ISPE | GAMP RECORDS AND DATA INTEGRITY

GOOD PRACTICE GUIDE:

Data Integrity – Key Concepts

ISPE | GAMP® RECORDS AND DATA INTEGRITY

GOOD PRACTICE GUIDE:

Data Integrity – Key Concepts

Disclaimer:
The *ISPE GAMP® RDI Good Practice Guide: Data Integrity – Key Concepts* provides detailed practical guidance to support data integrity within a regulated organization. This Guide is created and solely owned by ISPE. It is not a regulation, standard or regulatory guideline document. ISPE cannot ensure and does not warrant that a system managed in accordance with this Guide will be acceptable to regulatory authorities. Further, this Guide does not replace the need for hiring professional engineers or technicians.

Limitation of Liability
In no event shall ISPE or any of its affiliates, or the officers, directors, employees, members, or agents of each of them, or the authors, be liable for any damages of any kind, including without limitation any special, incidental, indirect, or consequential damages, whether or not advised of the possibility of such damages, and on any theory of liability whatsoever, arising out of or in connection with the use of this information.

ISBN 978-1-946964-11-3

Preface

ISPE GAMP® RDI Good Practice Guide: Data Integrity – Key Concepts explores areas presented in *ISPE GAMP® Guide: Records and Data Integrity* in further depth. This key concepts Guide incorporates tools such as Cultural Excellence and critical thinking skills into data integrity practices to aid companies in meeting regulatory requirements and expectations.

Numerous examples of good data integrity practices along with ways to identify risks and detect issues are included to assist organizations in developing or raising their data integrity awareness.

This Guide is positioned under the *ISPE GAMP® Guide: Records and Data Integrity,* and is aligned with *ISPE GAMP® 5: A Risk-Based Approach to Compliant GxP Computerized Systems.*

Acknowledgements

The Guide was produced by a Task Team led by Lorrie Vuolo-Schuessler (GlaxoSmithKline, USA) and Charlie Wakeham (Waters Corporation, Australia). The work was supported by the ISPE GAMP Community of Practice (CoP) and sponsored by Michael Rutherford (Syneos Health, USA).

Core Team

The following individuals took lead roles in the preparation of this Guide:

Sam Andrews	Integrity Solutions Ltd.	United Kingdom
Erika Ballman	Albemarle Corporation	USA
George Bass	GGB Services	USA
Ivan Diamond	Bio Products Laboratory Ltd.	United Kingdom
Robert Dillman	Eli Lilly & Co.	USA
Sophie Zhiyao Ding	Ernst & Young LLP	USA
George Evgrafov	PAREXEL International	Germany
Kira Ford	Eli Lilly & Co.	USA
Tami Frederick	Perrigo Company	USA
Elmar Harringer	CoProCo Ing.-Büro Harringer	Germany
Volker Hattwig	Coconeo Ltd.	Germany
Oliver Herrmann	Q-FINITY Quality Management	Germany
Paul Labas		USA
Heather Longden	Waters Corp	USA
Anthony Margetts	Factorytalk Co., Ltd.	Thailand
Barry McManus	Empowerment Quality Engineering	United Kingdom
Leslie A. Paul, MS	Perrigo Company	USA
Siegfried Schmitt	PAREXEL International	United Kingdom
Markus M. Schröder	Coconeo Ltd.	Germany
Doug Shaw	Azzur Group	USA
Rob Stephenson	Rob Stephenson Consultancy	United Kingdom
Michelle Vuolo	Sanofi	USA

Contributors and Reviewers

The Leads wish to thank the following individuals for their valuable contribution during the preparation of this Guide.

Nuala Calnan, PhD	Biopharm Excel	Ireland
James Canterbury	Ernst and Young	USA
Tom De Rudder	Novartis NTO Aseptics	Belgium
Morten Friis	Epista Life Science	Denmark
Isabel Munoz-Willery	NL42 Consulting | Paperless Lab Academy	Spain
Anders Vidstrup	NNIT A/S	Denmark

Subject Matter Expert Input and Review

The Team Leads wish to thank the following for their significant contribution to the document.

Monica Cahilly	Green Mountain Quality Assurance, LLC	USA

Regulatory Input and Review

Particular thanks go to the following for their review and comments on this Guide:

Gaye Camm	Therapeutic Goods Administration (TGA)	Australia
Stephen Grayson	Medicines and Healthcare products Regulatory Agency (MHRA)	United Kingdom
Karl-Heinz Menges	Regierungspräsidium Darmstadt	Germany

Special Thanks

The Leads would like to give particular thanks to Bob McDowall (R D McDowall Ltd., United Kingdom), Mark Newton (HeartlandQA (Eli Lilly – retired), USA), Maximilian Stroebe, PhD (GSK Vaccines, Netherlands), and ISPE Technical Advisor, Sion Wyn (Conformity Ltd., UK) for their efforts during the creation process of this Guide. The Team would also like to thank ISPE for technical writing and editing support by Jeanne Perez (ISPE Guidance Documents Technical Writer/Editor) and production support by Lynda Goldbach (ISPE Guidance Documents Manager).

The Team Leads would like to express their grateful thanks to the many individuals and companies from around the world who reviewed and provided comments during the preparation of this Guide; although they are too numerous to list here, their input is greatly appreciated.

Company affiliations are as of the final draft of the Guide.

600 N. Westshore Blvd., Suite 900, Tampa, Florida 33609 USA
Tel: +1-813-960-2105, Fax: +1-813-264-2816

www.ISPE.org

Table of Contents

1 Introduction

1.1 Background

The *ISPE GAMP® RDI Good Practice Guide: Data Integrity – Key Concepts* provides detailed practical guidance intended to support data integrity within a regulated organization.

In considering a holistic approach to data integrity, some simple assumptions can help to focus effort:

- Data integrity is usually lost early in the process, typically at data collection or during initial processing.
- Once data integrity is lost, it can never be restored; however, it is important to be able to detect the loss of integrity.
- Enterprise-level systems may provide effective controls, but issues can occur if the data is corrupted before it reaches the system (per first bullet).
- Every human interaction is a potential opportunity for a data integrity issue.
- Every interface is a potential opportunity for a data integrity issue.
- Data must be protected throughout the data life cycle including passing through interfaces between systems and stages.

These data integrity considerations need to be identified and the resulting risks mitigated to remove or reduce as many of the potential threats as possible. It is important to be able to identify when data integrity could be compromised so that methods of detection or prevention can be implemented.

There are four primary elements required to ensure data integrity at all stages:

- Cultural Excellence, embedding behaviors and leadership in support of data integrity.
- Quality Risk Management approach to designing and managing GxP data processes such that risks are assessed and mitigated, and controls are communicated and iteratively reviewed and improved (and encompassing the Data Integrity by Design paradigm).
- Technical and procedural controls are in place to prevent unauthorized changes.
- Evidence to provide assurance of data integrity in support of business processes.

1.2 Purpose

This Guide provides detailed guidance in four core areas:

- Data Governance
- Data Life Cycle
- Risk Management Approaches
- Critical Thinking

These core areas represent four of the six data integrity key concepts introduced in the *ISPE GAMP® Guide: Records and Data Integrity* [1]. Of the remaining two key concepts from that Guide:

- GxP Computerized System Life Cycles is extensively detailed in *ISPE GAMP® 5: A Risk-Based Approach to Compliant GxP Computerized Systems* [2].
- ALCOA and ALCOA+ are fundamental to data integrity approaches, with well-documented guidance contained in the WHO TRS No. 996 Annex 5: Guidance on good data and record management practices [3].

In addition, Section 2.3 of this Guide presents user behaviors supporting ALCOA+.

1.3 Scope

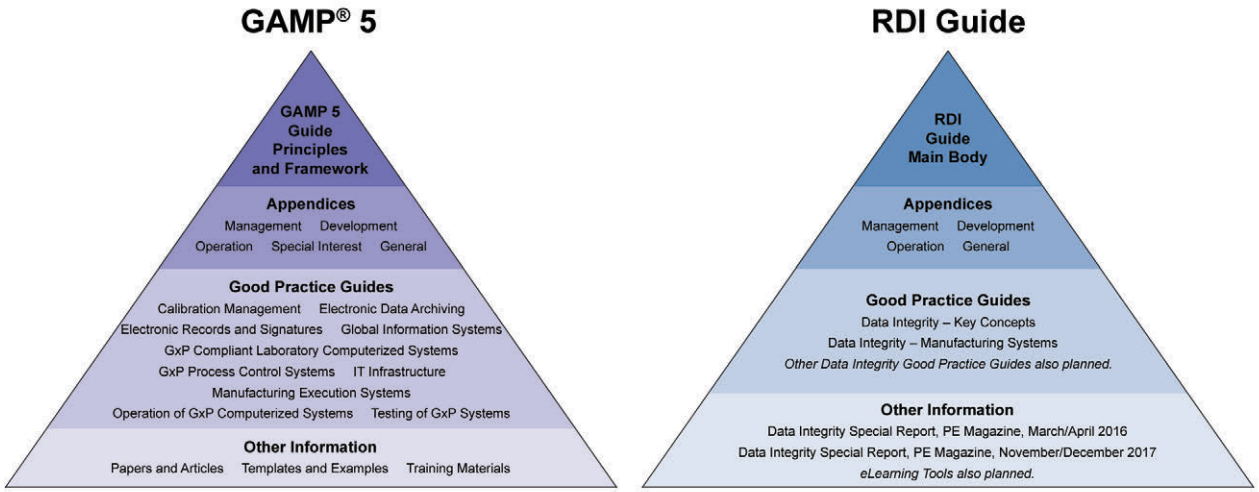
The scope of this Guide is data required under one or more GxP regulations (often referred to as the predicate rules). In preparing this Guide, the following regulatory guidances have been taken into account:

- WHO Technical Report Series No. 996 Annex 5: Guidance on good data and record management practices (2016) [3]
- FDA Data Integrity and Compliance with CGMP – Draft Guidance for Industry (April 2016) [4]
- MHRA 'GXP' Data Integrity Guidance and Definitions (March 2018) [5]
- EMA Questions and Answers: Good Manufacturing Practice – Data Integrity (August 2016) [6]
- PIC/S PI 041-1 (Draft 2) Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments (August 2016) [7]
- China FDA Draft Guidance on Drug Data Management Practice (October 2016) [8]
- Health Canada GUI-0001: Good manufacturing practices guide for drug products (February 2018) [9]
- TGA Data Management and Data Integrity (DMDI) (April 2017) [10]

1.4 Structure of This Guide

This Guide is positioned under the *ISPE GAMP® Guide: Records and Data Integrity* [1] and is intended to provide additional and specific detail on data integrity, in the same way as *ISPE GAMP® 5* [2] has Good Practice Guides that provide additional and specific detail on computerized systems validation.

Figure 1.1: Structure Basis of This Guide



The main body of this Guide is supplemented by appendices containing examples and case studies.

2 Data Governance

2.1 Data Integrity Culture

2.1.1 Introduction

The definition of data governance from MHRA is [5]:

> *"The arrangements to ensure that data, irrespective of the format in which they are generated, are recorded, processed, retained and used to ensure the record throughout the data lifecycle."*

Data governance encompasses the people, processes, and technology required to achieve consistent, accurate, and effective data handling. The *ISPE GAMP® Guide: Records and Data Integrity* [1] examines elements of the data governance framework including the technical, procedural, and behavioral controls that underpin data integrity. Appendix M3 [1] within that Guide discusses the impact of human factors on data integrity by highlighting the following considerations:

- Understanding and mitigating the impact of corporate and local culture
- Understanding the classification and underlying root cause of incidents (from minor lapses to acts, either intentional or unintentional) that result in material impact to patients and/or products and/or application integrity
- Implementing mechanisms to minimize human error rates
- Reducing motivation, pressures, and opportunities for data falsification and fraud
- Promoting impartiality in quality-related decision making
- Applying effective behavioral controls by influencing behaviors and attitudes

In other words, the *ISPE GAMP® Guide: Records and Data Integrity* [1] highlights key factors that should be considered relative to how an organization's culture can influence data integrity outcomes.

