Paper DH05

Utilising LSAF for data management at a Belgian biotech

Louella Schoemacher, OCS Life Sciences, ‘s-Hertogenbosch, Netherlands

ABSTRACT

While PC-SAS and Enterprise Guide are still widely used, they are slowly but surely being pushed aside by web and cloud-based solutions such as SAS Studio and SAS Life Sciences Analytics Framework (LSAF). LSAF is a hosted solution that is designed to be much more than just a programming environment. It is an integrated data repository and computing environment offering features such as audit trail and versioning out-of-the-box.

Over the past year OCS Life Sciences has led the implementation of LSAF at a Belgian biotech company. This encompassed the (technical) configuration, design of the folder hierarchy, associated permission and security model, defining best practices for the use of the system, creation of jobs and process flows, and much more.
The LSAF environment is now in production and being utilised by data managers, programmers and statisticians. In this paper we will share the most important learnings of our journey.

INTRODUCTION

OCS Life Sciences is a specialist CRO, based in ‘s-Hertogenbosch. The services provided within the life sciences industry are mostly clinical and statistical programming, biostatistics, and CDISC data conversion. OCS Life Sciences is part of OCS Consulting, providing more services such as managed services and technology solutions and support.

A perfect example of how these services complement each other can be found in the implementation of a customised LSAF environment for clients. For a Belgian biotech client this was a joined effort by consultants from OCS Consulting who managed the set-up, installation and maintenance of the system, and consultants from OCS Life Sciences involved in the folder structures, permission and privileges model and creation of jobs.

LSAF SET-UP

The implementation of LSAF started in March 2020. From that moment on, workshops and brainstorm sessions were organised to come to a tailored LSAF environment. These sessions were attended by the client’s end users, rather than an IT and/or QA employee. In this case, they were a lead biostatistician and the head of data management. From OCS, these sessions were attended by LSAF subject matter experts, knowing a lot about the system from a technical but also practical point of view.

DESIGN

The design of folder structures in LSAF is configurable and completely up to the client’s wishes. This means that the folder structure can be rather simply or highly complex. As visualisations help to understand the process, the folder structure was designed in an Excel file. Since this approach helped to simplify the process as one could clearly see how the structure would look like, the same was done for the permissions and privileges model (figure 1).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Type	level_1	level_2	level_3	level_4	level_5	level_6	level_7	level_8	level_9	level_10	system_admin	XstudyX_dm_readonly	XstudyX_stat_readonly	XstudyX_prg_readonly	XstudyX_dm	XstudyX_stat	XstudyX_prg
2	businessunit	XbusinessunitX										ARPCD/ARPCD	-R---/	-R---/	-R---/	-R---/	-R---/	-R---/
3	drug		XdrugX									ARPCD/ARPCD	-R---/	-R---/	-R---/	-R---/	-R---/	-R---/
4	indication			XindicationX								ARPCD/ARPCD	-R---/	-R---/	-R---/	-R---/	-R---/	-R---/
5	study				XstudyX							ARPCD/ARPCD	-R---/	-R---/	-R---/	-R---/	-R---/	-R---/
6	folder					data						ARPCD/ARPCD	-R---/-R---	-R---/-R---	-R---/-R---	-R---/-R---	-R---/-R---	-R---/-R---
7	folder					metadata						ARPCD/ARPCD	-R---/-R---	-R---/-R---	-R---/-R---	-R---/-R---	-R---/-R---	-R---/-R---
8	folder					production						ARPCD/ARPCD	-R---/	-R---/	-R---/	-R---/	-R---/	-R---/
9	folder						data_managers					ARPCD/ARPCD	-R---/-R---			-RPC-/-RPCD		
10	folder						statisticians					ARPCD/ARPCD		-R---/-R---			-RPC-/-RPCD	
11	folder						programmers					ARPCD/ARPCD			-R---/-R---			-RPC-/-RPCD
12	folder					staging						ARPCD/ARPCD	-R---/	-R---/	-R---/	-R---/	-R---/	-R---/
13	folder						data_managers					ARPCD/ARPCD	-R---/-R---			-RPC-/-RPCD		
14	folder						statisticians					ARPCD/ARPCD		-R---/-R---			-RPC-/-RPCD	
15	folder						programmers					ARPCD/ARPCD			-R---/-R---			-RPC-/-RPCD

Figure 1. Example of an Excel sheet containing the folder set up with permissions and privileges assigned per user group.

