AUDIT TRAIL REVIEW: A KEY TOOL TO ENSURE DATA INTEGRITY

An Industry Position Paper

April 2021

Version PR1

Authors: eClinical Forum (eCF) and the Society for Clinical Data Management (SCDM)

Table of Contents

1.0 - Abstract

In clinical research, data integrity and reliability of trial results are paramount, so the importance of the right policies, procedures, responsibilities, and governance cannot be overstated.

With an incremental use of electronic data capture modalities and the growing regulatory focus on data integrity on GxP records, audit trails which capture the **who, what, when, and why** of electronic data are a critical tool.

Beyond the reconstruction of the data events audit trails can also provide critical insights on **how** the data is being collected, leading to process improvements or lack of understanding of the protocol instructions, up to the rare cases of manipulation from data originators.

This paper outlines an industry perspective on maximizing the value of implementing the targeted, routine review of audit trails. It provides recommendations on risk-based use cases for audit trail review (ATR) and the corresponding desired reporting criteria, with suggestions on when to use visualizations and exception report listings to generate key, actionable insights. It contains practical implementation guidance, covering the people, processes and technology needed to execute ATRs effectively. The authors also take a deep dive into the technical aspects of what constitutes an audit trail, and how collecting and preparing audit trail data is fundamental to successful ATR capability.

This paper addresses the ATR process for sponsors, contract research organizations (CROs) and eClinical vendors. Investigational sites ATR responsibilities are out of scope for this paper.

As an industry, we are at the start of our routine ATR journey. At the eClinical Forum (eCF) and the Society for Clinical Data Management (SCDM), we recognize the need to learn, adjust, and contribute to a continued dialog on this topic across all stake holders.

Key words: audit trail, audit trail review, system access, use cases and visualizations.

2.0 - Acknowledgements

Lead authors:

- Linda King, Director, Data Science, Data Management- Global eCOA Capability Lead, Astellas Pharma, eCF Audit Trail Review Taskforce Chair, Past SCDM Board Chair
- Patrick Nadolny, Global Head, Clinical Data Management, Sanofi, SCDM Innovation Committee Chair
- Joseph Kim, Senior Principal, Slalom
- Miko Pietilä, Sr. Product Manager, Signant Health
- Steven Chartier, Sr. Director of Engineering, Parexel International
- Steve Young, Chief Scientific Officer, CluePoints
- Catherine Célingant, Executive Director, Data Monitoring and Management, Pfizer
- Diego Gayon, Associate Director, Clinical Data Management, Merck
- Jennifer Logue Nielsen, Competency Development Specialist, Data Management Process and Innovation, Novo Nordisk
- Shelby Abel, Director, Clinical Data Management, Alexion Pharmaceuticals, eCF Steering Committee Board Member
- Cinzia Piccini, Quality Consultant, Biometrics, Eli Lilly and Company

Content contributors and/or reviewers:

- Lordan Bulanadi, Sr. Product Manager, Cenduit
- Alice Phung, Executive Director, Clinical Programming, AbbVie
- Valdo Arnera, Scientific Advisor and General Manager ERT Geneva, ERT
- Ana Small, Associate Director, Data and Digital, Global Development Quality, Novartis
- Kathryn Engstrom, Data Scientist, Eli Lilly and Company
- Jo Rosser, Director, Global Study Operations Data Management, Amgen
- Jessica Millman, Associate Director, eCOA Business Lead, Bristol Myers Squibb
- Louise Lampe, Global Head, Central Data Science Solution, Boehringer-Ingelheim
- Michael Buckley, Enterprise Clinical Research Innovation Manager, Memorial Sloan-Kettering Cancer Center

Thank you to:

- The eCF Regulatory Team and Steering Committee review and feedback on the paper
- The eCF Conference Workshop participants who provided key insights into use cases
- The SCDM Executive Committee, Board of Trustees and associated organization's review, feedback and support of the paper

- The SCDM Leadership Forum (2019, 2020) for tackling the audit trail review topic and promoting a dialog across the industry between clinical data management leaders, regulators and technology vendors

eCF and SCDM would also like to acknowledge the many volunteers who have participated in the eCF ATR Taskforce/workshops and SCDM Innovation Committee and have contributed to forming the thoughts expressed in this position paper.

ABOUT THE eCLINICAL FORUM: The eClinical Forum (eCF) is a global, technology independent group representing members of industries engaged in clinical research. The eClinical Forum's mission is to serve these industries by focusing on those systems, processes and roles relevant to electronic capture, handling, and submission of clinical trial data. For further information visit the website at www.eclinicalforum.org. The eClinical Forum has sought out opportunities to promote electronic Clinical Trials since its inception in 2000. The cross-industry forum has a broad view of research with members - Sponsors, Contract Research Organizations (CROs), Technology vendors (both clinical research and healthcare), Academia, and Investigators - and with invited outreach opportunities with global Regulatory representatives.

ABOUT THE SCDM: Established in 1994, the Society for Clinical Data Management is a non-profit, international organization of over 2,400 members founded to advance the discipline of clinical data management. For further information on SCDM visit the website at www.SCDM.org

Our mission is to "Connect and inspire professionals managing global health data with global education, certification and advocacy" and vison to "Leading innovative clinical data science to advance global health research and development". For further information on Clinical Data Science visit the website at https://scdm.org/clinical-data-science/

3.0 - Introduction

3.1 - Preface

Since its first introduction in October 2020, the eCF and the SCDM received requests for clarifications to further guide organizations implementing ATR strategies. As a result, we are releasing this enhanced version as our final position paper.

This paper was written with two clinical research audiences in mind: development business professionals, including technology experts. The responsibility for ensuring data integrity through audit trails and ATR straddles both functions.

To properly process an audit trail for use in reporting, one first needs a deep technical knowledge of how it was constructed, and how to present it to the next user in a readable and process-enabled format. Once the audit trail is available, the business value proposition is about how to maximize its use in real-life risk scenarios, and how to derive actionable steps to address data integrity and/or study conduct concerns.

Audit trail is typically defined as a secure, computer-generated, time-stamped electronic record allowing for the reconstruction of the course of events relating to the creation, modification, or deletion of electronic GxP records. It is important to realise that audit trail datasets are commonly stored as several different "physical" datasets which may include, but are not limited to, clinical data and metadata audit trials, queries, as well as system generated logs. Beyond data processing system logs and audit trails, activity system logs have proven to be valuable in assessing process efficiency and activity patterns and are therefore in scope for ATR (see definitions on "Types of audit trail" within the Technology section for details).

It is also acknowledged that over the last few decades, technologies have evolved considerably, and systems have been designed with a variety of underlying architectures. As a result, this paper provides technical recommendations which should be used as guiding principles. When implementing ATR, each organization such as sponsors and CROs must tailor their strategy by considering the capabilities of their own specific technologies and the dependencies on their associated processes.

Both the technology and business sides of the ATR process are described here, so share this document across business and technology experts at your organization to help inform your ATR approach.

a) Trustworthy, Reliable Data

The publication of the Medicines and Healthcare products Regulatory Agency's (MHRA) final Data Integrity Guidance, in March 2018[1], was the culmination of several years of industry discussion on ensuring all data submitted to regulatory authorities is trustworthy and reliable.

[1] MHRA 'GXP' Data Integrity Guidance and Definitions. (2018). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/687246/MHRA_GxP_data_integrity_guide_March_edited_Final.pdf

One key message was the importance of creating a culture of data integrity through policies, procedures, responsibilities, and governance.

Adherence to International Council on Harmonisation (ICH) Good Clinical Practices (GCP) is a core tenet of the mission to achieve high data quality and integrity. Regulatory agencies including the MHRA, Food and Drug Administration (FDA), European Medicines Agency (EMA) and Pharmaceuticals and Medical Devices Agency (PMDA) have highlighted the role of routine electronic ATR in ensuring data integrity.

As noted in the SCDM Reflection Paper on the impact of the Clinical Research industry trends on Clinical Data Management: "The volume of data collected outside EDC has already eclipsed the volume of data collected on eCRFs and is growing at a much faster pace every year, fueled by the cries for patient centricity leading to the rapid adoption of eCOA, wearables, sensors and other eSource solutions.'[2] ICH E6 (R2) also strengthened the requirements for audit trails, and on the manufacturing side, ATRs have long been required.[3] More recent guidance has suggested that companies should "implement procedures that outline their policy and processes for the review of audit trails in accordance with risk management principles".[4]

As the risk-based processes and systems founded in manufacturing are being formally incorporated into the clinical regulatory environment through updates to Good Clinical Practice, so too are the requirements for audit trails and the expectation from regulatory auditors for audit trail reviews. When used routinely, they "may reveal incorrect processing of data, help prevent incorrect results from being reported and identify the need for additional training of personnel".[5] It is incumbent upon each organization involved in clinical research data management to incorporate the ATR practice into their data integrity strategy for GCP compliance.

This position paper will define how audit trail review can facilitate data integrity controls throughout the lifecycle of clinical data. It will examine 'what' should be reviewed, and the types of potential issues the methodology can detect. Five use case categories with 20 individual use cases and associated risk scenarios are defined in the context of risk to data integrity and site performance throughout the dataflow. We will also link the use cases and their associated scenarios to the characteristics of data integrity – "the degree to which data are complete, consistent, accurate, trustworthy, and reliable throughout the life cycle" – and identify the most efficient way to perform the review.

---

[2] SCDM, The Evolution of Clinical Data Management to Clinical Data Science A Reflection Paper on the impact of the Clinical Research industry trends on Clinical Data Management. (2019). https://scdm.org/wp-content/uploads/2019/09/SCDM-Reflection-Paper-Evolution-to-Clinical-to-Data-Science.pdf

[3] FDA, Data Integrity and Compliance With Drug CGMP Questions and Answers Guidance for Industry. (2016). https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-integrity-and-compliance-drug-cgmp-questions-and-answers-guidance-industry

[4] PIC/S, Good Practices For Data Management And Integrity In Regulated GMP/ GDP Environments. (2018). http://academy.gmp-compliance.org/guidemgr/files/PICS/PI_041_1_Draft_3_Guidance_on_Data_Integrity_1.pdf

[5] MHRA 'GXP' Data Integrity Guidance and Definitions. (2018). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/687246/MHRA_GxP_data_integrity_guide_March_edited_Final.pdf

b) ALCOA+

Data integrity is often associated with the term ALCOA+, or:

- Attributable
- Legible
- Contemporaneous
- Original
- Accurate
- The "+" refers to Complete, Consistent, Enduring, and Available.

Each data element is associated with an authorized data originator. Regardless the ALCOA+ principles, data might not be considered reliable if, for instance, is not originated by a non-qualified person, or by a non-calibrated instrument, or by a non-validated system.

In 1999, the Institute of Medicine defined high-quality data as "data strong enough to support conclusions and interpretations equivalent to those derived from error-free data". In 2018, MHRA defined data quality as "the assurance that data produced is exactly what was intended to be produced and fit for its intended purpose. This incorporates ALCOA."4 To ensure data quality, we need to identify and remediate issues in the data including, but not limited to systematic or significant errors in data collection and reporting at a site or across sites, or potential data manipulation or data integrity problems. ATRs are one of the key methods used to identify such errors.

This paper will offer practical advice, explaining that the implementation of ATR capability includes five components:

- process
- roles
- technology
- standards
- regulatory expectations

The authors will describe how maximizing the use of technologies such as reporting and visualization tools can facilitate the review of very large audit trail datasets by identifying actionable trends, meaningful outliers, and other areas of improvement while separating out ambiguous noise from the high volume of audit trail data. They will also discuss the best roles to conduct ATR and look at processes that document and file the review and any subsequent actions.

Ultimately, the successful implementation of effective and efficient risk-based data control and data review relies on industry and individual companies analyzing and prioritizing the critical aspects of data reliability as outlined in the MHRA and ICH E6(R2) guidance documents.

Sponsors bear the ultimate responsibility in ensuring data integrity and reliability of the trial results and, therefore, should determine the right level of ATR within studies based on the identification of critical data and processes and the associated risks.

3.2 - Importance of Audit Trail Review

The clinical research landscape is evolving. With adoption of new protocol designs such as adaptive or the collection of more eSource data, there is an ever-growing need to adapt practices to ensure that new drug applications contain the accurate, reproducible, reliable data that provides life changing, safe, efficacious therapies at lower costs, quickly.

To meet this tall order, industry is implementing risk-based processes and using various technologies to collect, process, and report data. As such, global regulators have seen approaches such as electronic data capture (EDC), electronic clinical outcome assessments (eCOA), including electronic patient reported outcomes (ePRO), electronic informed consent (eConsent) and digital health instruments, evolve.

According to the IQVIA4 Institute for Human Data Science, as of 2017, there were more than 318,000 Wellness Management apps and 340-plus consumer wearable devices tracking, measuring and monitoring health related parameters.[6] It is, therefore, not surprising that sponsors are increasingly using digital health technologies in clinical research, and leveraging apps to collect reported outcomes and other real-world data (RWD). As the market grows and as the industry increases adoption of technologies, there is also a growing trend for pharmaceutical companies to rely on third-party service vendors to provide these technologies – many of which may have less experience with the regulatory landscape, or expertise in the related requirements.

a) Regulatory Expectations

Regulators expect pharmaceutical companies to implement meaningful and effective data integrity risk-management strategies, based on strong data governance, collection technologies and operating models. This is how software developers can improve their products to better enable clinical investigators and sponsors to ensure the safety, efficacy and quality of their products, and fulfil their responsibility to protect the public health.

The reverberating message from the regulators is that data integrity and quality as well as reliability of trial results are of paramount importance. In the past few years, bodies have insisted, usually through major and critical inspection findings, that data collection system audit trails be available, convertible into generally readable format, and checked regularly in order to safeguard data integrity. At this time, however, ATR and audit trail oversight are not clearly described in regulations. While the agencies offer some description of their expectations of ATR in various reflection papers, they provide no clear instructions.

Without specific ATR regulatory guidelines with examples and/or use cases, pharmaceutical and biotech companies have struggled to appreciate the growing regulatory expectations on this topic. Regardless, companies could develop meaningful ATR processes using risk-based

[6] IQVIA, The Growing Value of Digital Health. (2017). https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health

approaches, that at least focus on critical data / data sets. They can realize the value of ATRs while balancing the resources required to develop and execute it.

b) Reconstructing Data Events

In clinical research, the ability to reconstruct data events is integrated into the overall quality system, and risk-based processes and technologies are designed to protect the integrity of data.

