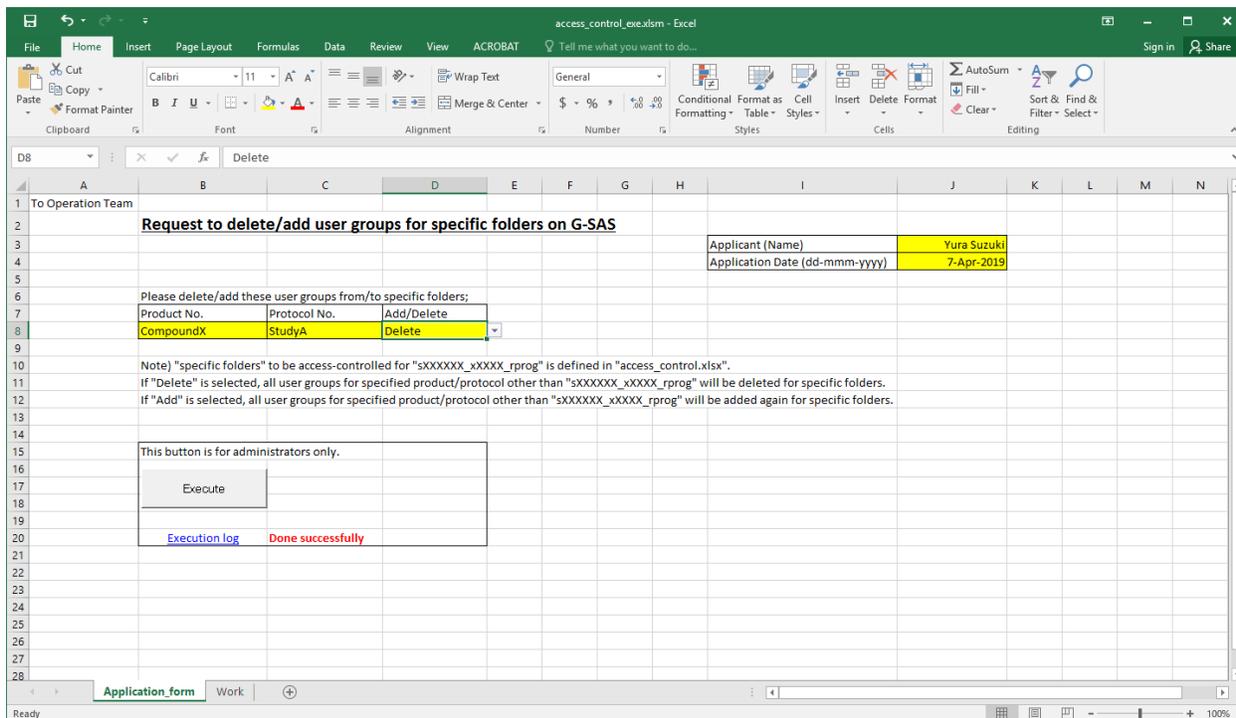
When the Project Team need access-restriction just before unblinding, Representative of the compound (e.g., Project Statistician or Data Scientist) should send a request to System Operation Team of G-SAS System including the following information;

- Compound name
- Study name
- Select “Delete” or “Add”. If the Project Team need access-restriction just before unblinding, “Delete” should be selected. On the other hand, after the study result is disclosed to all and additional resources is needed due to a lot of additional analyses, “Add” can be selected.

Macro Execution

System Operation Team fills in the form of spreadsheet based on the request from the Representative, and then presses the button of “Execute” (see screenshot of Display 1). Then SAS macro is executed. If

“Delete” is specified, user groups other than “(Compound name)_(Study name)_RPROG” (e.g., “(Compound name)_PROG”) will be deleted from the specific folders. If “Add” is specified, user groups other than “(Compound name)_(Study name)_RPROG” will be granted again.



Display 1. Access right of Compound X - after unblinding

APPENDIX shows the detail of the SAS macro and VBA code for access restriction.

