

THE DUAL ROLE OF ARTIFICIAL INTELLIGENCE IN DATA INTEGRITY: ADVANCING ASSURANCE OR INTRODUCING NEW RISKS

Abhishek Singhal¹ Dr. Garvendra Singh Rathore

Quality Specialist in Lupin Limited, Indore (M.P.)

Associate Professor in Pharmaceutics at L.B.S. College of Pharmacy, Jaipur

Abstract: - Data integrity (DI) is a foundational element of Good Manufacturing Practice (GMP), ensuring that all data generated throughout the Pharmaceutical Product Lifecycle are Accurate, Complete, Consistent, and Reliable. Global regulatory bodies—including the US FDA, MHRA, WHO, EMA, and PIC/S—converge on the principle that trustworthy data are essential for safeguarding Product quality, Patient safety, and Regulatory compliance. The ALCOA+ framework (Attributable, Legible, Contemporaneous, Original, Accurate, Complete, Consistent, Enduring, and Available) now serves as the universal benchmark for evaluating the credibility and robustness of GxP data across both paper-based and electronic systems.

This paper synthesizes the modern Regulatory landscape (FDA's DI Q&A^(1,2,3), MHRA GxP DI guide⁽⁴⁾, WHO TRS 1033 Annex 4⁽⁵⁾, PIC/S PI 041-1⁽⁶⁾, EU GMP Annex 11⁽⁷⁾, ICH Q9(R1)⁽⁸⁾) and translates it into a practical operating model: risk-based

data governance, computerized systems validation, audit-trail design and review, hybrid controls for paper–electronic environments, and culture. It closes with actionable implementation blueprints, inspection-tested controls, and metrics that quality leaders can adopt to reduce DI risk and demonstrate sustainable GMP compliance.

Modern pharmaceutical operations increasingly depend on complex digital ecosystems, requiring comprehensive data governance, validated computerized systems, secure audit trails, controlled user access, and risk-based oversight aligned with ICH Q9(R1). Despite advancements in automation and digitization, recurring inspection findings show that data manipulation, incomplete metadata, uncontrolled spreadsheets, inadequate audit-trail review, and poor documentation practices remain common challenges across the industry.

Moreover, this paper examines the scientific, regulatory, and operational foundations of

data integrity, integrating current global expectations with practical implementation strategies. It highlights the need for a sustainable organizational culture that promotes transparency, ethical behavior, technical competence, and continuous improvement. Ultimately, strong data-integrity systems not only ensure compliance but also strengthen scientific decision-making, manufacturing reliability, and the trust that patients and regulators place in pharmaceutical products.

Keywords: - Data Integrity (DI), Current Good Manufacturing Practice (cGMP), “G” serves as Good, “x” serves as a placeholder for various regulated domains (e.g., Manufacturing, Laboratory, Clinical) and “P” serves as Practices (GxP)

I. INTRODUCTION

Data integrity is critical throughout the cGMP data life cycle, including in the Creation, Modification, Processing, Maintenance, Archival, Retrieval, Transmission, and Disposition of data after the record retention period ends. System design and controls should enable easy detection of errors, omissions, and aberrant results throughout the data’s life cycle.⁽¹⁾

In the other side, Artificial Intelligence (AI) is rapidly transforming the operational landscape of Pharmaceutical Manufacturing, Quality Control, and Regulatory Compliance.

As the industry accelerates toward Digitalization, AI-driven systems are increasingly deployed to support data capture, pattern recognition, anomaly detection, and decision-making across GxP processes. These technologies offer unprecedented opportunities to strengthen data-integrity assurance by minimizing human error, enhancing real-time monitoring, and improving the consistency and completeness of critical records.

However, alongside these advancements, AI also introduces new and complex risks that challenge traditional interpretations of data-integrity principles. Issues such as Algorithmic opacity, Training-data bias, Model drift, and Cybersecurity vulnerabilities have the potential to compromise the reliability, traceability, and auditability of GxP data. Since regulatory frameworks were originally designed for deterministic, rule-based systems, integrating AI into GMP environments demands a re-evaluation of established controls, validation strategies, and governance structures.

