PhUSE EU Connect 2018

Paper TT02

SAS Life Science Analytical Framework (LSAF) implementation in a nutritional clinical research and development setting.

Hartweg Mickaël, NESTEC S.A., Lausanne, Switzerland
Justine Morlet, NESTEC S.A., Lausanne, Switzerland

ABSTRACT

This paper aims at providing an overview of the implementation process of SAS® LSAF system at Nestlé Clinical Development Unit. It highlights the clinical information technology (IT), data management and statistical perspectives.
The presentation covers the rationale of the implementation, the challenges met and the solutions deployed to overcome them, as well as the lessons learned (programming, documentations, guidelines and processes, trainings). The next steps towards a full integration of the system into the internal standard processes (for both insourced trials and outsourced clinical trials) conclude the paper.

INTRODUCTION

In the last decades, the progressive acknowledgment of the benefit of nutritional interventions to maintain healthy status and prevent diseases related to food intake led to a substantial growth of nutritional research. Nutritional companies, such as Nestlé, built groups dedicated to clinical research, development and substantiation of the communication on the health benefit of its nutritional products. The mission of such a unit is to conduct clinical research and run trials according to existing regulations, guidelines and methodologies. Many of those are coming from the pharmaceutical industry, which was the pioneer in human research. Reproducing one to one the ideas / structures developed by the pharmaceutical industry in a copy-paste fashion and applying them to nutrition research is not always ideal. However, the subjects' personal data storage and processing are important steps of the clinical operation process that need to comply with security, reproducibility and traceability rules as described in guidelines and regulations. This is why a data storage and analysis environment such as SAS LSAF is important when dealing with clinical data. In this paper, we present our implementation journey, by highlighting the benefit of such a solution in a nutritional research environment, as well as its burden and its limitations. We sincerely hope that this would benefit to those working in a similar environment.

THE INITIAL NEEDS

Nestlé clinical development unit (CDU) is responsible for designing and executing clinical trials. Its goal is to generate data based evidences for supporting the communication on the health benefits of Nestlé products and for generating knowledge that contributes to new discoveries and innovations by the Nestlé Research Center. Data management and statistical analysis are key components of the trials execution process. In 2016, the team started to look for a software solution to increase efficiency and quality of data management and statistical analyses. The improvement points identified were the need of processes that are more homogeneous, the need of a better system for the storage and archiving of clinical trial data, the need of a platform for sharing the data and programs with internal and external collaborators and the need of a standardized data extraction process from upstream environments (e.g. Medidata© Rave). The team decided in April 2016 to acquire SAS Life Science Analytical Framework v4.7 (LSAF). SAS LSAF is a web based software solution combining a controlled data repository (similar to Microsoft SharePoint) and a SAS programming environment. The "data repository" component brings the versioning, audit trail and access control requirements. The system also allows a direct link between eCRF software solution and the data repository via a connector designed on this purpose. SAS LSAF has a SAS-language coding environment and the possibility to automate programs execution. All the ingredients were available to kick-off the expected improvements.

DATA MANAGEMENT

From data manager perspective, the needs were to better control the data handling, storage and sharing with internal and external collaborators. Initially, each data manager extracted manually the data from the eCRF management solution. The statistician or any data processor requested the extraction to the data manager. The documentation of the request consisted in the email exchange, or a paper form completed for official extracts to prepare interim or final analyses. Secure, but inefficient solutions like password protected zip files plus email, or

Microsoft SharePoint served as data transfer media. This process was known to be resources and time consuming. By implementing the SAS LSAF solution, there is no more intermediates between the raw data from Medidata Rave and the central storage location. Access to the data by the users is controlled and documented via an extensive audit trail. Moreover, any individual having the access rights to do so could extract the data without any manual intervention of the data manager.

STATISTICAL ANALYSIS

From statisticians' (i.e. "data processors") standpoint, the benefit of the SAS LSAF environment is the homogeneity in the ways of working within the group, the possibility of sharing in a secured manner the data and the statistical programs with the external collaborators, as well as the efficiency in documenting the analysis process. Indeed, with the new General Data Protection Regulation (EU) 2016/679 ("GDPR") [1] that became enforced in May 2018, it is even more important to ensure a data-sharing environment that protects the subjects participating in a clinical trial and their personal data. Although the process of dossier submission for regulatory approval is less demanding for the nutrition industry compared with the pharmaceutical one, the reproducibility of statistical results and the documentation of the analysis process is there of equal importance, as it is in any research field. Implementing SAS LSAF, overcoming all the challenges and issues described further, allowed us to be ready for the future requirements on data protection and processing rules that applies to nutrition research.

