Digitally Secure Sequencing: Implementing Part 11-Ready NGS Data Security in a CRO Environment

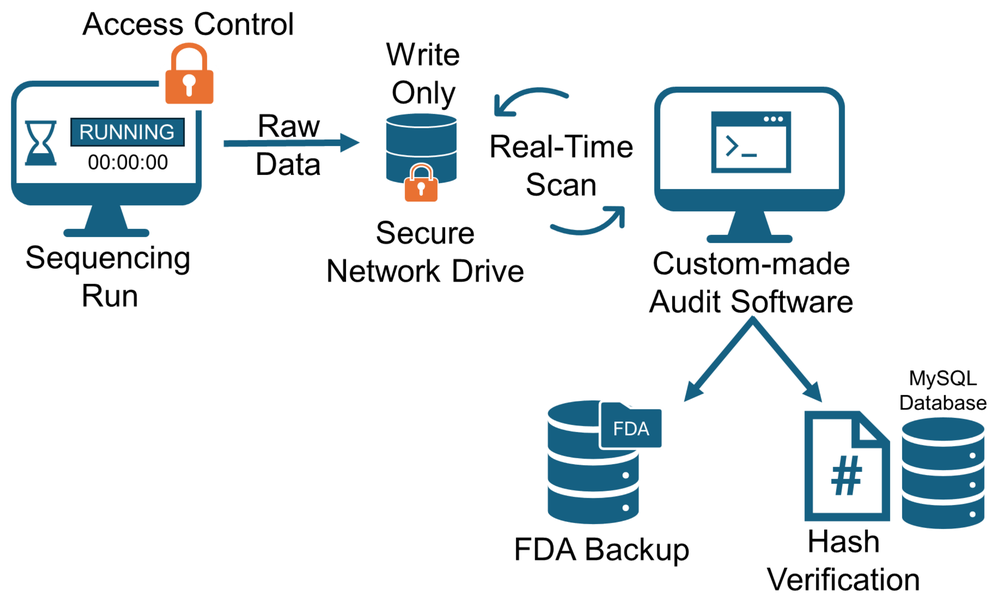
Apurva Vansadia, Joshua Holloway, David Zhou, Joshua Tran, Juan Hoyos, Ross Fu, Guanjun Liu, James Frost, Xuening Huang

INTRODUCTION

Data Integrity in Regulated Workflows

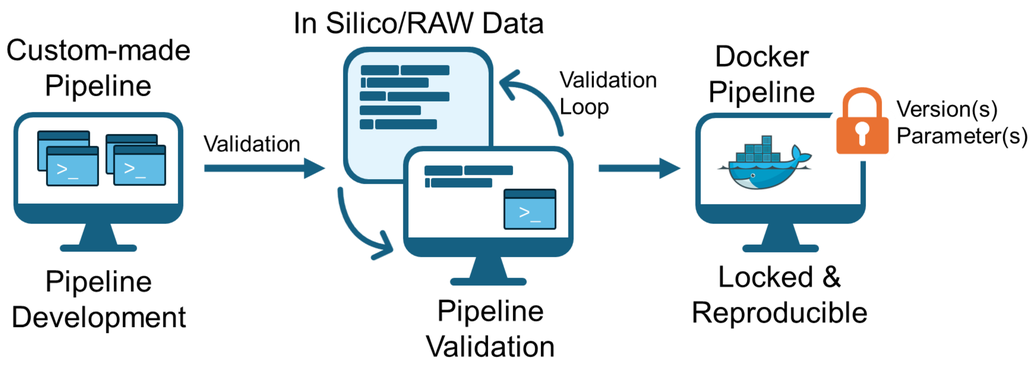
Next- and third-gen sequencing are essential in GMP, GLP, and clinical settings, with FDA support for applications such as gene editing analysis and adventitious agent detection. However, meeting data integrity and 21 CFR Part 11 requirements remains a major hurdle. Many sequencing instruments lack built-in audit trails, allowing raw data to be modified without detection. Since the FDA mandates secure, time-stamped audit trails to track all changes, this gap poses regulatory risk. Further complicating compliance, bioinformatics workflows generate dynamic outputs and involve complex, multi-step analyses. These pipelines often lack robust controls and typically require formal validation. As a result, true compliance demands an integrated solution—secure instrumentation, validated pipelines, version control, comprehensive audit trails, and rigorous quality checks.