A computer system-generated audit trail is one form of metadata that contains information associated with the data events. While it provides secure recording of data life-cycle steps, its risk-based review – either proactive or reactive – within the GCP setting should enable evaluation of compliance with critical processes and data.

c) Overseeing Processes

Beyond the reconstructing of data events, data and audit trail reviews (including system logs) have the potential to reveal significant information on the conduct of the trial, compliance to critical processes and process optimization needs.

In this climate, many pharmaceutical companies have started to explore how to define risk-based governance on audit trails oversight, expanding the objectives (*see table 1*) from the detection of fraudulent data to decision-making steps, incorrect processing of data, abnormal data, incorrect results to be reported, process improvements and training needs.

• investigation of data integrity issue • identification of suspicious justification and/or fraudulent data • identification of alternative source data implemented by sites • unauthorized accesses and data events • oversight on changes to critical data • process improvements based on trends • performance of users

Table 1: Potential objectives for audit trail reviews (See Appendix 3 for Use Cases and Risk Scenarios)

When data integrity or study conduct issues are discovered, reactive reviews of audit trails are part of the investigation. For prospective ATR, however, the leading obstacle is the scope of definition. A meaningful review must be based on risks, acceptance criteria, and tolerance limits that align with existing risk reduction activities and controls.

The foundational expectation of audit trails to independently secure recording of data events should be preserved. For both reactive and prospective reviews, other key factors, such as criticality, and information, such as format, might need to be included in order to fulfil the objectives of the review, without altering audit trail functionality.

4.0 – ATR Scope

4.1 – Systems and Data

As defined in ICH GCP E6(R2)[7], sponsors are responsible for the quality and integrity of all the data they use during clinical research and development. Therefore, all data types and systems used to generate and manage those data should be considered in scope of ATR.

Effective operationalization of ATR has become important to demonstrate, the integrity of data reported to the sponsors, to support product submissions and meet the heightened expectations of global regulators.

However, not all data have the same importance when it comes to demonstrating the safety and efficacy of a new drug, biologic, device, or diagnostic.

4.2 - Data Criticality and reliability

In the same way as sponsors and regulators have embraced risk-based approaches in the monitoring and data review areas, ATR techniques and frequencies should be based on data criticality and associated risks. To that end, the authors have elected to focus this paper on the higher risk systems with critical data, such as electronic data capture (EDC), electronic clinical outcome assessments (eCOA), and interactive response technologies (IRT). However, the need for ATR should be evaluated for all data supporting clinical development, patient safety, product quality, and regulatory compliance using justifiable risk-based approaches. Importantly, any ATR plans must incorporate the same rigor with respect to avoiding potential unblinding as for any other study data.

4.3 - Collaboration with Third-Party Vendors

Most third-party vendors involved in the generation and management of clinical development data are in scope of regulatory guidance on clinical research. This includes eClinical vendors providing data capture technologies in scope of ATR. Therefore, it is appropriate for sponsors to have expectations from eClinical vendors to enable ATRs and from CROs to have ATR embedded into their existing risk-based approach.

While sponsors retain overall responsibility for the quality and integrity of clinical research data, sponsors may transfer all or a portion of the ATR activities to an eClinical vendor or a CROs, and exercise appropriate oversight.

Sponsors should assess the ATR capability within the eClinical system and implement mitigations in case the system functionality and/or operational processes are not sufficient to enable optimum ATR.

ATR should be enabled to sponsor delegated personnel, auditors and inspectors, ideally on the data capture system side, close to the data and process source. The access to audit trails in live

[7] ICH E6 (R2) Good clinical practice. (2002). https://www.ema.europa.eu/en/ich-e6-r2-good-clinical-practice

systems should consider blinded and unblinded roles and have the ability to obtain a formatted extract/copy of the required audit trail data in a fully usable and, preferably, standard format.

4.4 - Use Cases

We have made the case for the importance of ATR – but where to start?

An audit trail is a rich dataset that can tell many stories, and our goal is to illustrate the most compelling ones in the form of use cases.

Our method:

- Guided by the MHRA Data Integrity Guidance[1] document, the team identified and discussed the data integrity risks within each component of the data life cycle
- Each risk was evaluated to determine the most appropriate actions and assessment tools
- Where audit trail was deemed to be the most appropriate approach, this was further categorized as a primary or secondary tool
- Data integrity risks where audit trails were the primary tool were grouped into five broad use case categories:

    1. System Access (including integrations)
    2. Data Changes
    3. Data Collection
    4. Reporting
    5. Device Concerns

Within each use case category, we identified multiple Use Cases with scenarios to illustrate the common data integrity risks that can be mitigated via ATR. For each data integrity risk, we list:

- The data type the risk is relevant for (‘applicable system’)
- The required data (‘sources needed’)
- Applicable creator of ATR tool ('Sponsor/Vendor')
    - *Where the source system enables the creation of ATR tools, we recommend to do so (e.g. EDC or eCOA Reporting modules with access to audit trail data). Those may automatically generate potential findings based on agreed upon criteria from the sponsor. Regardless of by who and where the ATR tool is created, sponsors need to interpret findings and take action as needed, in collaboration with the eClinical vendors where appropriate and/or necessary.*

      Note: As eClinical Solutions mature and ATR becomes an expectation across the industry, sponsors may expect eClinical vendors to provide standard ATR tools as part of their technology offering.
- Desired review capabilities (‘desired reporting criteria’)

The full set of use cases includes:*

- USE CASE CATEGORY 1: System access:

- four use cases and multiple scenarios including:
    - 1.1: access for the right role at the right time, including user role changes
    - 1.2: appropriate training time
    - 1.3: site and vendor performance regarding login data, data oversight, frequency of logins, timing of logins
    - 1.4: system login including system integration checks (e.g., source integrated with EDC)
- USE CASE CATEGORY 2: Data changes:
  - five use cases and multiple scenarios including:
    - 2.1: detecting modified/deleted data at an item, record, or patient level
    - 2.2: changes to inclusion/exclusion criteria, primary efficacy data, key secondary data, safety data and/or other critical data as defined in risk-based monitoring
    - 2.3: changes after key timepoints such as database locks or subject's study disposition status is marked as complete
    - 2.4: excessive changes within the database during the life of the study
    - 2.5: changes made long after initial data entry or source is obtained
- USE CASE CATEGORY 3: Data collection:
  - Seven use cases and multiple scenarios, mainly related to eSource:
    - 3.1: timing of the data collection:
      - 3.1.1: data not collected per protocol stipulated timing
      - 3.1.2: data collected / entered at implausible times
      - 3.1.3: data not collected contemporaneously
      - 3.1.4: data collected outside protocol required windows
    - 3.2: missing data such as visits, values, or Principal Investigator signatures
    - 3.3: data entry being influenced by the edit checks.
    - 3.4: data entry changes by patients
- USE CASE CATEGORY 4: Reporting:
  - two use cases and scenarios around data transfers and using the relevant system's logs (logs (see definitions on "Types of audit trail" within the Technology section for details):
    - 4.1: duplicate datasets
    - 4.2: data changed during migration from source to sponsor (e.g., corrupt data, dropped data, partially transferred data)
- USE CASE CATEGORY 5: Device concerns:
  - two use cases and scenarios:
    - 5.1: changes to the date/time stamp of eSource devices
    - 5.2: merging of data (e.g., multiple subject IDs, replacement devices)

**Note that not all use cases are applicable to all computerized systems. For the full set of use cases see Appendix 3*

The Use Case Categories, Use Cases, and scenarios listed out in this paper are intended as a baseline to kick off the ATR discussion and its implementation. We would encourage the clinical research community to continue exploring the scope of ATR and report back to the SCDM and/or the eCF on any identified new use cases.

5.0 - Position Statement

5.1 - Process and People

This position paper includes an initial list of data integrity and study conduct risks with Use Cases that can be mitigated using ATR.

Appropriately defining processes and roles plays a key part in ensuring consistency and accuracy in the execution of ATR. As part of their process documentation, sponsors (i.e., functional leaders, quality representatives and system administrators) are advised to include details of:

- How ATR is used to mitigate each risk
- How organizations will respond to ATR findings
- Who is responsible, accountable and informed in each scenario

The exact roles and responsibilities by function may vary in each organization; however, the core purpose and benefits of the ATR should remain consistent.

Each sponsor should continuously assess and identify potential additional risks and mitigations, looking to incorporate learnings and enhance their standard ATR plan.

a) Process Set-Up

The authors recommend sponsors to leverage existing processes and tools to define, document and execute ATRs before creating new ones. For most organizations, discussions associated with the study's quality management plan will provide the framework necessary to identify the risks that ATR can mitigate. At a study level, like all reviews, ATRs should be implemented based on their ability to mitigate risks associated but not limited to critical data and processes (e.g., TransCelerate's Risk Assessment and Categorization Tool (RACT)[8]. Sponsors should also consider defining standard ATRs to be conducted across all studies and augmenting them with study-specific reviews.

All ATRs for a trial should be documented in the study's data management and/or data review plans, including information on:

- What is to be reviewed (i.e., which audit trail and for which specific scenario)
- Who is responsible for reviewing each scenario (i.e., sponsor, vendor, CRO, as well as the role)
- How it will be reviewed and where the outcome will be documented (i.e., manual or automated checks or alerts)
- When and how frequently it will be reviewed
- Differentiation between standard and study-specific checks
- Considerations for blinding protections as applicable
- Clarifications on scenarios with partial or complete limitations for ATR

[8] TranCelerate, Risk Based Monitoring Solutions. https://www.transceleratebiopharmainc.com/assets/risk-based-monitoring-solutions/

Some ATRs, such as those used to mitigate inherent system risks instead of study-specific risks, may be performed consistently at the system, rather than at the study level (e.g., system's logs for system access, audit logs for integration of data sources into a data lake, etc.). For such cases, organizations need a mechanism to ensure that relevant results are communicated to the study team to enable timely corrective actions.

When using systems from a third party, sponsors should ensure the eClinical vendors can enable and/or perform the required ATRs. This includes capabilities related to the structure of the system's audit trail, as well as the eClinical vendor's existing ATR processes if available. Careful consideration should be given as to whether the eClinical vendor or the sponsor is in the best position to enable and/or perform the ATRs. Where the eClinical vendor is selected to perform some specific ATR activities, the outcome must be reviewed and shared with the sponsor in a timely manner to enable any necessary corrective action by the sponsor or delegate. Once organizations have defined and documented the process and responsible roles, the next step is building the specifications for any necessary tools (standard or study/program specific), such as exception reports or visualizations (*see Technology section*). These will include the scenarios being addressed, the impacted data and system(s), the desired reporting format and the testing and validation requirements. When a program or an indication includes a specific risk that can be mitigated via ATR, sponsors should aim for standard and consistent ATR across all studies within the program/indication.

b) Process Execution

The ATR process execution follows data through its data lifecycle, beginning when the trial's data collection activities start and ending at the final data archival. In other words, the process is performed throughout the trial execution.

The frequency of each ATR activity will depend on the criticality of the data it pertains to, the risk level associated with the potential issue it could identify, and the type of tool used. For example, a review of too frequent or suspicious data changes in the EDC system should be as contemporaneous as possible to enable timely corrective action. This strategy should be included in the ATR plan as documented in the data management or data review plan.

Upon the execution or application of each ATR tool, the reviewer will confirm signals and the presence of indicators that might suggest a data integrity issue, thus triggering a deeper analysis by the study team (e.g., a trigger for excessive changes at a given site or on a specific eCRF).

The study team should conclude their analysis with an interpretation of the scenario, perform a root cause analysis and generate a corrective and/or preventative mitigation plan. These decisions must be documented at a study level, and actions logged and tracked through to resolution. It is expected that mitigation plans can be measurable to verify efficacy after implementation.

Organizations (e.g., sponsors, CROs) are advised to keep ATR outcome and mitigation activities in their Trial Master File (TMF) as well as ensuring that a broader analysis at program or portfolio level is carried out. Mechanisms to ensure appropriate mitigation plans are identified and correctly implemented are an integral part of the ATR process.

The ATR study-level plan within the data management or data review plan is a living document that should be assessed by the study team during the trial. ATR continuous improvement allows for additional risks and tool enhancements to be identified and adopted at trial, program and portfolio level.

In case the ATR activity identifies a serious breach and/or persistent noncompliance on the part of an investigator/institution to GCP and or the trial protocol, the sponsor should notify the regulatory authority(ies) in alignment with GCP and site/trial monitoring expectations.

c) People

Different roles may be assigned to specific ATR use cases:

- For use cases pertaining to system access or site login activity on a CRO or eClinical vendor portal for example, the study team role that is accountable for granting, monitoring and removing access would be accountable (often a combination of Data Management and Study Management).
- For the use cases related to data collection, data changes, reporting and device issues, roles within Data Management and Monitoring, with data review skills and dealing directly with study affairs and study data, are best suited to perform ATR.
- Data stewardship roles responsible for data acquisition, quality, and management, have the most incentive for conducting related ATRs. They will be able to deliver an improved quality assurance on data after identifying and mitigating audit trail related issues.

A general rule of best practice is for team members to be responsible for the review tasks that most closely relate to the process step for which they are accountable or have oversight for. For example, for an ATR of a test migration of data, the data manager who focuses on oversight of user acceptance testing (UAT) should perform the ATR against the issue log. For the ATR of eCOA using a vendor centric data change process, a data manager with eCOA subject matter expertise or a study manager accountable for oversight of the vendor's data changes (e.g., through Data Clarification Forms) should be responsible for performing the quality review.

Other team members, of course, need to be involved in the process to ensure it is effective. When an audit trail uncovers a potential site compliance issue with the visit schedule, the Data Manager responsible for the ATR review should report the issue to the Study Manager, who in turn will ask the appropriate site-facing team members to further investigate. A third-party quality review may also be necessary, to further evaluate the issue, for example, from a compliance standpoint. If the issue is found in more than one study, the quality review team

accountable for the ATR capability can then escalate it to a program-level or capability-level lead to implement the appropriate mitigations across the portfolio.

