Paper DH04

Migrating to LSAF – The Danone Nutricia Research Journey

Paul Vervuren, Danone Nutricia Research, Utrecht, The Netherlands
Jean-Marc Ferran, Qualiance, Copenhagen, Denmark

ABSTRACT

In 2017, Danone Nutricia Research (DNR) defined their Digital and Analytics strategy to advance their clinical systems and data flow and to improve access to data. As part of this initiative, DNR selected the SAS® Life Science Analytics Framework (LSAF) as their Statistical Computing Environment. Migrating to LSAF from a SAS Server coupled with a file system was the opportunity for DNR's Clinical Data Science department to integrate further their processes across Data Management, Statistical Programming and Statistics, as well as to gain tighter control over access and user management. Following go-live, DNR continued to improve their usage of LSAF, its integration with other systems and developed more utilities to optimise tasks and processes. The implementation project was run over a period of one year (2018). This paper will elaborate on the implementation and migration strategy, the benefits of moving to LSAF, lessons learned and next steps in the journey of clinical data system development at Danone Nutricia Research.

INTRODUCTION

In 2017, Danone Nutricia Research (DNR) defined a 'Digital and Analytics strategy' for its Specialized Nutrition organisation. In relation to this, an elaborate global programme was started, with a number of interrelated projects. The migration to the SAS® Life Science Analytics Framework (LSAF), as a new Statistical Computing Environment for clinical study data, was one of these projects and was the first in the area of clinical research.

This Digital and Analytics programme presents the broader context that influences the LSAF project. Let's therefore consider its goals and approach. The goals of the programme were to:

1. Explore and apply the use of new digital technologies to advance clinical research and real-world evidence generation
2. Streamline the data flow and improve the interoperability of (clinical) research data to enable real-time information provision
3. Make data more accessible to a broader R&D population to generate new insights and research hypotheses
4. Protect DNR's data assets and ensure compliance with applicable regulations (including the GDPR), as well as quality requirements for clinical research
5. Prepare for responsible data sharing.

It was acknowledged that achieving this ambitious set of goals would require various technical, procedural and organisational adaptations, e.g. the implementation of systems, deployment of tools and services, new policies and work procedures, new roles and collaborations. To create a workable way forward, a roadmap was put in place with concrete objectives and results that were expected to be achievable in 1–2 years. The fundaments laid by this first stage would form the basis for successive stages. These were to encompass further integrations and extensions of data systems, the development of collaboration workspaces for data analytics, a full-blown data governance framework, provision of data services and catalogs, and common data capture processes and standards.

The LSAF project, while being part of a bigger programme, was run largely as an independent project. A second early-stage project, named the 'Virtualization & Visualization' project (V&V), existed under the umbrella of the Digital and Analytics programme. The V&V project aimed to create an infrastructure connected to LSAF that allowed the development, piloting and deployment of clinical data services that could be offered to users outside the Biometrics department, such as Clinical Study Researchers, Study Managers and Medical Monitors. For example, a 'Study Catalog' was developed, analogous to the concept of AstraZeneca's 'MetaDataHub' [1], as well as prototype dashboards for the ongoing review of CRF data loaded into LSAF. Jointly, these projects address most of the longer-term goals of the Digital and Analytics programme, with LSAF being linked mostly to points 2 and 4, and the V&V project to 3 and 5.

In this paper we will describe DNR's journey of migrating to LSAF, touching mainly on project management aspects, technical issues and the organisational change.

THE BIOMETRICS CONTEXT

In the introduction, we set the broader scene and global drivers of this project. Let's now look at the local context and needs of the intended LSAF core users. These are all members of the Biometrics department, consisting of data managers, statistical programmers and statisticians supporting clinical research in specialised nutrition. The user needs and requirements were as follows:

- **Computing performance**. The department's legacy Statistical Computing Environment was formed by SAS (server-based) as the statistical analysis software, with programming artefacts (datasets, programs, logs, outputs) filed on a local network system, providing storage, back-up and access management. Over the years the department had grown substantially, but computation power of the SAS server had not been adjusted. On top, there had been an increase in heavier jobs (more complex modeling, and larger, sometimes multi-study, datasets) from the Statistics team.
- **Managing the working environment**. Another challenge that emerged over time, as teams grew and data accumulated, was managing the digital work environment: how to keep an overview of files, folders, access, etc. Initially, procedural measures had been sufficient to manage these, in particular when the teams were still small and functions overlapped (data manager/ programmer, programmer/statistician), and when there were still a limited number of studies to manage. While clinical data generally is low in volume, the numbers of files involved in each study can be huge as a result of the variety of data captured, elaborate analyses with many outputs produced, and the different milestones in the life cycle of a study such as interim analyses, safety reviews, headline results reporting, post-hoc analyses, and publications. Also, these systems were not built specifically for managing clinical data and do not provide any specialised workflows and tools, e.g. for e-signing or access reporting. Manual processes and administration have thus been put in place, which are generally inefficient and error-prone. Finally, legal requirements for managing personal data have increased with the arrival of the GDPR.
- **Flow of data**. Linked to the previous point, the legacy environment was lacking advanced tools and capabilities to implement automated processes for data import and life-cycle management. As a result, data extraction from EDC systems, reconciliation processes, as well as downstream transfers, were done in a manual fashion. Versions of data tend to build up as their clean-up is also a manual process. Automated data flows, and better integration of work processes between Data Management on the one hand, and Programming & Statistics on the other, will help to better manage data, and is a requirement for real-time or near real-time data access for data monitoring and decision-making purposes. Moreover, future data acquisition scenarios (devices) are likely to render streaming data, for which API-led connectivity and automated data loading become a need.
- **Leveraging data standardisation and data integration initiatives**. There had been ongoing efforts in the department to implement standards, in particular the harmonisation of CRF templates (data acquisition standards), the adoption of SDTM terminology (variable names, controlled terms) in the EDC database, and the conversion to SDTM of a selection of completed studies to enable pooled analyses. To further extend and valorise these efforts requires better tools and capabilities to implement and manage data standards, as well as facilities for responsible sharing and broader use of pooled datasets.
- **Use of new analysis tools, specifically R**. In recent years R has become more and more prominent. Specifically, in our Statistics team, new team members have been using R as their primary statistical programming language (the same holds for many scientists in other departments). Also, with the growing complexity of study designs and advances in statistics, the need to explore and apply the latest statistical methodologies – available in R – has grown. This may be more prominent in nutrition research compared to pharma, given the smaller treatment effects, the greater variety of interaction factors, and the longitudinal nature of the data. The popularity and possibilities undoubtedly present great opportunities, but they also bring challenges: (1) the technical integration of R in a SAS-oriented environment, (2) the implementation of processes to ensure R processing (for non-exploratory purposes) is validated and reproducible, (3) the broader adoption of R, and collaboration between programmers (predominantly SAS users) and statisticians (predominantly R users).

SELECTION CRITERIA

The selection of a new statistical computing environment (SCE) was driven by the global consideration arising from the Digital and Analytics programme, as well as the more proximate context and criteria of the Biometrics team. There were basically three main considerations/criteria: 1. Limit complexity (one vendor, proven solutions), 2. Focus on end usage and value creation (customisable, off-the-shelf solutions; mature technology), 3. Connectable (strong API features). SAS LSAF presented a mature SCE system with strong API features which we could extend and integrate with SAS solutions for data federation, data integration and data visualisation (for 'V&V'). Combined, these covered most of the Biometrics', as well as the Digital and Analytics programme's, goals and needs, allowing for further growth and extension, with increasing usage and maturity of data services.

PROJECT PHASING AND PROGRESSION

The phasing of the project is given in Figure 1. From inception to full production mode, the project knew four phases, which together covered a period of about 16 months. Starting from the date the production system was made available by SAS, go-live was achieved in 10 months. At that moment, User Acceptance Testing (UAT) was completed, and the first users, working on legacy study activities, started working in the system.

ACTIVITY TRACKS

Four parallel activity tracks were set out: 1. Specification and Testing, 2. Business (process and user-oriented topics), 3. Data Migration, 4. Technical (covering authentication/security and integrations). The inception and planning of the project took a little less than three months, starting with a preparatory workshop provided by Qualiance and ending with the release of the LSAF 4.7 production environment by SAS, followed by LSAF training for the core implementation team.