1 Raw Data Generation and Protection



All instruments are configured with user access controls. Runs are programmed to automatically write data to a secure, write-only network drive, preventing user modification. Our internally developed, validated software monitors these directories and creates an immediate secondary "FDA backup" on a separate network location upon run completion. File hashes are used to confirm the integrity of the copied data.

2 Bioinformatics Pipeline Validation



Our bioinformatics team designs custom analysis pipelines tailored to specific client needs. Each pipeline is rigorously validated using synthetic in silico reads containing defined indels and mutations. Consistency is assessed by repeatedly analyzing both in silico and actual raw NGS data across multiple runs. Once validated, each pipeline is encapsulated in a Docker container to lock software versions and analysis parameters.

CORE PRINCIPLES

Controlled access and data protection:

Implementation of secure, role-based user access to instruments and protection of data from unauthorized modification or deletion.

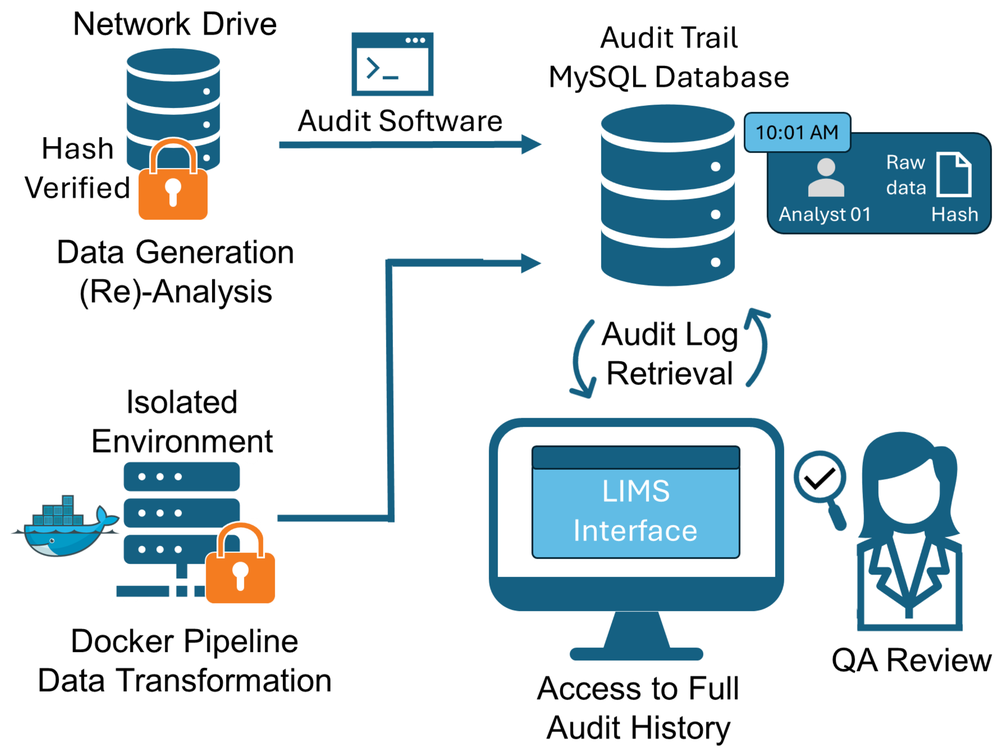
Validated, traceable pipelines:

Formal validation of all pipelines with locked, controlled data analysis environments to prevent undocumented manual intervention.