Global data-integrity expectations apply fully to AI-driven systems, even though regulators have not yet issued AI-specific GMP requirements. Existing standards—such as **“FDA’s Data Integrity guidance, MHRA’s GxP DI definitions, WHO TRS 1033 Annex 4, PIC/S PI 041-1, and Annex 11**

for computerized systems”—continue to govern AI adoption and risk mitigation. These frameworks highlight critical expectations: lifecycle validation, robust metadata control, audit trails, secure user access, and risk-proportionate oversight of all computerized systems used in GxP environments.

This section explores AI’s dual role as both an enabler and a potential threat to Data Integrity. It examines how AI can reinforce ALCOA+ principles while also creating new avenues for data-integrity failures if not managed through robust lifecycle validation, risk-based oversight, and strong human governance. By understanding this balance, pharmaceutical organizations can harness the benefits of AI responsibly while maintaining the high standards of trustworthiness required for patient safety and regulatory compliance.

AI as an Advancing Assurance of Data Integrity

AI technologies significantly strengthen data-integrity assurance when designed and validated appropriately. Automated data capture and real-time anomaly detection reduces human error—one of the most common sources of GMP data-integrity failures. AI-driven pattern-recognition systems can continuously monitor large volumes of QC, manufacturing, or environmental data and intelligently flag

deviations, unexpected trends, or outliers that may not be easily detectable through manual review.

These capabilities align with WHO Annex 4’s⁽⁵⁾ emphasis on ensuring complete, consistent, and accurate datasets across the data lifecycle and PIC/S PI 041-1’s⁽⁶⁾ promotion of risk-based governance and continuous monitoring mechanisms. By improving the **completeness** and **accuracy** of data, AI supports ALCOA+ principles and reduces operational variability. Moreover, Electronic and Hybrid systems governed under Annex 11 & Part 11 require validated and traceable data flows; AI-enabled automated workflows can enhance traceability, enforce contemporaneous recording, and simplify audit-trail analysis.⁽⁷⁾

AI also contributes to better decision-making in deviation investigations, change control, and CAPA analysis by rapidly processing historical datasets, identifying recurring patterns, and suggesting scientifically plausible Root causes. These uses reflect the core purpose of FDA’s guidance on data integrity: generating reliable and accurate data upon which GMP decisions can be confidently based.

AI as a New Risk Vector in Data Integrity

Despite its transformative potential, AI also creates new and complex risks that traditional GMP data-integrity controls were not originally designed to address. These

risks arise from Algorithmic opacity, Training-Data Vulnerabilities, Model Drift, Cybersecurity Exposure, and the Possibility of over-Reliance on automated decision-making.

The Figure and their explanations of Risk Vectors which may impact in Data Integrity Principles are mentioned as below: -



Fig: - 01

1. Opacity and Explainability Challenges:

AI models—especially deep learning approaches—often operate as “black boxes,” producing outputs without clear, reviewable scientific rationale. This is problematic under GMP expectations and the ALCOA+ principles of **Attributable**, **Accurate**, and **Original**, which require traceable and scientifically justified records. Regulators such as the MHRA and PIC/S stress that data must be interpretable and reconstructable to support quality decisions and inspection transparency.^(2,8)

2. Training-Data Integrity Risks:

WHO Annex 4 and FDA DI guidance emphasize that source data must be trustworthy, accurate, and properly governed. AI systems trained on incomplete, biased, or poor-quality datasets may amplify errors, jeopardizing the integrity of downstream outputs. Models built on corrupted or incomplete datasets violate the Complete and Consistent attributes of ALCOA+ and undermine the reliability of automated decision support. (5)

Model Drift and Lifecycle Management:

Any computerized system governed under Annex 11, AI models require lifecycle validation and periodic review. Over time, model performance may degrade (drift), leading to erroneous or inconsistent

predictions. Annex 11 and PIC/S PI 041-1 emphasize that GMP systems must remain in a validated state throughout their lifecycle and undergo periodic reassessment, particularly when new risks emerge. AI therefore demands continuous monitoring, version control, and change-control governance similar to other critical computerized systems.^(6,9)

Cybersecurity and Manipulation Exposure:

AI systems frequently involve cloud-based architectures, integrated data pipelines, or APIs, increasing exposure to data tampering or unauthorized access. WHO Annex 4 and Annex 11 discuss the importance of secure data handling, access controls, and protection against manipulation, all of which apply strongly to AI-driven environments. The integrity of both training and operational datasets must be safeguarded. (5,9)

Responsibility and Human Oversight:

