

Backup vs. Archive: What's the Difference and Why You Need Both

INTRODUCTION

In **June of 2020**, the U.S. FDA issued a 483 warning letter to a company after an inspection for a failure to exercise sufficient control over computerized systems noting: "Missing raw data files associated with recovered solvent testing were observed in folders on the local hard drive of the operating system connected to the GC instrument. Your firm indicated that the files appear to have been deleted."¹

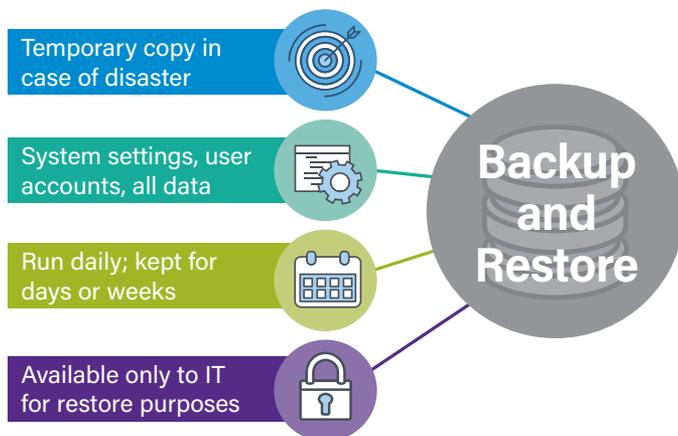


Figure 1. It is critical to run frequent backups to ensure assets can be restored in case of data loss from the original system.

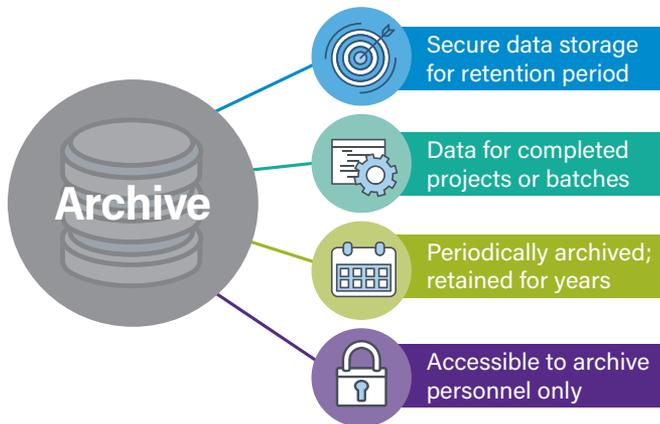


Figure 2. Archiving provides secure storage of regulated data through the mandated retention period.

FDA REPORTS ON BACKUP AND ARCHIVING

2018 FDA Inspection

Report: "Backup data shall be assessed to verify successful completion of the backup process"... "This practice is not being performed."²

In **May 2015**, the FDA issued a 483 warning letter to a company after an inspection where its back up strategy was called in to question stating that "without complete, accurate, reliable, or retrievable raw data about the HPLC system's qualification, you lacked complete assurance that the system was operating as intended"...³

Today, laboratory-based organizations face a wide variety of unaddressed data management challenges, and yet, ultimately, the scientific data is the currency with which they trade. Proper data management may not pay shareholders but it fundamentally defines the integrity of the organization and it's purpose for existing. Being the cheapest, the fastest, or the most definitive is desirable, but it is all meaningless if the data is untrustworthy.

Undeniably, along with the continual advancements in analytical technologies comes the ability to generate vast amounts of data. In order to extract the most value from this information, organizations must evolve their data management practices. This change in approach has a direct impact on backup and archiving methodologies.