Within LSAF, it is up to the client to decide who can access which data and what they are allowed to do with it. In order to come to such a model, user groups have been set up based on function and studies one should have access to. Each of these user groups have different permissions and privileges. In LSAF the permissions control access to data, where the privileges control access to features. Examples of permissions are; read-only, write and delete permissions. Examples of privileges are create SAS session, manage standards, manage user accounts. During the set-up of the permissions and privileges model, a consultant from OCS Consulting was involved who would become the admin of the system for this client. Eventually, the Excel files are used in jobs to assign or revoke permission and privileges or set up folders according to the designed structure. Whenever these Excel files are updated by the client, the OCS admin will be alerted to run the job that was created to perform this action.

USER ACCEPTANCE TESTING AND SOFTWARE DEMONSTRATIONS

After the design of the folder hierarchy and permission and privileges model in Excel, the next step in the process was to test if LSAF met the client's requirements. The user acceptance testing (UAT) was based on the test script and user guide of SAS and complemented with tests based on the client's own user requirements. This new phase also encompassed the involvement of other employees. From the client's side, this was a subject matter expert on validation and IT manager. Consultants from OCS Life Sciences were involved in executing the test script. Before starting UAT, they were being trained to properly execute the test script according to the client's standards. All findings were then communicated to the client's subject matter expert on validation who would decide on the appropriate actions for the finding. After the test script had passed validation, all end users had to be trained on how to use LSAF. This training was provided in the form of different demonstrations of LSAF for different types of users. The demonstrations were given by one of the OCS Life Sciences consultants who was also involved in the UAT. Examples of the different topics were data exploring, programming, and the creation of jobs.

LSAF IN DAY-TO-DAY ACTIVITIES

After the folder structure was designed, the permissions and privileges model were set up, the system passed the UAT, and the end users were trained, it was time to start utilising LSAF. It was October 2020 that I, the author of this paper, started working as a consultant for the client's data management department. Until then, a lot of data storage and exploring was executed on internal drives. Over the past year we gently started switching from that internal drive to LSAF. After the first data was stored, the development of jobs facilitating the data managers in their QC activities started. In the next paragraphs, the use of LSAF as both an integrated data repository and programming environment will be discussed in more detail.

PROCESS FLOW DATA TRANSFER

With several trials running, lots of data is being received by data management. As mentioned in the previous paragraph, this was all stored on an internal drive before LSAF was implemented. After the implementation of LSAF there was a new, clean environment. In order to keep this new environment as structured and clean as possible, we had to come up with a solution to find a controlled process to both upload and move data to the designated folders. The solution currently used is that of a process flow (figure 2).

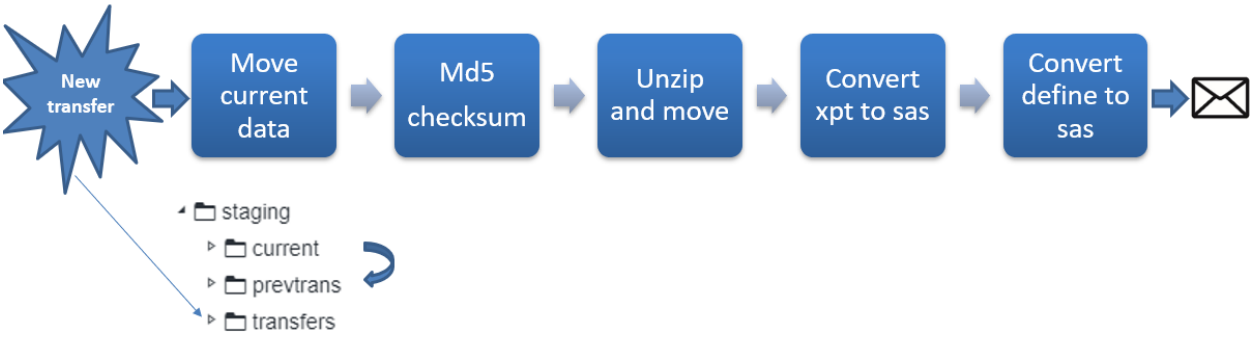

Figure 2. Process flow data transfer.