5.2 - Technology

a) Data Maturity Model

Technology is essential in our quest to discover data-driven insights from audit trails.

First things first, like any other data, the audit trail data needs to be of sufficient quality and maturity to generate meaningful insights. Understanding how the data will be ultimately compiled and how raw data of indeterminate quality is prepared into data of known quality that is ready for analysis will help us understand the appropriate level(s) at which ATR can and should be conducted.

Figure 1 depicts this data transformation from raw data to insights as a generic Data Maturity Model, with hierarchical levels, each one of which should be achieved before moving onto the next. It is also important to realize that the applicability and effort spent at each level may vary from company to company and must be tailored to each raw data.

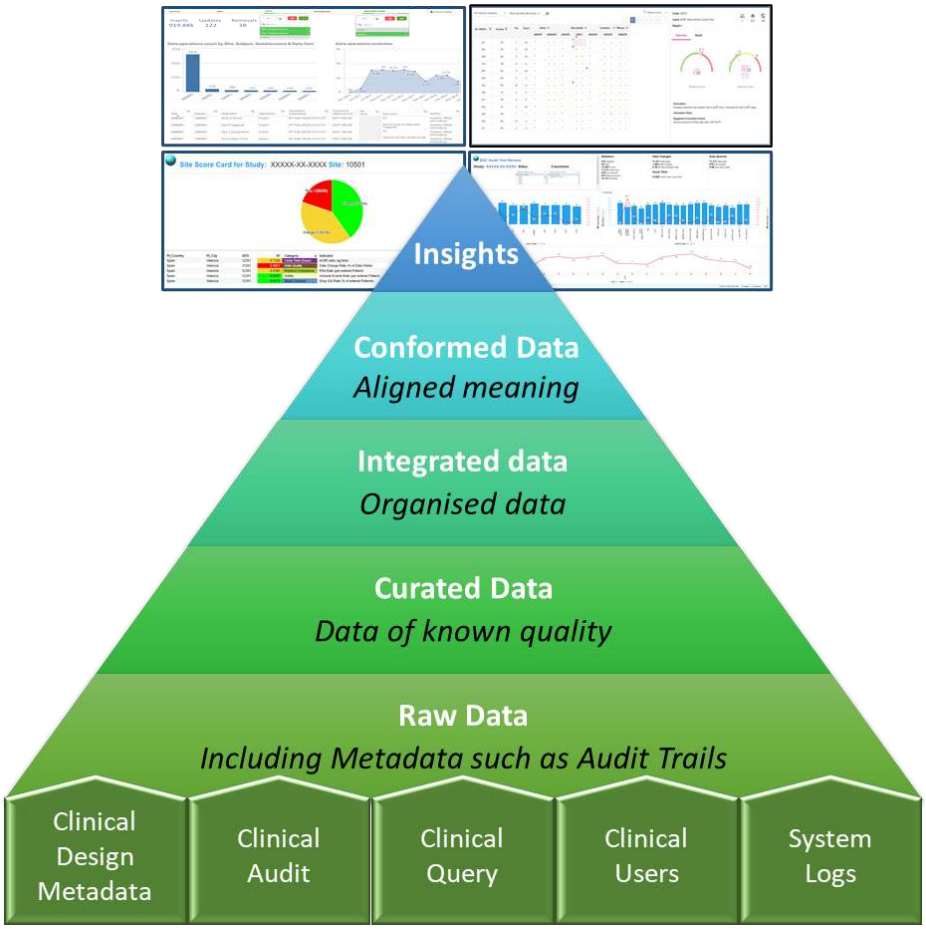


Fig 1. Data Maturity Model

Raw Data

First, we need to start with a comprehensive set of **raw data** (i.e., not limited to the audit trail) to generate insights. This includes the raw data itself, and its associated metadata, which is the data about the data, such as our audit trail. We cannot generate insights when we do not have the raw data containing the correlative measures we wish to unearth. With EDC and eCOA systems, the raw data might be an updated Form-Item value and its associated metadata that typically includes the fully qualified timestamp of the update. Data from other systems such as diagnostic equipment might not contain a complete audit trail and may only provide the date of last update for the entire record (i.e., Diagnostic Procedure). Last, it is critical to prevent access to raw data that could compromise the integrity of the clinical trial by exposing potentially unblinding data.

Curated Data

The next step is **curated** data which we consider as data of a known quality that must be measurable. We must understand the areas of our data with quality weakness, and how that may prevent us from achieving our insights. An EDC or eCOA form may only have some of the required fields populated (i.e., having missing data). For example, metadata may have its time stamp without the associated time zone as a result of a device replacement or reset issue. A small percentage of missing data may not compromise an insight, but a significant proportion of missing data may make an insight unattainable until data quality rises to a certain, measured threshold.

Integrated Data

Third, we have **integrated data**. This involves joining the many single datasets into a single master dataset in a way that does not introduce duplicates nor exclude data. Industry standards such as Clinical Data Interchange Standards Consortium's (CDISC) could be used to allow for easier interpretability of the integration outcome. By having a data model that supports common data keys, data relationships, and data grains (i.e., right level of data specificity), we will extend the width and depth of our dataset to gain all the information needed for our analysis. Without integrated data, some insights may not be discoverable. As an example, audit data, may need to be the overlay of Clinical Audit data (Form-Item grain) with a companion User Access dataset. We must be able to successfully join these two datasets to understand which Clinical Personnel was updating which Subject Form-Item.

Conformed Data

Fourth, we have **conformed data**. This is possibly the most challenging level in our Data Maturity Model, as it exposes the most risks when mistakes happen. In data conformance, we ensure that any data column we are combining from two different sources have the same meaning. It is not enough to simply overlay data with the same column names. If EDC, eCOA and IRT systems all provide patient status and we have not aligned the status values, we may alter the patient's status meaning when we combine the information together. Worse, should the same EDC system and IRT systems provide patient blood pressure and we combine the data

with the same integration key, we may be combining blood pressure supine with blood pressure sitting – this is easily done if both columns are simply labelled "patient blood pressure". It may be more challenging to identify meaningful data patterns when we have created high data variance for each subject simply by joining and combining data that is not the same.

Data-driven Insight

Last, we have **data-driven insight**. When we have completed each layer correctly, we are in a good position to unearth data-driven insights. Whatever our hypotheses are, we have the data that we believe is necessary to reach a conclusion on it, we know this data is of a known, measured quality, is complete, is unique/non-duplicative, and it has its true meaning. Now human, automated and AI-based solutions can identify the trends and correlations necessary to discover and prove a result.

b) Types of Audit Trail

The guidelines from the International Council for Harmonisation (ICH) on Good Clinical Practices (GCP), E6(R2) Section 4.9. Records and Reports stipulates: "The investigator/institution should maintain adequate and accurate source documents and trial records that include all pertinent observations on each of the site's trial subjects. Source data should be attributable, legible, contemporaneous, original, accurate, and complete (i.e., ALCOA+). Changes to source data should be traceable, should not obscure the original entry, and should be explained if necessary (e.g., via an audit trail)."[9]

As being ultimately accountable for the quality and reliability of the trial results, sponsors need to ensure that all trial contributors adhere to all regulations expectation including investigators. Beyond site monitoring ATR can enable compliance with ALCOA principles. But what does it take for our data to be attributable, legible, contemporaneous, original, and complete? It takes that "one thing" called an audit trail, right? Well, it is actually a bit more complicated than that as one "logical" audit trail dataset is typically stored as several different "physical" datasets. Thus, an audit trail includes but may not be limited to the following:

- **Clinical Design Metadata** – includes the base study design and its full change history. Study-specific clinical trial systems (e.g. EDC and eCOA) have an initial study specific design (or set-up, configuration) released prior to first patient being enrolled. This design often undergoes major and/or minor revisions until database lock which is often but not always resulting from protocol amendments.
- **Clinical Audit** – When Form-item values are modified, its attributes are modified and/or its design is modified. Many form-item values are updated more than once before the form is locked from Investigator change. Additionally, form-items values are enriched with attributes such as timestamps of significant reviews such as Source Data Verification (SDV),

[9] Guideline for good clinical practice E6(R2). (2015). https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5_en.pdf

investigator electronic signature and medical review. Lastly, the system pushes study design revisions to specific CRFs (e.g., incomplete CRFs, to site based on IRB approval dates, etc.). So, CRFs can have multiple values updated, can go through many types of review, and have multiple metadata revisions. As a result, the volume of clinical audit data may become the largest component of audit trail.

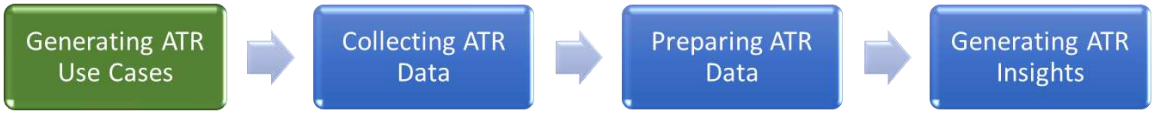
- **Clinical Query** – typically combined with clinical audit, every query raised on a form or form-item, its associated attributes and lifecycle.
- **Clinical Users Administration** – every clinical user, their role(s), and permissions to read/write/update clinical forms. This includes any modifications to users, their roles, and permissions.
- **Clinical Authentication Logs** – every time a clinical user attempts access to the system, regardless if their access was successful.
- **Clinical User Data Access Logs** – every time a clinical user is presented with a clinical form, regardless of whether they modify form values. This functionality may be available only for sensitive data such as unblinding or potentially unblinding data.

  Note: If user roles and access privileges within the eClinical system do not have functionalities to track and prevent access to sensitive data such as potentially unmasking data, then these logs should be reviewed to ensure compliance with data access restriction procedures.
- **System Event Logs** – every time a "non-human" event occurs including updates to the software and updates to the software configurations.
- **System Import Logs** – a record of every time a data set is imported into the system including trial specific data such as subject and site lists and global data such as country codes and drug descriptions.

There is no requirement that the audit trails must be constructed as exactly as the eight components listed above and most likely, audit trail for data collection systems used in a single study will vary dramatically. Some components may not even apply to specific data types such as sequenced data from sensors and wearables which are usually, not being organized by forms. Regardless of the source system, the relevant and adequate components must be present in the audit trail to ensure that data is attributable, accurate, legible, contemporaneous, original, and complete. Typically, several of these components must be joined to enable our ATR use cases. As an example, we need to join clinical audit and clinical user access logs to identify which specific site personnel are modifying specific forms with abnormally high frequency.

Note: It is critically important to consider blinding and privacy requirements when sharing audit trail from the above list as those could expose sensitive information. So, only the relevant information required to support ATR should be used from those.

c) Generating ATR Cases



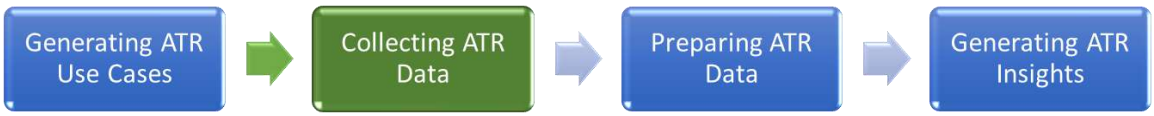
Technology facilitates the collection, preparation, and presentation of source data in order to support the efficient review of key data integrity risks. Understanding and defining ATR use cases is the first step to designing a solution that provides specific and actionable answers such as what, when, by who, and why data has been updated or removed.

The following guidelines summarize key system design principles to support an efficient and flexible approach to ATR:

- In general, a source data system (e.g., EDC, eCOA, IRT, etc.) that 1) contains all required audit trail data to support the ATR use cases and 2) provides the necessary tools to enable effective review and analyses should include the ATR solutions. This will not only facilitate a more pro-active review and detection of issues, but it can also reduce the effort of configuring and transferring complete audit trails into another system enabling ATR. Additionally, it will enable the relevant system users (e.g., audit trail reviewer) to address issues more simply at the source. Examples of Use Case scenarios include the detection of sites not accessing the source system routinely, or sites deleting and/or modifying more data than an established change threshold.
- Some ATR Use Cases require data from more than one source system to be combined. In such situations, each source data system should enable the transfer of its audit trail data to a centralized system or data warehouse to support the required analyses, usually at the sponsor- or CRO-level. While a source system enables some, but not all, of the desired monitoring capability for a given ATR use case, sometimes a centralized solution enables the additional desired monitoring. Where this is the case, it may be appropriate to leverage both systems, taking advantage of the timely reviews in the source system, such as real-time checks of incoming source entries, and complement the ATR with a more complete and holistic solution in the centralized system. Examples of Use Case scenarios more likely to be performed in a centralized system include the assessment of SAE reporting timeliness by site in EDC versus timeliness reporting to the sponsor Pharmacovigilence (PV) department, and the comparison of roles between the source systems and a Clinical Trial Management System (CTMS).

The following recommendations on collecting, preparing, and presenting audit trail apply to audit trail reporting in either a source and/or centralized system.

d) Collecting Audit Trail Data



Since the release of 21 CFR Part 11[10], it is well understood that source clinical data collection systems must be validated and maintain an audit trail for GCP transactions that change electronic records, including their creation, modification, and removal.

The recommendations below focus on recording and storing complete audit trails in systems such as EDC, eCOA, IRT and others, in line with existing regulations:

- The system should not delete or obscure previously recorded audit trail information
- The system should prevent user modifications to, or deletion of, the audit trail information
- The system should record complete audit trail records at the data item and form level including:
  - The creation, modification, or removal of a record, including original and new data value
  - Identification of the data element or field that was changed (e.g. record ID, field name, user account, etc.)
  - Date and time stamp records with consideration for time zones and daylight-saving changes
  - The identification (name or user ID) of the person who made the change
  - The reason for change to identify why data was changed
- The system should store the user access history including:
  - Addition/removal of users
  - Changes of roles and role permissions

It is worth noting that as the clinical data gathering process continues, the audit trail will grow proportional to the data volume.

