

Data Load Automation

Jennifer H. Desai, Janssen R&D, Raritan, NJ, USA

ABSTRACT

Data transfers are essential aspects throughout the duration of a clinical trial. The current Janssen process of manual data transfers with its external vendors is very time consuming. A solution to enhance the manual data transfer process is with data load automation. This paper will describe Janssen's current manual data transfer process, and how data load automation can be implemented to create a more efficient and effective method for data transfers between Janssen and its vendors.

INTRODUCTION

Data is the backbone of medicine. All medications, from pills to vaccines, exist due to clinical research and the collected and examined data from clinical trials. Therefore, it is evidently important to produce clean data from clinical trials in order to advance in the medical world.

However, within Janssen, it can be difficult to collect and analyze data from clinical trials at a reasonable rate. From the data transfer agreements to granting the appropriate access for our vendors to the correct folders in our data repository system, the entire manual data transfer process can become a hassle. Therefore, manual data transfers remain to be an unpleasant experience for both Janssen and its vendors.

A solution to enhance the manual data transfer process is with data load automation. Data load automation is when source data is delivered through an automated process from external vendors directly into Janssen's SAS data repository platform, called Life Science Analytics Framework (LSAF).

WHAT IS A DATA TRANSFER AGREEMENT?

During the study start-up phase, the data streams and vendors that will be utilized in the study are identified. Once the database is set up and the study goes live, it begins the conduct phase. During this phase, trial specific data transfer agreements (tsDTAs) are created between Janssen, the study's contracting research organization (CRO), and the vendor.

Data transfer agreements (DTAs) are specifications that outline the expectations of the file transfers between Janssen and each of its vendors, per study. Janssen and its study CRO work together with the vendor to gather the necessary data needed to complete the tsDTA template. The tsDTAs are created per each data stream and each vendor in the study. These tsDTAs consist of trial specific information, such as visit/visitnum combinations, test and test codes, transfer file type, frequency of transfer, etc. The tsDTAs are living documents and can be updated throughout the study as needed.

There is a method of transfer section in every tsDTA. This section specifies the folder path and locations of where the vendor will need to manually drop the data file in LSAF, Janssen's data repository platform. This section also states the frequency of the data transfers, which can be daily, once a week, biweekly, once a month, etc.

WHAT IS SAS LSAF?

SAS® Life Science Analytics Framework (LSAF) is Janssen's data repository tool. This tool is used to share data between our CROs and vendors. As defined on Janssen's LSAF portal, LSAF affords a singular analytics foundation for the efficient development, execution, and management of the analysis and reporting of clinical studies through a hosted environment. LSAF provides the ability to integrate analytic applications and other SAS solutions into the framework to deliver the power of analytics to more roles within the organization.

Furthermore, SAS specifies that LSAF is worldwide accessible and can be used at any time and any location. It is a uniform platform for various departments and vendors that all contain the same configuration, same user interface, same central repository for data, SAS programs, output files, etc.

LSAF simplifies sharing & transferring of data in a secure way. Typical types of data sharing and transferring include when CROs upload data packages to Janssen Data Management (DM) via LSAF, when external vendors share their data with Janssen DM and other vendors via LSAF, and when Janssen DM shares the data with statistics and programming for analysis via LSAF.

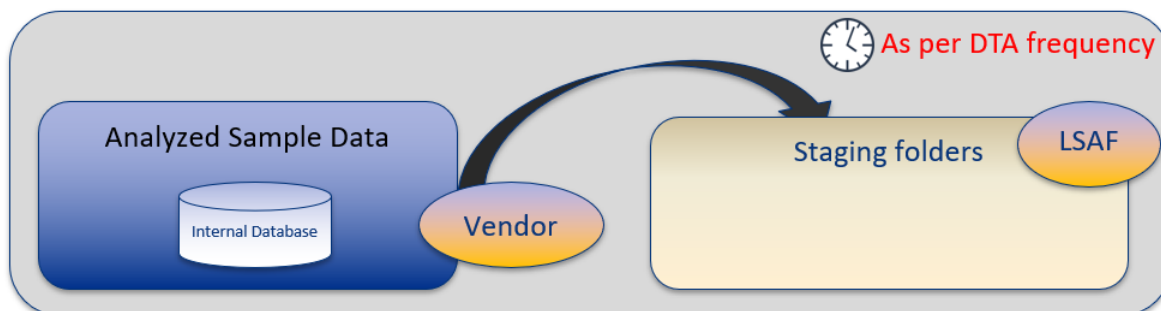
Additionally, LSAF helps streamline and automate processes. The tool contains automated process flows that consist of automated job running, it has central security management (users, groups, folders), it can process large volumes of data, and it has an integrated SAS environment (statistics analysis, SDTM QC, scheduling jobs, etc.).