GMP requires human accountability. PIC/S PI 041-1 and MHRA emphasize that quality decisions cannot be delegated to automated systems without oversight. AI recommendations must be reviewed, justified, and approved by qualified personnel. Without defined roles and proper oversight, AI use may violate the

“Attributable” principle by masking human decision responsibility. (6,9)

II. EXPANDING ROLE OF AI UNDER A GMP DATA

INTEGRITY FRAMEWORK:

As AI matures, it is poised to play a larger role in GMP operations, provided that deployment is aligned with established regulatory expectations. The revised ICH Q9(R1) highlights improved risk-based decision-making and reduction of subjectivity—areas where AI can be a powerful tool when appropriately validated and governed. A robust AI-GMP governance model must therefore incorporate:

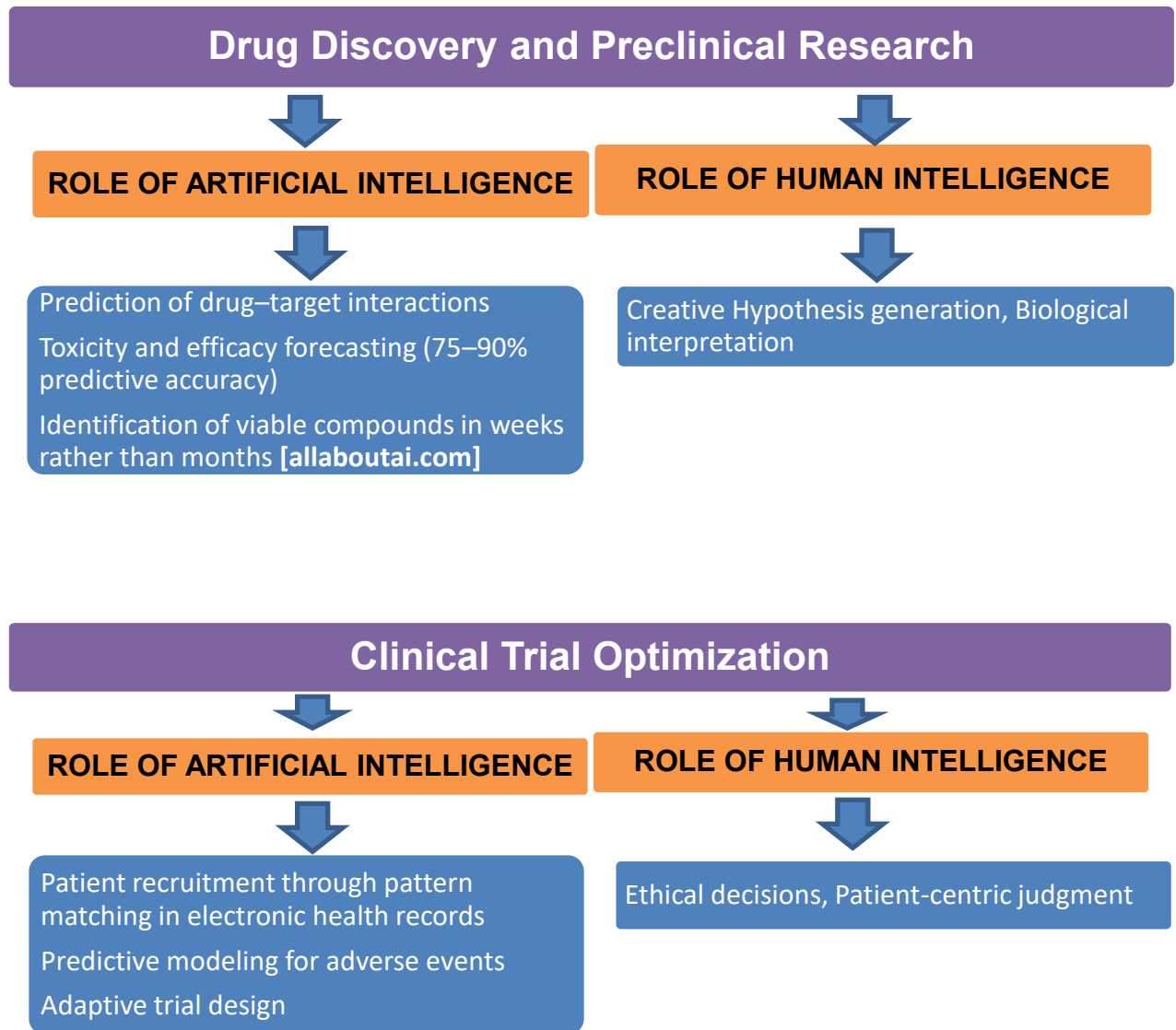
- **Risk-based validation** consistent with ICH Q9(R1) and Annex 11 requirements.
- **Strong data governance** to ensure training and operational datasets meet ALCOA+ criteria.
- **Continuous performance monitoring** to detect drift and maintain validated state.