To readily support comprehensive review during regulatory inspections or periodic ATR during the maintenance phase of the study, audit trails should be available in a live system in a human-readable format, as well as in a searchable export that allows integration into data repositories or the use of third-party visualization tools. In some cases, source systems including devices retain audit trails through audit log or abbreviated/coded summary which can hinder holistic ATR. In such a case, audit trail data should be transcribed or transformed into a human-readable format through a validated process.

Finally, upon decommissioning, a complete audit trail should be included as a part of the final study archives. These archives are typically provided as a PDF, which can represent electronic documents in a manner that preserves their static visual appearance over time. However, ATRs should not only be stored in a static PDF format, as the process requires a dynamic interrogation of data. In addition to a static format, audit trails should be added to final archives in XML, CSV, SAS or other ASCII text files formats that can be read with basic text

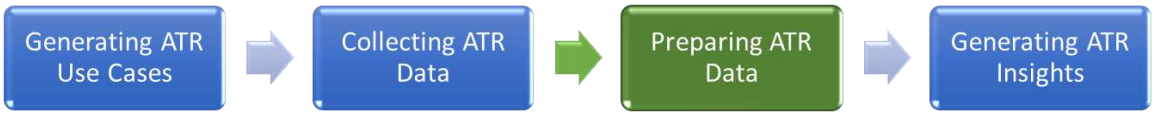
[10] Guidance for Industry Part 11, Electronic Records; Electronic Signatures — Scope and Application. (2003). https://www.fda.gov/regulatory-information/search-fda-guidance-documents/part-11-electronic-records-electronic-signatures-scope-and-application

editors, support the access requirements of the long retention period, and can be imported to other applications to support the dynamic analysis of the data. Moreover, the decommissioning process needs to prevent access to potentially unblinding data until its access does not represent a risk to the integrity of the clinical trial anymore.

There are situations that require different approaches to data reliability confirmation. For example:

- Data collected through sensors and wearables devices are typically not modifiable or removable, and audit trails from these "time sequenced" source systems typically only consist of a list of the times when new data was collected
- Some diagnostic equipment devices (e.g., ECG machines) may not contain audit trails or may not allow for their extraction if available

e) Preparing Audit Trail Data for Analysis



To help organizations meet the visualization requirements of ATR use cases, this next section will summarize the processing and preparation of audit trail data records into a structured format.

The authors of this paper recommend the following common format:

- Consistent variable/data field naming across all audit trail records
- Code list mapping: Map raw data values to interpretable data values (e.g., visit names vs. visit codes)
- Consistent treatment of numbers within their numerical value space, separate from numbers within a character value space. This will better support calculations and analysis
- Consistent date and time stamp formats, including the time zone used
- Consistent treatment of null (the absence of data) versus empty string (a string of zero length)
- An internally generated key (also known as a surrogate key) to facilitate the tracking of records that may require additional processing
- A familiar, recognizable identifier (also known as a natural key) to support effective filtering and analysis as well as traceability back to the original data source
- Show the type of the data change (e.g., addition, deletion, or update)

The source system data can be transferred to a centralized system in real-time, where possible via an Application Programming Interface (API), or as a batch data transfer. Because of the huge volume of data in source systems, we do not recommend the repeated batch transfer of cumulative data. Rather, the source system should support incremental batch transfer of data

between bookmarks, i.e., data that has been collected and/or modified since the last data request.

When data is being onboarded for the first time, the source system should support the full transfer of data from time 0 to the current time. Additional helpful features during the transfer process include:

- Configurable batch size
- Configurable retry frequency and retry delay
- Lossless compression technology

When transferring data from source to central repositories, it is highly desirable for the transfer to support data validation. Two such examples, core CDISC's ODM-XML and HL7's FHIR XML, are both vendor-neutral, platform-independent standard formats that provide optimized structure to transfer data from source to external systems, and centralized data warehouses.

Quoting CDISC.org:

> *"ODM-XML is a vendor-neutral, platform-independent format for exchanging and archiving clinical and translational research data, along with their associated metadata, administrative data, reference data, and audit information. ODM-XML facilitates the regulatory-compliant acquisition, archival and exchange of metadata and data. It has become the language of choice for representing case report form content in many electronic data capture (EDC) tools."*

Quoting HL7.org:

> *"FHIR offers Interoperability out-of-the-box: base resources can be used as is, but can also be adapted as needed - which happens a lot - for local requirements using Profiles, Extensions, Terminologies and more … Support for RESTful architectures, seamless exchange of information using messages or documents, and service-based architectures … A human-readable serialization format for ease of use … Ontology-based analysis with formal mapping for correctness."*

ODM-XML is a widely used in eCOA, EDC, IRT, eTMF and CTMS systems to transfer data with complete audit trails. Some ODM extensions may be vendor specific and therefore require data source specific adjustments. FHIR-XML is often utilized to exchange clinical data with health systems that have the ability to semantically enrich ODM data. Thus, FHIR-XML provides an "efficient and effective data exchange that ultimately will allow clinicians to switch their focus back to decision-making and evidence-based medicines"[11]. Most importantly, both standards include Schema Definitions (ODM's AuditRecord and FHIR's AuditEvent) which allow the receiving system to validate audit trail data before ingestion to a central repository.

[11] Hugo Leroux and Al, Towards achieving semantic interoperability of clinical study data with FHIR, Sep 2019. https://pubmed.ncbi.nlm.nih.gov/28927443/

Where web service APIs are supported in the source and receiving systems, both XML formats provide the transfer medium for the real-time data integration. Where they are not supported, both XML files can be delivered via secure file transfer protocols (such as sFTP or FTPS).

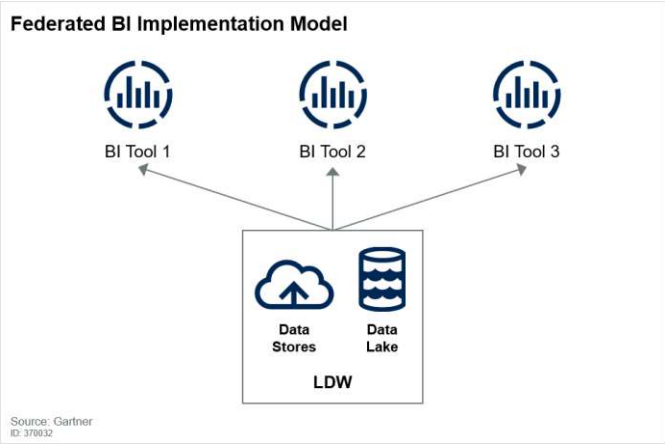
Receiving Data

The receiving and ingestion of audit data from multiple vendors is itself a complex task that requires a security-first, robust, high performing, agile infrastructure. Author recommendations include:

- Security planning to ensure the full transmission from vendor to consumer has been made as secure as possible
- Continuous monitoring of data arrival so that it can moved within managed service as timely as possible
- Automatic validation of data before it is ingested or pipelined into downstream processes
- An idempotent interface. This is a design principle that ensures any identical request has the same result as the first time it was presented, i.e., if we import a record, and then import that exact same record the following day, we expect the system to show one update and not two.
- Retention of raw data to facilitate reconstruction of all transformations from source system versus downstream processing tools

Federated BI Implementation Model

After bringing a copy the data from source data systems (i.e., Clinical Data Collectors) into sponsor's internal store for downstream processing such as ATR, we need to become custodians of our data in support of our mission to gain business insight. Becoming a custodian of our data is no small undertaking. As we prepare for this responsibility, one option is to establish what Gartner calls a Federated BI [Business Intelligence] Implementation Model[12] pictured below.



[12] Reference Architecture to Enable Real-Time Self-Service Analytics" Soyeb Barot, Daren Brabham, 9 July 2019. https://www.gartner.com/document/3947274 (Gartner subscription required)"

It starts with the premise that no one BI tool will satisfy all the complex needs of different business users executing widely different use cases. Even when we constrain the use cases to ATR, the needs of the analyst curating the data to verify authenticity can vary. As our needs and roles vary, no one BI tool will do the whole job. Rather we ideally need one model with a single Logical Data Warehouse (LDW) capable of feeding many BI tools.

With one single data store built from one architecture, this does not imply the LDW itself is a simple entity. Rather, it is a scalable and extensible infrastructure where different users may readily establish dedicated zones for their tasks. Using their dedicated zones, with appropriate data masking and security, users can start datasets from the most granular, atomic data of any recently transferred audit trail or build a new data set from existing snapshots. Users can operate on their datasets and curate them to resolve data quality issues that may be corrected via the application of rules and macros.

Our model must support the creation of a shared metadata library. Modern toolsets not only ingest data and place it in a secure store, they analyze it and look at its structure. For example, they use XML element names and CSV headers to better understand data. When we ingest audit trails from different systems, we will observe similar names and/or similarly shaped data values such that we recognize and differentiate information that is clearly a string from datetime. Data profiling tools can suggest the meaning of our data and write this into our metadata library. After automated analysis, we may find some suggested mappings where a human is still more powerful than algorithmic identification. And where we know better than the machine, we can play the role of data steward and override profiling recommendations.

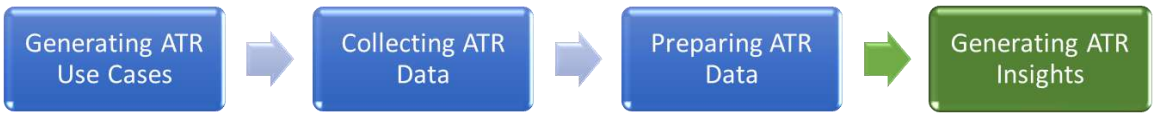
The model must support data governance processes and associated tooling to ensure any data that enters our system is treated with the same confidence as we have in our bank transactions. As we operate on data, we are investing our valuable time. The result datasets we create must be secure and well-maintained. The metadata we steward must be treated as an extension to our secure and well-maintained dataset. Users must trust that they can perform their data preparation with 100% confidence their time is well invested. Realizing that as much as 80% of our effort to gain any audit trails insight may be spent in data preparation, the leverage of data governance technology and keeping our trust are crucial to our success.

The model should support real-time or near-real-time data ingestion. Audit trails from today may be more valuable than audit trails from last year. This is not to imply that older data has little value, rather to affirm that many forms of analysis need relatively fresh data. This is especially true where data correction must happen at source and we must support rapid reingestion of corrected data. Fresh data means putting tools in the hands of decision makers, rather than expecting them to raise service requests and suffer the inevitable delays of non-self-service data.

Lastly, the model must support multiple visual interfaces during data preparation to:

- help make data easier to understand and enable data preparation processes
- help assess audit trail completeness
- help understand data consistency between one audit trail snapshot and the last
- help understand data uniqueness
- help understand data timeliness for data sets gathered at different points in time
- show where data has been joined and related integration issues

f) Presenting Audit Trail Data



Audit trail data should be presented in a manner that is easily understandable to a reviewer and enables the efficient identification of key data integrity risks. Simple-to-use interfaces that consider how users interact with data make audit trail reports actionable.

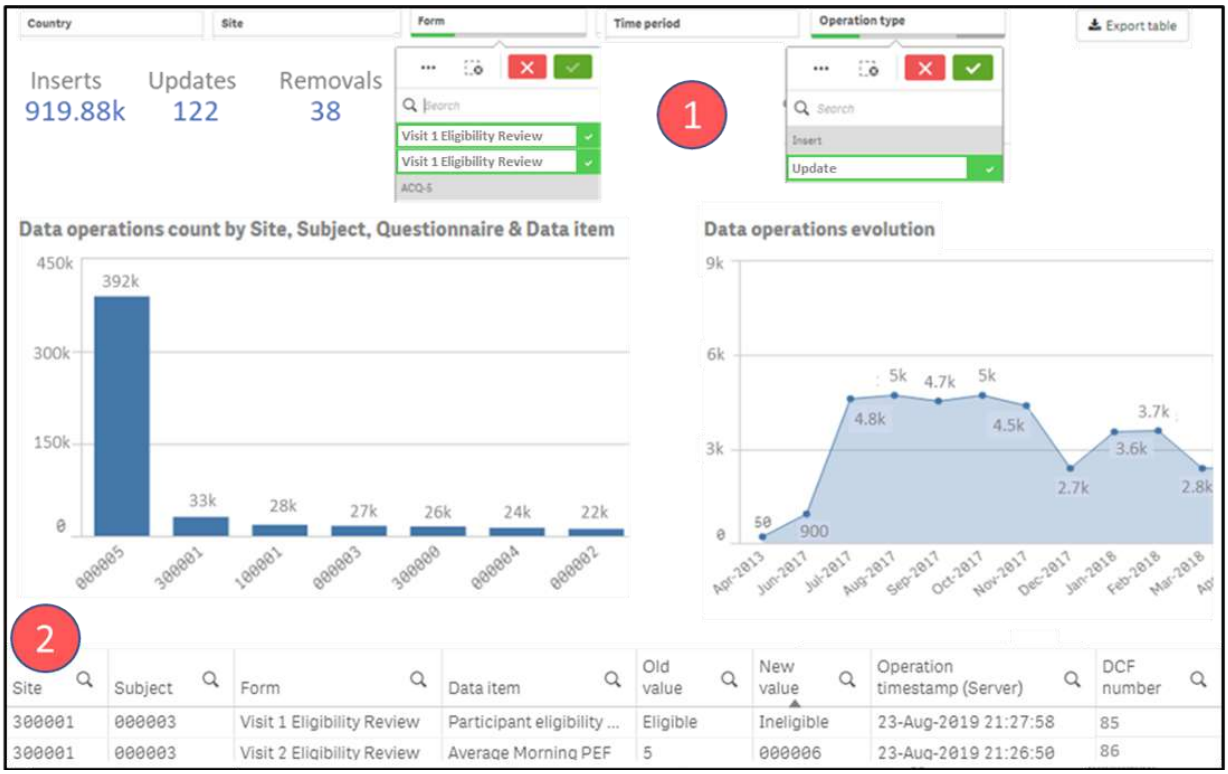
The recommendations below focus on presenting audit trail data in a format that supports the ability to reconstruct the chronology of "who, what, when, and why" for all data entries for the whole life of the record, and support the ATR use cases:

- Audit trails should clearly show the "before" and "after" values for data changes, as well as when data is soft-deleted (i.e., non permanently removed) or no longer visible
- Reports should facilitate dynamic review and interaction with data through built-in filtering and sorting options, and other functionalities
- Data visualizations should be used to translate high volumes of data into summary information that is more readily interpreted
- Analytics, including statistical methods, can be used to identify outliers and anomalies from large amount of data
- Reports should support the ability to export complete audit trails to widely used formats for data interrogation such as XML or .csv format
- Reports should support the configuration of normalized thresholds that can be used to push notifications highlighting key signals, for example when defined limits are exceeded (e.g., excessive changes, changes to critical data items)
- Data dictionaries with the detailed description of each data item/variable's name and other training materials that summarize how to use reports to review the most common ATR use cases can be used to help reviewers interpret the data
- Unblinding considerations and the reviewers' roles need to be taken into account when making audit trail reports available

Here are a few examples of how vendors and sponsors have used data aggregation, analytics, and visualization options to support efficient ATR across eCOA, EDC and IRT audit trails:

g) Presenting Audit Trail Data through Dashboards

Example 1: eCOA data entry and data change report

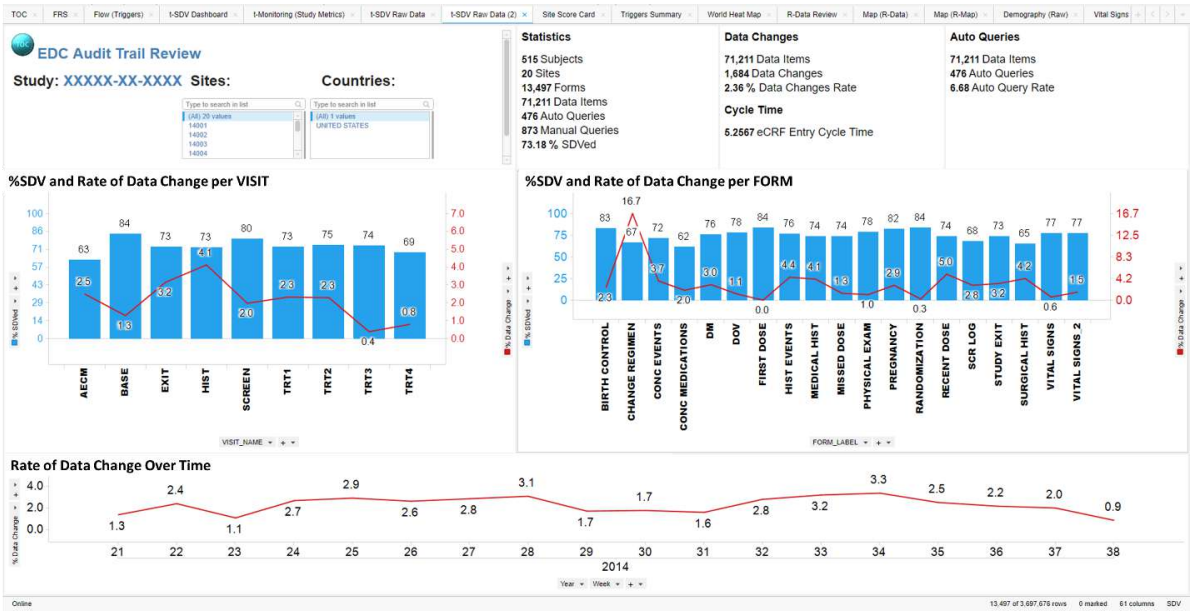


In the eCOA ATR dashboard example above data visualizations and built-in filtering options were used to highlight data changes on critical data (see Use Case Category 2 in appendix 3).

The graph on the left is used to show an abnormally high number of data updates or removals at one site compared to others (use case 2.4) and the graph on the right shows the timing of data changes at the monthly or daily level.

Filters can be used to dynamically analyze changes on critical data such as eligibility questionnaires that are used to support the eligibility scoring to determine the patient inclusion or exclusion in the study (use case 2.2). In the example above, the user selects the eligibility questionnaires and update operations in the step #1 and then reviews the details of the change in the data table below in the step #2.

Example 2: EDC data entry and data change report



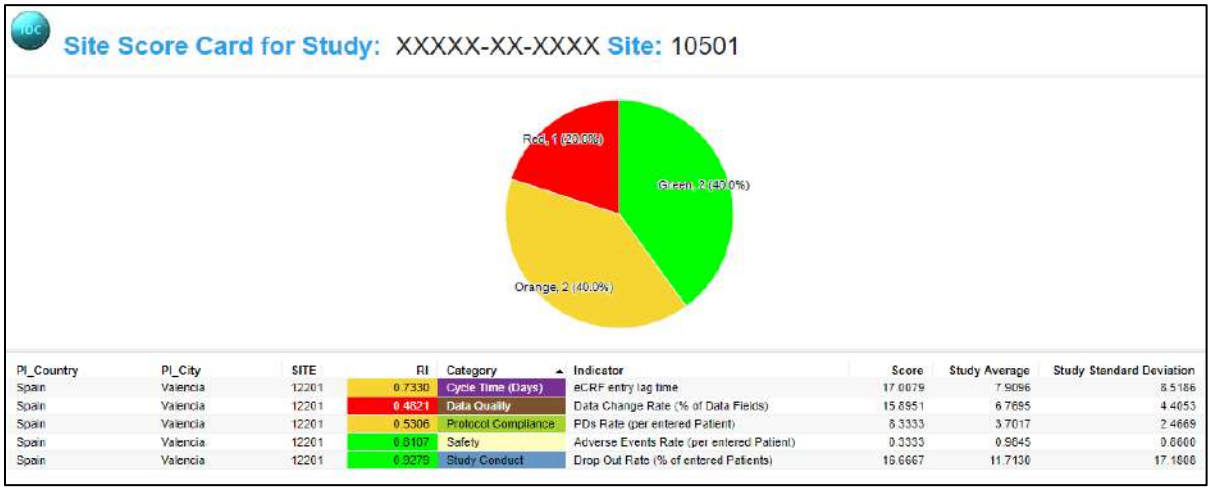
In this example, the audit trail data from EDC was integrated into a sponsor's central repository using incremental ODM-XML data transfer to assess data changes (see Use Case Category 2 in appendix 3). The audit trail data is collected and presented across clinical studies in the same way to standardize ATR.

This first visualization shows typical study statistics, such as number of subjects, sites and queries, as well as high level audit trail-based statistics, such as rate of data changes, data entry cycle time and percentage of source data verification, at the top. Bar graphs show the rate of data changes per visit and enable the identification of excessive changes (Use case 2.4) or excessive changes to critical forms (Use Case 2.3). The percentage of SDV, shown in the blue bar, is used as a reference to assess the level of on-site monitoring in relation to the rate of data change.

Lastly, the bottom graph shows the evolution of data changes over time. It supports the identification of episodic and non-systemic signals that would not result in exceeding an alert threshold for the overall study rate. As an example, the rate of data change may go up by the end of the study, but not result in the overall study average increasing enough to cause concern. The dashboard can be filtered by site and/or countries to allow more thorough investigations.

Like the first example for eCOA, this approach requires an active ongoing ATR to detect potential issues in the absence of threshold or alerting mechanism. To remediate this limitation, the same sponsor implemented a site score card strategy, based on standard deviation, to identify sites of interest (*see example 3*).

Example 3: Cross source site score card report



PI_Country	PI_City	SITE	RI	Category	Indicator	Score	Study Average	Study Standard Deviation
Spain	Valencia	12201	0.7330	Cycle Time (Days)	eCRF entry lag time	17.0079	7.9096	8.5186
Spain	Valencia	12201	0.4821	Data Quality	Data Change Rate (% of Data Fields)	15.8951	6.7685	4.4053
Spain	Valencia	12201	0.5306	Protocol Compliance	PDs Rate (per entered Patient)	6.3333	3.7017	2.4669
Spain	Valencia	12201	0.8107	Safety	Adverse Events Rate (per entered Patient)	0.3333	0.9645	0.6660
Spain	Valencia	12201	0.9279	Study Conduct	Drop Out Rate (% of entered Patients)	16.6667	11.7130	17.1868

This site score card extends the capabilities of the dashboard in example 2 by:

- Using additional data sources beyond audit trail, such as the protocol deviations and study data
- Using simple statistically based key risk indicators (KRIs) to identify and categorize site deviations from study mean:
  - Green - Site KRI with +/- 1 standard deviation from study mean
  - Amber - Site KRI with +/- 2 standard deviation from study mean
  - Red - Site KRI beyond 2 standard deviation from study mean
- Providing a more holistic view of site quality by looking at multiple categories (i.e., cycle time, data quality, safety, study conduct, and protocol compliance) which is a recommended best practice.
- Allowing the review of Site KRIs for **all use cases** in relation to other study KRIs

In this example, the dashboard has identified a signal that highlights a higher rate of data changes in the audit trail data. The potential seriousness of the signal is provided to the reviewer via the KRI score. The dashboard can then be used to drill down into the site's figures, by visits, forms, and trends over time, to investigate signal further.

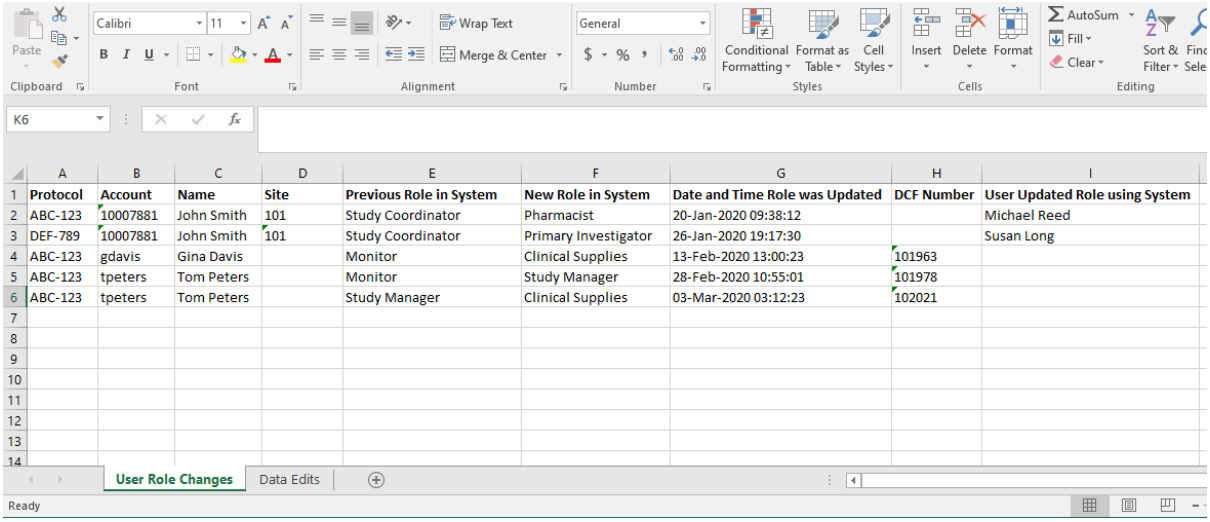
Additionally, this method could be used as a way to drive focus on potential signals on a smaller proportion of KRIs rather than reviewing all of them. Assuming a normal distribution, 68% of the KRIs would be within one standard deviation driving a targeted review of 32% of those outside that threshold. When using multiple KRIs, it would also be possible to prioritize sites with the most Red and/or Amber KRIs as potentially indicating a concerning trend.

Finally, the audit trail signal is contextualized alongside other categories that show the site in question also had signal for protocol deviations and cycle times, all pointing to site execution challenges. As stated earlier, ATR is but one of many tools to mitigate possible data integrity issues.

h) Presenting Audit Trail Data through exception reports

Exception reports, or simple extracts in an Excel or CSV file format, can be a good start in helping the reviewer to identify potential risks or data integrity issues, such as items related to system access concerns and changes to critical data (*Use Case Categories 1 and 2*)

Example 4: IRT User Role modification report

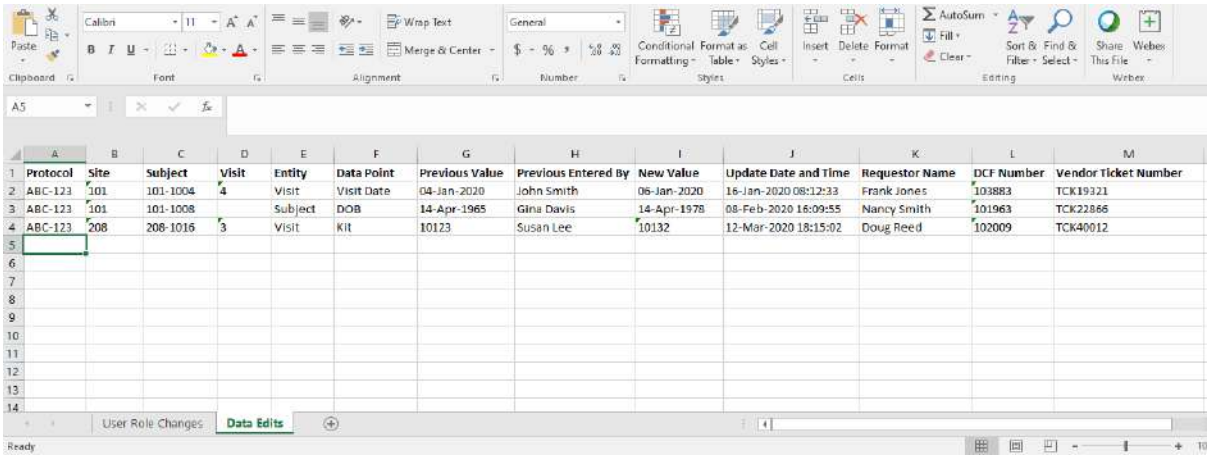


	A	B	C	D	E	F	G	H	I
1	Protocol	Account	Name	Site	Previous Role in System	New Role in System	Date and Time Role was Updated	DCF Number	User Updated Role using System
2	ABC-123	10007881	John Smith	101	Study Coordinator	Pharmacist	20-Jan-2020 09:38:12		Michael Reed
3	DEF-789	10007881	John Smith	101	Study Coordinator	Primary Investigator	26-Jan-2020 19:17:30		Susan Long
4	ABC-123	gdavis	Gina Davis		Monitor	Clinical Supplies	13-Feb-2020 13:00:23	101963	
5	ABC-123	tpeters	Tom Peters		Monitor	Study Manager	28-Feb-2020 10:55:01	101978	
6	ABC-123	tpeters	Tom Peters		Study Manager	Clinical Supplies	03-Mar-2020 03:12:23	102021	

This example shows an exception report that contains a list of all users in a system whose user roles changed at any time within a given study or protocol (Use case 1.1). This can aid in investigating concerns related to access and compliance by ensuring authorized users are performing the right updates or transactions in the system. There may be users who were inadvertently assigned to the incorrect role or user group. This method can also be used to review the blinding and unblinding roles within a sponsor organization, and appropriate access to system and data privileges.