After execution of this macro, the necessary log file is automatically stored. By using the different user group (e.g., “(Compound name)_(Study name)_RPROG”) for the access-restricted area and by executing this macro, we can easily manage access control for highly confidential data with real treatment code, as shown Figure 5. All Project Team members of Compound X are added to “CompoundX_PROG” for preparation of Study A first. Additionally, Core Members of Study A are added to “CompoundX_StudyA_RPROG” so that they can produce final deliverables with real treatment code and they can see the study results immediately to consider the strategy of next study or to prepare submission for the approval of the Compound X. Just before unblinding, the user group of “CompoundX_PROG” is deleted from the access-restricted areas of the Study A and only restricted personnel can access the information which contains real treatment code. Programmer C and D are able to continue working on other study (e.g., Study B) for Compound X, because they have already been added to CompoundX_PROG and there are no restrictions for folders/files of Study B.

regulatory agencies of various countries including FDA and PMDA. I'm very happy this paper gives an idea of collaborating statistical programming activities globally with secure environment.

REFERENCES

- [1] Yuichi Koretaka et al. 2014. "Automatic Access Control via SAS: More Efficient and Smart." *Proceedings of the PharmaSUG 2014 Conference*, San Diego, CA : PharmaSUG.
Available at <https://www.pharmasug.org/proceedings/2014/AD/PharmaSUG-2014-AD06.pdf>.

ACKNOWLEDGMENTS

The authors wish to acknowledge colleagues from Japan who belongs to Biostatistics Center, Global CSV QA at Shionogi & Co., Ltd., and Shionogi Digital Science for their contributions in the G-SAS System Upgrade Project. And also from US, colleagues who belongs to Biometrics Department, IT Administration Department and Global CSV QA at Shionogi Inc. for their suggestions to the G-SAS System Upgrade Project Team.

Also, the authors thank Accenture PLC for supporting implementation of the upgraded server, and thank System Communication Assist co., Ltd. for validating the SAS macro to restrict access rights.

CONTACT INFORMATION

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

Yura Suzuki
Shionogi & Co., Ltd.
+81-6-6485-5094
yura.suzuki@shionogi.co.jp

Yoshimi Ishida
Shionogi Digital Science
+81-6-6209-8004
yoshimi.ishida@shionogi.co.jp

Malla Reddy Boda
Shionogi Inc.
+1-973-307-3350
Malla.Boda@shionogi.com

Yoshitake Kitanishi
Shionogi & Co., Ltd.
+81-6-6485-5094
yoshitake.kitanishi@shionogi.co.jp

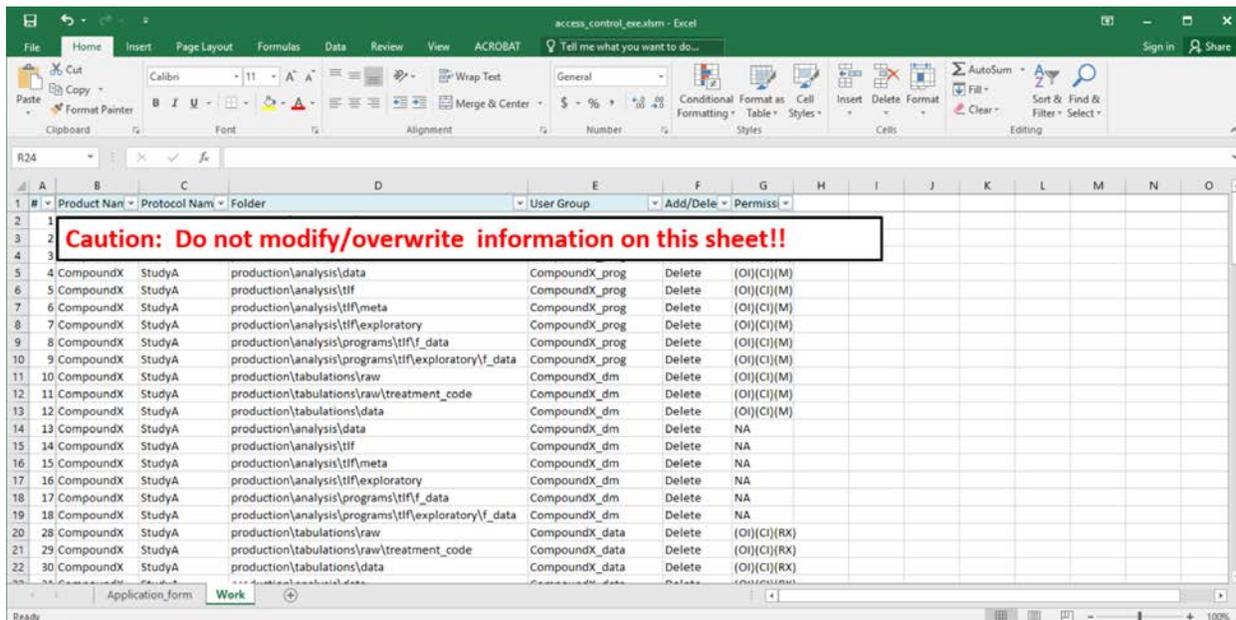
SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