In this Chapter, the behavioral controls are examined in more detail in the context of a cultural excellence approach.

The *ISPE Cultural Excellence Report* (issued in April 2017) [11] aligns with and complements these key concepts, stating that:

> *"Culture can be described using many different terms, but the key is to define, emphasize, and support the demonstration of desired behaviors and results."*

The *ISPE Cultural Excellence Report* [11] explores the term "cultural excellence," proposing that within any given organization there is not a separate quality culture, safety culture, data integrity culture, etc. Rather, one primary corporate or organizational culture exists that influences the behaviors and actions giving rise to quality, data integrity, and safety outcomes that matter to the patient and the business.

The *ISPE Cultural Excellence Report* [11] recommends moving away from the traditional "culture of compliance" toward a "culture of excellence," stating:

> *"This signals a shift from reliance solely on regulatory compliance to an emphasis on continuous improvement in which there is deep understanding throughout an organization of the elements critical to product quality."*

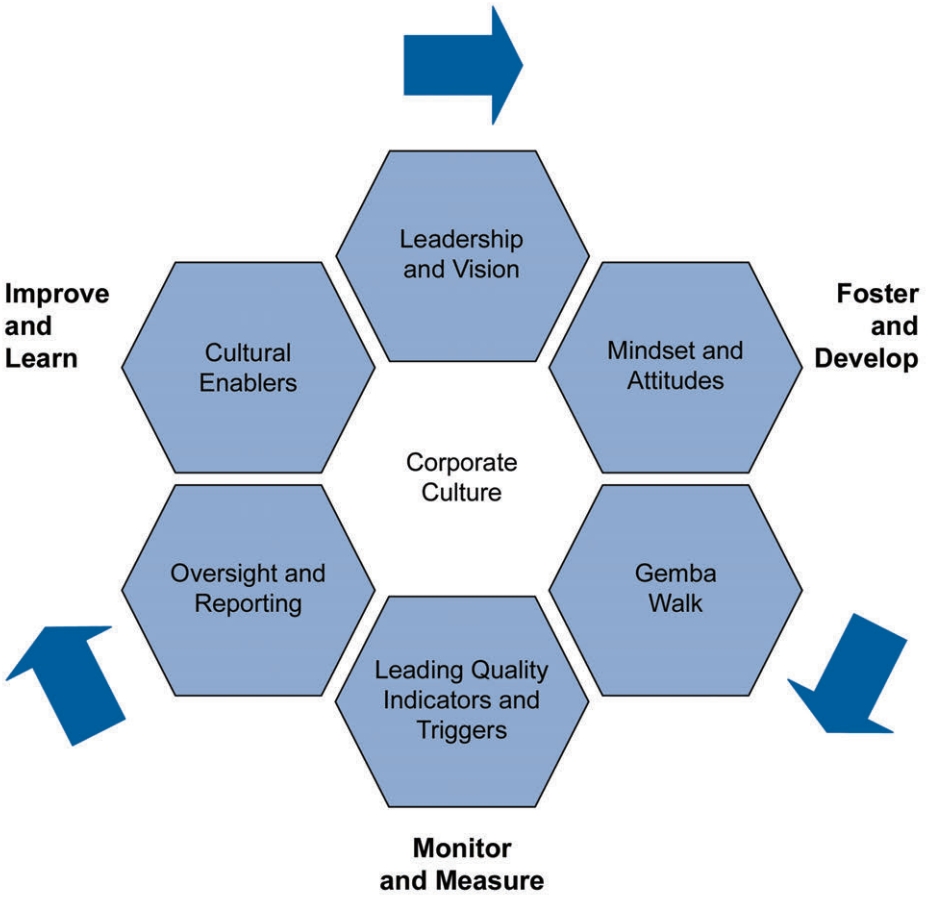
The cultural excellence framework in Figure 2.1 focuses on six dimensions that together foster, develop, monitor, measure, learn, and ultimately improve an organization's culture of excellence [11]. They are:

- Leadership & Vision
- Mindset & Attitudes
- Gemba Walk
- Leading Quality Indicators (LQIs[1]) & Triggers
- Management Oversight & Reporting
- Cultural Enablers

All are discussed in more detail later in this Section.

The *ISPE Cultural Excellence Report* [11] outlines a series of practical and powerful approaches, practices, and tools to support implementation of the cultural excellence framework and promote behavioral change. The following Section on data integrity culture demonstrates how some of these approaches, practices, and tools can be applied to ensure the integrity of the data an organization produces.

Figure 2.1: Six Dimensions of Cultural Excellence Framework [11]



[1] LQIs are a variant of quality metrics that include a measure of behavioral attributes, and are discussed in more detail in the *ISPE Cultural Excellence Report* [11].

2.1.2 Leadership and Vision

2.1.2.1 The Importance of Leadership to Data Integrity

Leadership plays a key role in establishing a healthy corporate culture. Attitudes toward safety, quality, and operational excellence can be shaped, both positively and negatively, by leader actions and behaviors. Leaders can also set the desired tone and direction relative to data integrity practices and expectations.

2.1.2.2 Leadership Model for Data Integrity

The first step for success is to ensure that corporate and site leadership are aware of, and fluent with, the recent regulatory findings and concerns around data integrity, and expectations for best practices. Only then are they likely to successfully influence their organization toward needed actions.

A clear data integrity direction or plan enables the entire organization to understand the desired state and the importance of compliance with the requirements, enabling employees to work in alignment with the specified goals and standards.

Leaders must clearly articulate the importance of data integrity as a regulatory standard for patient safety as well as a business success factor. Leaders must share the message broadly and frequently within the organization; this can be accomplished both formally and informally. It is essential, however, that leadership returns often to the message to maintain agreement and reaffirm its importance.

2.1.2.3 Modeling Leader Behavior

Leaders must make decisions consistent with the company quality expectations. As patient safety, product quality, and data used to support regulated decisions are paramount, actions must align with the company's data integrity plan, and data integrity risks and gaps should be viewed as potential patient safety risks.

If data integrity violations arise, all aspects must be considered to make the right decisions for patient safety, product quality, and data used to support regulated decisions. As leader actions are scrutinized, it is vitally important that leaders understand the impact of their words and actions; this will help ensure that the right messages related to the data integrity plan cascade throughout the organization.

With respect to data integrity, leadership values of transparency and humility are important. Being open to understanding the vulnerabilities within any given system and transparency about improvements needed are necessary to improve and develop data integrity cultural maturity.

Often executive-level leaders are not the technical Subject Matter Experts (SMEs) as it relates to data integrity. Site leaders, therefore, must enable those within the technical areas (IT, Laboratories, Quality, Operations) to coach and mentor the SMEs in the desired leader roles and empower them with the tools and resources to execute data integrity action plans effectively.

Leaders at all levels should clearly establish acceptable data integrity practices and show, by their actions, sincere and serious efforts to improve the overall data integrity maturity.

Leaders who employ both formal and informal communication can gain more exposure and reach different organizational levels more effectively. Their behavior helps to shape the thoughts and actions of other employees; therefore, leaders must actively model the principles, values, and vision desired by the company, personalizing the overall data integrity message with their own leadership style to ensure credibility.

Leaders can also promote an environment that is open to change, one in which ideas to improve quality and data integrity compliance are welcome and in which employees are not afraid to voice data integrity concerns. This emphasis on "Speak Up" is key, as employees who discover vulnerabilities should not be fearful to raise and address issues.

Many companies provide anonymous phone lines that allow employees to share confidential concerns of quality, safety, or other topics. Leaders should actively seek to develop Speak-Up cultures (viewed as ideal for enabling cultural excellence) in which employees feel comfortable sharing all concerns, including those that relate to data integrity.

2.1.3 Key Leader Actions

Leaders must consistently promote and seek within their organization information and suggestions to drive improvement.

Additionally, leaders should invest in their own development and evaluate their own effectiveness in order to determine areas in which they can improve as leaders.

They need to vigilantly monitor and display key performance metrics that hold the organization accountable for continuous improvement goals. (See more in Sections 2.1.6 Key Performance Indicators, and 2.1.7 Management Oversight and Review.)

Vigilance requires monitoring the entire organization to assess and reassess the ongoing process of data integrity maturity, not just as a remediation project or initiative.

2.1.3.1 Shaping Data Integrity Mindsets and Attitudes

As discussed in greater detail in the *ISPE Cultural Excellence Report* [11], leaders can influence mindsets and attitudes by:

- Clarifying behavioral expectations for employees
- Monitoring behavioral performance
- Providing positive cultural learning experiences to reinforce desired behavioral expectations
- Assessing cultural performance to address gaps

Cultural excellence is an ongoing commitment by leaders and individuals to model desired behaviors, promote transparent, proactive ownership of quality by all, and hold themselves and others accountable to standards.

2.1.3.2 Behavioral Criteria and Improvement Actions

The *ISPE Cultural Excellence Report* [11] includes a Cultural Excellence Assessment Tool, which is designed to help organizations assess the maturity of 21 desired key behaviors as part of their quality culture maturity program. This tool provides a behavior-based framework to understand, assess, and develop excellence in quality culture within organizations as described below.

Accountability

Employees consistently see quality and compliance as their personal responsibility. This is key to applying self-behavioral controls in the area of data integrity.

Establishing clear individual accountability for data integrity compliance is a foundational step in helping to shape the collective mindset for cultural excellence. Accountability should be communicated consistently through job descriptions, onboarding practices, GxP training, and performance goals, and be supported by coaching, capability development programs, rewards, and recognition. Leaders should hold themselves and others accountable for performing to quality and compliance standards.

This concept of employee accountability does not supersede or replace the need for management leadership and accountability for overall quality.

Ownership

Employees must have sufficient authority to make decisions and feel empowered to do their jobs well.

Individual ownership of quality and compliance is a primary driver for shaping the quality mindset. When individuals are fully engaged, empowered, and taking action to improve product quality, patient safety, and the quality of data used to support regulated decisions, organizations typically benefit from continuous improvement and faster decision making.

Action Orientation

Employees committed to cultural excellence regularly identify issues and intervene to minimize potential negative effects on quality and compliance.

Establishing the expectation that individuals demonstrate a proactive orientation helps shape the quality mindset and foster cultural excellence. Leaders should promote and leverage proactive efforts (e.g., risk assessments, Gemba walks (see Section 2.1.5), employee suggestions) to reinforce support for the desired behavior. Additionally, it is important that rewards and recognition be aligned to support these efforts, rather than reinforce reactive "fire-fighting" efforts.

Speak Up

Employees are not afraid to speak up, identify quality issues, or challenge the status quo for improved quality; they believe management will act on their suggestions.