The data managers receive monthly transfers from their vendors and upload the transfers from their mailbox, SharePoint or Box accounts, directly onto LSAF. These transfers are usually zip-files that contain SDTM datasets in xpt format, the define.xml and the annotated CRF. The data transfer has a specific naming convention in order to run the process flow correctly on each data transfer.

The client has chosen to store these monthly transfers in a folder called “staging”, which is divided into three subfolders: “current”, “prevtrans” and “transfers” with each of these having their own subfolders (figure 2). The “current” folder is used to store the latest, unzipped data. The “prevtrans” folder is used to store all previous unzipped data with subfolders per previous transfer. The zip-files that are received, are uploaded by the responsible data manager directly into the “transfers” folder. Once the transfer is uploaded, the data manager sends out a request the run the process flow.

The process flow starts off with a job that moves data that is in the “current” folder to the “prevtrans” folder, to make sure nothing is overwritten and hence all data is being retained. Then, the md5 checksum job generates an md5 code for the zip-file as stored in the “transfers” folder. This md5 code can be used to check if the content of the zip-file on LSAF matches the content of the zip-file that was sent by the vendor. Once the md5 checksum job has run, the process flow moves to the next job. This third job unzips the data in the “transfers” folder and moves the unzipped data package to the “current” folder. In case the zip-file is password protected, a .txt file containing the password should be uploaded with the same name as the corresponding zip-file and then the process flow will use the password to unzip the data transfer.

In the fourth step of the process flow the datasets in the transfer, that are usually in xpt format, are being converted to SAS datasets. The final step of the process flow reads the define.xml file and converts it to both Excel format and SAS datasets. The Excel format define.xml mimics the look and feel of a Pinnacle21 converted define.xml. For each sheet in the Excel format, a separate SAS dataset in being created, being: Study, Datasets, Variables, ValueLevel, WhereClauses, Codelists, Dictionaries, Methods, Comments and Documents.

A user has the option to subscribe to the process flow, indicating if the process flow was completed and if there were any warnings or errors. This notification can either be an internal message in LSAF or an e-mail based on the users’ preferences.

CREATION OF JOBS FOR DATA MANAGEMENT

During the past year, several jobs have been created for data management. These jobs can be divided into three subcategories:

1. Jobs for administrative tasks
2. Jobs for data format conversion
3. Jobs for data review

1. Jobs for administrative tasks

The jobs created for administrative tasks can perform the following actions: transfer data from one folder to the other, set-up a new folder structure or assign and/or revoke permissions and privileges from certain user groups or users.

There are different jobs for transferring data from one folder to another. Although these jobs move data to different target folders, the set-up is the same. The process consists of two jobs. The first job requires the user to indicate what the location of the items to be copied currently is. If the first job runs, it generates a metadata file in Excel format that lists content of the original folder. In this file, the user needs to indicate with crosses in the copy column which files should be copied (figure 3). Then the second job can be run. This job reads in the metadata file and subsequently copies the items indicated with a cross to the designated folder.