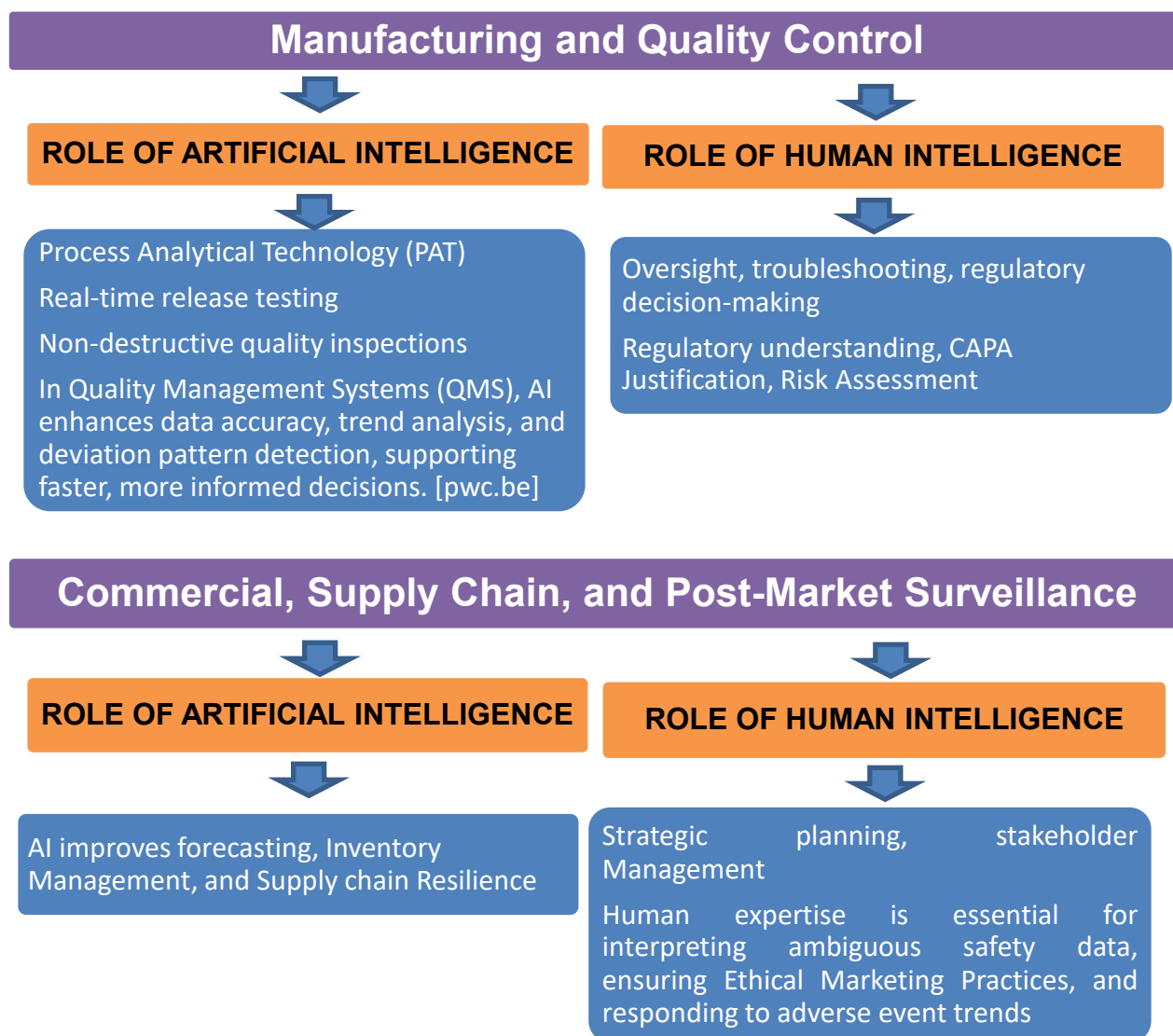
- **Audit trails and traceability** for all model inputs, outputs, and version updates.
- **Human-in-the-loop review** for quality decisions, ensuring accountability. (6,7,8)

AI's expanding role does not replace human judgment but augments it. When appropriately governed, validated, and monitored, AI becomes a powerful enabler—strengthening the integrity, reliability, and robustness of pharmaceutical data. However, without proper controls, AI can introduce novel risks that compromise the foundational principles of GMP compliance.