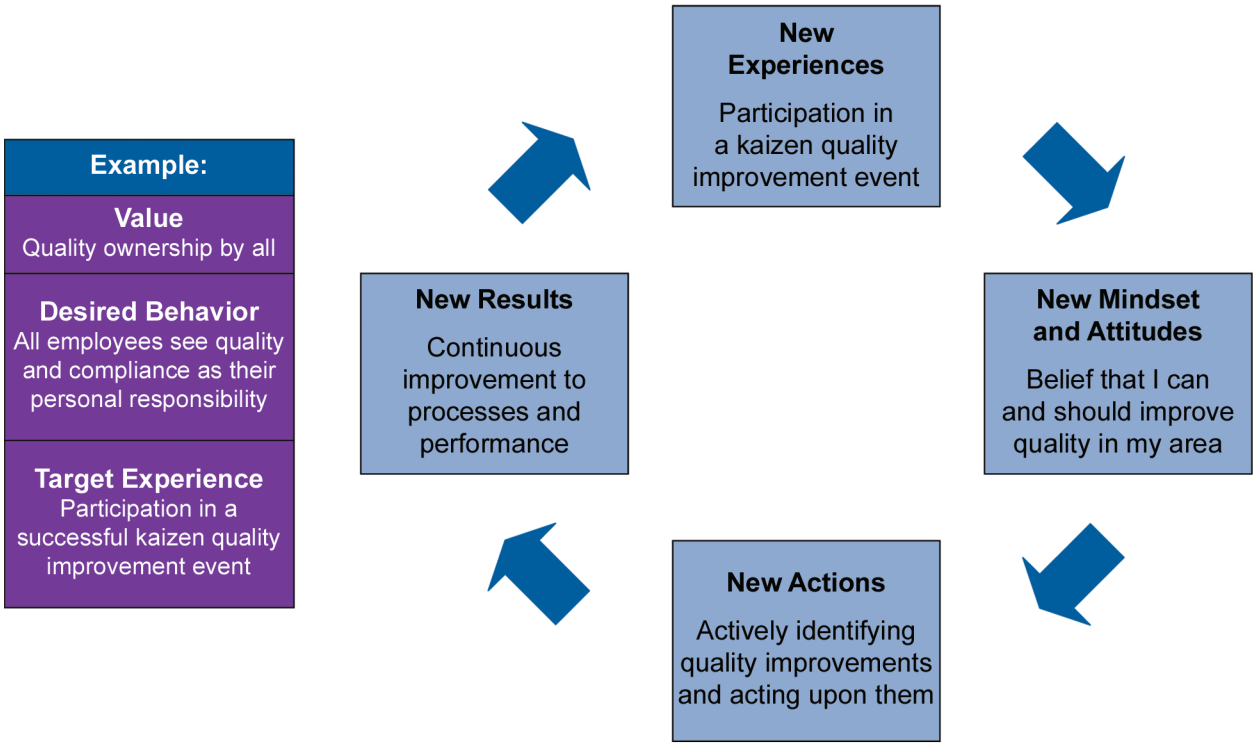
Empowering individuals to speak up and raise quality issues helps foster the quality mindset. Leaders should support this by modeling the desired behavior, building trust, and creating an environment in which individuals feel comfortable raising quality issues, engaging frontline personnel in problem solving, and involving employees in continuous improvement activities.

There are geographical cultures that may inhibit employees from speaking up; the impact of local cultures is discussed in more detail in the *ISPE GAMP® Guide: Records and Data Integrity, Appendix M3* [1].

2.1.4 Mindset and Attitudes

A company's ability to monitor and shape mindset and attitudes can greatly improve results and performance. Figure 2.2 shows an experience-based process in which mindset influences attitudes, which influence behaviors, which directly influence actions, which influence results, which ultimately influence performance. This becomes a continuous improvement loop as new results and performance further form desired mindsets.

Figure 2.2: Shaping Quality Mindset: Experience-based Approach [11]



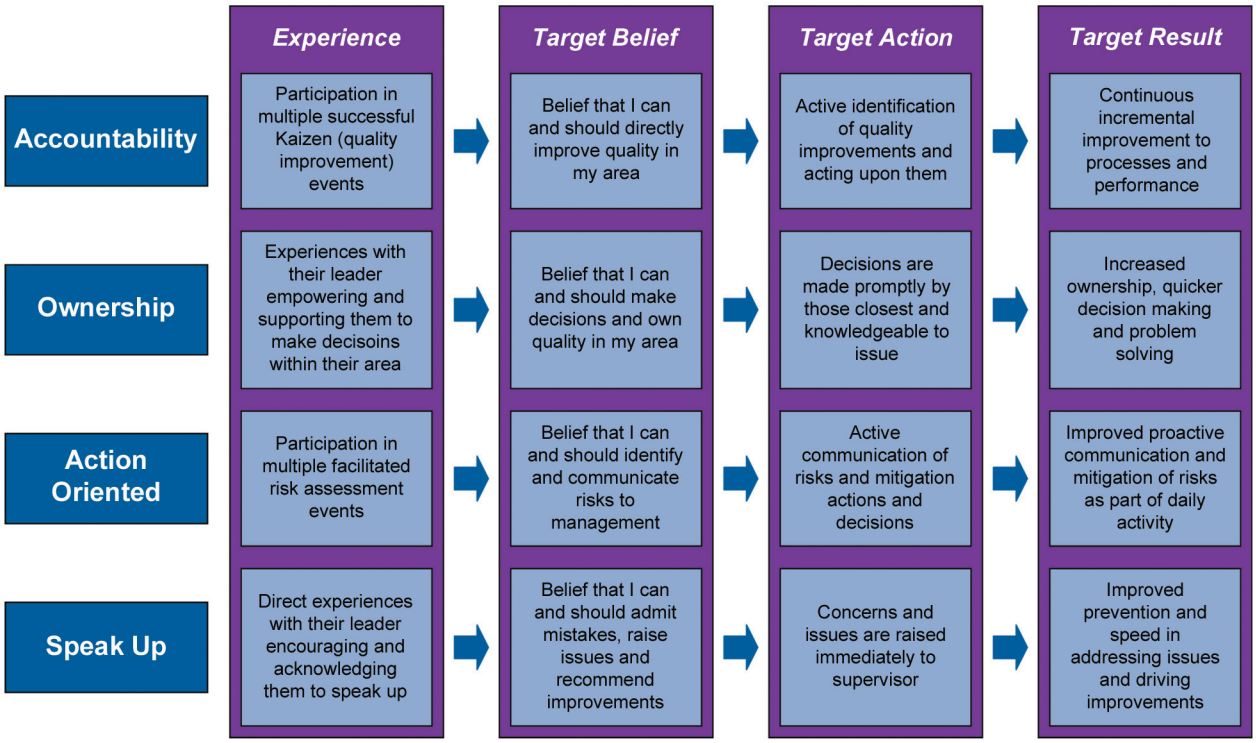
Individual leader actions and behaviors clearly contribute to site and corporate culture. There are commonalities among industry leaders related to behavior, acttions, and traits that facilitate greater employee engagement, attainment of site goals, and a corporate culture of excellence.

To more effectively shape the desired mindsets and attitudes leaders should:

- Share a vision about the importance of data integrity frequently and broadly within the organization, including case studies, regulatory observations or citations, and other shared industry knowledge.
- Demonstrate decision making and behaviors that align with the company's stated quality vision, including specific data integrity action items.
- Value operational excellence above a focus on regulatory compliance for compliance sake.
- Shape employee experiences and mindsets through formal and informal quality discussions where site metrics are reviewed, data integrity action plans are revisited, and quality issues are raised.
- Establish Gemba walks as a best practice activity for the shop floor, laboratories, and other functional areas; develop Gemba guidelines or checklists to aid a data integrity walk through.
- Develop key site metrics, implement leading quality metrics, and monitor proactive measurements to drive continuous improvement.
- Provide organization-wide structural enablers to support improvement and inspire an environment of continual learning.
- Challenge the organization to drive for excellence and create a culture in which patients and employees benefit.

Figure 2.3 lists experience-based examples across the key areas of accountability, ownership, action orientation, and Speak Up that can help shape mindset and attitudes, emphasizing targeted, positive experiences that can assist in delivering the desired actions and results.

Figure 2.3: Shaping Quality Mindset: Experience-based Examples [11]



Management awareness of mindset and attitudes, combined with an understanding of their effect on behavior, results, and performance, can increase an organization's capability to improve cultural performance over time.

2.1.5 Gemba Walks

Engaged employees build strong processes and systems that are continuously improved. These provide sustainable growth, reduce costs, increase profitability, and create a happy, collaborative team.

This Section contains examples of cultural enablers to enhance the data integrity and cultural excellence of the system through key performance and key behavioral indicators via the use of a Gemba approach. It also provides a sample road map that may be used for each key role in the system to standardize the plan for process improvement, role-based behaviors, training, and sustainability/control.

Enablement requires the development of Leading Behavioral Indicators (LBIs) at all levels within an organization. Cultural excellence culminates with all six cultural dimensions operating at peak maturity: leadership and vision, mindset and attitudes, Gemba walks, LQIs and triggers, oversight and review, and cultural enablers [11]. Each of these elements demonstrate a direct linkage to data integrity. The following provides an overview of the power of Gemba.

2.1.5.1 Gemba Definition

The Japanese term Gemba means "actual place." Jim Womack, author of *Gemba Walks*, expands this definition to call Gemba the place in an organization "where humans create value." [11]

Gemba is a well-defined element of the Lean process improvement concept and, as such, an accepted operational excellence tool in industries that have adopted Lean principles. The Toyota Production System, which originated the Lean concept, has used Gemba walks for decades. Within the pharmaceutical industry, however, the concept of Gemba has not yet been widely implemented [11].

It is important when establishing a Gemba walk to develop LBIs that align with the desired performance targets. Employees understand that we act upon what we measure.

Gemba walks **are** [11]:

- An enabler for cultural change in management style and philosophy
- A role-modeling opportunity for leaders
- A way to empower personnel
- An enabler for continuous improvement through problem solving with the people who experience them
- An opportunity to find the root cause of issues, spot waste and quality risks, and for leaders to remove obstacles
- A coaching/mentoring opportunity to build and/or enhance capabilities and behaviors, and recognize and reinforce desired behaviors
- An enabler for communication of site priorities/challenges and how the unit´s performance contributes to the overall success of the site
- An opportunity to learn from the shop floor, foster quality mindset; encourages informed decision making for leaders
- An opportunity for the operators/technicians to show their pride and excellence in their jobs

Gemba walks **are not** [11]:

- An audit (neither quality/compliance nor environmental health and safety)
- A general complaint or venting session
- A debate to defend individual viewpoints without facts
- A troubleshooting exercise in which participants focus exclusively on areas with (technical) issues

2.1.5.2 Gemba Walk Results

In the pharmaceutical industry there may be complaints that supervisors and management rarely make time to go out onto the shop floor or into the laboratories where they could interact with employees and observe what is really going on.

Gemba walks demonstrate a visible commitment from the leadership to all members of the organization. They allow site leadership and supervisors to spread clear messages using open and honest dialogue, to coach and mentor employees, and to get a good indication of the progress of behavioral change at all levels.

Gemba empowers employees, as their contributions to site results are recognized and their ideas for continuous improvements are heard and acted upon.

As the key purpose of Gemba is to identify continuous improvement opportunities, it is critical to record commitments and agreed actions. One of the easiest ways to do this is to display the agreed actions on visual boards in the area, promoting local ownership of progress and providing a central point for the Gemba to commence [11].

Cultural enablement of the leader, supervisor, and technician level employees can be demonstrated through standardized Gemba walks. Appendix 1 demonstrates an example of a Gemba checklist for the laboratory, focusing on the behaviors and activities in support of data integrity.

2.1.6 Key Performance Indicators

Metrics can serve as a powerful tool in driving behaviors in both positive and negative ways. As discussed in the *ISPE GAMP® Guide: Records and Data Integrity* [1], metrics that encourage fraudulent practices (pressure, opportunity, and rationalization) can create data integrity issues.

However, well-designed metrics help raise awareness of best practices, identify issues, and ultimately improve data integrity within the organization. Within a cultural excellence framework supporting data governance, Key Performance Indicators (KPIs) may be referred to as LQIs or LBIs.