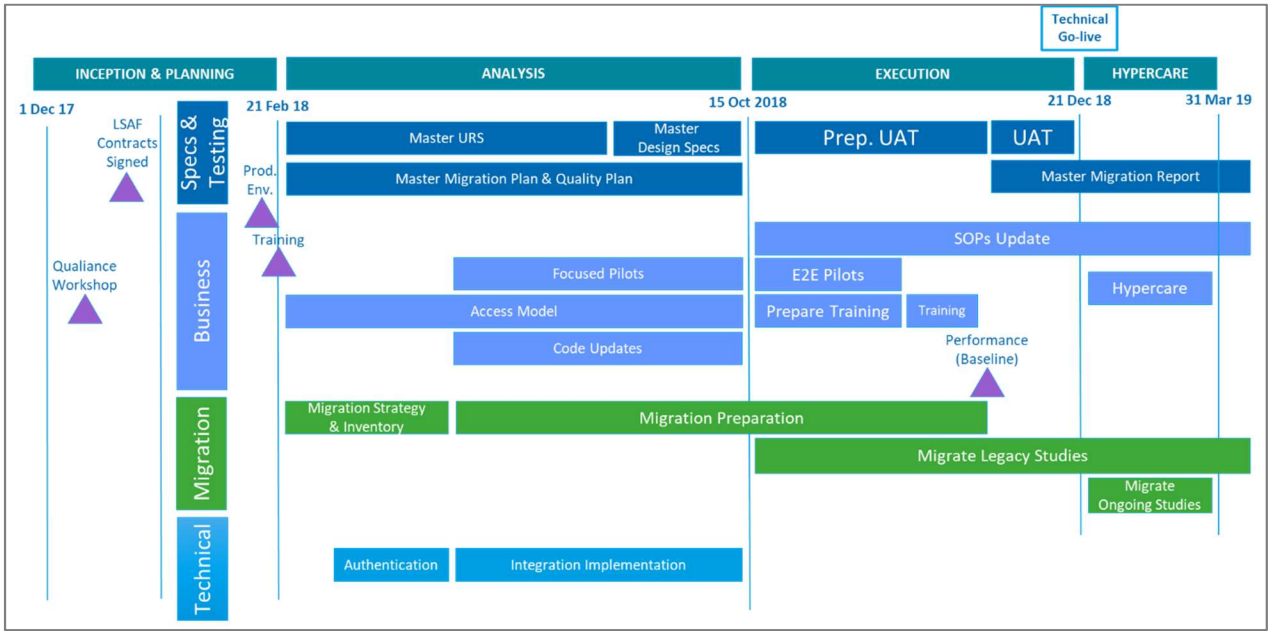

Figure 1 – Project Plan & Timelines

PREPARATORY WORKSHOP

During the workshop an assessment was made of the current state of processes within Data Management, Programming and Statistics of DNR. With the LSAF 4.7 capabilities and constraints in mind, an impact assessment was performed, for example on the data loading and transfer processes, folder structure, version control, and productivity tools (e.g. log checker). Moreover, the deployment strategy was considered; that is, whether we would take a 'big bang' approach or an iterative one, starting with a subset of users (e.g. 10 of 50) as early adopters and introducing the rest after a first cycle of learning and improvement (we chose the first option). We also addressed system integrations, in particular of the EDC and Safety Database systems (both external to DNR), and for running R. Several follow-up queries were formulated to aid in further planning, e.g. the availability of APIs or web services for the interfacing systems. Finally, we assessed change management, based on implementations elsewhere, considering both advantages and pain points felt by users. The workshop also helped to get a picture of the nature and extent of the various activities and the organisational impact. Further, it helped to onboard the implementation core team, and to create further awareness and buy-in within the Biometrics management team.