Additional fields or data points can be added, such as account deactivation or termination date, or the date and time of the last transaction or login by the same user account. The report can also be extended to highlight single users being assigned to multiple roles.

Example 5: IRT Data Change Report



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Protocol	Site	Subject	Visit	Entity	Data Point	Previous Value	Previous Entered By	New Value	Update Date and Time	Requestor Name	DCF Number	Vendor Ticket Number
2	ABC-123	101	101-1004	4	Visit	Visit Date	04-Jan-2020	John Smith	06-Jan-2020	16-Jan-2020 08:12:33	Frank Jones	103883	TCK19321
3	ABC-123	101	101-1008		Subject	DOB	14-Apr-1965	Gina Davis	14-Apr-1978	08-Feb-2020 16:09:55	Nancy Smith	101963	TCK22866
4	ABC-123	208	208-1016	3	Visit	Kit	10123	Susan Lee	10132	12-Mar-2020 18:15:02	Doug Reed	102009	TCK40012

Another example is a report that displays the data edits that have occurred in the system for a given study or protocol (Use case category 2). Though some systems already have robust restrictions or validation rules to ensure the data is being collected according to the protocol, exception reports can look at processed or submitted manual data edits, such as those made by an IRT, lab, or eCOA vendor per an authorized manual data change request process. This is particularly useful when examining critical data points, such as visit dates and IP dispensing.

i) Applying KRIs And Risk Thresholds to detect likely issues

Reports and visualizations facilitate effective and efficient review of audit trails by translating high volumes of detailed data into summary information that is more readily understood and interpreted (KRI Methodology apply to all use cases).

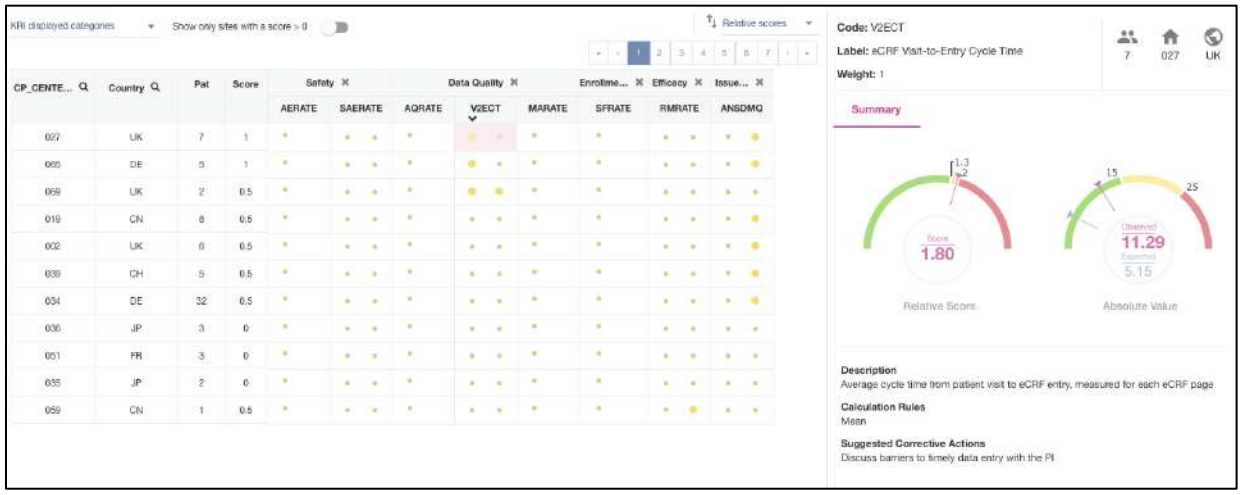
A further improvement to the review process involves the use of risk thresholds, which can rapidly alert reviewers to areas of concern in the audit trail data. For many of the audit trail use cases presented in this paper, KRI metrics, identified and derived from the relevant audit trail data, can enable timely issue identification. As mentioned above, a commonly used KRI metric is the average time from the date of patient's visit to entry of that data into the study EDC system, or the "Visit-to-eCRF Entry Cycle Time". It means that sites that exceed a pre-defined threshold can be flagged as "at risk", and in need of further review and possible intervention.

Two types of risk thresholds, "absolute" and "relative", can be considered for any given KRI. An absolute threshold refers to a discrete value, determined by the study team or organization, which represents a tolerance level beyond which the site, country, or study, is considered at-risk. Using "Visit-to-eCRF Entry Cycle Time" as the example, it is not uncommon for study teams to set an absolute risk threshold of 15 or 20 days. Sites found averaging more than this will be flagged as "at-risk" for this metric.

A relative threshold represents a statistical comparison of each site's observed KRI metric value to the overall study trend across all other sites in the same study. A probability score, representing the likelihood that the site's current difference from the study trend occurred by random chance, is computed from this statistical comparison. Example 3 used a simple statistical methodology (i.e., deviation from the mean), but a more advanced relative threshold (p-value) can be applied. It is commonly set to 0.05, representing a less than 5% likelihood that the observed difference happened by chance.

Note: For direct data capture systems such as eCOA entry is contemporaneous and as such as a pre-defined threshold may be less of value than the relative threshold.

Example 6: Relative and Absolute Risk Threshold for a KRI



Example 6 depicts the use of both a relative and absolute risk threshold for a KRI, and how the two thresholds provide alternate risk perspectives for a site on the "Visit-to-eCRF Entry Cycle Time" metric. The KRI dashboard on the left includes a row for each site in the study presenting current risk status across multiple KRIs. The KRI labelled "V2ECT" (Visit-to-eCRF Entry Cycle Time) is selected for the first site in the list, which displays two dots representing the risk statuses based on each threshold type – relative and absolute.

The relative threshold dot is amber, indicating an "elevated" risk level, while the absolute threshold dot is green to indicate a low risk level. The right-hand window in the picture displays more detail related to the selected KRI result.

The gauge on the right depicts the assessment of this site's average eCRF entry cycle time based on absolute thresholds, which have been set by the study team to 15 days (elevated risk level) and 25 days (high risk level). The site's current average cycle time is measured at 11.3 days, which is less than 15 days and therefore in a low-risk range (green). The gauge on the left depicts the assessment of this site's average cycle time based on the relative thresholds, which in this example are set to 1.3 (p-value = 0.05) for elevated risk and 2.0 (p-value = 0.01) for high risk. The site's current cycle time of 11.3 days is more than twice the current overall study average of 5.1 days, which results in a relative score of 1.8 (p-value ~ 0.02) which is in the elevated risk range since it is higher than 1.3.

This approach of using advanced KRIs, including the use of a dashboard to quickly identify risk situations across a number of audit-trail related KRIs, represents a very effective and efficient starting point for ATR. Visualizations and reports such as those presented here can work in tandem with a KRI dashboard, enabling a "drill-down" from this entry point to enable each risk alert to be further interpreted and characterized.

It also aligns very well with the broader risk-based approach to quality management (RBQM) our industry is rapidly adopting. Various audit trail review use-cases can be thought of as standard risks for which a set of KRIs can be defined and incorporated into an organization's full

standard KRI library. These could then be implemented along with all other KRIs into a single centralized monitoring solution (KRI dashboard, visualizations, etc.) as part of the overall data review process to enable a more comprehensive review of study risks in one place.

5.3 Vendor Considerations

Other considerations relating to ATRs is the use of technology provided by eClinical vendors in a clinical trial. eClinical vendors commonly provide EDC, eCOA/ePRO, and IRT systems or software as a service (SaaS). As a key stakeholder, eClinical vendors have an integral role to play in enabling and potentially driving the future of ATRs requirements beyond the use cases included in this paper.

But until ATRs become a routine activity enabled by most if not all eClinical systems, how can sponsors best ensure that the technology suppliers meet their ATR expectations? For instance, sponsors need to ask if the audit trail can ever be turned off and if yes, how is it detected and documented? Also, can any changes be made outside of the audit trail, for example by an IT Administrator of the system? If yes, under what SOP, what authority, where is it documented and who is involved (e.g., a Quality organization representative has to sign off on the change control documentation, a secondary IT organization representative has to verify the change made and that only that change was made, etc.)? These key fundamentals of the execution of the audit trail should be documented.

The ATR conversation should be initiated during vendor selection, with the sponsor expectations being described in the original Request for Information (RFI) or the Request for Proposal (RFP).

During the qualification process, sponsors should assess the vendor's level of knowledge and capability, as this could vary considerably. The considerations that could be discussed with a vendor are:

a) Format of the audit trail

Details around the format of a particular audit trail impact how ATR is performed in a clinical trial. Beyond ATR in the scope of this paper, eClinical systems need to meet other stakeholders' needs including investigational sites:

- Can sponsors and clinical sites access the audit trail in the system in real time during the study, or is it supplied as a report or data dump?
- Are dashboards and other visualization tools utilized?
- How consumable (i.e., human readable) is the audit trail, and does it make sense (i.e., easy to interpret)?
- Is the audit trail easily accessible by the Principal Investigator? At the form/raw data level? In aggregate?
- Do Principal Investigators have the ability to review access rosters at their sites as part of their oversight?

Flat files should be in searchable format like .CSV. Audit trails supplied as a 'data dump' may be acceptable as they could be loaded into a sponsor visualization tool.

The source system data should also be transferrable to a centralized system in real-time, where possible via an API, or as a batch data transfer to enable time sensitive use cases.

b) Level of Detail in the Audit Trail and Types of Audit Trails

Is the level of detail at the item or form level? Are audit logs (one of the types of audit trails) also available and maintained, for example, to demonstrate user access? How long are audit logs retained? Audit logs typically provide information about user activities, such as dates and times of access, and in some cases, what sensitive data, modules, applications the user has accessed while logged on. It is key that logs are maintained for sufficient time to be able to demonstrate control of the system. For example, if inappropriate access was given to a user, audit logs can demonstrate if a user had been unblinded (e.g., a user such as an independent blinded assessor.) A question to ask is when the audit trail is initiated on the device or in the system.

Note: Note: Sponsors should consider the impact of real-time feedback to site and/or patients that may influence the response prior to the data being saved (e.g. feedback on inclusion/exclusion criteria, protocol deviations, etc.). The edit check specification should define browser-side, server-side and offline edit checks such that real-time feedback is only allowed when acceptable.

c) Frequency of Review

When some of the ATR use cases are delegated to vendors by the sponsor, how often should the audit trail be reviewed by the vendor? What is their process? What is the scope of the ATR service they offer? A risk-based approach that considers the use of the data from the source systems, and its relation to primary and secondary endpoints and patient safety, should be considered.

d) Roles and Responsibilities

Whenever possible, ATR should occur at the source (i.e., in the data capture system). There are several questions to consider in defining and setting up ATR. What are the vendor's delegated responsibilities and what is the sponsor's accountability? Who is best positioned to perform the ATR? Could and should the vendor check for unusual audit trail's data patterns? If so, does the process meet sponsor expectations or can it be defined by the sponsor upfront? Is the process for audit trail review manual, semi-automated or fully automated? What is the communication plan to sites and sponsors, and how will any issues be resolved? And finally, how will reviews be documented?

Ongoing ATRs may result in additional reviews and checks of that audit trail by the sponsor and/or vendor, depending on the results of their review.

e) Inspections

In the case of inspection, which typically happens after a study is closed, how will inspectors access the audit trail? Will it be possible to access the audit trail in the system, and are the audit trails dynamic, i.e., active and searchable, within the online system?[13] Is a searchable export of the audit trail in .csv available? Has the system audit trail functionality been validated as a part of system validation? Sponsors should take a risk-based approach to inspection readiness.

A sponsor would then need to determine how robustly the vendor's capabilities meet their ATR expectations from a system and services point of view. This can be done by obtaining input from multiple stakeholders such as Quality Assurance, Data Management, and IT, or through a vendor audit. Any gaps should be discussed with the vendor, and mitigations put in place.

If sponsor expectations regarding inspection support cannot be met, a risk-based decision concerning the use of the eClinical vendor's technology must be made. The purpose of the data being collected and if it is primary or secondary endpoint data should be considered along with how the data is being collected. For example, is it a device that simply transmits data and does not allow for data changes?

Current vendors can enable and in some cases facilitate the process of ATR by following the Use Cases and best practice recommendations described in the paper. Note, all audit trails provided should meet the expectations laid out in the paper.

5.4) Risks and Limitations to Adequately Conducting a Routine Audit Trail Review

Of all the data that exists today, it is estimated that at least 90% has been generated in the last few years.[14] Clinical research data generation is increasing exponentially within existing modalities (e.g., EDC, eCOA, IRT) and with new approaches that automatically generate immense volumes of data by the second (e.g., wearables, sensors). In conjunction, the complexity of efficiently and effectively reviewing audit trail data increase significantly. While in some cases, only aggregated data would be considered critical, identifying risks and issues within this massive data volume may still be beneficial to reveal atypical data patterns. The successful execution and alignment of people, processes and technology is absolutely essential.

In this paper, we have discussed extensively the need for a risk-based approach to routine ATR. However, the industry as a whole has been slow to adopt such methodologies in other domains, such as risk-based monitoring. Among many challenges, a recent industry meeting highlighted the lack of clear regulatory guidance as one of key barriers to the implementation of risk-based approach to monitoring.[15] Similarly, a clear regulatory guidance and direction is needed for risk-based ATR.