The *ISPE Cultural Excellence Report* [11] leverages the "ABC" model of behavioral science outlined by Leslie Braksick [12], which states that:

A. Antecedents trigger behavior

B. Behaviors are followed by consequences

C. Consequences determine if behaviors will recur

Braksick's work also included the consequence rule that states that consequences have a four times greater impact on behavior than antecedents, yet organizations often leave consequences largely unmanaged [12]. As such, how management responds to metrics has a significant impact on the behaviors espoused, and therefore, the results achieved.

To create well-designed metrics that will drive the desired behaviors, Braksick presents an IMPACT model for developing behavior-based LQIs or LBIs [12].

The IMPACT model [12] requires the following steps:

- **I**dentify the desired quality-improvement goal
- Establish the appropriate **M**easure to deliver the goal
- **P**inpoint the "desired" behavior to deliver the goal
- **A**ctivate the **C**onsequences to motivate the delivery of the goal
- **T**ransfer the knowledge across the organization to sustain the performance improvement

The *ISPE Cultural Excellence Report* [11] presents a tool based on Braksick's ABC and IMPACT models [12] called the LQI design tool. This report highlights that the strength of the tools not only come from pinpointing behaviors that matter, but also from designing positive consequences to deliver the desired results [11].

See Appendix 2 for an example of the IMPACT Tool [12] applied to data integrity.

2.1.7 Management Oversight and Review

Per ICH Q10 [13]:

> *"Leadership is essential to establish and maintain a company-wide commitment to quality and for the performance of the pharmaceutical quality system."*

Robust oversight and review, engaging both management and employees, reinforces a strong culture of excellence by demonstrating transparency, fostering trust, and facilitating dialogue. All of this enables learning, brings attention to issues so they can be addressed, and highlights best practices so they can be replicated.

One manner in which leadership can achieve this is by instituting a corporate data integrity program. The *ISPE GAMP® Guide: Records and Data Integrity* [1] highlights the following key implementation considerations for a corporate data integrity program:

- A documented rationale
- Executive sponsorship and governance process
- Management accountability
- Levels of training
- Metrics to measure performance
- Program reporting to communicate progress
- Audit and assessment processes

See Appendix 3 for a case study of one organization's corporate data integrity program.

Companies that have a healthy or mature quality culture are often excellent learning organizations that value and share knowledge across the organization. This learning characteristic is a guiding principle for effective management oversight and review programs as they:

- Align quality objectives
- Monitor for continuous improvement
- Ensure leadership involvement
- Engage in effective external party oversight and reporting

2.1.8 Cultural Enablers

Finally, enabling cultural excellence requires more than just setting goals and KPIs. As emphasized in the preceding Sections, actions and behaviors must also be aligned across the organization.

To that end, the *ISPE Cultural Excellence Report* [11] includes a powerful road map tool that uses a systems approach to identify possible improvements and the necessary associated behaviors.

The Culture and Continuous Improvement Capability Road Map is a tool designed to perform informal continuous improvement within a work area, to provide technical expertise to support cultural excellence, to assess training needs, to align KPIs to role-based behavioral indicators, and to provide information for activities and controls in a structured roadmap [11].

This road map tool encompasses the following key elements: [11]

- Assuring appropriate mentorship and training for all staff
- Building a roadmap for improvement and sustainability to assure cultural excellence
- Using a Plan-Do-Check-Act-Monitor approach to enhance current systems and programs for each role
- Developing ***Role-Based Behavioral Indicators*** modeled for sustainable data integrity

Appendix 4 provides an example using this tool as a template.

Ultimately, cultural excellence recognizes quality not as an operational burden or compliance requirement, but as a necessity that allows companies to make decisions that best benefit the patient. The ISPE Cultural Excellence Report [11] is an additional resource available to organizations that seek to define, emphasize, and support the demonstration of desired behaviors as a means to consistently deliver enhanced quality outcomes.

2.2 Roles and Responsibilities

Within any organization it is important that senior management establishes and supports an organization built on quality and integrity. They must define the appropriate roles and responsibilities and assign appropriate personnel.

Data integrity must be driven into every aspect of the organization and form the basis of everything the organization does.

Table 2.1 lists the typical roles and responsibilities in an organization. These may not align with job titles in many organizations because often individuals assume multiple roles within the company. In the table below, the roles should not be limited to a focus on an individual system but rather apply across the data life cycle and the wider business process.

Table 2.1: Typical Roles and Responsibilities

Role	Responsibilities
Senior Management	• Establish and maintain the data integrity assurance culture throughout the organization • Set corporate data integrity policy and standards • Ultimately responsible for quality throughout the organization • Allocate appropriate resources to support and sustain good data integrity management • Provide oversight, risk management, and monitoring for data integrity • Define and communicate personnel responsibilities relevant to data integrity practices • Commit to quality process improvement and modernization
Chief Data Integrity/ Governance Officer	• Representative of Senior Management responsible for defining and implementing the strategical approach to organization-wide data integrity assurance • Responsible and accountable for organizational data integrity initiatives • Depending on the scale and complexity of the organization this may be fulfilled as a full-time role, responsibilities assigned as part of an individual's role, or through a committee

Table 2.1: Typical Roles and Responsibilities (continued)

Role	Responsibilities
Process Owner	• Typically, a senior member of the functional unit using the system • Ultimately responsible for the business process or processes being managed • Controls access to the system, in conjunction with the system owner • Responsible for ensuring that the computerized system and its operation is in compliance and fit for intended use in accordance with applicable Standard Operating Procedures (SOPs) • Responsible for providing adequate resources to support development and operation of the system, including continually trained personnel in operations, oversight, and process documentation, as well as financial resources to support the system • Accountable for the existence of appropriate SOPs for operation, change control, and periodic review, and that these SOPs are followed • Accountable for continuous assessment of the system and responding to reported or uncovered findings, observations, or data integrity risks with appropriate remediation and controls • May also be the System Owner and/or Data Owner, depending on the size and complexity of the organization
System Owner	• Ultimately responsible for the availability, and support and maintenance of a system, and for the security of the data residing on that system • Supports the Process Owner to assure continuous assessment of the system and to respond to reported or uncovered findings, observations, or data integrity risks with appropriate remediation and controls • Controls access to the system, in conjunction with the process owner • Acts on behalf of the users and is typically from IT or Engineering functions • Responsible for the availability, support, and maintenance of a system, and for the security of the data residing on that system • Responsible for ensuring that the computerized system is supported and maintained in accordance with applicable SOPs • Accountable to ensure IT support SOPs exist and are followed, provide for training of support staff, and promote system life cycle management including change management, system upgrades/replacement, and availability of system inventory and configuration management
System Support and Administration Personnel	• Support system operations in alignment and compliance with applicable policies and procedures • Responsible for maintaining system access controls - those administering user ID and passwords should not also have responsibility for, or report through, those responsible for the data within the system (segregation of duties) • Depending on the size and complexity of the system and/or organization this may also be the System Owner
Database Administrator	• Responsible for internal and external controls over databases • Depending on the size and complexity of the system and/or organization this may also be the System Owner **Internal Controls** • Assures installation qualification procedures include internal security controls e.g., database password maintenance, system and user administration controls • Assures back end database changes and data migrations are made in accordance with established procedures that include traceability to source data and ALCOA+ principles (see *ISPE GAMP® Guide: Records and Data Integrity, Section 1.5 Key Concepts* [1])

Table 2.1: Typical Roles and Responsibilities (continued)

Role	Responsibilities
Database Administrator (continued)	**External Controls** • Establishes procedures to assure the inclusion of data governance principles to protect databases from external corruption, including risk management procedures on SQL inserts, and other external data corruption methods • Assures that database maintenance can identify between internal and external threats or changes
Data Steward	• Typically a member of the functional unit using the system • Responsible for tactical coordination and implementation of data usage, management, and security policies as determined by data governance initiatives • Responsible for ongoing data review, data audits, and/or continuous improvement initiatives
Data Owner	• Typically a member of the functional unit using the system • Ultimately responsible for the integrity and compliance of specific data at various stages of the data life cycle in accordance with applicable policies and SOPs • May also be the Process Owner
Quality Function	• Responsible for independent oversight and review to assure integrity of data throughout the data life cycle • Accountable for quality policy and standards • Provides current regulatory agencies and industry guidance to the organization • Approves key documents and records including procedures, standards, specifications, plans, and reports • Reviews critical to quality data, whether paper or electronic, as part of QMS activities including qualification and validation, change management, and batch release • Audits systems, processes, governing documents, and records to assure data integrity
Subject Matter Expert	• Individuals with specific expertise in a particular area or field • Responsible for planning and defining verification strategies, defining acceptance criteria, selection of appropriate test methods, execution of verification tests, and reviewing results • Provide expertise in their area or field to determine, define, and implement data integrity measures that prevent, detect, and respond to data integrity concerns
End User	• Responsible for collecting, analyzing, reviewing, reporting, and using data and information in a manner that accurately, truthfully, and completely represents what actually occurred in either paper or electronic format
System Developer	• Provides best practices and identifies technical limitations during all phases of the system life cycle • Responsible for developing the system in accordance with established procedures so that all user requirements are met • Responsibilities vary depending on systems and services provided (e.g., COTS vs SaaS Cloud)

2.3 Good Documentation Management Practices

Good Documentation Practices (GDocP) are the combination of technical, procedural, and behavioral controls needed to ensure that GxP data meets the requirements of ALCOA+ for data integrity, irrespective of the format (paper or electronic) [1].

2.3.1 Technical Controls and Procedural Controls

The regulatory expectations for technical and procedural controls for both paper and electronic records to meet ALCOA+ are comprehensively and clearly discussed in WHO TRS No. 996 Annex 5 Appendix 1 [3], and therefore, are not reproduced here.

Additionally, specific guidance on establishing and defining data integrity requirements for computerized systems is described in the *ISPE GAMP® Guide: Records and Data Integrity* [1] and covered in further detail in Appendix 6 of this Guide.

2.3.2 Behavioral Controls

Table 2.2 lists the behaviors needed to ensure GDocP in addition to the technical and procedural controls documented in the WHO TRS No. 996 Annex 5 Appendix 1 [3]. In isolation, these behaviors are not sufficient to ensure data integrity.

Table 2.2: Behaviors Supporting ALCOA+ Requirements

ALCOA+ Requirement	Behaviors Essential to Support the Requirement
Attributable	Data must be attributable to the person or system generating the data, or performing an activity. Users shall choose passwords that are not obvious and are sufficiently complex that the password cannot be easily guessed by a colleague. Users shall keep their passwords secret from other users, and if they do keep a record of their password it must be in a secure location not accessible to others. Ensuring data is attributable can be complicated when an automated process is started by one shift worker and continues to run autonomously through other shifts under the supervision of other operators. Clear timesheets, or similar records, are needed to document the shift changes, and to maintain a record of which operators made interventions to the process during that extended operation.
Legible	Where a user is required to record a reason for change, for example in an audit trail, the reason recorded should clearly explain why the change was necessary and reference any supporting documentation such as an approved change request.* This will facilitate reconstruction of the event and understanding of the motivation for the change such that a reviewer or auditor can decide if the change was procedurally or scientifically justified. Any errors in recording data shall be conserved in the original documentation, with a clear identification of the error or reason for change; the original data must remain visible; and the correct data must be recorded along with the user name, date, and time. Where entries are handwritten, they should be clear and readable by anyone, and written in indelible ink. *Audit trail requirements are inherent within Attributable, Legible, and Complete. Different regulatory guidances discuss audit trails in each of these requirements.