CURRENT MANUAL DATA TRANSFER PROCESS AND ITS CHALLENGES

The current Janssen process for manual data transfers with its external vendors can be an unpleasant experience for both Janssen DM and for its vendors.

Once a tsDTA is signed and finalized, the data transfers can begin. However, before they can begin, the logistics must comply. For example, Janssen DM will have to individually grant access to their study in LSAF for each vendor and the vendor's individual team members. Before any vendor personnel can receive access, they need to complete the appropriate trainings. This entire process is time consuming and the coordination can become frustrating. Also, when team members at the vendor changes, Janssen DM will need to revoke/grant access as necessary. Hence, manual data transfers remain to be a bothersome experience for both Janssen and its vendors.

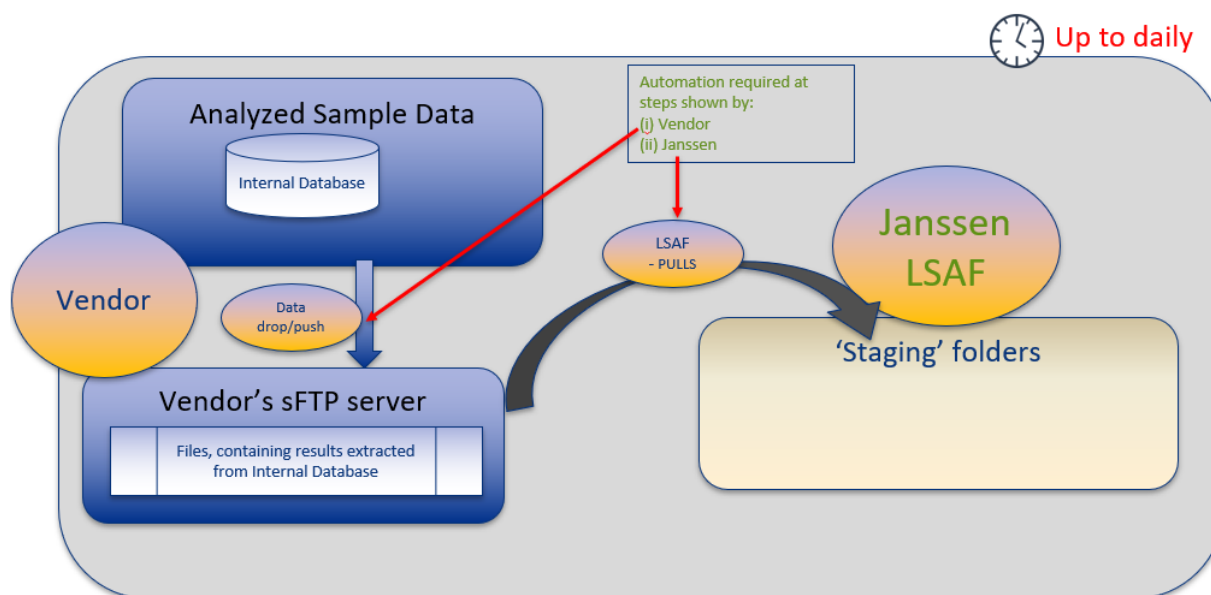
The manual process for data transfers is as per the below diagram. The vendor will take the analyzed sample data from their internal database, manually log into LSAF, and upload the data into the Janssen study folder location, as specified in the tsDTA. This manual data transfer occurs as per the agreed frequency in the tsDTA.



In contrast, with automated data transfers, there is no need for the vendors to have each team member trained nor granted access to LSAF.

NEW DATA TRANSFER PROCESS WITH DATA LOAD AUTOMATION

Data load automation is when source data is delivered through an automated process from external vendors directly into Janssen's SAS data repository platform, LSAF, via its PULLS feature. Per the diagram below, the vendors will drop their data onto their servers and allow Janssen access to their servers. Once the vendors drop their data onto their servers, Janssen can automatically go into them and pull the data into our data repository, LSAF. This process is time efficient for the vendors as they do not have to manually upload the data into our data repository, and, it is time efficient for Janssen as we can program the data pulls to have an automatic daily pull schedule. Also, this removes the extra hassle of granting the vendors individual access to the specific folders in our data repositories.



