tetrascience | The Scientific Data Cloud Company

Ensuring data integrity with TetraScience

FACT SHEET

Scientific data undergoes a journey from its initial generation to its interpretation and analysis. At each step, the information it contains must be fully preserved. Otherwise, the data and any results from analytics or artificial intelligence (AI) can't be trusted. The entire value chain relies on high-fidelity data, from discovery to commercialization. It's critical for data-based decision-making, data-driven insights, regulatory compliance, operational efficiency, and product efficacy, quality, and safety.

Biopharmaceutical organizations need a robust approach to maintaining data integrity. The value, volume, and variability of their scientific data demand it. Here, we review the ALCOA++ principles and describe how the Tetra Scientific Data Cloud™ upholds them to ensure data integrity for its users.

ALCOA++ principles for data integrity

ALCOA++ is a set of principles and guidelines used in the life sciences and other regulated spaces. The acronym, which has expanded over the years (hence the pluses), represents ten key tenets for ensuring the integrity of data throughout its lifecycle.[1-3]

	Principle	Description
A	Attributable	Data should be attributable to a source (human or program) that created, modified, or reviewed it. Actions like data entry, changes, approvals, and movements should be credited. With accountability, potential errors and discrepancies can be quickly corrected.
L	Legible	Data must be easily readable and understandable throughout their lifecycle. This includes maintaining consistent formatting in electronic records and avoiding abbreviations or other jargon that may introduce ambiguity. Legibility ensures accurate interpretation.
C	Contemporaneous	Data should be recorded in a timely manner, as soon as possible, after the event or observation occurred. This recording helps minimize the risk of information loss or distortion. It also supports accurate and reliable documentation for critical events.
O	Original	Data should be captured without alterations, manipulations, or unauthorized edits. Steps should be taken to avoid any unauthorized modifications that could compromise data accuracy or reliability. Original data provides a reliable, reusable source of information for analysis, audits, and regulatory compliance.
A	Accurate	Data must be complete and free from errors or omissions. It should reflect the true values, observations, or results obtained during data collection or processing. Accuracy ensures that the data can be trusted for decisions downstream.
+	Complete	All relevant data is captured, including any necessary metadata.
	Consistent	Data recording practices are uniform and standardized across different systems, instruments, or operators.
	Enduring	Data is preserved over time in accordance with retention requirements.
	Available	Data can be retrieved when needed.
+	Traceable	Data can be tracked throughout its entire lifecycle.

Corporate Headquarters | 177 Huntington Avenue, Suite 1703, Boston, MA 02115

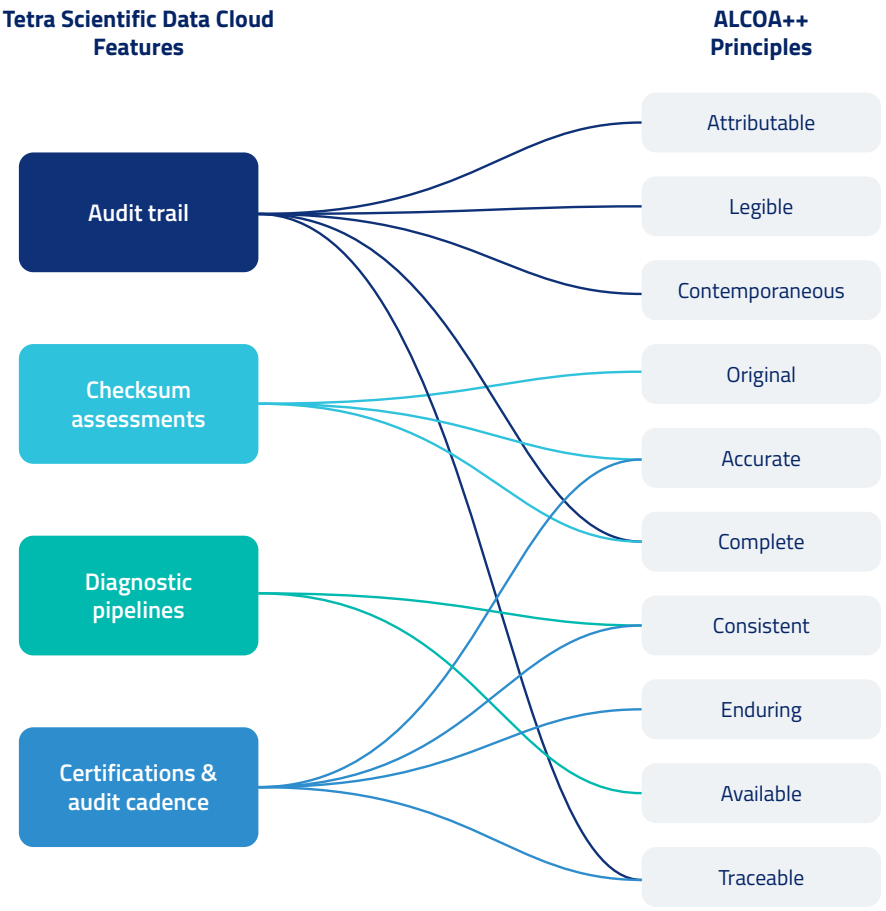
Atlanta | 8000 Avalon Boulevard Suite 247, Alpharetta, GA 30009