PLANNING, FROM START TO GO-LIVE

In the weeks following the workshop, the activity and milestone planning was drafted. The target date for go-live was initially set to mid-September 2018, with the realisation that this was probably an optimistic projection. Indeed, we had

to adjust go-live to a later date in the course of the project. There were a few weeks of delay, for logistical reasons, in the planning of the core team training and kick-off meeting between DNR and SAS. A larger delay was caused by the Focused Pilot activities, which took much longer than initially thought. Summer holidays also slowed down progress to some extent. This led us to postpone the target date for go-live to the end of November 2018. During the execution phase, workload and pressure were increasing on all tracks. Training for users had to be prepared, which was a logistical challenge (meeting rooms, accommodating users in Utrecht and Singapore) but also required significant preparation (developing slides, exercises). The End-to-End Pilot resulted in substantial updates to the folder structure and permission model and some other changes. The assumption had been that the End-to-End Pilot would be largely a confirmatory exercise, so the updates had to be absorbed as extra work. UAT had to wait for this to be completed. Finally, data migration was executed in this phase. While the technical steps of the migration steps could be executed efficiently and fast, dependencies between the study programmers (who needed to give input to the process and verify successful delivery in LSAF) and the migration team (coordinating the work and executing the migration) caused the overall legacy migration process to be stretched. As a result of all of this, a few extra weeks had to be added at the end; yet, we still managed to reach go-live within the first year.

RISK-BASED REQUIREMENTS ANALYSIS & TESTING

Requirements were gathered across functional areas and assessed in terms of risk. Risk was determined as a function of probability, impact and non-detectability, and requirements were then classified into three categories:

Risk Level	Testing Level	Example
Low Risk	No Testing*	Typically, processes or sub-processes involving standard functionalities with low impact
Medium Risk	Single Testing	Typically, processes or sub-processes involving standard functionalities with medium impact or custom development with low impact
High Risk	Extensive Testing	Typically, processes or sub-processes involving standard functionalities with high impact or custom development with medium to high impact

*: It must be noted that SAS extensively tests standard functionalities as part of their OQ, and our core team also used them extensively during the course of the project when piloting the different processes and developing custom code. This was also considered in our assessment of probability and non-detectability.

Requirements were also assessed in terms of priority. Such priority determined whether requirements are conditional to the system going into production:

Priority	Condition
Must	Must pass UAT without workarounds in order to go into production
Important	Can pass UAT with workarounds in order to go into production
Desired	Could improve User's Acceptance. Not compulsory to address to go into production
Nice-to-have	Only to be addressed if time and resources allow

The analysis of risk of each requirement enabled us to focus on the 30% of requirements that were classified as Medium and High Risk while documenting the rationale of our risk assessment and assigned testing level across all requirements. The time spent upfront during the analysis phase of the project in risk assessment ensured that we had determined objectively the level of testing and could focus our efforts on the most critical requirements.

PILOTING & PROCESS DESIGN

The core team analysed the processes in scope of the LSAF migration, as well as processes peripheral to LSAF and where integration with other systems was needed. The diagram below (Figure 2) details the different processes at stake for the project.