---

[13] SCDM The Evolution of Clinical Data Management to Clinical Data Science – Part 2 (2020)

[14] Bernard Marr, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read," *Forbes*, May 21, 2018, Forbes.com

[15] "Improving the Implementation of Risk-Based Monitoring Approaches of Clinical Investigations," https://www.fda.gov/drugs/news-events-human-drugs/improving-implementation-risk-based-monitoring-approaches-clinical-investigations, *The U.S. Food and Drug Administration (FDA)*, July 17, 2019

Establishing robust change management and execution strategies are also important. For example, a key consideration is that appropriate audit trail review guidance is included in data governance policies or equivalent at both the highest and lowest levels of the organization.[16]

For general considerations with regards to the ability and inability to detect and monitor data integrity via ATR, we have specified several risks, limitations, and mitigations in Appendix 4.

6.0 - Conclusion

Data integrity is paramount to clinical research, and ensuring data is reliable and credible demonstrates our commitment to patients and the trust they place with us. The industry strives to create a culture of data integrity through policies, procedures, responsibilities, and governance.

This paper is intended as the first step in an evolving journey to maximize the use of routine ATR as one aspect of ensuring data integrity. In writing this document, we mapped out areas of risk to data integrity along the data lifecycle and identified 5 key Use Case Categories containing 20 Use Cases for ATR where data may be incomplete, inconsistent, or inaccurate, or where access controls are lacking. Each trial or vendor system can identify others by evaluating their risks, based on data integrity principles. To effectively execute the Use Cases set out here, the authors suggest a common format for the very large audit trail datasets, coupled with visualizations and exception reporting. These critical tools are based on statistical models to inform thresholds for identifying actionable trends and outliers.

Tools alone cannot improve data integrity. It takes an understanding of the data domain and good data science judgment combined with thorough data content reviews, and most importantly, cross-functional and cross-partner collaboration.

Within sponsor companies, clinical data science, clinical study management, quality and IT functions must work together to identify and correct issues based on root causes with a focus on prevention. Vendor companies should incorporate audit trail checks natively into their systems wherever possible, and partner with sponsor companies on integrations into data warehouses where ATRs can be performed across systems.

ATR is just one data integrity tool, but it is an important one. Increased dialog across stakeholders – regulators, sites, eClinical vendors, sponsors and CROs – is needed to embrace its potential.

We invite you to continue the dialogue on ATR with the SCDM Innovation Committee and the eCF by participating in their respective working groups, webinars and events.

Together, we can improve on the use of ATR as a key tool in the quest to achieve data integrity.

[16] MHRA 'GXP' Data Integrity Guidance and Definitions. (2018). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/687246/MHRA_GxP_data_integrity_guide_March_edited_Final.pdf

7.0 - Appendices

Appendix 1: Definitions and Acronyms

AE	Adverse Event
ALCOA+	attributable, legible, contemporaneous, original and accurate; += complete, consistent, enduring and available
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
AT	audit trail
ATR	audit trail review
BI	Business Intelligence
CDISC	Clinical Data Interchange Standards Consortium
CFR	Code of Federal Regulations
cGMP	current Good Manufacturing Practice
CRF	Case Report Form
CRO	contract research organization
CSV	Comma Separated Values
CTMS	Clinical Trial Management System
DCF	Data Clarification Form
DCR	Data Clarification Request
eCF	eClinical Forum
eCOA	electronic Clinical Outcome Assessment
eCRF	electronic Case Report Form
EDC	electronic data capture
EMA	European Medicines Agency
ePRO	electronic Patient Reported Outcome
eTMF	electronic Trial Master File
FDA	Food and Drug Administration
FHIR	Fast Healthcare Interoperability Resources
FTPS	extension of File Transfer Protocol
GCP	Good Clinical Practice
GDP	Good Distribution Practices
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act of 1996
HL7	Health Level 7 International
I/E	Inclusion/Exclusion
ICH	International Council for Harmonisation
ICH E6 R2	ICH E6 (R1) integrated addendum

ID	identification
IRB	Institutional Review Board
IRT	interactive response technologies
IT	Information Technology
KRI	Key Risk Indicator
LDW	Logical Data Warehouse
MHRA	Medicines and Healthcare products Regulatory Agency
ODM	Operational Data Model
PDF/A	Portable Format Document
PI	Principle Investigator
PIC/S	Pharmaceutical Inspection Co-operation Scheme
PMDA	Pharmaceuticals and Medical Devices Agency
PV	Pharmacovigilance
QbD	Quality by Design
QTL	Quality Tolerance Limit
RACT	Risk Assessment and Categorization Tool
RBM	Risk Based Monitoring
RBQM	Risk Based Quality Management
RFI	Request for Information
RFP	Request for Proposal
SAE	Serious Adverse Event
SAS	Statistical Analysis System
SCDM	Society for Clinical Data Management
SD	Standard deviation
SDV	Source Data Verification
sFTP	secure File Transfer Protocol
UTC	Universal Time Coordinated
XML	Extensive Markup Language

Strategy Document Definitions (from MHRA, FDA and CDISC)

Data integrity is the degree to which data are complete, consistent, accurate, trustworthy, and reliable, and that these characteristics of the data are maintained throughout the (data) life cycle. The data should be collected and maintained in a secure manner, so that they are attributable, legible, contemporaneously recorded, original (or a true copy) and accurate. Assuring data integrity requires the appropriate quality and risk management systems, including adherence to sound scientific principles and good documentation practices. *MHRA 'GXP' Data Integrity Guidance and Definitions, March 2018*

For the purposes of this [FDA Data Integrity for CGMP, Q&A December 2018] guidance, **data integrity** refers to the completeness, consistency, and accuracy of data. Complete, consistent, and accurate data should be attributable, legible, contemporaneously recorded, original or a true copy, and accurate (ALCOA).

The **data lifecycle** is defined as all phases in the life of the data, including raw data, from initial generation and recording, through processing, including transformation or migration, use, data retention, archive/retrieval and destruction. *MHRA 'GXP' Data Integrity Guidance and Definitions, March 2018*

The **audit trail** is a form of metadata containing information associated with actions that relate to the creation, modification, or deletion of GXP records. An audit trail provides for the secure recording of life-cycle details such as creation, additions, deletions or alterations of information in a record, either paper or electronic, without obscuring or overwriting the original record. An audit trail facilitates the reconstruction of the history of such events relating to the record regardless of its medium, including the "who, what, when and why" of the action. *MHRA 'GXP' Data Integrity Guidance and Definitions, March 2018*

For purposes of this [FDA Data Integrity for CGMP, Q&A December 2018] guidance, **audit trail** means a secure, computer-generated, time-stamped electronic record that allows for reconstruction of the course of events relating to the creation, modification, or deletion of an electronic record.

From DRAFT FDA Data integrity for CGMP April 2016, an **audit trail** is a chronology of the "who, what, when, and why" of a record. Electronic audit trails include those that track creation, modification, or deletion of data (such as processing parameters and results) and those that track actions at the record or system level (such as attempts to access the system or rename or delete a file). CGMP-compliant record-keeping practices prevent data from being lost or obscured (see §§ 211.160(a), 211.194, and 212.110(b)). Electronic record-keeping systems, which include audit trails, can fulfill these CGMP requirements.

Data quality is the assurance that data produced is exactly what was intended to be produced and fit for its intended purpose. This incorporates ALCOA. *MHRA 'GXP' Data Integrity Guidance and Definitions, March 2018*

There are multiple other related terms defined in the MHRA 'GXP' Data Integrity Guidance and Definitions, March 2018. Here is the complete list:

Data; raw data (source data); metadata; data integrity; data governance; data lifecycle; record and collection of data; data transfer/migration; data processing; excluded data; original record and true copy; computerized system transaction; audit trail; electronic signatures; data review and approval;

computer system user access /system administration roles; data retention (archive, backup); file structure; validation for intended purpose; and IT suppliers and service providers.

December 2018

Please clarify the following terms as they relate to CGMP records:

a. What is "data integrity"?
For the purposes of this guidance, *data integrity* refers to the completeness, consistency, and accuracy of data. Complete, consistent, and accurate data should be attributable, legible, contemporaneously recorded, original or a true copy, and accurate (ALCOA).5
5 These characteristics are important to ensuring data integrity and are addressed throughout the CGMP regulations for drugs. For *attributable*, see §§ 211.101(d), 211.122, 211.186, 211.188(b)(11), and 212.50(c)(10); for *legible*, see §§ 211.180(e) and 212.110(b); for *contemporaneously recorded* (at the time of performance), see §§ 211.100(b) and 211.160(a); for *original or a true copy*, see §§ 211.180 and 211.194(a); and for *accurate*, see §§ 211.22(a), 211.68, 211.188, and 212.60(g).
6 For examples of record retention periods, see §§ 211.180 and 212.110(c).
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's retention period ends.6 System design and controls should enable easy detection of errors, omissions, and aberrant results throughout the data's life cycle.

b. What is "metadata"?
Metadata is often described as data about data because is the contextual information required to understand data. A data value is by itself meaningless without it. It is structured information that describes, explains, or otherwise makes it easier to retrieve, use, or manage data. For example, the number "23" is meaningless without metadata, such as an indication of the unit "mg." Among other things, metadata for a particular piece of data could include a date/time stamp documenting when the data were acquired, a user ID of the person who conducted the test or analysis that generated the data, the instrument ID used to acquire the data, material status data, the material identification number, and audit trails.
Data should be maintained throughout the record's retention period with all associated metadata required to reconstruct the CGMP activity (e.g., §§ 211.188 and 211.194). The relationships between data and their metadata should be preserved in a secure and traceable manner.

c. What is an "audit trail"?
For purposes of this guidance, *audit trail* means a secure, computer-generated, time-stamped electronic record that allows for reconstruction of the course of events relating to the creation, modification, or deletion of an electronic record. For example, the audit trail for a high-performance liquid chromatography (HPLC) run should include the username, date/time of the run, the integration parameters used, and details of a reprocessing, if any. Documentation should include change justification for the reprocessing. Audit trails include those that track creation, modification, or deletion of data (such as processing parameters and results) and those that track actions at the record or system level (such as attempts to access the system or rename or delete a file).
CGMP-compliant record-keeping practices prevent data from being lost or obscured and ensure that activities are documented at the time of performance (see §§ 211.68, 211.100, 211.160(a), 211.188, and 211.194). Electronic record-keeping systems, which include audit trails, can support these CGMP requirements.

This guidance also has additional definitions on systems etc and sections on who should review audit trails, etc.

CDISC definition from CDISC Glossary v14.0 of **Data Integrity**

A condition of data reflecting the degree to which the data are complete, consistent, accurate, trustworthy, and reliable at any given time as well as consistently so maintained throughout the data life cycle. NOTE: The data should be collected and maintained in a secure manner, so that they are Attributable, Legible, Contemporaneously recorded, Original (or a true copy) and Accurate (ALCOA). Assuring data integrity requires appropriate quality and risk management systems, including adherence to sound scientific principles and good documentation practices. (After MHRA Guidance on "GxP data integrity") See also ALCOA, ALCOA+, traceability (data). Compare to data quality.

CDISC definition from CDISC Glossary v14.0 of **Data Quality**

A dimension of data contributing its trustworthiness and pertaining to accuracy, sensitivity, validity, and suitability to purpose. Key elements of data quality include attribution, legibility (decipherable, unambiguous), contemporaneousness, originality (i.e., not duplicated), accuracy, precision, completeness, consistency (logical, not out of range), and those who have modified the data. NOTE: Scientists may reasonably trust data that are accurate (high quality) that have also been reviewed by investigators and protected from unauthorized alteration (high integrity). See also ALCOA, data integrity.

Appendix 2: The MHRA's 10 Principles for Data Integrity [17]

Principle 1: The organization needs to take responsibility for the systems used and the data they generate. The organizational culture should ensure data is complete, consistent and accurate in all its forms, i.e. paper and electronic.

Principle 2: Arrangements within an organization with respect to people, systems and facilities should be designed, operated and, where appropriate, adapted to support a suitable working environment, i.e. creating the right environment to enable data integrity controls to be effective.

Principle 3: The impact of organizational culture, the behavior driven by performance indicators, objectives and senior management behavior on the success of data governance measures should not be underestimated. The data governance policy (or equivalent) should be endorsed at the highest levels of the organization.

Principal 4: Organizations are expected to implement, design and operate a documented system that provides an acceptable state of control based on the data integrity risk with supporting rationale. An example of a suitable approach is to perform a data integrity risk assessment (DIRA) where the processes that produce data or where data is obtained are mapped out and each of the formats and their controls are identified and the data criticality and inherent risks documented.

Principle 5: Organizations are not expected to implement a forensic approach to data checking on a routine basis. Systems should maintain appropriate levels of control whilst wider data governance measures should ensure that periodic audits can detect opportunities for data integrity failures within the organization's systems.

Principle 6: The effort and resource applied to assure the integrity of the data should be commensurate with the risk and impact of a data integrity failure to the patient or environment. Collectively these arrangements fulfil the concept of data governance.

P**rinciple 7:** Organizations should be aware that reverting from automated or computerized systems to paper-based manual systems or vice-versa will not in itself remove the need for appropriate data integrity controls.

Principle 8: Where data integrity weaknesses are identified, companies should ensure that appropriate corrective and preventive actions are implemented across all relevant activities and systems and not in isolation.

Principle 9: Appropriate notification to regulatory authorities should be made where significant data integrity incidents have been identified.

Principle 10: The guidance refers to the acronym ALCOA rather than 'ALCOA +'. ALCOA being Attributable, Legible, Contemporaneous, Original, and Accurate and the '+' referring to Complete, Consistent, Enduring, and Available. ALCOA was historically regarded as defining the attributes of data quality that are suitable for regulatory purposes. The '+' has been subsequently added to emphasize the requirements. There is no difference in expectations regardless of which acronym is used since data governance measures should ensure that data is complete, consistent, enduring and available throughout the data lifecycle.

[17] MHRA 'GXP' Data Integrity Guidance and Definitions. (2018). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/687246/MHRA_GxP_data_integrity_guide_March_edited_Final.pdf

Appendix 3: Use Cases and Risk Scenarios

Below is the list of use cases compiled by eCF and SCDM Members which provide robust and concrete examples for consideration but are not intended to be complete. Each study team needs to consider the relevant risks and the specificity of the data captures systems being used.