On the basis of Aforesaid, a formal assessment is performed to address the Role of Human Vs Artificial Intelligence to Facilitates the requirement of Current Pharmaceutical Industry with respect to GMP and regulatory requirement.

Assessment of AI Tools and Human Strategies are analyzed in the different fields of Pharmaceutical Industries to determine the best and effective techniques which can be used in future for GMP Compliance and anticipation in Regulatory Bodies: -





Moreover, **Human Strengths and Irreplaceable roles** are also evaluated where AI cannot be reached for advancing Assurance in Pharmaceutical Industry.

These are discussed as below: -

❖ **Ethical, Regulatory, and Strategic Decision-Making**

Experts emphasize that **AI cannot operate autonomously in a highly regulated sector**. R&D leaders agree that a “human-in-the-loop” or “human-at-the-

core” approach is necessary to set strategic parameters and ensure regulatory alignment, particularly when algorithmic decisions carry patient-safety implications.⁽¹⁰⁾

❖ **Scientific Intuition and Complex Judgement**

Human scientists integrate intuition, tacit knowledge, and contextual

understanding capabilities where AI lacks. This is essential across:

- Hypothesis generation
- Clinical risk–benefit assessment
- Handling ambiguous or novel scientific findings

❖ Oversight of AI Systems

Regulators and industry groups stress that:

- AI requires validation, monitoring for drift, and governance
- Humans must supervise model decisions to maintain GMP compliance

III. CHALLENGES AND RISKS OF AI ADOPTION

For the Adoption of AI tools in Pharmaceutical domain, It is critical to manage the GMP regulations and DI Principles. There are many risk factors are coming for implementation of AI tools. Further, detail evaluation of Challenges and Risk Factors are summarized as below: -

• Data Integrity & Quality Concerns:

AI relies on complete, accurate, and unbiased datasets; however, only 11% of pharma companies report having AI-ready data infrastructures, making reliability a major constraint. (10)

• Regulatory and Compliance Risks

51% of organizations cite regulatory constraints as the primary barrier to AI scaling, due to concerns around validation, accountability, and transparency. (10)

• Model Drift and Opacity

AI models may become unreliable over time. Deep learning systems often lack explainability, challenging their acceptance in critical GMP decisions.

• Workforce Skills Gap

An industry-wide skills gap persists, requiring upskilling in data literacy, AI governance, and digital QMS operations. (11)

IV. THE FUTURE: HUMAN-AI COLLABORATION

Industry studies predict that the next phase is not AI replacing humans but augmenting them:

- AI handles data-heavy, repetitive, or predictive tasks
- Humans maintain governance, ethics, and contextual interpretation

Pharma leaders emphasize that successful companies adopt “human-at-the-core” governance frameworks to ensure AI remains trustworthy and compliant. (10)

The future of DI will increasingly involve AI-assisted compliance, such as:

- Automated audit-trail review using anomaly detection
- Intelligent deviation triage and CAPA recommendation
- Predictive DI metrics
- Digital twins for continuous manufacturing DI risk assessment

Regulators are also discussing:

- algorithm transparency
- lifecycle AI validation
- governance frameworks for autonomous systems

Although formal AI-GMP legislation is still emerging, the principle is clear: AI must enhance, not replace, human decision-making—and must fully comply with existing DI controls until explicit guidelines are published.

V. SUMMARY

Artificial Intelligence (AI) is emerging as a transformative technology within the pharmaceutical industry, offering significant opportunities to strengthen data integrity through automation, advanced analytics, anomaly detection, and enhanced decision-making. By reducing manual interventions and enabling continuous monitoring of critical process and quality data, AI can support compliance with ALCOA+

principles and improve the reliability, accuracy, consistency, and completeness of GxP records.