Of course, data volume is only one part of the story. There are a number of contributory factors that make the plot far more complex, including:

- The need to manage raw lab data under such regulations as 21 CFR part 11, Annex11, ISO17025, and the Food Safety Modernization Act (FSMA), among others
- Ensuring potential audits can be readily addressed by optimizing data integrity, searchability, and accessibility
- Accomplishing all of this with IT budgets that are flat or declining

Given the requirements described above, organizations desire storage products that provide reliability, long-term retention, searchability of data, and low total cost of ownership, without losing the ability to respond quickly in an audit. In this situation, traditional back up is not sufficient to meet these needs and this white paper will explain how an archive strategy can:

- Reduce backup and recovery times
- Remove manual intervention and variability
- Minimize exposure during an audit
- Establish a process around the full data lifecycle

BACKUP vs. ARCHIVE

A classic backup application takes regular snapshots of data in order to provide a means of recovering records that have been deleted or destroyed. Most backups are retained only for a few days or weeks as later backup images supersede previous versions. The best way to think of backup is as a short-term insurance policy against an unforeseen disaster; backups help recover information and processes in current use in case they are interrupted, corrupted, or lost.

Archives serve a very different purpose to backups. They preserve inactive information as required by regulations or company policies. An archive is designed to provide fast search and access to years of information and as a result can aid in the discovery of information not currently in use, in case they become useful again to prevent duplicate work or meet an unanticipated regulatory need. In science-driven industries, results and data integrity can be challenged at any time and even inactive documents may need to be retained for many years.

Archived records can exist outside the traditional backup cycle for a long period of time because by comparison the data is quite static. Meanwhile, the regular backup is protecting live data that is changing on an everyday basis. That does not mean you hold records forever – the best archive solutions also allow you to manage data and documents that are no longer required. This is a critical point – information that should have been deleted could represent a risk to the business given that all data contained in the backup is subject to inspection. But if an effective archive solution is in place, data can be automatically flagged and destroyed according to regulations or company policies.



42% identified archiving and extracting data as an obstacle in their labs.⁴

Table 1. Backup and archiving at a glance.

	Backup	Archiving
What is it?	Protection for mission critical systems and live data	Searchable records of inactive data in a "steady state"
Why use it?	Recovery – backup restores systems after data loss, interruption, or disaster	Searching – allows interrogation of data for regulatory inspections and data investigations
What does it contain?	Several snapshots of the live system(s) captured on a time basis	One single repository of historical data indexed and quickly searchable

BACKUPS ARE FOR DISASTER RECOVERY - ARCHIVES ARE FOR DATA SEARCHING

IS AN ARCHIVE NECESSARY?

A backup is not an archive. If you try to use backups as an archive to support an audit, you will soon see a few reasons why this is not recommended:

There are too many backup copies

Backups help to recover systems so you take multiple snapshots of the same data. Explaining which one is valid over another and which is actually the raw data to an auditor can be challenging. In contrast, an archive provides a single "official" indexed record.

You cannot search a backup

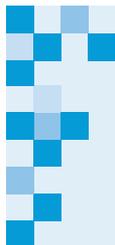
In order to search a backup, you have to restore the whole thing. In contrast, an archive gives you the ability to search surgically – quickly getting to the data you need and restoring that and that only. Backups treat all the data in the same way and there is no way to decide what needs to be kept and what does not.

Backups can increase the risk to your organization

If you use your backup as an 'archive', all data will be categorized based on the date you backed it up. Separating data for legal hold and/or managing the lifecycle of individual data sets is impossible and opens the business up to unnecessary risks. With an archive you can make this distinction.

"Legal hold" increases your storage

Continuing from the previous point, if you have one file under legal hold in your backup then it means you have to keep the whole backup and that wastes a lot of storage. An archive allows you to flag only the documents needed and delete the rest when the time comes (e.g., data end of life, mergers, and acquisitions).



Professionals spend over 500 hours annually reviewing and routing files and another 150 hours looking for incorrectly filed documents. It costs \$120 to search for a misfiled document, and, if you can't find it... it costs approximately \$250 to recreate a lost document.²

THE IMPACT OF ARCHIVING SPANS SCIENCE, OPERATIONS, AND COMPLIANCE

A good archiving solution will automatically determine if data is in use or idle and then move that data from expensive high performance storage to more economical archive storage. Furthermore, indexing that archive and its metadata allows for swift search and retrieval when it is really needed without IT assistance, and legal hold will protect that data from accidental deletion or loss.