Audit trail monitoring and review:

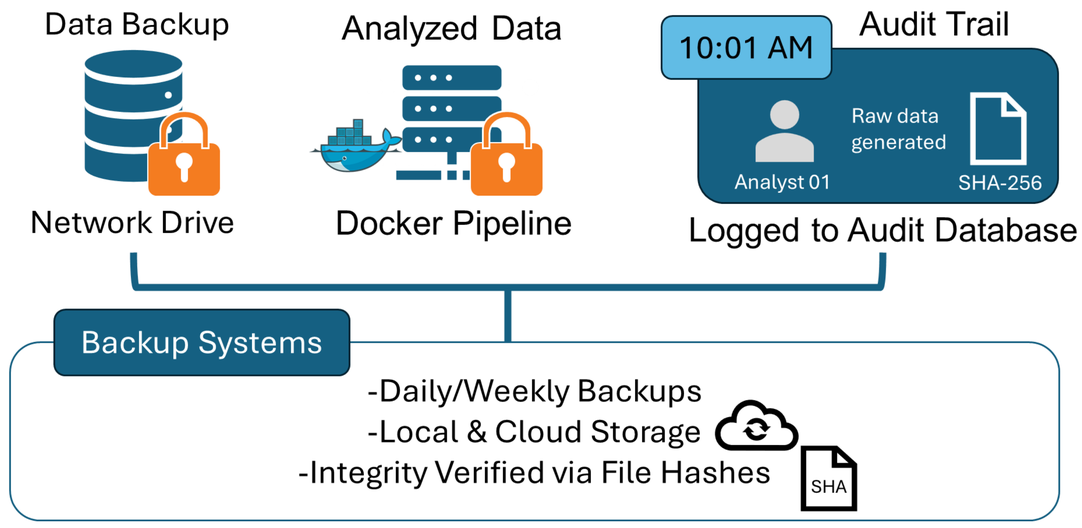
Utilize internally developed software to monitor data generation from all NGS instruments, track offline data processing on servers, and provide a centralized GUI for quality review

3 Audit Trail and LIMS Integration



Avance implemented an automated, database-driven workflow to eliminate manual data handling. All data generation and analysis are verified with SHA-256 hash and logged in a validated MySQL database. All data transformation are performed through pre-validated, automated pipelines operating in a non-interactive environment. A custom LIMS interface provides QA personnel with access to a complete, searchable audit trail.

4



Ensures traceable, recoverable, and verifiable data meeting GxP.

CONCLUSIONS

End-to-End NGS Data Integrity for FDA-Regulated Workflows

Ensuring NGS data integrity requires more than secure instruments—it demands end-to-end control. Avance combines write-protected data generation, SHA-256–verified backups, and validated Docker pipelines. Custom software logs all activity in a validated MySQL database, while a centralized LIMS gives QA real-time access to a full audit trail across the workflow.

Avance's Experience with NGS in GxP

Since 2013, we have implemented NGS technologies in GxP-regulated environments, meeting evolving regulatory standards through validated, fit-for-purpose workflows. We were the first to implement GMP NGS on Illumina MiSeq and offer GMP adventitious agent detection. Key milestones include GLP amplicon sequencing for CRISPR and ZFN trials, bioinformatics development, and GMP NGS for plasmid and viral ID. We launched GUIDE-seq and rhAmpSeq, supported COVID-19 vaccine trials with a GLP integration assay, and now offer GMP single-cell DNA-seq. New Platforms include PacBio, ONT, 10x, Element, and Illumina.

ACKNOWLEDGEMENTS
Edward Wang, Edward Chen, George Smith, Wesley Reader, Suchan Niroula, Ye Tao, Jawid Fatih, Hadir Madrigal, Thy Nguyen, Aldrin Lugena, Rahul Neupane, Apurva Vansadia, Chunmei Qu, Jenny Nguyen, David Wall, Elena Rondon, Oneida Ibarra, Loretta Olamigoke, Lechandra Gee, and James Otto.

www.avancebio.com
(832) 968-8698
9770 West Little York Road, Houston TX 77040