Despite these advantages, AI adoption introduces new challenges that may adversely affect data integrity if not adequately controlled. Key concerns include algorithmic opacity, training-data bias, model drift, cybersecurity vulnerabilities, and limitations in explainability and traceability. These risks highlight the need for robust governance frameworks, lifecycle validation, periodic model review, and comprehensive data-management practices to ensure that AI-generated outputs remain accurate, reliable, and scientifically justified.

Current regulatory expectations established by the FDA, MHRA, WHO, PIC/S, EU GMP Annex 11, and ICH Q9(R1) remain fully applicable to AI-enabled systems. Although specific global GMP regulations for AI are still evolving, existing data-integrity requirements provide a clear foundation for the validation, monitoring, and control of AI applications used within regulated pharmaceutical environments. Consequently, organizations must apply risk-based approaches to AI deployment, ensuring that computerized systems remain in a validated state throughout their operational lifecycle.

Importantly, AI should be viewed as a complementary technology rather than a replacement for human expertise. Critical

activities such as scientific interpretation, ethical evaluation, regulatory judgment, quality-risk assessment, and final decision-making continue to require human oversight and accountability. A human-in-the-loop governance model is therefore essential to maintain transparency, traceability, and compliance while maximizing the benefits of AI-driven insights.

In conclusion, the future of pharmaceutical data integrity lies in the successful integration of advanced digital technologies with strong quality systems and human expertise. Organizations that establish effective data governance, maintain regulatory compliance, invest in workforce competency, and implement rigorous AI oversight mechanisms will be better positioned to achieve sustainable compliance, improve operational excellence, and enhance patient safety. The most effective model for the future pharmaceutical industry will be a collaborative Human-AI partnership, where artificial intelligence augments human capabilities while preserving the scientific rigor, ethical responsibility, and regulatory accountability required within GxP environments.

Artificial Intelligence has the potential to revolutionize pharmaceutical data integrity by enhancing data reliability, process efficiency, and compliance oversight.

However, its successful implementation depends on robust governance, validated systems, continuous monitoring, and sustained human supervision. The future regulatory landscape will likely favor a balanced Human-AI collaboration model in which technological innovation and scientific accountability coexist to ensure product quality, patient safety, and regulatory trust.

References

- [1] <https://www.fda.gov/media/119267/download>
- [2] FDA. Data Integrity and Compliance With Drug CGMP: Questions and Answers (Dec 2018). <https://www.fda.gov/media/119267/download>
- [3] FDA. 21 CFR Part 11 — Electronic Records; Electronic Signatures and Part 11 Guidance (Scope & Application). <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-A/part-11>
- [4] MHRA. GxP Data Integrity Guidance and Definitions (Mar 2018). https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjn9Y67O6VAxVYb2wGHVEhKQYQFn_oECA0QAQ&url=https%3A%2F%2Fassets.publishing.service.gov.uk%2Fmedia%2F5aa2b9ede5274a3e391e3

- 7f3%2FMHRA GxP data integrity guide March edited Final.pdf&usg=AOvVaw1UHvfQ5bKPHhPBEmyYle5A&opi=89978449
- [5] WHO. Guideline on Data Integrity (TRS 1033, Annex 4, 2021). https://cdn.who.int/media/docs/default-source/medicines/norms-and-standards/guidelines/inspections/trs1033-annex4-guideline-on-data-integrity.pdf?sfvrsn=6218a4e6_4&download=true
- [6] PIC/S. PI 041-1 Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments (Jul 2021). <https://picscheme.org/docview/4234>
- [7] EU GMP. Annex 11 — Computerised Systems (and evolving texts). <https://www.gmp-compliance.org/guidelines/gmp-guideline/eu-gmp-annex-11-computerised-systems>, https://health.ec.europa.eu/document/download/40231f18-e564-4043-94de-c031f813d38b_en
- [8] ICH. Q9(R1) Quality Risk Management (2023). https://database.ich.org/sites/default/files/ICH_Q9%28R1%29_Guideline_Step4_2022_1219.pdf
- [9] https://assets.publishing.service.gov.uk/media/5aa2b9ede5274a3e391e37f3/MHRA_GxP_data_integrity_guide_March_edited_Final.pdf
- [10] [thedhcgroup.com](https://www.thedhcgroup.com)
- [11] [pharmtech.com](https://www.pharmtech.com)
- [12] [mckinsey.com](https://www.mckinsey.com)
- [13] OECD. GLP Data Integrity (Series No. 22) (2021). <https://one.oecd.org/document/ENV/CBC/MONO%282021%2926/en/pdf>