Table 2.2: Behaviors Supporting ALCOA+ Requirements (continued)

ALCOA+ Requirement	Behaviors Essential to Support the Requirement
Contemporaneous	Users shall record data at the time it is generated and shall use the time shown on a networked computer or official company-maintained clock to provide the date and time if manually recorded (instead of using a personal wristwatch that may not be synchronized to company time). Users must always record the current date and time, irrespective of when the task should have been completed. Where the complexity or limitation of the process physically prevents a user from recording contemporaneously (for example a user is working with biohazardous substances within an isolator), the WHO TRS No. 996 Annex 5 [3] and PIC/S PI 041-1 (Draft 2) [7] make allowance for the use of a scribe as follows: *"…the supervisory recording should [must] be contemporaneous with the task being performed and should identify both the person performing the observed task and the person completing the record. The person performing the observed task should [must] countersign the record wherever possible, although it is accepted that this countersigning step will be retrospective. The process for supervisory (scribe) documentation completion should be described in an approved procedure which should also specify the activities to which the process applies."*
Original	Data must be recorded directly into the authorized media. Users must record the first reading/measurement/calculated result, even where that value may be deleterious to the overall pass/fail status of the task. In the event that a user makes an error in completing a paper form or the form sustains damage in some way, the original paper form must be kept and cannot be discarded for a new form. Users must not take unauthorized copies of blank controlled forms with the intent of using them to replace any forms containing mistakes or undesirable results. Consideration should be given to incorporating watermarks and iridescent inks into forms to allow easy detection of unauthorized photocopying, over and above the procedural requirement for controlled issuance of forms. All documents should have a unique identification number, a form design that provides sufficient space for manual data entries, and make clear what data should be recorded and where. Users must take care to preserve the original record, which would be the first location in which the information is recorded. For example, when manually writing down a balance reading the user must write it directly into their laboratory notebook or controlled form; a temporary recording onto a scrap of paper would make the scrap of paper the original record. All unused forms should be accounted for, and either defaced and destroyed, or returned for secure filing, with access to the (electronic) templates strictly controlled. Reviewers need to ensure they review the original records (or true copies thereof) including complete data. For dynamic data, the review should be done on the electronic data. See Appendix 5 for the regulatory definitions of these terms.

Table 2.2: Behaviors Supporting ALCOA+ Requirements (continued)

ALCOA+ Requirement	Behaviors Essential to Support the Requirement
Accurate	Users must rigorously follow written procedures governing their daily tasks and apply their training and skills to ensure that any data they generate is accurate. Care must be taken when manually writing or typing data to avoid transcription errors; self-checking of recorded or calculated data should be encouraged as part of routine working practices. In a laboratory environment this would include ensuring that equipment is calibrated and has the correct range and resolution for the intended measurements, that any reagents are within their expiry date, and that the test method and parameters are appropriate for the sample under test. In a clinical study the data must represent the real facts of the study, including accurate measurement of patient vital signs using calibrated blood pressure monitors, weigh scales, etc., and recording all responses.
Complete	Complete data relies on the user collecting and preserving all of the original data, derived data and results, metadata, and audit trail. Complete data requires the user to collect all records created or modified and not selecting only those records that meet specifications. Where multiple results are generated, the user must ensure that all results are formally addressed and reported. All operator actions must be recorded and maintained. The user must follow a clearly defined procedure to invalidate any results where there is sound scientific justification to do so, and must preserve all data, metadata, and audit trail entries related to the results, invalidated or not. A user must not delete electronic GxP data or destroy GxP paper records that are necessary to preserve the data, content or meaning as required under the predicate rules. Such data can only be deleted at the end of the retention period (excepting data under legal hold), under a controlled procedure.
Consistent	Due care and attention must be given to any task undertaken by the user that generates GxP data. The user may develop a consistent set of key phrases for reasons for change and error corrections to facilitate searching and reviewing. Users reviewing data must first determine “can the data be trusted” before making an approve/reject decision on the data as compared to its governing specification. This may involve examining audit trails and metadata, verifying operational sequences using the time stamps recorded, searching for duplicate or modified data, and checking the original data upstream of the activity (e.g., checking the sample preparation data recorded in the laboratory notebook as part of reviewing a chromatography result, or comparing the Electronic Health Record to the data entered in the eCRF system).

Table 2.2: Behaviors Supporting ALCOA+ Requirements (continued)

ALCOA+ Requirement	Behaviors Essential to Support the Requirement
Enduring	Data must be recorded in a permanent, maintainable form for the retention period. Procedures should confirm that archived data, including relevant metadata, is available and human readable. Any GxP data created electronically must be stored in a location that is backed up. For this reason, it is not recommended to store GxP data on local computer hard drives. Users of controlled forms and worksheets must take care to protect the forms throughout their use and then to return them to the original issuer or archivist for safe archival for the retention period.
Available	A user must ensure that all of the GxP data they have generated, whether recorded on paper or electronically, is filed in a logical and labeled secure location for ease of retrieval. It is good practice that filenames for electronic data are chosen to reflect the contents of the file to facilitate search and retrieval.

2.3.3 Overview of Controls

Table 2.3 outlines the holistic approach, encompassing People, Process, and Technology considerations, essential to support the data integrity key concepts.

Table 2.3: Mapping Data Integrity Key Concepts against the People, Process, and Technology Aspects of the Holistic Approach

Key Concept	People	Process	Technology
Risk Management Approach	Data Integrity risk management and assessment process in place	Data integrity risk assessment documented, approved, and periodically reviewed for both systems and records	System configuration optimized to reduce data integrity risks via automated scheduling of backups, audit trails always on, etc.
Data Governance	Cultural excellence inherent in leadership and vision, with strong employee engagement	System training matrix includes data integrity and computer system validation, and evidence of system stakeholder training records	Granular access controls to limit access to system functionality according to responsibility and competency
Data Life Cycle	Data Owners and Data Stewards trained and in place across the data life cycle for GxP records	System records and data life cycle defined, including Creation, Processing, Review-Use, Retention, Destruction	Data audit trails are linked to GxP data and operational activities

Table 2.3: Mapping Data Integrity Key Concepts against the People, Process, and Technology Aspects of the Holistic Approach (continued)

Key Concept	People	Process	Technology
ALCOA+	Data Integrity policy and ALCOA+ principles incorporated within system stakeholders, supplier support services, contractors job descriptions	System records and data classified including primary record, metadata, master data, system access, security procedures, and evidence of control practices in place	System security includes unique user identification and passwords with expiration
Critical Thinking	Periodic Data Integrity audit process established with quality function, system stakeholders and Process Owner	Data audit trail procedure and routine review practices in place supporting GxP decisions and processing	Data audit trails can be accessed and sorted for review purposes. Technical system logs are linked to system updates, changes, backups
GxP Computerized System Life Cycle	Supplier/service provider assessment and ongoing service level agreements. Computerized system validated as fit for purpose and all system users trained.	System incident management procedure in place including assessment of accident/ malpractice Routine use and system administration SOPs in place and followed	System validation status and testing correctly reflects system changes, use, updates, and shows that the system is fit for purpose

2.3.4 Considerations for Moving from Paper to Electronic Records

The guidance documents issued by the regulators [3, 5, 7] promote the use of computerized systems and integrated technical controls to reduce data integrity risks.

This can be perceived as a driver to move from paper to electronic records; however, simply transferring a paper process "as is" to an electronic equivalent (the so-called "paper on glass" approach) may carry some of the paper system weaknesses into the electronic system, and may not achieve all of the efficiency gains and benefits that a computerized system is able to offer.

Key activities in planning the paper to electronic transfer include:

- Start by defining the business process, identifying the critical records generated within this process, and determining the corresponding data life cycle for those records. (See Section 3.2 on Data and System Life Cycle Interrelationships for more detail.)
- Define the scope of paper records that would be replaced with electronic records (in one or more computerized systems).
- Ensure all of the data sources are identified for the existing paper records and that the record routings and approval processes are captured in a data flow diagram (e.g., search for those forgotten spreadsheet calculations). It may be useful to reanalyze the business process to determine if all of the record routing and approvals are required.

- Consider the controls and procedures required for data security and confidentiality. (See Section 2.4 and *ISPE GAMP® Guide: Records and Data Integrity* [1] for more information on data classification.)
- Ensure all interfaces (between computerized systems and also between business processes) are captured.
- Strive to minimize manual processes as they pose the highest potential risk.
- Consider, in a multidisciplinary team, how the process for the paper records could be optimized, such as:
    - Merging similar forms (for example, one for each product or test type) into a single form with embedded logic to offer fields determined by the selected type.
    - Interfacing forms to equipment such as balances to automatically document weights to ensure the first (original) readings are captured, any subsequent readings are audit trailed, and transcription errors are eliminated.
    - Interfacing forms to equipment, reagent inventories and/or raw material inventories to ensure the equipment used is in calibration and properly maintained, and the reagents and raw materials are within shelf life.
    - Simplifying the approval routing and implementing email notification to ensure timely approvals.

Fundamental to this process is ensuring that the team has all of the critical stakeholders represented including: business users, GxP regulatory experts, IT systems experts, and configuration and architecture experts.

Also important to this process is visible management commitment, as well as management support for the time and resources needed to perform this analysis. The time invested in the planning stage to define and optimize the existing business process will yield returns in the efficiency of the electronic process, and the knowledge gained will drive the specification of the requirements for the computerized systems to be implemented.

2.4 Data Classification

2.4.1 Data Classification Definition

Data classification directly affects how organizations understand and manage business processes at the most basic level. It is a fundamental element in data protection and a crucial element in enterprise data management and data governance.

Data classification is the process of assigning a value or rating to data. Data classification includes identifying and rating sensitive databases, tables, or columns, identifying restricted or confidential information in database storage, etc.

2.4.2 Why Data Classifications?

With an increase in big data and the proliferation of data from emerging technologies (such as the Internet of Things and connected devices), current challenges for organizations include:

1. Lack of data inventory and data owners
2. Lack of understanding of data, its value, and the different levels of protection needed
3. Lack of framework that mandates different levels of control implementation based upon a structured approach

Well-defined classification of data within a framework can assist in resolving such challenges.

2.4.3 Data Classification Objectives

Data classification as the first step toward data protection includes objectives such as:

- Classifying data to allow a tiered protection scheme and handling
- Encouraging proper labeling and handling of sensitive data for compliance
- Preventing unauthorized access to sensitive information
- Compliance with legal and regulatory requirements

2.4.4 Data Classification Framework

The overall goal is simple: to create a data classification framework that enables the organization to identify, label, and protect sensitive data in different databases. Such a framework consists of several components, including:

1. Data classification policies – define scope, responsibilities, and other governance requirements
2. Data classification scheme – determines tiers of data and associated levels of protection. Three or four-tier schemes are common, although five tier ones exist, especially in industries that produce intellectual properties.
3. Data labeling guidelines – give instructions for labeling and enables automated protection tools
4. Data handling guidelines – provide specific requirements for different data classes based on classification level
5. Data classification mapping – keeps current mapping that links data types or data sets to classification levels

To identify sensitive structured or semi-structured data in each type of database, data classification exercises can start with scanning or eDiscovery with Data Loss Prevention tools, or a walk through with the business and IT to identify locations that might hold sensitive data. Appendix 8 contains an example of a four-tier classification system.

The concept of classification and detailed information for applying it are comprehensively covered in ISO/IEC 27001 Information Security Management [14]. **Note:** the standard uses the term “information” in place of the term “data” used throughout this Guide.

2.4.5 Creating and Maintaining the Data Classification Framework

There is no “one size fits all” approach to building a data classification framework or system. Decisions should be made based on the type of organization and the overall data protection strategy. At a minimum, a high-level policy specifying the requirements for protecting sensitive data should exist and clearly link to the data classification policy.