	A	B	C	D	E	F	G	H
1	copy	path_name	name	size	last ModifiedBy	lastModified	date LastModified	path
2		/ae.xpt	ae.xpt	7040	User	Tue Nov 03 09:44:51 GMT 2020	11/3/2020 9:44	/clinical/compound/indication/study/dm/current/ae.xpt
3	X	/cm.xpt	cm.xpt	13920	User	Tue Nov 03 09:44:52 GMT 2020	11/3/2020 9:44	/clinical/compound/indication/study/dm/current/cm.xpt
4		/da.xpt	da.xpt	5120	User	Tue Nov 03 09:44:52 GMT 2020	11/3/2020 9:44	/clinical/compound/indication/study/dm/current/da.xpt
5	X	/dm.xpt	dm.xpt	3840	User	Tue Nov 03 09:44:52 GMT 2020	11/3/2020 9:44	/clinical/compound/indication/study/dm/current/dm.xpt
6		/ds.xpt	ds.xpt	4000	User	Tue Nov 03 09:44:52 GMT 2020	11/3/2020 9:44	/clinical/compound/indication/study/dm/current/ds.xpt
7		/ex.xpt	ex.xpt	4240	User	Tue Nov 03 09:44:52 GMT 2020	11/3/2020 9:44	/clinical/compound/indication/study/dm/current/ex.xpt
8		/ie.xpt	ie.xpt	2560	User	Tue Nov 03 09:44:52 GMT 2020	11/3/2020 9:44	/clinical/compound/indication/study/dm/current/ie.xpt
9		/mh.xpt	mh.xpt	7920	User	Tue Nov 03 09:44:52 GMT 2020	11/3/2020 9:44	/clinical/compound/indication/study/dm/current/mh.xpt
10		/relrec.xpt	relrec.xpt	1920	User	Tue Nov 03 09:44:53 GMT 2020	11/3/2020 9:44	/clinical/compound/indication/study/dm/current/relrec.xpt
11	X	/sc.xpt	sc.xpt	4000	User	Tue Nov 03 09:44:53 GMT 2020	11/3/2020 9:44	/clinical/compound/indication/study/dm/current/sc.xpt
12	X	/se.xpt	se.xpt	3840	User	Tue Nov 03 09:44:54 GMT 2020	11/3/2020 9:44	/clinical/compound/indication/study/dm/current/se.xpt
13	X	/sv.xpt	sv.xpt	3280	User	Tue Nov 03 09:44:54 GMT 2020	11/3/2020 9:44	/clinical/compound/indication/study/dm/current/sv.xpt

Figure 3. An example of a metadata file listing the content of a folder, containing crosses in the copy column to indicate which items should be copied.

For each of the default folder structures in LSAF (e.g. for a study or for the storage of a general job) jobs have been developed that create a default folder structure and assign the correct privileges and permissions. These jobs make use of the Excel file discussed in section “Design”.

2. Jobs for conversion

In order to use all the data uploaded to LSAF, three jobs were created to convert different data types to SAS datasets.

These jobs are:

- Convert Excel to SAS
- Convert xpt to SAS
- Convert define.xml to SAS

Both the “Convert xpt to SAS” job and the “Convert define.xml to SAS” job are used in the process flow, but can also run separately. Finally, also a job was created to convert SAS to Excel in order to easily share data with others that do not have access to SAS in any form.

3. Jobs for data review

Next to moving or converting data, there were also jobs created that focus on checking the content of the transfers and that facilitate the data managers in their data review.

The first data review job created was the “compare datasets” job. This job compares two datasets which can be either SDTM datasets or the converted define.xml datasets. This job can produce three outputs: one output lists the variables in both the old and the new dataset. Another output shows all duplicate records, if any. The third output shows all the differences between the old and new version of the dataset. The datasets are compared based on the key variables that are defined in the define.xml. If there is no define.xml available (yet), the example key variables listed in the CDISC SDTM Implementation Guide are being used. If differences between the old and new datasets are found, records of the old version and new version are displayed and a new record is added to indicate the changes that are indicated by crosses and highlighted red (figure 4).

Record Information	Dataset	Unique Subject Identifier	Dictionary-Derived Term	Start Date Time of Adverse Event	Domain	Serious Event	Action Taken with Study Treatment	Causality	Outcome of Adverse Event	Trtmnt Given	Standard Toxicity Grade	End Date Time of Adverse Event
Record appears both in NEW and OLD dataset, difference found in value	OLD	TEST-01	Headache	2019-12-11	AE	N	DOSE NOT CHANGED	UNLIKELY RELATED		Y	2	
Record appears both in NEW and OLD dataset, difference found in value	NEW	TEST-01	Headache	2019-12-11	AE	N	DOSE NOT CHANGED	UNLIKELY RELATED	RECOVERED/ RESOLVED	Y	2	2019-12-12
Record appears both in NEW and OLD dataset, difference found in value	DIFF	TEST-01	Headache	2019-12-11					X			X

Figure 4. Differences found in “Outcome of Adverse Event” and “End Date Time of Adverse Event”, when comparing an old version and new version of the AE dataset, highlighted with crosses and a red background.

