21 CFR Part 11 Compliant Data Replication Strategies for LC-MS/MS & HR-GC/MS Data Acquisition Systems

Ike D. Tabani, Scott R. Serl, Mikhail Amchislovsky and Vittalie Pappof *Innovative Automation*, Sacramento, CA 95823

Robert A. Bethem, Robert S. Mitzel, Stanley Murakami, Jim Lehman, Paul B. Woolley, Cindy Gilman and Dale Schoener *Alta Analytical Laboratory, El Dorado Hills, CA 95762*

Introduction

21 CFR Part 11 has enacted several mandates to ensure the security, integrity and authenticity of electronic records and signatures. The law outlines controls for ensuring that electronic records and e-signatures are trustworthy, reliable, and compatible with FDA procedures. The ultimate goal has been to make electronic records and e-signatures as verifiable and traceable as their paper counterparts. File based instrument data acquisition systems (IDS) pose a particular compliance challenge in that data are often stored to local workstations within individual files that may be subject to deletion, modification or a loss of traceability once the files are moved. One solution to this problem is to mirror IDS data files in real-time to a central data repository or instrument server via TCP/IP. The instrument server drive array may be organized into instrument specific partitions with WORM attributes. Access to the instrument server may be controlled using a LIMS system or off-line analytical processing programs. Following this approach, we describe the advantages of this data management strategy within the CRO environment including enhanced file traceability, detailed logging of file uploads, improved data protection and greater facilitation of reporting operational metrics.

Real Time Data Replication Strategies

Replication is the process of creating a second copy of data on a remote disk-array. There are hardware-based solutions where the disk-array will transmit data from one disk-array to another, and there are software-based solutions where software that resides on a host will transfer data from one disk-array to another. There are many types of replication as well, but two are most predominant, and of these, one is much more widely used. The two types are synchronous and asynchronous replication, and synchronous replication is used only in special circumstances when a second exact copy of the data is a must.

Synchronous Replication
This form of replication adds a degree of latency to the overall data protection process. However, it does ensure that each disk-array has the exact same copy of the data. How does it work? When the application writes a piece of data, the replication engine intercepts the write and sends it to the primary disk-array as well as the secondary disk-array. Only when both systems have confirmed that the write has made it to disk does the replication engine inform the application that it is okay to accept another write and process it. For a high transactional and real time acquisition applications, this process can slow things down and in certain cases may interfere with acquisition itself. Additionally, if this type of replication is required, bandwidth for transferring data becomes an issue, and increasing the size of the pipe between the two disk-arrays is often necessary, and quite expensive.

Asynchronous Replication
Asynchronous replication is the most common type of replication because of the nature in which it operates. IIDMS/FRMS™ is an asynchronous replication product that runs under Windows 2000/3 Server™ and Windows 2000/XP™ operating systems. Figure 1 shows the overall process of asynchronous data replication using IIDMS/FRMS. In the first step a user changes or creates a file on a local workstation or file server. As application data is written to the primary data store, the replication engine grabs that write and sends it to the secondary disk-array. File changes are captured at the operating system level tagged with a GUID and then placed within MSMQ™ based messaging queue for processing. Replication parameters may be set such that the data writes are buffered locally in data queues and sent at certain times or when sufficient bandwidth is available. As this happens, and unlike synchronous replication, the application can continue to operate and acquire data, without waiting for a response from the secondary disk-array to acknowledge that the write was successful. The ability to move a preset amount of data at a given point in time during a day coupled with the ability to utilize only a portion of the bandwidth pipe between replication sites are all cost-effective design features to consider when implementing data replication.

Snapshots
Alternatively, a time based snap shot replication can also be used to construct the chronological history and life of a data file. One technology getting a good deal of attention today is “Snapshots”. A snapshot is a picture of what a file system looked like at a given point in time. A snapshot contains a list of pointers to the files as they exist on a given file system. When a file in the file system changes, a copy of the original file is moved to the snapshot, thereby maintaining the same file subset that was created at the time of the snapshot. This technique is called “copy on write”.

Figure 2 shows a graphical representation of these replication strategies. In this poster we describe an asynchronous replication solution using IIDMS/FRMS™ file replication and mirroring software (Innovative Automation, Sacramento, CA).