Classification guidelines (e.g., labeling and handling guidelines) and compensating controls must be linked to each classification level. Data classification processes should be defined for consistent and repeatable execution.

More importantly, data changes over time, which means its value, sensitivity, and protection needs change as well. Data or database owners must be involved to keep database data classification guidance up to date.

2.5 Gap Assessments as Part of a Corporate Data Integrity Program

The importance of performing data integrity assessments as part of the corporate data integrity program to understand the state of data integrity controls is described in the *ISPE GAMP® Guide: Records and Data Integrity, Appendix M1* [1].

2.5.1 Assessment and Remediation Planning

An approach to a data integrity system assessment and remediation program is presented below. The program may be executed at a local departmental level or may be part of a company-wide assessment and remediation initiative:

- Create an assessment and remediation program team
- Develop program tools and procedures
- Provide training in data integrity and ALCOA+ principles to all involved personnel
- Appoint or confirm existing area/department System Owners and Data Stewards
- Review System Inventory to determine critical GxP systems (and to confirm completeness)
- Perform a high-level risk ranking exercise to establish overall priorities for review
- Perform data integrity risk assessment
- Identify mitigation and control actions
- Develop a detailed risk-based Assessment and Remediation Plan
- Track and review progress
- Perform ongoing monitoring of compliance and data integrity controls against changing requirements and evolving working practices

In terms of timescale for remediation of data integrity risks, it is important to consider the criticality of the business process and the data and the mitigating control strategies currently implemented, and then consider the following:

- EU regulations [15] first required an audit trail in 1991
- US regulations [16] first required an audit trail in 1997
- UK MHRA 2018 guidance [5] requires “demonstrated progress” toward addressing system deficiencies

2.5.1.1 Assessment and Remediation Program Team

As described in the *ISPE GAMP® Guide: Records and Data Integrity, Appendix M1* [1], a data integrity program depends on senior management support and leadership; thus, the remediation program team needs to be formed at that level within the company.

Suggested members minimally include: the Chief Data Officer, a senior Quality representative, a senior business representative, and at least one data integrity SME. The team is responsible for setting strategies and direction for the data integrity initiative, and for providing the drive and resources to enable the assessment and remediation program to succeed.

2.5.1.2 Develop Program Tools and Procedures

A set of standard processes and associated templates is required to ensure a compliant and sustainable outcome from the assessment and remediation program, which will need to be developed by the program team or their delegates. The following procedures are required:

- Management of computerized system inventory
- Records and data integrity analysis
- Risk assessment process in line with quality risk management approaches
- Identification and management of remedial actions and controls

2.5.1.3 Provide Training

The provision for training in good record and data integrity practices to all personnel potentially involved in system assessment and remediation is performed as part of the overall record and data integrity governance initiative as described in the *ISPE GAMP® Guide: Records and Data Integrity* [1].

This training is critical to the success of the program to ensure it is not a checkbox activity but a meaningful evaluation of the computerized systems in light of an understanding of the business process supported by the systems and the necessary controls required.

2.5.1.4 Define Roles and Responsibilities

The program team is responsible for the oversight of the project including determining the priorities.

Lower organizational-level teams are likely to perform the assessment/analysis and remediation tasks. To help facilitate the program in each area or department, it is important to have a local SME in records and data integrity (the Data Steward) who can consult with other Data Stewards and remediation program team members, to ensure a consistent approach to the performance of the analysis and the identification of remedial controls. Each system has a system owner who must work closely with the process owner to ensure the ongoing integrity of data produced in that system.

A full description of Roles and Responsibilities is contained in Section 2.2.

2.5.2 Identify and Prioritize GxP Systems

2.5.2.1 Review System Inventory to Determine Critical GxP Systems

As described in EU and PIC/S Annex 11 §4.3 [17, 18], there should be a system inventory for the site/department that lists all the systems in use and summarizes critical information about each system including its GxP impact:

> *"An up to date listing of all relevant systems and their GMP functionality (inventory) should be available.*
>
> *For critical systems an up to date system description detailing the physical and logical arrangements, data flows and interfaces with other systems or processes, any hardware and software pre-requisites, and security measures should be available."* [18]

If the system inventory is incomplete or outdated, there is a risk that one or more systems requiring remediation may be overlooked. Consideration should be given to walking through the business process in each area and following the data through each activity, to ensure that all systems are captured and compiled into an updated, current, and accurate system inventory.

It is important to evaluate systems for their impact on the data integrity of the entire business process. A holistic review or walk through of the business process can capture paper-based records and/or hybrid situations where the level of manual intervention increases the potential for data integrity risk.

Additionally, it will identify areas where there is the misconception that printing the data for retention and archiving is an acceptable alternative to securing and maintaining dynamic data as electronic records, as required by the regulations.

2.5.2.2 Establish Review Priorities

A risk ranking exercise considering factors relevant to the overall record and data integrity within each system is performed. To ensure consistency, conduct the assessment needs according to a set of criteria agreed to by the program team.

The assessment ratings for each system are determined and the systems ranked in priority according to their total scores. Weighting factors may be used to reflect the relative importance of the factors if appropriate.

Factors to consider in this exercise are:

- Business criticality
- Overall system GxP impact
- Impact on patient safety, product quality, or data supporting regulated decisions
- Impact on distributed/commercial product
- Original validation approaches (e.g., were the data integrity requirements adequately considered and verified?)
- Level of known data integrity issues
- Frequency of use of system
- Other considerations as appropriate (e.g., audit findings)

The system ranking resulting from this exercise allow the systems to then be divided into groups, for example, high priority, medium priority, and low priority. These groupings are used to determine the timeline and resource allocation for assessment and remediation within the overall site or company-wide Assessment and Remediation Plan.

2.5.2.3 Perform a Data Integrity Gap Analysis

A record and data integrity analysis is performed to identify systems requiring remediation to address risks in the system or process.

This Section discusses common controls and issues within a system that could lead to increased risk to patient safety, product quality, and data integrity. Identified risks are documented as part of the gap analysis.

An effective way of achieving this in a formal and robust manner is to prepare a standard list of verification checks in a questionnaire format; this could be based on *ISPE GAMP® Guide: Records and Data Integrity, Appendix D1* [1].

The results of these evaluations should be formally documented to verify both the existence and the adequacy of the current controls. Documented evidence that the verification has been performed (for example during the original system validation) should be referenced and evaluated.

For systems where this verification is not available it may be necessary to produce evidence that the system meets the verification check, that is, testing as part of the mitigation activities.

Security and Access Control

Some systems may have a limited number of user accounts (due to application limitations rather than a financial reluctance to purchase user licenses); such a restriction can necessitate the use of generic passwords and shared accounts, preventing actions from being attributed to an individual. Where the installed version of the application may restrict the number of user accounts and access levels, it is expected that where available, users should upgrade their systems to a version that provides the necessary functionality or consider replacing the system.

Another typical shortfall is a limited number of available role profiles resulting in inadequate segregation of duties, such as between user and administrator. Where there are limited personnel available, one person may have to fulfill the administrator role in addition to their routine tasks. While that person may therefore be granted both user and administrator roles, it is important that they always log in with the role appropriate to the task: user role for routine tasks, and administrator role for administration duties [7].

In many cases the administrator account is fixed with all privileges assigned, including the ability to generate data within the system, in contradiction of the principle of least privileges and segregation of duties. There have been numerous regulatory citations of lower access-level users being assigned administrator access [19, 20] resulting in inappropriate granting of delete privileges.

It is important to understand the business process supported by the computerized system and ensure that individuals are granted access only to the functionality and/or data appropriate to their job role. Often, the focus of the roles and responsibilities definition is on write and change rights; however, for research organizations, clinical systems, and systems containing unblinding or Personally Identifiable Information (PII) data, read access needs to be as strictly limited as write access.

Physical security is a regulatory requirement [17, 18] and also an area of potential failure by a company during implementation. The storage media for original data, reported results, metadata, and audit trails must be physically isolated (e.g., server in a locked server room not in the main office area), with additional logical security to prevent remote access to the storage media for the purposes of deleting files at the operating system level.

Standalone systems are particularly vulnerable since there is a temptation to store regulated data locally on the hard drive, which is then vulnerable to accidental or malicious attack. Data stored on an accessible hard drive can be subject to deletion, renaming, and re-use as part of data manipulation.

Backup and Restore

Standalone systems are especially prone to lack of data backup given that this typically requires manual user intervention to connect suitable media and start the backup of the hard drive.

Some applications provide an option not to include the audit trail in the backup or archive copy. Where such an option is present, it should be clear that all GxP-relevant audit trail(s) should to be captured in all backups and archive versions.

Another common failing, created during system implementation, is to store backup files on the original server or on a separate server located in the same server room. In the event of a server room disaster, both the original files and the backup files are lost with an associated loss of regulated data. During system implementation or remediation, the backup scripts should be set up such that the backup files are automatically copied to a remote location [7].

Audit Trails and Logs

Systems may have some or all of the following issues regarding data audit trails:

- GxP-relevant activities not captured
- Merging entries of GxP-relevant changes with general event log entries
- Insufficient information captured in an entry, missing any of: user identification, date and time, a description of the action, and the new and old values
- Overwriting (first in, first out) of audit trail entries
- Inability to read, sort, search, or filter audit trail entries for review
- Entries are not independently generated and secure from alteration

In addition to audit trails on GxP-relevant activities, systems may also be deficient in capturing technical system logs of important system activities, such as:

- Configuration changes
- Backup activities
- User logons

These types of technical system logs need to be reviewed for any indicators of potential data integrity risks as part of a periodic review, data integrity audit, or an investigation.

The WHO guidance [3] gives clear instruction that the content, independence, and review capability of the audit trail is more important than the particular term used to describe it in a software application.

Validation

Some systems originally implemented as supporting a non-GxP process may have been put into use without adequate computerized system validation and data integrity controls.

Failure to validate a system for its intended use leaves potential data integrity risks across the business process, as there is no documented evidence that the technical controls function correctly and that, as a whole, the system is fit for purpose.

Archival

Data is often left to build up in the system without a clear archiving process. Storing files indefinitely in the original system for the retention period brings two main concerns:

- **Backup time and cost:** The time taken to backup the system increases over time with the increase in quantity of records in the system.

Backup storage has a cost associated with it, and backup files become very large as they contain copies of inactive data that may not have been modified or viewed in years. So long as this inactive data continues to reside only in the original system, it must be included in the backup. (The regulators are clear that a temporary backup is not an acceptable archiving solution [4].)

Additionally, there is an increased risk of data loss if the time to complete a single backup operation exceeds 24 hours (due to the sheer quantity of data), since it is then no longer possible to complete a daily backup of data within the system.

- **Original System Performance:** As the quantity of files stored within the original system increases, system performance may decrease. Also, searching for data may take longer or may not work as the quantity of files to be searched increases.

Generic Issues across the Data Life Cycle

The data life cycle is discussed in Section 3.2. Each life cycle stage may present potential record and data integrity issues for systems.