As described earlier in this paper, the last step of the process flow encompasses the conversion of define.xml to both an Excel file and SAS datasets. The latter can be used in this compare datasets job as well, making it possible to not only compare old and new versions of SDTM datasets, but also old and new versions of define.xml.

The second job for data review is the “create data profiles” job. Not to be mixed up with patient profiles, this job shows if data is available per domain (figure 5). The job creates one data profile per subject. On the y-axis the domains are plotted, if possible subcategorised into different vendors. On the x-axis the visits are plotted. The Disposition (DS) and Exposure as Collected (EC) domain are plotted automatically, the other domains allowed are findings domains that can be specified based on the data managers’ needs.

Colour coding is applied in order to show whether data was available for a specific visit/domain/vendor combination. A green circle indicates that for all records in that visit/domain/vendor combination a result was found (i.e. --STRESC variable was filled). A blue triangle indicates that records with a result (i.e. --STRESC variable filled) and status “not done” (i.e. --STAT = “NOT DONE”) were found. The red square indicates that for that visit/domain/vendor combination only records were found where the status was “not done” (i.e. --STAT = “NOT DONE”). If no record exists for that visit/domain/vendor combination nothing is shown. To illustrate; in the example below no records exist for the Vital Signs (VS) domain on visit 2, whereas the Immunogenicity Specimen (IS) domain at visit 3 contains records where --STRESC is filled and records where the status equals “NOT DONE”.

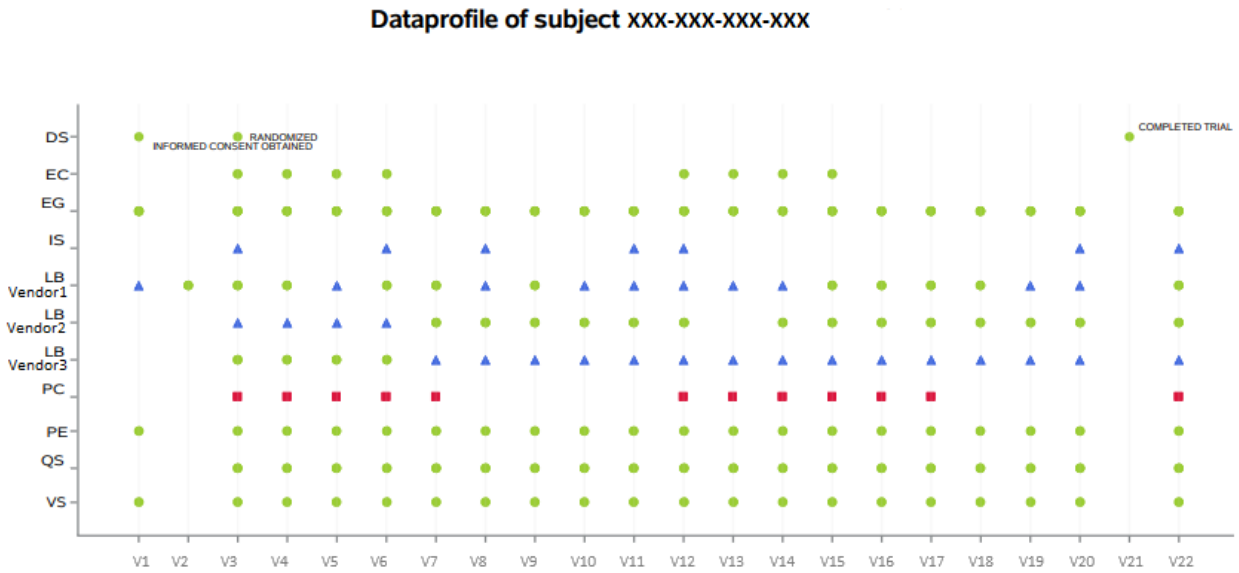

Figure 5. An example of a data profile showing data availability per domain, per visit and if applicable per vendor.