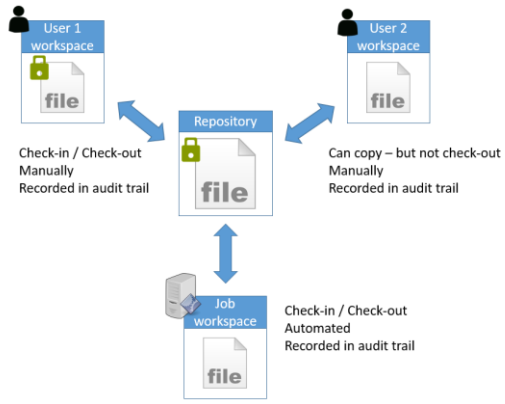
INFORMATION TECHNOLOGY NEEDS

The cloud and web based nature of the tool was also a key driver. Fitting more within current IT trends and strategies in term of implementation, daily support and infrastructure. A web-based solution allows to get rid of the installation and licenses renewal steps, by transferring IT responsibilities directly to the solution provider. This responsibilities transfer has direct impacts in term of internal resources and linked costs.
However in the case of "Software as a service", it is requiring to think beforehand about and then to set up appropriately the governance model and service level agreements, in order to make sure that the provider deliverables are aligned with the users' needs.

SOFTWARE PHILOSOPHY

Without going in depth into the details, it is important to understand the basic principles behind SAS LSAF, and the vocabulary that goes with it, to follow the rest of this paper.
The central data storage place, called "repository", is accessible by all users (those having access rights). There is a strong control process on the repository, which consists in registering the details of each actions performed into an audit trail called "manifest". A system of versioning control makes sure that all the files stored are archived, and that each new version(s) of the same document supersedes the previous ones.
As it is the case in Microsoft SharePoint, any files to be processed (data, programs, outputs and documents) must be checkout of the repository before being processed. If a user is willing to write a new line of code into a program for instance, the program is first checkout in the user's own workspace and locked in the repository. The user workspace is not controlled, neither accessible to other users – it is personal and does not include the audit trail functionality.
Once the modification made, the user must check-in back into the repository the new version of the file. This removes the lock on the file into the repository and the new version supersedes the previous one.
For automated updates, for instance executing a SAS program and producing a new version of a table in the repository, there is a new type of programs, called jobs. The job checks-out / runs the programs / checks-in back all needed material in an automated fashion. In a way, the job "acts" as a virtual user.
For a SAS program to run properly, the files paths allocation must be dynamic. Indeed the physical location is not the same if you are running the program into your user workspace, via a job into the repository or if another user is running the same program into his own workspace.
All of those new steps makes the automated audit trail possible, but it increases as well the number of steps before having a set of programs ready for the production of TLFs.

Figure 1 – Basic principle of SAS LSAF



CHALLENGES AND SOLUTIONS

The Nestlé team recognizes that implementing and using such an environment comes with many challenges. The key to handle them has been to set up the appropriate governance model between Nestle CDU and SAS team. Management team decided that a group of three members from Nestlé, comprising a clinical IT specialist, a user from data management group and a user from biostatisticians group would perform the governance. This team interacts with the SAS team on a regular basis, to share needs and issues from the Nestlé users and to monitor the implemented solutions.

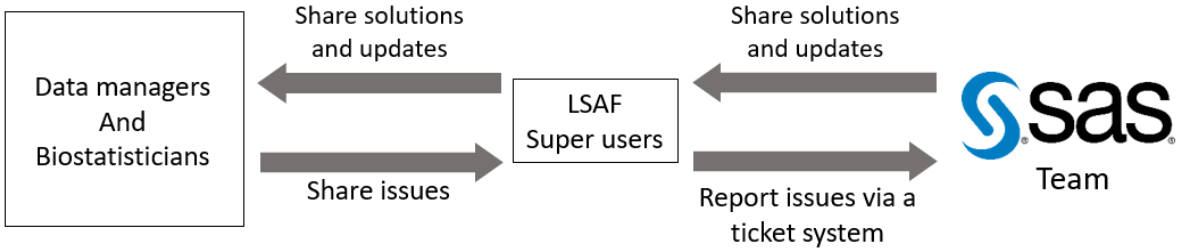
DAILY ROUTINE

Once the set-up was done and the "go live" became effective, the two super users for both data management and biostatistics developed a daily routine. This routine, described below, ensures efficient communication from the users (data managers and biostatisticians) to the SAS team and back.