APPENDIX

DETAIL OF THE VBA AND SAS CODE FOR ACCESS RESTRICTION

As shown Display 1, System Operation Team fills in the form of spreadsheet and press “Execute” button. After that, target folders and restricted user groups are automatically specified on the separate sheet as Display 2. Usually, System Operation Team don’t need to update any information of this sheet unless there are no additional restricted folders.



Display 2. Information of target folders and restricted user groups

After pressing “Execute” button of Display 1, VBA code as shown Output 1 runs to execute the SAS macro.

```
Sub CommandButton1_Click()
    retval = Shell("""[Drive]\Program Files\SASHome\SASFoundation\9.4\sas.exe" "
    -config ""[Drive]\Program Files\SASHome\SASFoundation\9.4\SASV9.CFG" "
    -sysin [Program Path]\[Program Name].sas
    -log [Log Path]\[Log Name].log")
End Sub
```

Output 1. Output from VBA code to execute SAS macro

SAS macro imports the spreadsheet of Display 2 and converts it to a SAS data set, see Table 1.

Column name of Excel spreadsheet (Display 2)	Variable of SAS data set (data set name: perm_00)	Description
Product Name	product	Compound name
Protocol Name	protocol	Study name
Folder	folder	Folder path Actual folder path is: [root path]\[product]\[protocol]\[folder].

Column name of Excel spreadsheet (Display 2)	Variable of SAS data set (data set name: perm_00)	Description
User Group	group	User group name (e.g., (Compound name)_PROG, (Compound name)_(Study name)_RPROG)
Add/Delete	status	"Delete" or "Add" If the Project Team need access-restriction just before unblinding, "Delete" should be selected. On the other hand, after the study result is disclosed to all and additional resources is needed due to a lot of additional analyses, "Add" can be selected.
Permission	permission	If "Add" is selected in "Add/Delete" column of the spreadsheet, permission which is specified in the "Permission" column is applied when ICACLS command is executed.

Table 1. Relationship between spreadsheet and variable of SAS data set

By using the SAS dataset of Table 1, the following SAS code runs to delete/add user groups from/to the folders:

```

data perm_01;
  set perm_00; *Converted data set from Display 2;
  length fullpath $2000 command $3000;
  fullpath = cat("[root path]", trim(product), "\", trim(protocol), "\",
trim(folder));
  if upcase(status) = "DELETE" then command = cat("icacls ", trim(fullpath),
"/inheritance:d | icacls ", trim(fullpath), " /remove:g ", trim(group));
  if upcase(status) = "ADD" then do;
    if upcase(permission) ^= "NA" then do;
      command = cat("icacls ", trim(fullpath), " /inheritance:d | icacls ",
trim(fullpath), " /grant:r ", trim(group), ":", trim(permission));
    end;
    else if upcase(permission) = "NA" then delete;
  end;
run;

*Create batch file to confirm the detailed log of command prompt after
execution;
filename txt "[Path]/command.bat";
data _null_;
  set perm_01;
  file txt dsd;
  put command;
run;

```