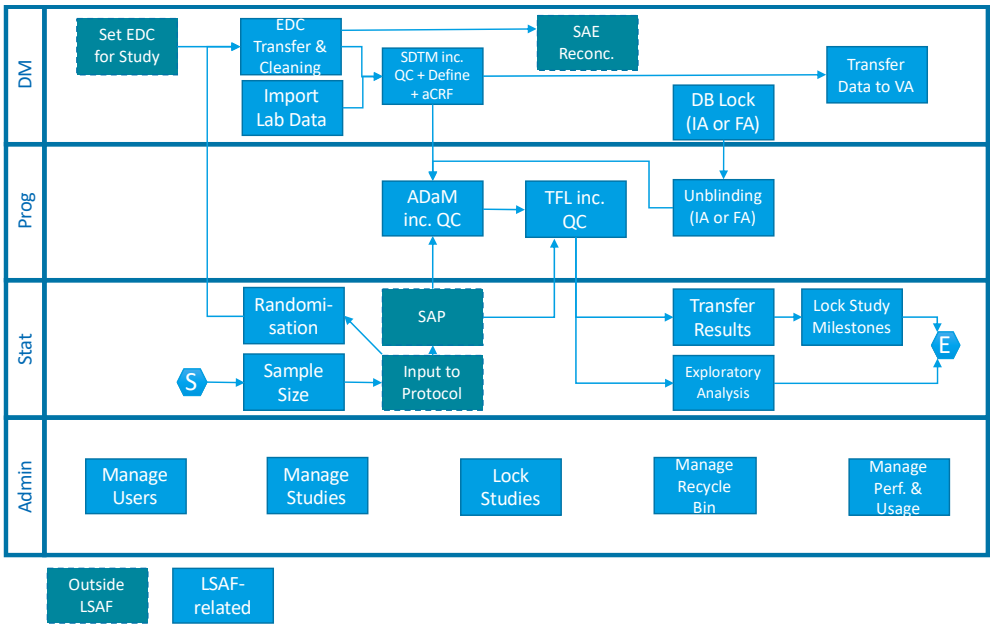

Figure 2 – DNR Processes in Scope of LSAF Migration

Each process was prototyped through individual Focused Pilots and reviews conducted by the entire core team. Each Focused Pilot was then updated and an impact analysis conducted to determine technical and functional impact to supporting utilities and procedures, respectively. Once all processes were piloted and associated utilities developed, two End-to-End Pilots were conducted for two different studies. For one of these pilot studies, an interim analysis was also simulated. Each End-to-End Pilot started from study creation and each process was conducted iteratively, or in parallel, right through to locking of studies and conduct of exploratory analysis (Figure 3).

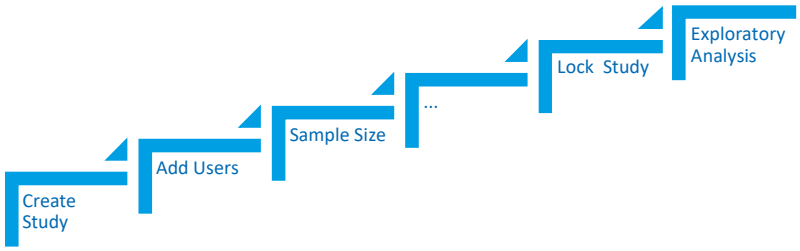

Figure 3 – Focused & End-to-End Pilots

The conduct of two End-to-End Pilots prior to UAT enabled us to get final confirmation on both the access model and the processes and associated utilities we had developed in almost real settings. There was a special focus on verifying that the access model both secured the level of confidentiality required across the different milestones and group of users and provided enough permissions to conduct tasks and access necessary data. It was possible for us to detect issues prior to UAT and then have draft scripts and test code from the pilots to reuse when developing final UAT test cases.

STUDY MIGRATION

A Study Migration Strategy document was developed during the analysis phase and it was agreed not to migrate all studies but select only legacy studies with potential for re-use on top of ongoing studies. It was also agreed to have the study teams select folders and data to be migrated and not by default consider all data, files and folders within a study on the legacy systems. These two decisions meant that more analysis and collaboration with study teams was necessary as preparation but this would lead to a lighter and easier-to-navigate structure in LSAF in Production. All study folders in legacy systems would then be locked and available for consultation or later use.

The initial analysis resulted in migrating the data of:

- 26 ongoing studies
- 33 legacy studies.

100+ legacy studies were considered out of scope and remain on the legacy system drive.

A dedicated sub-team was assigned to Study Migration in terms of coordinating input required from study teams and defining and testing scripts to support the migration, re-mapping and QC of the transfers. The figure below describes further the full process we designed to perform migration of each study in scope:

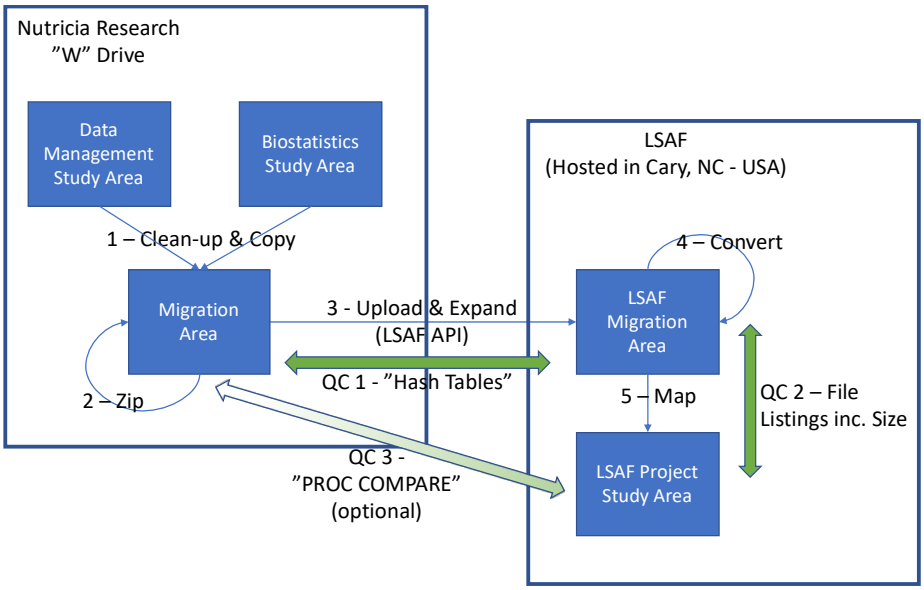

Figure 4 – Study Migration Process

The Study Migration process was a six-step process (Figure 4) involving three types of QC activities:

- QC of transfer from the legacy drive ("W" Drive) to LSAF through comparison of hash tables
- QC of remapping of files and folders within LSAF through comparison of file listings including size
- Final QC of datasets using a SAS "PROC COMPARE" between the original datasets on the legacy drive and the final ones that were converted and remapped in LSAF

This extensive focus on QC ensured neither data loss nor data integrity issues. Performing different types of QCs for each major transfer enabled us to track at the source any potential issues and take any corrective action, as the issues were identified.

Scripts were developed to support or automate each step of the process, and to generate reports for documentation.

The Study Migration sub-team were able to start migrating legacy studies during the Execution phase of the project in parallel to other activities (End-to-End Pilots, UAT, etc.). On the other hand, ongoing studies that also required remapping and would be completed in LSAF at a later date, using the new processes, were migrated in batches following completion of UAT. This also enabled us to onboard users iteratively into LSAF, as their ongoing studies were made available in the new SCE.

DEVELOPMENT OF UTILITIES

During the Focused Pilots and based on the analyses of the different processes, it was decided to develop different utilities to automate certain tasks. LSAF comes with a complete set of APIs as Java classes or SAS Macros [2], which we were able to leverage to custom develop utilities fitting with the defined access model and DNR's processes.

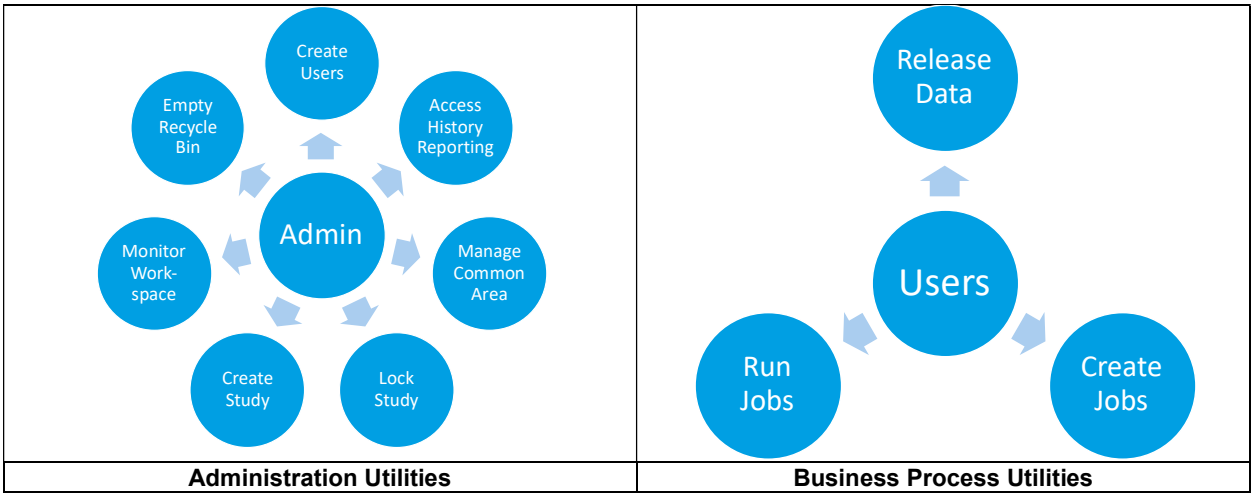

Figure 5 – Administrative and Business process related utilities developed during the project