The third job created for data management data review is the adverse event and concomitant medication timeline. In the client's study portfolio, there are a couple roll-over studies. In the parent study a subject can suffer from an adverse event or take concomitant medication which can be ongoing at the end of the study. If this is the case, and this subject participates in the roll-over study, it is expected that these ongoing adverse events and concomitant medication are also recorded in the roll-over study. In order not to search for these adverse events and concomitant medication in multiple datasets, an adverse event concomitant medication (AECM) timeline was created. This timeline (figure 6) visualises all adverse events and concomitant medication that were ongoing at the end of the parent study and shows if they were also recorded in the roll-over study. Concomitant medication is indicated by a purple colour. The adverse events are colour coded according to their toxicity grade, where green represents toxicity grade 1, yellow represents toxicity grade 2, orange represents toxicity grade 3 and red represents toxicity grade 4.

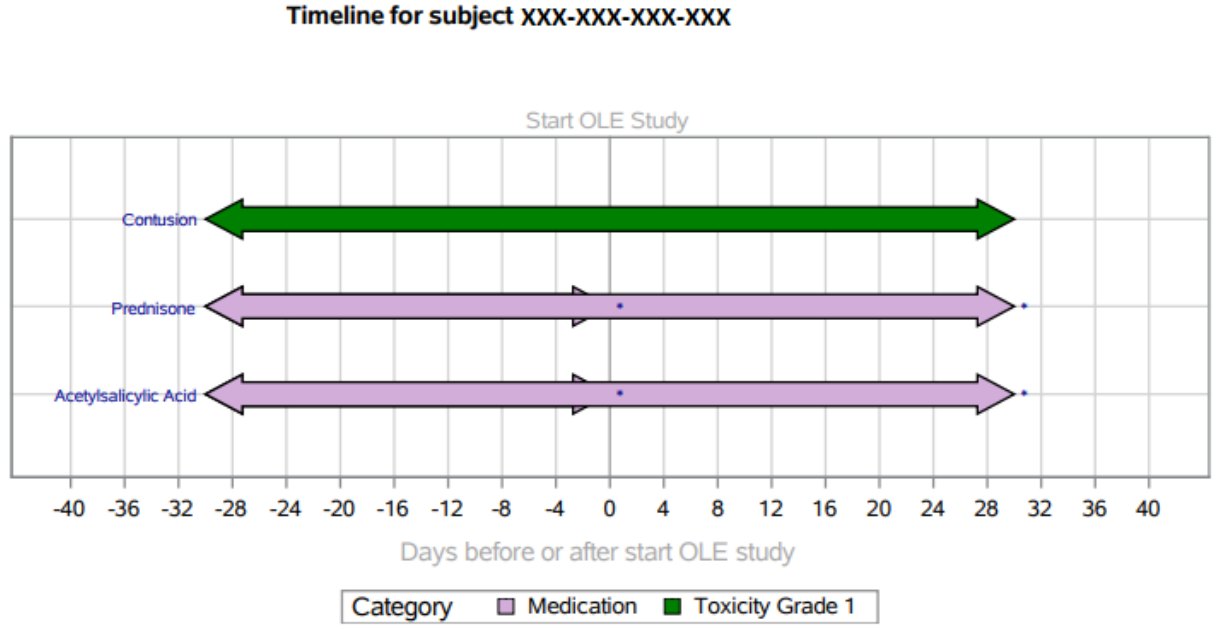

Figure 6. An example of an AECM timeline, showing adverse events and concomitant medication that were ongoing at the end of the parent trial and are also documented in the Open Label Extension (OLE) study.

The final job described in this paper is the job that creates weekly listings for the clinical trial team. For some particular studies, questionnaires are collected at each visit. These questionnaires are important for monitoring the disease course. For that reason, these questionnaires need to be evaluated more frequently. The job created lists the questionnaire outcomes and highlights an increase in score red and a decrease in score green (figure 7).