In order to collect feedbacks from the users, it was decided to implement a so call: "LSAF – Issues and solutions to share" file, which consists in a matrix, open to all the users, where the issues encountered can be described in details. SAS LSAF superuser tries to solve the issue internally, and if it is not possible, escalates the issues to SAS team via a ticket system. Once a solution is found and put in place, the SAS team informs (via the ticket system, but as well during regular meetings) the superuser team on the solution. The super user then documents the solution with detailed descriptions in the shared file. The team decided to set-up regular meetings between superuser and the internal users in order to share the latest information.

The routine set-up allows fluid execution of those tasks. However, this come with additional need of resources.

Figure 2 – Dedicated escalation process



In addition to the standard support to resolve programming/use issues or misunderstandings, the implementation of SAS LSAF added an administrative layer on both super users, who became responsible to manage users' accounts, passwords and rights.

DOCUMENTATION DEVELOPMENT

The team rapidly identified the need to revise existing documentations (e.g. Standard Operating Procedures), and to develop new operating procedures.

The first document developed by the superuser team was a "role and responsibility" RACI matrix to define each activities allocation. The team wrote a list of activities dedicated to SAS LSAF maintenance and usage, followed by several discussions in order to find the most appropriate team members to execute the task. The data managers became responsible for the Medidata Rave to SAS LSAF connector implementation, for each of their trials. The team decided that each functional superuser would be responsible for its own pool of users support, including the

administrative tasks (password reset, accounts blocked, programming issues, and so on...). This exercise helped the team to realize that SAS LSAF would not replace on a one-to-one fashion the previous system in place, which was a combination of SAS local licenses and various data storage solutions.
Once the roles and responsibilities clarified, the team wrote a guidance for trial creation within the system. This activity is the first step towards the complete integration of the trials into a SAS LSAF-driven process. It ensures that all trials are set-up in appropriate similar way, and thus information are stored homogeneously into the SAS LSAF repository. The document covers the "by-default" options and ensures that the naming convention and the accesses rights are set-up according to the planned role and responsibility matrix. Training sessions and continuous support from the super users to the team followed the release of the guidance.

The biostatistician super user wrote a programming guidance, for the internal users. This programming guidance is not a repeat of the SAS trainings and documents, but rather a summary, focusing on Nestlé user daily tasks rather than covering the full possibilities of the system. The writing of this guidance required a good understanding of the system by the writer to make it simple for the users.

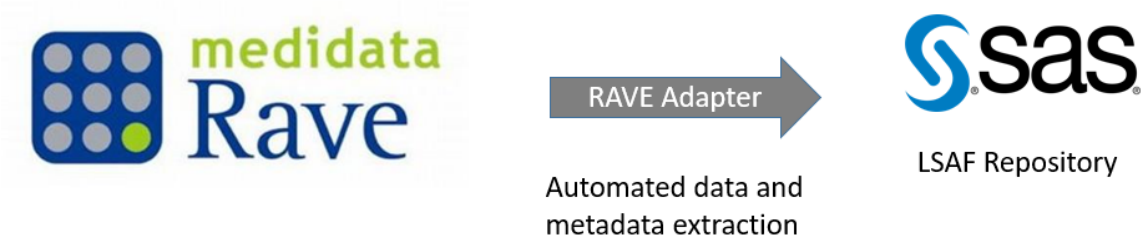
As Nestlé CDU clinical trial portfolio is quite large, the team set-up a rollout plan to decide which ongoing trials would be migrated to the new environment, depending on their operational phase related to statistical analysis milestones. The new studies are by default integrated into SAS LSAF.

IMPROVEMENT OF EXISTING TOOLS

In parallel to the documentation development, the current CDU tools had to be adapted in order to be usable within the SAS LSAF environment. This allowed the team to re-evaluate their usage and to think about potential improvement.
The new source of data physical location drove the changes. The team revised all macros (including data monitoring toolkit and biostatistics' standard programs), as the data were previously stored in local drives or internal servers and are now located on a cloud-based server.