In particular, it was possible to automate most administrative tasks to ensure full control over the actions to be performed, as well as to minimise human error. Figure 5 summarises the two sets of utilities the team developed to support both administrators and end users.

PROJECT DOCUMENTATION

The project followed a Computer Software Validation (CSV) methodology. The different aspects of the project were documented in key documents across phases and milestones, as listed in the table below:

Specifications & Testing	Master User Requirements Specifications
	Master Design Specifications
	Traceability Matrix
	UAT Test Cases and Execution Documentation
Quality Activities	Quality Plan
	Master Migration Plan
	Master Migration Report
Study Migration	Study Migration Strategy
	Study Migration Plan
Roll-out	Roll-out Plan

These documents in their initial version constitute the backbone of the implementation and will serve as the basis for future change control of LSAF at DNR.

In addition, supporting documents were developed to document the piloting phase, the migration of each study (in the form of checklists), and the implementation and usage of utilities (through user guides). A number of SOPs were impacted; these were updated, and related training material developed.

ROLL-OUT AND ADOPTION PROCESS

After go-live there was a phased adoption by the users. For ongoing studies, data migration was performed after go-live, to ensure that there would be no or very limited impact on critical study timelines. For ongoing studies close to or just after database lock, migration was postponed, allowing key milestones to be delivered in the legacy environment. Thus, users working exclusively on completed studies moved fully to LSAF after go-live, whereas others continued to work in the legacy environment for some time, and mixed usage also occurred. By the end of March 2019 all users (and their studies) had migrated to LSAF.

During the training period it became clear that upon go-live users would probably be greatly helped or even require a high level of support, i.e. more than on-demand support by expert users. For that we devised the concept of 'hypercare', which consisted of a series of planned Q&A sessions, the usage of a dedicated ticketing system for reporting and managing issues and requests, and sharing (additional) documentation, plus ad hoc support from the core implementation team. The first Q&A meetings were held as walk-in sessions for all three disciplines (Data Management, Programming, Statistics), but it turned out that each discipline had different needs. Follow-on meetings were therefore organised separately. These included sessions with presentations by the trainers of the earlier end-user training. Thus, the second wave of Q&A meetings had the nature of a refresher training. At this stage we were able to conclude that the earlier training had limitations. Firstly, while the timing suited those who started working in the system soon after go-live, for others there could be a two-month gap before starting to do production work in the system. Secondly, the training had an operational focus, going through all the main work processes. Attendees were expected to familiarise themselves beforehand in the basic concepts and functionalities by reading an introduction manual and viewing some of the videos prepared by SAS. Limited time and attention had been given to the basics in the actual training. Some catching up was therefore needed during the roll-out phase.

CONCLUSION

The implementation of LSAF was a substantial effort and the introduction of this new Statistical Computing Environment brought significant changes to the DNR Biometrics team's way of working. It was a successful journey, though, for several reasons: go-live was achieved within timelines, all study data selected for migration was transferred without data integrity issues or loss of records, and there was no meaningful impact on the continuity of study activities. There were hurdles, for sure, but in general user adoption was successful. Clearly, there was a learning curve after the training. When users started to use the system for daily production activities on their studies, many needed further familiarisation, as well as additional support during the first few months. Yet, in a survey held 8 months after roll-out, close to 90% of the users expressed that their experience working with the system was better or much better than when they first started using it. In the same survey, users identified different benefits achieved with the introduction of LSAF, such as Access Control, Traceability of Actions, Establishment of a Standard Folder Structure and Better Data Flow. Successful migration to LSAF was further illustrated by the fact that the legacy system was decommissioned within three months after go-live, and only continued to be used in this period for studies in a critical reporting phase.