tetrascience

tetrascience.com

Achieving ALCOA++ with the Tetra Scientific Data Cloud

TetraScience supports the entire scientific data journey by replatforming and reengineering raw data into AI-native data. Data integrity is a must throughout this process. That's why the Tetra Scientific Data Cloud is designed with ALCOA++ in mind. Below is a list of select features and how they collectively support all ALCOA++ principles.



Audit trail

The Tetra Data Platform (TDP) has an audit trail feature that automatically logs user and system actions that impact files and data. Admin users can access the audit trail page, filter and view records, and download them to a CSV file for inspections or further analysis. Audit trail records can be accessed but not altered.

- **Attributable:** The name, type, and IP address of the user or system who initiated each action are recorded.
- **Legible:** The audit trail page provides an intuitive interface for admin users to view, search, filter, and download records.
- **Contemporaneous:** Actions are logged when they occur and appear on the audit trail page shortly thereafter.
- **Complete:** All actions that may impact data are recorded, including changes or deletions of any raw or processed data (or metadata).
- **Traceable:** A complete audit trail history for any data can be viewed and exported.

Corporate Headquarters | 177 Huntington Avenue, Suite 1703, Boston, MA 02115

Atlanta | 8000 Avalon Boulevard Suite 247, Alpharetta, GA 30009

tetrascience

tetrascience.com

Checksum assessments

Data collected and uploaded to TDP undergoes checksum assessments to verify data integrity. Our agents and connectors interrogate the size, shape, and data type of an incoming file or data package. Upon ingestion into TDP, the values are compared against those scanned, retrieved, or generated locally.

- **Original & Accurate:** Data will only be uploaded to TDP if its checksum and other properties match the original.
- **Complete:** Checksum information is stored and associated with the incoming data.

Diagnostic pipelines

TDP uses diagnostic pipelines to detect issues or anomalies within the system. They check for any "stuck" or timed-out workflows, too many active workflows, diagnostic lambda errors, or workflow errors. They also monitor whether a pipeline hasn't been triggered for a certain period. Diagnostic pipelines run automatically every 3 hours and after platform updates; they can also be triggered manually. All alerts are sent via AWS's Simple Notification Service (SNS), which can be received by subscribing to the SNS topic or using CloudWatch.

- **Consistent:** Diagnostic pipelines monitor data workflows for anomalies to ensure they operate consistently.
- **Available:** They flag errors, bottlenecks, and inefficiencies within the platform to enable swift intervention and promote the timely availability of data.

Certifications & audit cadence

TetraScience maintains a rigorous quality organization to ensure scientific data is secure, protected, and reliable. Core to this effort is a well-documented set of policies, standard operating procedures (SOPs), work instructions, and templates. These policies align with industry standards and are regularly evaluated, revised, and approved.

- **Accurate:** Quarterly internal audits and monthly customer audits are conducted to evaluate the platform's performance, policies, and safeguards for data fidelity.
- **Consistent:** We adhere to industry best practices for data security (SOC 2 Type II and ISO/IEC 27001:2013) and quality management (ISO 9001:2015).
- **Enduring & Traceable:** TetraScience supports GxP, 21 CFR Part 11, and Annex 11 compliance, including their guidelines for data retention and traceability.

References

1. U.S. Food and Drug Administration, *Data Integrity and Compliance With Drug CGMP: Questions and Answers; Guidance for Industry*, FDA-2018-D-3984, 2018, https://www.fda.gov/media/119267/download.
2. Pharmaceutical Inspection Convention/Pharmaceutical Inspection Cooperation Scheme, *Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments Draft (PI-041)*, 2021, https://picscheme.org/docview/4234.
3. European Medicines Agency, *EMA Draft Guideline on Computerized Systems and Electronic Data in Clinical Trials*, 2021, https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guideline-computerised-systems-electronic-data-clinical-trials_en.pdf.

Corporate Headquarters | 177 Huntington Avenue, Suite 1703, Boston, MA 02115
Atlanta | 3280 Peachtree Road NE, 7th Floor, Atlanta, GA 30305

tetrascience
tetrascience.com