Use Cases: Data integrity risk	Risk scenarios where audit trail is primary tool	Applicable system(s)	Source(s) needed	Ability to create AT report?	Examples of desired reporting criteria
Use Case Category 1: Access concerns					
1.1: Unauthorized user-access/ Changes by right role, right person	Assigned to wrong user group (e.g., site vs sponsor vs CRO) OR Site/Sponsor user assigned to wrong site	EDC, IRT, eCOA	CTMS and user access list to the source system(s)	Sponsor	Cross-reference audit trail for unauthorized users with the sponsor/CROs CTMS listing and site users to verify roles and site assignments
	Deactivations not performed in requisite time following notification of site staff changes				Compare the date of employment termination or reassignment with date of last access after termination date
	User access data entry rights are wrong (e.g. pharmacovigilance group assigned to write access roles)				Exception report of users, their roles, and their access date and any changes to access (initial entry and changes)
	User roles changed inappropriately during study conduct				Exception report of users, their roles, and their access date and any changes to access (initial entry and changes)
1.2: Access without training	PI completed scales prior to rater training	EDC, eCOA	Training database for sites Rater or patient training	Sponsor	Exception report – cross-reference who had access list, rater training date list, and date of first completed scales

Use Cases: Data integrity risk	Risk scenarios where audit trail is primary tool	Applicable system(s)	Source(s) needed	Ability to create AT report?	Examples of desired reporting criteria
	Sponsors or sites getting system access without training	EDC, IRT, eCOA, vendor database when DCFs are created	Training database for sites		
	Patient did not receive tool training		Rater or patient training		
	Access to Sponsor roles				
1.3: User login and activity	Sites not accessing database routinely as per protocol requirements or study expectations and/or users never logging in	EDC, IRT, eCOA	logins in source system	Vendor	Compare last access to the web portal to the subject's final questionnaire of the visit (patient entered or clinician entered) or other protocol defined or training defined requirement (e.g., review diary data in portal prior to visit,etc.) and look for no activity by users
	PI or study coordinator not entering data or reviewing data in a timely manner	EDC, IRT, eCOA	logins in source system: * if eCOA relates to portal views ** if EDC needs AT to track views, or *** if IRT needs AT to track views	Vendor	* Overall eCOA login compliance: express as a percentage the daily logins, with details of who accessed each day and when, then compare to site's patient activity ** Cycle time for transcription activity into EDC by visit/form *** IRT: login activity outside expected business hours
1.4: System login and activity	Super user role activity is higher than expected	EDC, IRT, eCOA		Vendor	Exception listing report for number of super users and listing of activity

Use Cases: Data integrity risk	Risk scenarios where audit trail is primary tool	Applicable system(s)	Source(s) needed	Ability to create AT report?	Examples of desired reporting criteria
	System integration and how it's recorded		audit trail in source system(s)		Review system logs for data movement and any integration failures (what, when, why)
1.5: Lack of PI access at site	PI cannot review or sign-off data or cannot unblind in an emergency situation(e.g., due to PI was never granted access by vendor, PI access was revoked due to inactivity, or PI never took required training)	EDC, IRT, eCOA	audit trail in source system(s)	Vendor	Report or dashboard to identify a minimum number of PI's at each site w/ access (combine with access and patient activity reviews above)
Use Case Category 2: Changes					
2.1: Incomplete data- data deleted	Data item deleted without query or explanation (e.g., removed AEs impacting analysis and label	EDC, IRT, eCOA	audit trail in source system(s)	Vendor	Report (listing, dashboard) for inactivation or deletion at field-, form-, record-, or patient-level
	Mass deletes or changes/updates by system non-user (e.g., IT admin super user role)				
	Frequent deletions within a subject record or by a user				
	Record- or patient level-deleted data				
2.2: Changes to inclusion/exclusion (I/E) criteria, primary efficacy, key secondary	Any changes to I/E criteria or eligibility scoring by authorized user could impact patient safety by enrolling ineligible patient	EDC, IRT, eCOA	audit trail in source system(s)	Vendor	Interactive dashboard or listing of all changes to any records used in I/E criteria, eligibility scoring or primary efficacy, key secondary efficacy, safety or critical data.

Use Cases: Data integrity risk	Risk scenarios where audit trail is primary tool	Applicable system(s)	Source(s) needed	Ability to create AT report?	Examples of desired reporting criteria
data, safety data and/or other critical data as defined by risk-based monitoring	Any changes to primary efficacy, key secondary, safety or other critical data could impact analysis				
2.3: Inconsistent data- critical data changes after key timepoints such as data locks or subject's study disposition status is marked as complete	Unexpected changes to critical data since lock, i.e., changes to efficacy forms, safety forms	EDC, eCOA IRT (if using to collect CRF-like data)	audit trail in source system(s)	Vendor	Compare pre- and post-lock incremental audit trail snapshots within reports, and listings of all post-lock critical data changes or changes after subject disposition is marked as complete
	Changes not reported in DSURs				
2.4: Excessive changes of critical data	Multiple changes of critical data on a given form	EDC, IRT, eCOA	audit trail in source system(s)	Vendor	eCOA: Compare each vendor's ATR to changes requested by sites (e.g., DCR reporting integrated with ATR) and look for changes made by vendor outside of DCRs; additionally look for volume of DCR changes per site for trends EDC: Calculate average changes per field and standard deviations (SD) (suggest 3 SDs to start) to determine 'excessive' threshold This report will provide excessive changes outliers to enable trend review, and filter out background noise of normal EDC change behavior
	High number of data change at one site compared to others				
	Multiple changes in patient reported data as requested by site				

Use Cases: Data integrity risk	Risk scenarios where audit trail is primary tool	Applicable system(s)	Source(s) needed	Ability to create AT report?	Examples of desired reporting criteria
2.5: Timing of changes after initial data entry or source is obtained is longer than expected	Timing of change to original value made after original entry is longer than changes made by other sites to same field	EDC, eCOA, IRT	audit trail in source system(s)	Vendor	Establish threshold for timing of changes and look for sites making changes above that threshold
Use Case Category 3: Data collection concerns					
3.1: Data not collected per protocol timing	Questionnaires/Scales to be conducted prior to any protocol procedures	EDC, eCOA, IRT (if using to collect CRF-like data)	audit trail in source system(s)	Vendor, Sponsor	Compare start time of procedures (in EDC or eCOA) and end date time of completion of questionnaires (eCOA)
3.2: Electronic Data not collected contemporaneous to event	Site using paper backups and using eCOA system as transcription tool	eCOA (using EDC data as additional reference source) IRT (if using to collect CRF-like data)	audit trail in source system(s)	Vendor, Sponsor	Compare procedure start time and completion of questionnaires end date/time (*see duration use case 4 - below*). Very short data entry times may indicate a paper questionnaire being entered retrospectively --- Compare EDC visit dates with eCOA questionnaire dates
	Data changed after established allowable recall window				
	Site or patient is entering data retrospectively (i.e., data completion or entry date vs actual collection date/time has a gap)				

Use Cases: Data integrity risk	Risk scenarios where audit trail is primary tool	Applicable system(s)	Source(s) needed	Ability to create AT report?	Examples of desired reporting criteria
3.3: Data collected at suspicious timing	implausible entry date/time stamp for time zone of the site or patient	eCOA, EDC, IRT	audit trail in source system(s)	Vendor	AT visualization to identify patterns of data entry times within / across sites --- AT report listing includes date and time in the user's local time, as well as days of the week.
3.4: Varying completed durations	Inaccurate data: too long (aggregate at patient, site, or study level) or too short to complete (at patient, site or study level) Site performance: site has statistically different duration times from other sites or site reports long durations not supported by ATR	eCOA, IRT (if using to collect CRF-like data)	audit trail in source system(s)	Vendor	Identify and flag high/low outliers by comparing the calculated duration (between start and end time of data entry) for each questionnaire to protocol expected duration and/or comparing to average completion times and then calculate standard deviations (SD) (suggest 1 SD to start) for threshold of low and high duration times
3.5: Reporting time is outside of required timing	SAE report to site vs report to sponsor is outside of required window	EDC	PV data and EDC audit trail	Sponsor	Using report exception listing of date/time stamps, identify if any unreported, late data entry vs SAE occurred, and when site was aware of SAE
	Data entry date vs data collection dates are too far apart (not meeting established entry criteria for timeliness)	EDC	EDC and other sources (at site, IRT, labs etc)	Sponsor	Using listing report for data entry date/time stamps and compare to other sources to determine timeliness
3.6: Missing data	PI signature missing	EDC, eCOA (as applicable)	audit trail in source system(s)	Vendor	Use date/time stamp of EDC entry vs login information to demonstrate ongoing review and sign off close to collection of data.

Use Cases: Data integrity risk	Risk scenarios where audit trail is primary tool	Applicable system(s)	Source(s) needed	Ability to create AT report?	Examples of desired reporting criteria
	Missing visits/values	IRT (if using to collect CRF-like data)			Using audit trail for data entry date/times and expected dates, calculate compliance and identify missing values or visits in a dashboard.
3.7: Behavior by site is changed by sponsor checks	Use of browser side edits before data is submitted may influence Principal Investigator's or Study Coordinator's data entry (i.e., if an Inclusion/Exclusion criteria check, they may make changes prior to submitted data to ensure inclusion into the study)	EDC	audit trail in source system(s)	Vendor	Review volume of edit checks and changes by PI's prior to submission of data and look for trends in critical data, e.g., inclusion/exclusion criteria and eligibility changes.
3.8 Patient making multiple changes to questionnaires or diaries prior to submission of responses	Excessive data changes by patient prior to submitting may indicate a lack of understanding of questions or poor design of diary etc.	eCOA	Audit trail in source system(s)	Vendor	Review changes by patients prior to submission of data to identify trends
Use Case Category 4: Reporting concerns					
4.1: Duplicate datasets	Duplicate datasets –different data but same create date	Datasets	Audit logs, audit trails (date/time stamps)	Vendor, Sponsor	At data transfer: *compare checks prior to sending data to identify duplicate records * compare checks at sponsor to identify duplicate datasets errors. Look for matching contents with different creation dates, or different content data labeled as the same.

Use Cases: Data integrity risk	Risk scenarios where audit trail is primary tool	Applicable system(s)	Source(s) needed	Ability to create AT report?	Examples of desired reporting criteria
4.2: Changes during data migration	Data changes (corrupt data, dropped data, partially transferred data) occurring during data migration from vendor to sponsor or within sponsor organization	Datasets	Audit logs at various data flow points	Sponsor	Audit trails of transfer logs can be used to verify that the right amount of data was moved Internal transfer programs are validated and automated External transfers are checked to ensure data was received as per predefined specifications Future case: blockchain will be able to track changes to files (distributed ledger)
Use Case 5: Device concerns					
5.1: Date/timestamp inaccurate	Dead battery or technical malfunction can lead to errors relating to the in-device timestamp	eCOA	audit trail in source system(s)	Vendor	Report exception listing for changes on date/time stamp updates or changes by vendor. Can be included in change report above.
5.2: Merging of subject data	Subject ID is merged due to site change or device change (e.g., replacement devices (multiple device IDs under one subject) or one device ID for multiple subjects)	eCOA	Audit trail in source system(s)	Vendor	Listing to verify audit trail of merged subject IDs

Appendix 4: Risks, Limitations, and Mitigations Considerations

Risks	Limitations	Mitigations
Effective and timely implementation of a risk-based approach	• Adoption and uptake can be slow if existing approach already exists • Return on Investment may take years and may be challenging to measure	• Establish clear regulatory guidance and direction on a risk-based approach for ATR • Establish robust change management and execution strategies • Applying centralized monitoring principles and approaches with upfront key performance indicators established
New technologies (e.g. wearables and sensors) do not have clear regulatory requirements/guidelines on the what and how to review the audit trail	• Roles and responsibilities challenges include managing mutually conflicting or dependent conditions	• Sponsors and vendors should be aware of the limitations of the technological capabilities and continue to collaborate with regulators to define requirements for ATR
Lack of available automation capabilities and lack of how best to use automation for ATR	• Due to data parameters such as volume and readability, achieving the appropriate scalability without automation isn't possible • When analyzing perceived anomalies, it can be challenging to determine which signals are useful	• Focused utilization of technology to facilitate ATR (e.g. combination of visualizations, statistical analytics tools, issue management system)
Failed implementation of technology (e.g. selection of a vendor or developing an ATR tool in-house)	• Complexities with cross-functional coordination • Building value-added ATR requirements (e.g. reconstruct data integrity events, managing too much data to review, readable AT) • Poorly organized data (e.g. various data sources) creates incorrect output to deal with (or incomplete output)	• Clearly understand ATR capabilities of the system(s), such as customization based on needs • By building or adding processes for ATR, addresses regulatory and gaps in data integrity management

Risks	Limitations	Mitigations
Increased complexity challenges as additional source systems are considered (e.g. CTMS, TMF, wearables, etc) in the clinical development lifecycle	• Consider data privacy requirements (e.g., GDPR, HIPAA) • Source audit trail data quality varies • Cannot combine all audit trail data	• Establish a centralized reporting platform • Clearly define and differentiate technical (e.g. running reports by IT) and business process (e.g. data manager reviews)
People: Inadequate training or a skillset leading to ineffective ATR that adds no value	• Additional tasks requiring additional time for existing roles • Finding the right combination of data management skills and statistics skills is challenging	• Designed training for study manager for reviewing analytics • Upskilling data analysts and/or statistical programmers for ATR
Process: Which function, or company (e.g. sponsor or vendor), owns and manages the ATR process?	• Complexity of processes changes with the size of organizations (e.g. larger organizations have more specializations)	• Define the frequency of routine review • Have a clear delineation of roles and responsibilities internally and across partner organizations (sponsor, CRO, site, technology vendor etc.)
Defining appropriate thresholds (e.g. What are the thresholds, how to manage the change of thresholds)	• Limited experience • What is the baseline threshold and when to change threshold challenges	• Apply statistical concepts to identify outliers • Apply KRIs, QTLs, and QbD critical to quality factors • Be adaptive with thresholds as more data is accumulated
Creating an organizational culture that ensures data is complete, consistent and accurate, as per the principles of data integrity from MHRA guidance	• Organizational change management (e.g. culture) is more qualitative and difficult measure effectiveness	• Ensure an appropriate data governance policy or equivalent is in place at both the highest and lowest levels of the organization