Planning and preparation are key in any complex project. The early workshop we held was important in that sense. The team had never conducted a migration of this kind, and the workshop helped to get an understanding of the effort that was required, the phasing of the project and types of activities that needed to be undertaken, the resources and expertise required, and the impact on processes. Having the experience on board of someone who worked extensively with the system and who had been running similar migration projects was key to planning and preparation, as well as to running the project effectively, utilising the system's potential and solving issues efficiently. The project team was another success factor. At the start of the project, none of the project members had specific experience with the system. By the end, core team members had provided training and guidance to the Biometrics team, designed and executed study data migration, developed and implemented the access model and various tools and utilities, performed administrative tasks and supervision, and more.

LSAF is now the central repository of data and code for clinical studies at Danone Nutricia Research, and all ongoing and new studies are conducted in the system. The replacement of our legacy systems and associated procedures by LSAF represents a major step in DNR's Digital and Analytics strategy, in particular in the areas of interoperability, data protection, compliance and quality. Moreover, the system has brought enhanced computing performance, better management of the Statistical Computing Environment, and improved data flow, thus fulfilling major needs of the Biometrics team, too. Further enhancements will come from continuous review cycles conducted as part of the LSAF Access Management and Administration procedures that we established, as well as from ideas and requests brought forward by users. Finally, DNR is exploring the next version of LSAF, which should also bring various improvements, as well as new features.

REFERENCES

[1]: The MetaDataHub – Integrating, Searching and Presenting Study Descriptive Information. Elisabeth Nyman, Kerstin Forsberg & Michail Doulis. Paper TT07. PHUSE EU Connect 2018.
[2]: LSAF APIs: https://support.sas.com/kb/57/862.html.

ACKNOWLEDGEMENTS

The authors would like to thank the core team members of our LSAF implementation project: Herman Ament, John Hendrickx, Roel Hobo, Gertjan van Maaren, Elleny Balder, Annemarie Wagemans, Hanneke Lankheet, Patrick Varney and Steven Ting. We would also like to thank Josefien Vendrik for SOP reviews and advice on quality system matters. Finally, we are indebted to Jan van der Mooren and Egbert Biesheuvel for sponsoring the project and offering overall guidance and support.

ABOUT THE AUTHORS

Paul Vervuren is a Senior Project Leader at Danone Nutricia Research (DNR), where he is currently responsible for the LSAF project, as well as a Clinical Data Warehouse project. He is also involved in various projects and working groups at DNR, e.g. SDTM Implementation, GDPR, Data Transparency, Clinical Study Data Sharing and Data Visualization. Paul has had different technical and leadership roles within Clinical Data Science/Biometrics throughout his career. A central theme has been to achieve high-quality process improvements and innovations through design thinking and collaboration. He has presented various papers about his work at PHUSE conferences since the first event in 2005. In recent years, Paul has been co-leading the PHUSE Single Day Event in Utrecht, the Netherlands.

Jean-Marc Ferran is a Life Sciences Entrepreneur based in Copenhagen, with 15+ years of experience in the Life Sciences industry. Prior to starting his company, Qualiance, he worked as a Statistician, Standards Manager and as Director of Statistical Programming at Novo Nordisk and Ferring Pharmaceuticals. During his career, Jean-Marc has worked on many Statistical Computing Environment Implementations in business and technical positions and is also a user of these systems. He has a unique 360-degree experience spanning from working as an end user, supporting customers as an implementation partner, to developing trusted solutions alongside software vendors. In particular, he has worked with LSAF (previously SAS Drug Development) since 2005 and has since been following carefully the development of the product. Jean-Marc also supported Danone Nutricia Research as Project Co-Lead during their LSAF implementation, which is related in this paper.

CONTACT INFORMATION

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

Author Name	Paul Vervuren
Company	Danone Nutricia Research
Address	Uppsalalaan 12, The Netherlands
City/Postcode	3584CT
Email:	paul.vervuren@danone.com
Web:	www.danone.com, www.nutriciaresearch.com

Author Name	Jean-Marc Ferran
Company	Qualiance ApS
Address	Copenhagen, Denmark
Email:	jmf@qualiance.dk
Web:	www.qualiance.dk