Test	Subject	Location	Screening	Visit1	Visit2	Visit3	Visit4	Visit5	Visit6
Assessment1	Test001	Head	1	1	1	0	1	0	0
	Test001	Neck	2	2.3	2.3	2	2	1	1
	Test001	Eyes	0	0	0	0	0	0	0
	Test001	Mouth	2	2	2	1	2	1	1
	Test001	Ears	2	2	2	2	0	0	0
	Test001	Nose	0	0	0	2.3	0	0	0
	Test001	Teeth	1	1	1	2.3	0	0	0
	Test001	Tongue	5	5	5	5	2.3	2	2
	Test001	Throat	2.3	2.3	2.3	2.3	2	0	0
Assessment1 Subscore	Test001		15.3	15.6	15.6	16.9	9.3	4	4

Figure 7. An example of a weekly report for the clinical trial team with all score increases highlighted in red, and decreases in green.

After development, these jobs were validated as well. Within the clients' LSAF environment, there are separate folders for development, validation and production. In order to copy the files needed for validation in a controlled and traceable way, a copy job was used. As described earlier in this paper, this job allows for copying all items needed to the validation folder. Within the LSAF environment one is able to version files, indicating if it concerns a major or minor update. If different versions of an item exist, for example due to rework after validation, there is the option to compare different versions and have an overview of the differences. When validation is finished, within LSAF there is also the ability to sign-off the files validated. The signatures in LSAF are 21 CFR part 11 compliant making them valid signatures. All these steps performed during the validation process can be found back in an audit trail making all steps in the development and validation process traceable. Once validated, the jobs were moved to the production folder, again making use of a copy job.

DATA STANDARDS

LSAF provides the option to upload client's data standards and NCI or client's controlled terminology to the system. In this case one version of the client's library and latest NCI controlled terminology was uploaded for all studies, but one could even upload different standards and versions of controlled terminology. When uploading the data standards to LSAF they become available in SAS datasets. Since the process flow converts the define.xml to SAS datasets as well, we were able to also build a job to compare the study metadata to the library and vice versa. This job produces a report to indicate which metadata matches the study library and which metadata is specific for the study.

DEMO AND DOCUMENTATION

After the UAT, demonstrations on how to use LSAF were given. However, most questions usually arise when really working with the system. The same applied to utilising LSAF. Over the past months more demos were given on data management specific items such as uploading new data transfers, running jobs (for example SAS to Excel), and data browsing in LSAF. The questions raised and topics discussed in these demos were bundled into a wiki page for the data managers to enable them to quickly look up their topic. Moreover, the wiki contains a guide on how to best use LSAF and which jobs are available for data management.

WHAT'S NEXT

At the client's side the next steps involve continuing to make the switch from a traditional network drive to LSAF. It also means being involved in the creation of new jobs to keep all processes controlled and structured, and for the data managers to be involved in the creation of jobs in order to support their data review. For SAS it means developing new features and improving current features with new versions having options for R integration, improved collaboration efforts, and improved search functions. For OCS, the next steps would be to keep helping the client in the development of new jobs and to keep performing administrative tasks such as adding users, revoking rights, monitoring performance, and running jobs.

CONCLUSION

One important learning in this implementation process is the benefit to have subject matter experts during each phase of the LSAF implementation to really optimise the process. To have these subject matter experts in one company and also have short communication lines with SAS ensured that this process could smoothly transition from one phase into the next. Another benefit of having subject matter experts available is the fact that, because they understand the process so well, they are able to simplify it for clients which helps to determine the best strategy. With regards to the use of LSAF for data management, the system has shown so far to function well as a structured data repository. With all data stored into LSAF, the jobs already created and the jobs to be created in the future can be used to manage, review, process and transfer data in a structured, traceable and easy way.

ACKNOWLEDGMENTS

First, I would like to thank the client for the opportunity to present about our journey of implementing LSAF and all learnings from the past year. Second, I would really like to acknowledge my colleagues who helped me finalise this paper by providing constructive feedback and encouragement along the way. In particular, I would like to mention Lieke Gijsbers for proof-reading and Caro Sluijter and Stephan van Driel for their help and useful feedback when preparing the presentation that goes with this paper.

RECOMMENDED READING

Caro Sluijter, “Automation of Process Flow Instances”, Oct 2021.

CONTACT INFORMATION

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

Louella Schoemacher
OCS Life Sciences
Ruwekampweg 2G
‘s-Hertogenbosch / 5222 AT
+31 (0)73 523 6000
sasquestions@ocs-consulting.com
https://ocs-consulting.nl/ocs-life-sciences/