The new step of automatic data extraction required to develop a completely dedicated process. A communication stream between Medidata ODM Adapter and Medidata Rave EDC (Electronic Data Capture) Interface enables the data extraction from Medidata Rave to SAS LSAF repository. The Medidata ODM Adapter provides interfaces to web services that support the secure transfer of data. The Medidata Rave EDC Interface contains a set of validated programs. Those programs, as well as the job linked to the programs, reside in SAS LSAF. When running, this job performs the data extraction. User must specifies only few parameters: study number, environment, datasets names – in case only a specific subset of datasets needed.
This process is very powerful – as of today and to our knowledge– as it contributes to a significant reduction in the data managers' workload. Indeed, once in place, this data extraction from Medidata Rave to SAS LSAF takes less than a minute to trigger, and then a couple of minutes to run (when the user can continue working on different tasks).

Figure 3 – Data extraction flow



In addition to the new data extraction process described above, the "data quality review" process had to be adapted. The purpose of the Data Quality Review is to monitor the clinical data in order to check the data quality. The data manager presents to the trial core team some graphical representation of the data cleaning progresses, source data verification and blinded descriptive statistics about simple measures related to the trial intervention. By reviewing those materials, the team helps the data manager to identify current data issues and foresee upcoming ones.
The data management superuser developed a standard job in SAS LSAF from previous codes available. The advantage of using one standard job is to get an homogeneous way of doing the data quality review between the studies as well as keeping, once again, a full audit trail of which data has been reviewed, when and how.

TRAININGS

SAS LSAF has many advantages, but the interaction with the user is not its strength. Superuser team performed extensive information and training sessions in order to give to the users the confidence to use the tool in an appropriate and effective way. The SAS team provided beforehand two trainings for general knowledge and superuser awareness.
The Nestlé superusers put in place a newcomer training session, as the usage of such a programming environment is new and not implemented in the curriculum of universities. Even those familiar with SAS programming language benefit from the training sessions, as SAS LSAF recquires understanding of the control mechanism as well.

ACCEPTANCE BY THE TEAM – CHANGE MANAGEMENT

Implementation of new solutions are coming with challenges, complexities and demand users to leave their comfort zone. It has been peculiarly the case with SAS LSAF. Such a shift, to be successful, requires a change management strategy. A change management strategy ensures a smooth transition of the users from established comfort zone to the new environment. The Nestlé change management strategy consisted in a list of actions to set-up during each phases of the transition. During the implementation phase, representatives of each user teams (data managers and biostatisticians) were involved and regular updates provided to the teams. It helped them to get familiar with the tool terminology. It helped as well the superuser team to identify the early concerns, anticipate them and plan for solutions focused on user. During the intermediate phase, the superuser team put in place a relaxed rollout plan granting to user sufficient adaptation time. The user receives constant support from the superuser team who acted as ambassadors off the new environment and promoting the long-term advantages to the teams. Super user team conducted many internal open discussions, formal and informal to achieve this goal.

OVERALL PERFORMANCE

As highlighted above, the onboarding phase is not without challenges. The superuser team has to define a plan targeting both newcomers and employees who never used a similar system. The learning of few additional programming features, the creation of an automated audit trail are just few examples that require adaptation. This means that those who plan resources must allocate more time than one would expect to the transition phase. Before a user starts to feel autonomous, he needs to undergo several trainings sessions. Even then, because of the lack of intuitiveness of the system, the autonomous users have to come back regularly to superuser team. This counter-balance the efficiencies coming from the system and counts negatively when thinking about overall performance.
On the positive side, not repeating ourselves about the efficiency coming from the data extraction process, it is important to note that once SAS LSAF production environment go live is done, and superuser team familiar with the tool and its responsibility, it takes only a few minutes to create an account and thus give access to a new user. As soon as the superuser creates the account, the appropriately trained user is able, via any web browser, to connect and start working. However, once again, this strength is as well a weakness when the internet connection is slower than usually. The working experience is immediately impacted and the user is very quickly frustrated to wait for each letter to appear when writing a piece of program. In addition, web based means that without internet connection, there is no way to work at all. This is an issue for users travelling a lot and used to work in the planes, trains, or any transport mode where the internet connection might be slow or non-existent, which is the case of some of the CDU users.

NEXT STEPS FOR NESTLE CDU

We acknowledge that many of the SAS LSAF options are not (fully) used as of today by Nestlé team. However, superusers created a mid-term and long-term plan in order to leverage each of unused but known options and achieve full potency of the system.