Table 2.4: Generic Issues with Systems

Life Cycle Stage	Possible Issues and Deficiencies
Data Creation (Generation, Capture and transmission)	• Lack of secure, individual access control • Inability to support granular accounts enabling segregation of duties for system user, administrator, QA, etc. • Lack of control over date and time stamp settings and format • Inability to deal with multiple time zones • Inadequate audit trail capability on data creation, modification, or deletion • Manual transcription between systems due to lack of interface capability • Interfaces that lack data encryption and fail-safe mechanisms • Unvalidated data transfer between systems • Inability to preserve content and meaning of records during transmission
Data Processing	• Processing methods not version controlled, and earlier versions not retained
Data Review, Reporting, and Use	• Lack of data review capability (including searching and filtering of audit trails) • Inability to support exception reporting processes • Lack of electronic signature capability • Lack of traceability between original data, processed data, and reports • Lack of flagging or reporting invalid and/or atypical data • Lack of ability to trend or perform analytics
Data Retention and Retrieval	• Lack of backwards/forwards compatibility for records from earlier software versions • Lack of automated archiving ability • Lack of checksums or other integrity checks on archived records • Reliance on a manual backup process
Data Destruction	• Reliance on manual deletion at the end of the retention period • Inability to deal with data inadmissible for destruction due to legal hold requirements

These technical failings can be exacerbated when combined with inadequate procedural controls, such as:

- Failure to apply segregation of duties to user roles
- Procedural definition of printouts to be the original data
- User-defined equations created without version control or verification
- No defined record retention period
- Data disposal process inadequate, not procedurally described, and/or not documented

Paper Records

Paper records are unsuitable for complex data as the dynamic nature of the data is lost in the static format (see Section 3.1 for a more detailed discussion), so relying on a printed record in place of secure electronic archiving is unacceptable for chromatography, spectroscopy, video images, electronic Case Report Form (eCRF) with clinical data, and other dynamic data.

Report data that is recorded by hand or transcribed by typing has a high dependency upon the individuals performing the task, which increases the data integrity risk.

System Support

Service and technical support for operational computerized systems is crucial to maintain efficient and effective operations. Service support is often provided by the external supplier that has developed the system or has been involved with the implementation.

In the case of older or outdated systems, ongoing service support from the supplier can be problematic due to a number of factors including:

- Limited support for previous versions
- Discontinued support
- Loss of knowledge, skills, and experience due to personnel change
- Supplier no longer exists

Obsolescence

At some point in its operational life, the hardware within an older system may become obsolete such that it can no longer be maintained in a functional state or cannot support the installation and operation of the latest software version or operating system. This eliminates upgrade as a mitigation option and introduces additional risks around hardware failure with no possibility of like-for-like replacement.

Personnel

Change of company management and staff can have an impact (positive or negative) on the potential risks to patient safety and product quality within a system that relies upon the right behaviors and procedural practices used to mitigate a lack of safeguards and technical controls.

Evaluation of working techniques and associated levels of competency supporting the process can identify risk factors that can compromise data integrity, such as:

- Inappropriate or overly complex workflows
- Outdated customs and practices
- Lack of records and data integrity or ALCOA+ understanding
- Excessive focus on productivity and yield

3 Data Life Cycle

3.1 Data Definitions and Requirements

3.1.1 Introduction

The objective of this Chapter is to appreciate the variety of terms used to describe regulated data, and to explain the terminology that will be used in this Guide. It is important to understand that there is both commonality and variation among global GxP regulations and regulatory guidance.

The terminology defined here is an essential precursor to the detailed examples around applying the data life cycle contained in Section 3.2.

3.1.2 Data Definitions in Regulations

Table 3.1 summarizes which regulations define and use which data term. A full listing of the comparative regulations and terminology is contained in Appendix 5.

Table 3.1: Data Definitions in Regulations

Regulatory Reference/Term	Raw Data	Complete Data	Source Data	Critical Data
US FDA GLP (21 CFR Part 58.3k) [21]	Yes	No	No	No
OECD GLP (GLP No. 1) [22]	Yes	No	No	No
OECD GLP (GLP No. 17) [23]	Yes[1]	No	No	No
EMA/CHMP/ICH Tripartite GCP[2] (Guideline for good clinical practice E6(R2)) [24]	No	No	Yes	Yes, with partial definition[3]
US FDA Guidance for Industry (Electronic Source Data for Clinical Investigations) [25]	No	No	Yes	No
US FDA GMP (21 CFR Part 211.194(a)/211.188[4]) [26]	No	Yes, with list of requirements	No	No
EU GMP (EudraLEX Volume 4 Chapter 4) [27]	Yes, no definition[5]	No	No	Yes, no definition[6]
UK MHRA 'GXP' Data Integrity Guidance and Definitions [5]	Yes	No	Yes, synonymous with raw data	Yes, no definition[7]

Notes:
1. Also defines Derived Data.
2. Also uses the term Source Documents.
3. Used in context: *"The sponsor should develop a monitoring plan that is tailored to the specific human subject protection and data integrity risks of the trial."; "monitoring of critical data and processes"*
4. Uses the term complete information in place of complete data
5. Used in context: *"For electronic records regulated users should define which data are to be used as raw data"*
6. Used in context: *"A system should be in place to indicate special observations and any changes to critical data."*
7. Used in context: *"The approach to reviewing specific record content, such as critical data and metadata…"*

3.1.3 Original Data

For paper and electronic processes, original records of an activity must be captured and retained for the applicable retention period, with the retained format of the records preserving both the content and meaning.

Within this Guide, and based upon the precedent set in the *ISPE GAMP® Guide: Records and Data Integrity* [1]:

- A regulated record is a collection of regulated data (and any metadata necessary to provide meaning and context) with a specific GxP purpose, content, and meaning, and required by GxP regulations. Records include instructions as well as data and reports.

- The terms "original data" and "original record" will be used interchangeably to describe the first recording of original observations and/or acquired data before any processing.

Original data includes the metadata that describes the attributes of other data and is required to maintain GxP content and meaning [1] and that forms an integral part of the original record [5].

The term "original record" is not specific to any one GxP (see Table 3.1 above for the variety of terms used) and is consistent with the definition of "Original" within ALCOA+, as documented in WHO TRS No. 996 Annex 5 §9.5 [3]:

> *"Original data include the first or source capture of data or information and all subsequent data required to fully reconstruct the conduct of the GXP activity."*

For electronic records, the concept of first capture creating an original record that is then preserved has less relevance since the electronic record lacks the physical presence of a paper record.

A true copy of an electronic record that preserves content and meaning is identical to the original record, whereas a true copy of a paper record is typically distinguishable from the original while still preserving the content and meaning of the original (e.g., a photocopy of a filled page in a laboratory notebook can be a true copy of the original page, but the laboratory notebook page will always be the original, contained within the notebook and bearing the original handwritten recording of the data).

The requirements for true copies are discussed in Section 3.1.4.1, where the key import is that all of the content and meaning must be preserved in the true copy.

3.1.4 Data Requirements

3.1.4.1 Original Record and True Copies

The original record is the first capture of data into an authorized recording medium (paper, photograph, or electronic records including image or spectral files). Maintaining the original record for the retention period may not always be practical, with records such as:

- Thermal printouts, which degrade over time

- Electronic data where the original record alone is not sufficient; there must also be a backup copy of the electronic data

- Records in systems where large volumes of electronic data are created, the ability to archive true copies of the data into offline storage becomes increasingly important

In such situations, a true copy can be retained in lieu of the original records if the copy is equivalent to the original in that it preserves the GxP content and meaning of the original record. Such a true copy ensures that predicate rule requirements are fully met, and the ability to fulfill regulatory obligations preserved.

The various GxP regulations again use a variety of terms, this time for copies of original records: true copies, certified copies, verified copies, or exact copies as documented in Appendix 5. The regulations also discuss static and dynamic data where the static data can be used and understood "as recorded" directly to make a decision (on patient safety, product quality, or regulatory approval), but dynamic data requires additional processing to derive a meaningful value (reportable result) that can be used for decision making.

Regardless of the terms, the content and meaning of the copy must be the same as the original record, including preserving the ability to interpret and process the record for dynamic data.

Table 3.2 lists a variety of original records with examples of permitted true copies for each. Note: in all cases, the true copies must:

- Have been through either a manual verification (e.g., signing and dating the photocopy) or validation (verifying the consistent creation of a true copy via a controlled and automated process)
- Be secured against unauthorized changes
- Be protected throughout the record retention period

Table 3.2: Examples of True Copies of Original Records

Original Record	Nature	Examples of True Copies
Paper notebook or paper form with handwritten values, observations or information	Static format	Photocopies, microfilm, scanned to secure/protected PDF
Electronic file containing a list of values from a data logger – no further processing required	Static format	Paper printout, conversion to secure/protected PDF
Electronic file containing a list of values from a data logger – statistical analysis and trending required	Dynamic format	Electronic record containing all logged values, with the ability to interpret or export the data for analysis and trending
Electronic file containing 3D or complex data requiring processing or trending	Dynamic format	Electronic record containing all raw data, metadata, and supporting data, and giving the ability to further interpret the data in the same way as with the original record

3.1.4.2 Good Clinical Practice Special Requirements

Good Clinical Practice (GCP) requirements include additional provisos unique to GCP: that the investigator must have access to the source data at all times, and that source data cannot reside only in a storage location that is under sole control of the trial sponsor.

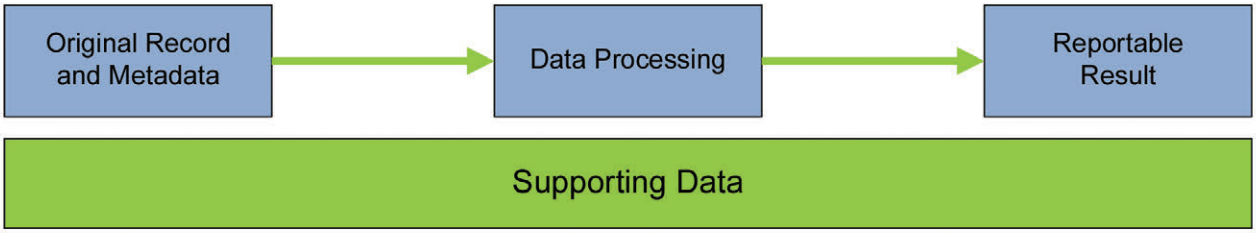
3.1.4.3 Supporting Data

An original record contains metadata, which provides the context and meaning to the data and forms an integral part of the record [5]. When reviewing the original record, the regulatory expectation is that the original data, including metadata, be included in the review when needed to establish the integrity of the data [3, 4]. This is reflected in MHRA 'GXP' Data Integrity Guidance and Definitions §6.15 [5]:

> *"Summary reports of data are often supplied between organisations (contract givers and acceptors). It must be acknowledged that summary reports are limited and critical supporting data and metadata may not be included."*

Figure 3.1 shows how some original records may need to be processed to gain a reportable result.

Figure 3.1: Components of Data



Underneath this process is the supporting data comprising records that could impact the data processing and/or are used to ensure that the system is fit for its intended use and is under control (e.g., effective change control), maintained, and calibrated, as appropriate. Table 3.3 contains examples of reportable results and supporting data for a process control system and a Chromatography Data System (CDS). In a manufacturing environment, supporting data may apply across multiple batches or even be batch independent (see Appendix 18 for examples).

Table 3.3: Original Records, Metadata, Reportable Results, and Supporting Data

System Type	Data			
	Original Record	**Metadata**	**Reportable Result***	**Supporting Data**
Process Control System	• Process values from sensors • Operator-entered parameters	• Audit trails • Date/time stamp • User ID • System ID	• Critical Process • Parameters (feeding into batch record)	• Sensor location • Warnings and alarms • Sensor calibration data • Cleaning records • Alarm settings • Batch run paper checklist
Chromatography Data System	• Acquired channels from detectors • Operator-entered sample information and method selection	• Audit trails • Date/time stamp • User ID • Method ID • Instrument ID	• Critical Quality Attributes (amount, purity, impurities, feeding into Certificate of Analysis	• Instrument method • Acquisition method • Processing method • Method history • System configuration settings • Calibration and qualification data • Method validation data • Invalidated out of specification results • Laboratory notebooks
*Referred to as Derived Data in OECD GLP No. 17 [23].				