For example, NuGenesis™ Scientific Data Management System (SDMS) allows scientific data generated in your laboratory to be accurately and automatically captured, indexed, and securely stored in a compliance-ready environment immediately after its creation or change. Often this is combined with the Empower™ Chromatography Data System (CDS) to manage inactive LC data.

Managing the mad panic urgency of regulatory inspections can disrupt IT groups and scientists in the execution of their daily project duties. Archives prove their worth during the first regulatory inspection, but even without any such requirement, they pay back quickly by simplifying and reducing the IT burden on backup processes.

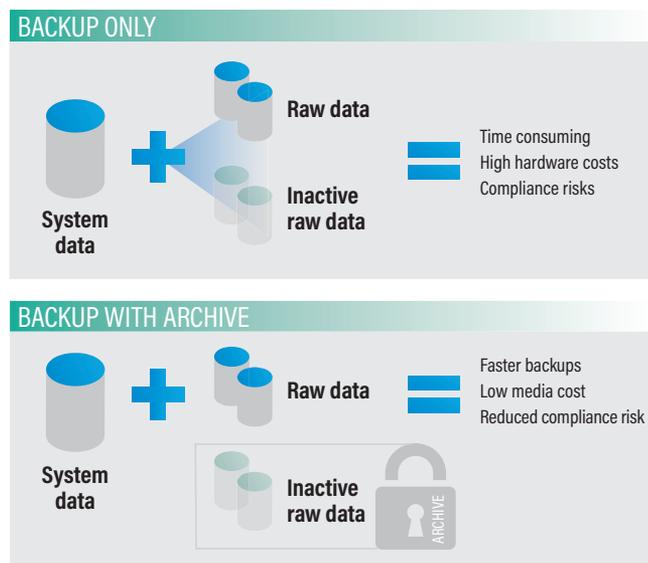


Figure 3. Backup efficiency gains by using an archive solution. Without archiving, much of the time, bandwidth, and storage spent on backup is simply wasted.

CONCLUSION

Backups and archives perform separate functions but the capabilities of each one help the other work better and more efficiently.

Implementing an archive is an efficient, comprehensive approach to managing and protecting laboratory data. Science-driven industries can use an archive in addition to backup solutions to address the growing data volume, regulatory requirements, and technological complexity found in the contemporary laboratory environment.

When an archive solution is in place backups run faster, consume less time, energy, and system resources, which means better protection for mission critical systems such as Empower.

References

1. U.S. Food and Drug Administration. Warning Letter. [Online] June 17, 2020, <https://www.fda.gov/inspections-compliance-enforcement-and-criminal-investigations/warning-letters/vega-life-sciences-private-limited-604469-06172020>.
2. U.S. Food and Drug Administration. [Online] Accessed September 2020, <https://www.fda.gov/files/drugs/published/Compliance-Record-483--Dr.-Reddy's-Laboratories-Ltd.--March-9--2018.pdf>.
3. U.S. Food and Drug Administration. Warning Letter. [Online] May 27, 2015, <http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/2015/ucm448433.htm>.
4. IQPC Laboratory Informatics: Current Trends & Predictions for 2015 [Online], May 10, 2012, <https://www.pharma-iq.com/informatics/whitepapers/laboratory-informatics-current-trends-predictions>.

Waters

THE SCIENCE OF WHAT'S POSSIBLE.™

Waters, The Science of What's Possible, Empower, and NuGenesis are trademarks of Waters Corporation. All other trademarks are the property of their respective owners.

©2020 Waters Corporation. Produced in the U.S.A. October 2020 720005528EN LM-PDF

Waters Corporation
 34 Maple Street
 Milford, MA 01757 U.S.A.
 T: 1 508 478 2000
 F: 1 508 872 1990
www.waters.com