DATA LOAD AUTOMATION TECHNICAL SPECS FOR SET UP PROCESS

Setting up data load automation is a one-time implementation process per each vendor. To proceed with data load automation the vendor must meet certain requirements. These requirements include having a compatible server that can technologically communicate with LSAF, such as an sFTP server. Next, the vendor will need to create a root folder for only Janssen studies and be able to provide a username and a non-expiring password with read only access to the root folder. Within that root folder will contain all the Janssen studies the vendor provides a data transfer for. Further, the vendor will need to allow traffic via a port 22 into their server. If the vendor meets all these requirements, they will need to provide Janssen IT with their sFTP host and folder names, and their IP address and IP owner (if they are not the owner as Janssen will need to document this). When these technical specifications are provided to Janssen IT, they will begin their whitelisting process and once a connection is active between SAS and the vendor, data load automation transfers can begin via the process displayed in the above diagram.

The vendor and their individual team members will not all need access to LSAF nor will need to take the LSAF trainings. However, it is recommended that at least one or two members from each vendor still have access to LSAF in case of any automation issues, so they can resort to a manual upload if needed as a backup procedure.

The method section on the tsDTAs will say automated data transfers instead of manual data transfers and it will include the vendor's sFTP server location and folder details. The structure, format, naming conventions, etc. will remain the same, as the only change is the method of transfer, from manual to automated. The LSAF pulls are programmed to automatically run daily, even if the vendor does not push data daily. The data will only be pulled into LSAF if there is a new file in the vendor's server.

Data load automation can be set up with all data types including, but not limited to, laboratory, electrocardiography (ecg), medical imaging, etc. Because of data load automation, source data is now readily available more frequently, so it supports Real Time Learning and Decision Making (RTLDM), as well as improving Study Data Tabulation Model (SDTM) conversion timelines.

For Janssen, we have set eligibility criteria for vendors who we want to implement data load automation with. The inclusion criteria include that the vendor can automate data transfers, the vendor has an appropriate server (sFTP, MBOX, etc.), the vendor meets all technical requirements (allowing traffic via a port 22, landing at root folders, etc.), the vendor allows Janssen read access to their servers, and that the vendor has multiple studies with Janssen and planned future studies. We will not implement data load automation with vendors who cannot automate data transfer, vendors who do not generate their own data (CROs), vendors who do not have the appropriate servers, vendors who do not allow sponsors access to

their servers, and if Janssen only utilizes the vendor's services for 1 or 2 studies and/or has no future plans to work with the vendor.

LIMITATIONS OF DATA LOAD AUTOMATION

A limitation to this data load automation process is that not all vendors can implement automated data transfers, especially smaller vendors, such as wet labs or universities. With the vendors who cannot automate, our counter solution is to increase their manual data transfer push frequency. For example, for the smaller vendors who manually transfer their data into LSAF once a month, we will like to increase it to twice a month, weekly, etc.

Another limitation is the potential contractual or financial impact in our relationship with the vendors. However, because data load automation benefits the vendors themselves as well, they have been supportive in its implementation.

CONCLUSION

Overall, data load automation is a solution to enhance vendor data transfers. It helps deliver data on a more frequent schedule. Since the data is available more frequently, it is continually looked at and reviewed by all parties involved in the clinical trial, such as data managers, statisticians, programmers, study responsible physicians (SRP), study responsible scientists (SRS), etc. As a result, data load automation ultimately aids in delivering cleaner data for submissions. Hence, implementing data load automation between Janssen and its vendors would create a more efficient and effective data transfer process throughout the duration of a clinical trial.

REFERENCES

SAS Life Science Analytics Framework. A singular analytics foundation for clinical research to increase efficiency. Fact Sheet. (2019). Retrieved from <https://www.sas.com/content/dam/SAS/documents/product-collateral/fact-sheets/en/life-science-analytics-framework-107108.pdf>

CONTACT INFORMATION

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

Jennifer H. Desai, MPH
Janssen Research and Development
920 US 202 South
Raritan, NJ, 08869
908 927 4994
jdesai12@its.jnj.com
<http://linkedin.com/in/jenniferdesai>

Brand and product names are trademarks of their respective companies.