SCHEDULING

SAS LSAF has a scheduling option, which makes possible the planning of programs execution without any human intervention. The current version available at Nestlé allows to configure a date / time (single run or recurrence), which is of a limited usefulness (for us). It allows running regularly, for instance, data extract making sure that the latest data are always available for the data processors. However, Nestlé trials are not large enough to benefit from the small running time saved by this way (a few minutes).
An interesting development would be to schedule program execution not on a pre-specified date and time, but on a pre-specified condition fulfilled by the data recorded itself. For instance, to trigger the run of a statistical interim analysis the day the XXth subject information was source data verified. An additional improvement would be to allow SAS LSAF system to inform the user via email alert that the run execution ended, or even better to mail the results automatically.

DASHBOARDING

Another interesting development is to improve reporting capabilities in SAS LSAF. Currently, those capabilities are only available in upstream systems (e.g. Medidata Rave at Nestlé CDU).
The idea would be to combine all metadata sources into SAS LSAF. For instance, we could get the queries list and any information about the cleaning process analyzed via SAS LSAF analytic tools. It would be then possible to use those data to build monitoring reports. For the time being, metadata about the queries (i.e. number of opened queries, closed queries …), as well as about the source data verification (i.e. number of pages verified and still to

pending for verification by the monitor) are not part of the data extraction. Once in place, the idea would be to track the cleaning of the data via user's defined listings and fine-tuned reports using SAS LSAF.

EXTERNAL USERS

Another expected development would be to grant access not only to internal Nestlé users, but as well to contract research organization currently subcontracted by the Nestlé CDU. This would replace the traditional secured data transfer processes and would allow performing statistical analyses without getting any data out of the controlled repository.
However, this would come with new challenges: first, the collaborators not familiar with SAS LSAF would recquires the same extensive training session than internal users (as described above). Second, the time estimate, and thus the budget, requested by the sub-contractor to complete the activity would for sure increase.

OTHER ANALYTICAL SOLUTIONS

One of the most important and expected improvement would be the implementation of an R [2] platform linked to SAS LSAF data repository. As many of our internal and external collaborators may use other statistical software (R being the most used one, but thinking as well about a Python environment), it is important for us to allow access to such programming environments. The access should be as easy as the access to the SAS programming window already embedded into SAS LSAF.
A temporary process has been set-up by superuser team to allow single user to analyze the data outside SAS LSAF with an alternative software like R. However, this process is against the principles of the data repository control and superuser team will disable it as soon as a better-integrated way to do will be set-up.

CONCLUSION

SAS LSAF implementation in a nutrition research context showed us that such a software solution comes with many advantages on long-term efficiency, sensitive data protection and security. Web based storage and analytical solutions combined seems to be a good alternative to old fashion processes mixing local software licenses and various type of data repositories. However, as shown, one must anticipates and addresses several challenges before such an environment reaches its full potential and becomes an alternative. A cloud-based system brings new tasks and requires more resources during the implementation and the onboarding phase. Especially, the change management does not need to be underestimated. It would be easier for customers if IT providers would include the user-friendliness of the tool as an axis of development of equal importance than its capabilities.
Nestlé CDU team considers that taking up these challenges is worth the effort in order to prepare for the future, to continue improving our ways of working, to increase efficiency and to accelerate the clinical operation process thanks to innovative solutions.

REFERENCES

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)
OJ L 119, 4.5.2016, p. 1–88

[2] R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

ACKNOWLEDGMENTS

The authors thank the internal reviewers Maria Laura Gosoniu (CDU Biostatistics Team Lead), François Caijo (CDU Data Management Team Lead) and the whole CDU team for their supports and comments, which helped us to write this paper as completely and clearly as possible.
We thank as well the SAS team, for their continuous support on this long journey to implement SAS LSAF at Nestlé. Big thanks to the biostatistics and data management teams for their engagement and commitment throughout the process.

RECOMMENDED READING

SAS LSAF system fact sheet:
https://www.sas.com/content/dam/SAS/en_us/doc/factsheet/life-science-analytics-framework-107108.pdf

CONTACT INFORMATION

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

Mickaël Hartweg
NESTEC S.A.
Email: Mickael.Hartweg@rd.nestle.com

Justine Morlet
NESTEC S.A.
Email: Justine.Morlet@rdls.nestle.com