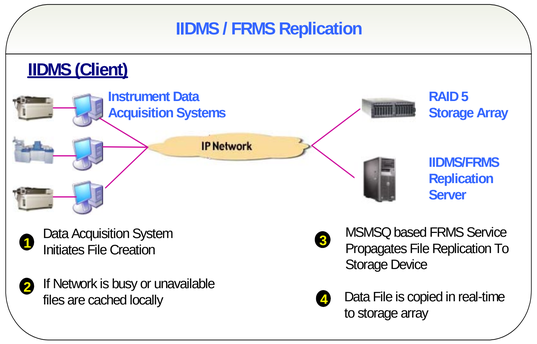

Figure 1

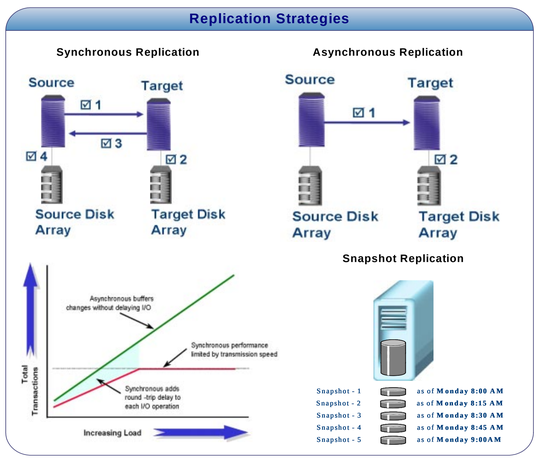

Figure 2

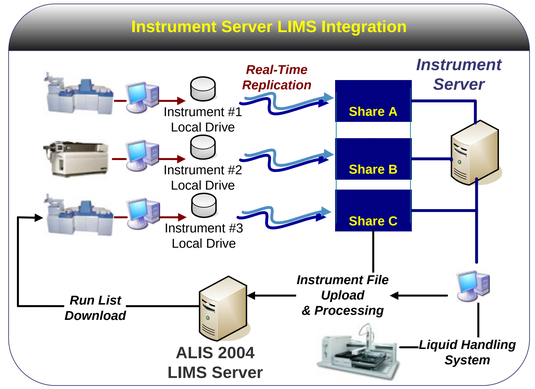

Figure 3

LCMS Instrument Network

Figure 4 shows a schematic of the LCMS instrument network similar to one deployed at Alta Analytical Laboratory (El Dorado Hills, CA). The availability of a central repository of instrument data enables users throughout the enterprise the ability to perform various laboratory functions including LIMS data processing, off-line IDS data review and analysis as well as remote data access. Access to instrument data is granted through a controlled secure environment where users may not delete nor modify files stored on the instrument server. In the future, availability of a data replica may require that users reexamine the definition of raw data (i.e. IDS local files versus replicated files) as well as reconsider appropriate procedures for reprocessing analytical data.

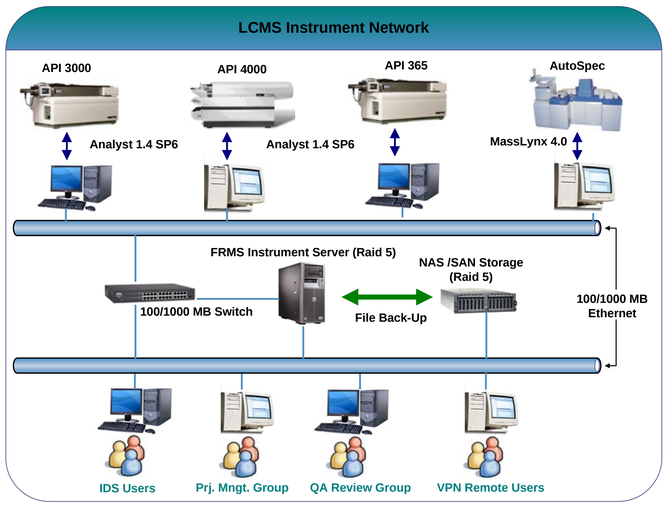

Figure 4

GLP / Part 11 Requirements and Replication Instrument Server

A Windows 2000/2003 server with a drive array running RAID 5 can serve as the secondary disk-array shown in Figure 1. The instrument server drive array is divided into multiple partitions in order to replicate data from one instrument to a single storage area within the array (Figure 3). This greatly facilitates integration with other programs used to process raw acquisition files such as a laboratory information management system (LIMS) or an off-line instrument data system (IDS). For example, in the LCMS instrument network shown in Figures 3 and 4, the ALIS2004™ instrument file upload module enables the user to select files for upload from the instrument server based on the instrument specific assigned drive partition. This approach provides significant control over post acquisition raw data file traceability, a critical 21 CFR Part 11 compliance requirement. A summary of 21 CFR Part 11 and GLP requirements is provided in Table 1. Figure 5 shows 21 CFR Part 11 required metadata for each replicated file for each instrument data system. In addition, this arrangement also allows the calculation of drive space consumption rates by instrument as well as operational metrics such as instrument usage based on user definable parameters.

Table 1. 21 CFR Part 11 Compliance Strategy

GLP/Part 11 Requirement	Instrument Server w/ Replication
Preservation of Raw Data	Delete is blocked at Instrument Server
Data Traceability & Reconstruction	Instrument file point of origin is recorded
Disaster Recovery	Data replica maintained in real-time
Instrument Performance Qualification	Instrument files scanned at upload for anomalies *
Change Control	File modifications detected by analysis of file attributes *
Adequate Resources are Provided	Disk space consumption & instrument utilization metrics tracked *

* *Feature requires integration with LIMS*

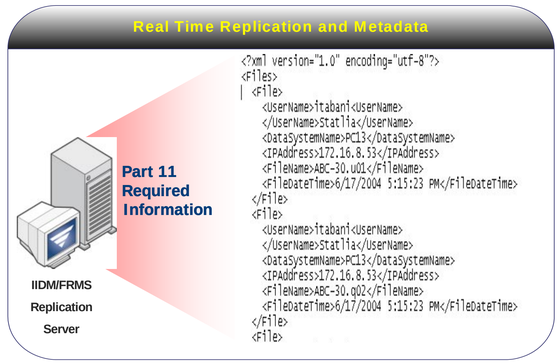

Figure 5

Conclusions

The use of data replication tools such as IIDMS/FRMS™ has proven to be a successful strategy for managing instrument data files generated from a laboratory instrument network (Figure 4). This approach has greatly facilitated 21 CFR Part 11 good electronic record keeping procedures particularly the management of instrument raw data files. In order to take full advantage of the data replication features it is essential that laboratory’s LIMS be fully integrated with the instrument server. In addition to extending this approach to other instrument platforms, future efforts will involve the integration of instrument server with an electronic laboratory notebook software and optical drive arrays to efficiently off-load data to near-line optical storage.

An electronic version of this poster is available at: http://www.21CFRPART11Compliance.com/postergallery.htm

ASMS 2005