3.1.5 GxP Impact/Data Criticality

Throughout this Guide, the GxP impact of a record or data is a measure of the extent to which it can affect product quality, patient safety, or a regulatory decision. Data criticality expands on this to combine the importance of the decision to be made based on the data [5, 7]. The risk to the data is related to its vulnerability to unauthorized manipulation and the ability to detect any such manipulation. The rigor of the controls established around those risks should be commensurate with the criticality of the data.

Whether data is recorded electronically or manually, it is important that controls are established to prevent and detect manipulation. The use of validated exception reporting processes [5] are fundamental to such detection without adopting a forensic approach to routine data review (which is clearly stated as not a regulatory expectation [5]). Identifying, understanding, and preserving data with GxP impact is further discussed in Section 3.2 Data and System Life Cycle Interrelationships.

3.1.6 Data within the Data Life Cycle

The data life cycle presented within the *ISPE GAMP® Guide: Records and Data Integrity* [1] consists of five phases:

1. Creation
2. Processing
3. Review, Reporting, and Use
4. Retention and Retrieval
5. Destruction

If one looks at a complex business process, including manual and computerized processes, some of these phases may become more complex:

- During the data creation phase, data can be collected electronically from a computerized system or via data recording from a manual process.
- The approach to data review, reporting, and use must be risk-based and consider all relevant supporting data including original paper and electronic records or true copies thereof.
- During the data retention and retrieval phase there may be a need to transfer or migrate data from one system to another, or from one format to another. It is important that any such data movement be performed so as to preserve GxP content and meaning.

For further information around these topics, see the following in this Guide:

- Chapter 2 Data Governance
- Section 3.2 Data and System Life Cycle Interrelationships
- Section 4.3.6 Preserving Existing Data

3.2 Data and System Life Cycle Interrelationships

3.2.1 Introduction

Regulated companies using computerized systems to create, process or store data with GxP impact have three life cycles to consider: the data life cycle, the records life cycle, and the computerized system life cycle.

The scope of the records life cycle extends beyond data with GxP impact and will, therefore, not be covered within this Guide; it is discussed in the *ISPE GAMP® Guide: Records and Data Integrity, Appendix M7* [1].

The data life cycle and the computerized system life cycle form two of the key concepts in the *ISPE GAMP® Guide: Records and Data Integrity, Section 1.5* [1].

This Section will build on these two life cycles, and then look at the relationships and synergies between them.

The objective of this Section is to explain the relationships between the life cycles, to allow the reader to understand how to map the life cycles in their organization to identify their data integrity risks, and then use that knowledge to drive the definition of system requirements and the scope of verification activities.

Understanding and leveraging the life cycles provides a systematic process for the assessment, control, communication, and review of risks to data integrity. This should be a continuous process throughout the computerized system life cycle from concept to retirement, and ongoing for those GxP records within their retention period.

This approach should be applied to any business process creating, processing, using, or storing GxP-regulated data.

3.2.1.1 Data Life Cycle

The *ISPE GAMP® Guide: Records and Data Integrity* [1] introduces the data life cycle as follows:

> *"All phases in data life cycle from initial data creation, capture, and recording through processing (including transformation or migration), review, reporting, retention, retrieval, and destruction should be controlled and managed in order to ensure accurate, reliable, and compliant records and data."*

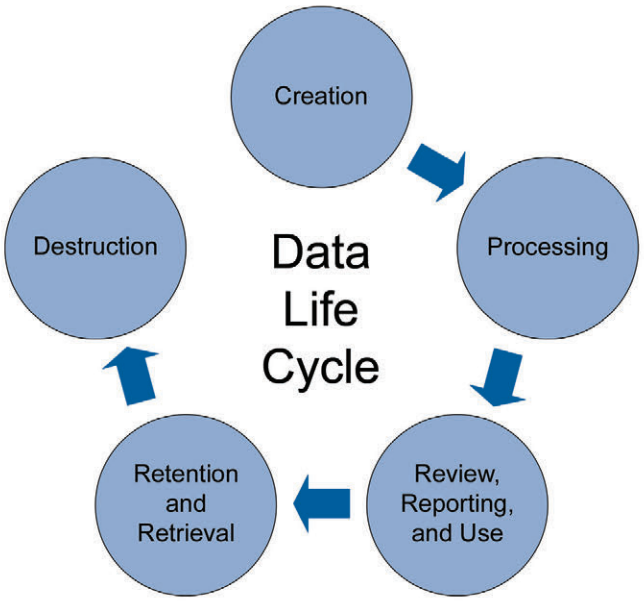
Data, in this context, is data that directly supports a GxP process or decision; hence, the data life cycle is managed at the business process level.

There are typically two representations of the data life cycle: a linear format, which reflects the similarity to the records life cycle, and a cyclic portrayal.

Figure 3.2: Data Life Cycle in Linear Format [1]

The cyclic representation is the main one used in the *ISPE GAMP® Guide: Records and Data Integrity* [1].

Figure 3.3: Data Life Cycle in Cyclic Format [1]



Irrespective of the diagrammatic representation used, the key points in the life cycle are [1]:

- **Creation:** Data capture or recording should ensure that data of appropriate accuracy, completeness, content, and meaning is collected and retained for its intended use.

- **Processing:** Data is processed to obtain and present information in the required format. Processing should occur in accordance with defined and verified processes (e.g., specified and tested calculations and algorithms), and approved procedures.

- **Review, Reporting, and Use:** Data is used for informed decision making. Data review, reporting, and use should be performed in accordance with defined and verified processes and approved procedures. Data review and reporting is typically concerned with record/report type documents. Second person reviews should focus on the overall process from data creation to calculation of reportable results. Such reviews may cross system boundaries and include the associated external records and may include verification of any calculations used. The data reporting procedures should contain the complete data set and define the data handling procedures, and ensure the consistency and integrity of the results.

- **Retention and Retrieval:** Data should be retained securely. Data should be readily available through the defined retention period in accordance with defined and verified processes and approved procedures. Retention periods vary by record type and applicable regulation, and some records e.g., validation and qualification records, need to be retained for the life of the system or process.

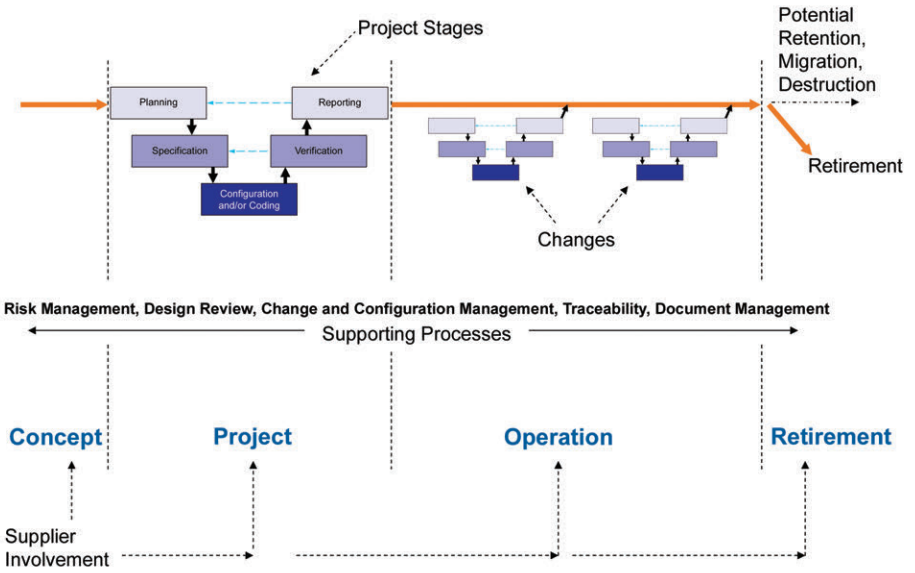
- **Destruction:** The data destruction phase involves ensuring that the correct original data is disposed of after the required retention period in accordance with a defined process and approved procedures. Data should only be retained beyond its retention period if there are legal requirements.

A more detailed discussion of the data life cycle is contained in the *ISPE GAMP® Guide: Records and Data Integrity, Section 4* for Data Life Cycle and *Appendix M4* of the same Guide for Audit Trail Review [1].

3.2.1.2 Computerized System Life Cycle

Computerized system life cycles have been thoroughly documented in *ISPE GAMP® 5* [2], and therefore this Section only contains a brief outline.

Figure 3.4: Project Stages and Supporting Processes within the {Computerized System) Life Cycle (according to ISPE GAMP® 5) [2]



A computerized system life cycle provides a systematic approach through four phases [2]:

- **Concept:** During the concept phase, the regulated company considers opportunities to automate one or more business processes based upon business need and benefits. Typically, at this phase, initial requirements will be developed and potential solutions considered. From an initial understanding of scope, costs, and benefits, a decision is made on whether to proceed to the project phase. The concept phase may be initiated as part of a remediation plan for an existing system. (See Section 4.3 for further details.)

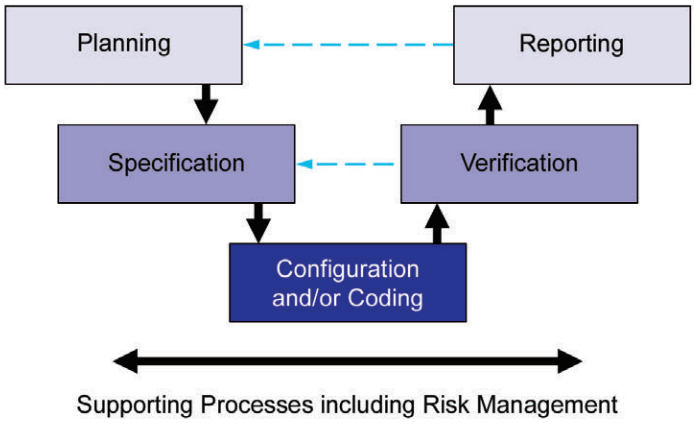
- **Project:** The project phase involves planning, supplier assessment and selection, various levels of specification development, configuration (or coding for custom applications), and verification leading to acceptance and release for operation. Risk Management is applied to identify risks and to remove or reduce them to an acceptable level.

- **Operation:** System operation, typically, is the longest phase and is managed by the use of defined, up to date, operational procedures applied by personnel who have appropriate training, education, and experience. Maintaining control (including security), fitness for intended use, and compliance are key aspects. The management of changes of different impact, scope, and complexity is an important activity during this phase. Another aspect is the completion of a periodic review to determine if the system is still fit for purpose.

- **Retirement:** The final phase is the ultimate retirement of the system. It involves decisions about data retention, migration, or destruction, and the management of these processes.Retiring the system has major implications about how the data created in the system during its operational life remains available, enduring, and readable throughout the data retention period.

This Chapter focuses primarily on the GAMP Specification and Verification approach in the project phase of the computerized system life cycle as an important mitigation of risks related to data integrity within the data life cycle. Applying this approach is comprehensively discussed in *ISPE GAMP® 5* and is not duplicated here [2].

Figure 3.5: A General Approach for Achieving Compliance and Fitness for Intended Use (Specification and Verification Approach from ISPE GAMP® 5) [2]



3.2.2 Business Process Definition

The starting point for all activities within a regulated company must be to define and understand the business process. A business process is a group of related activities or tasks. A data flow diagram shows how data flows through a system or process.

There are various standards and tools available[2] to do business process mapping; in this Chapter simple block diagrams [28] are used for illustrative purposes only. The deliberate use of block diagrams here, in place of more detailed mapping tools, provides the freedom for a block to represent a task/activity, a computerized system, or some combination of both.

For each level of definition, basic examples are shown for a manufacturing process, a Quality Control (QC) laboratory process, and an eCRF process.

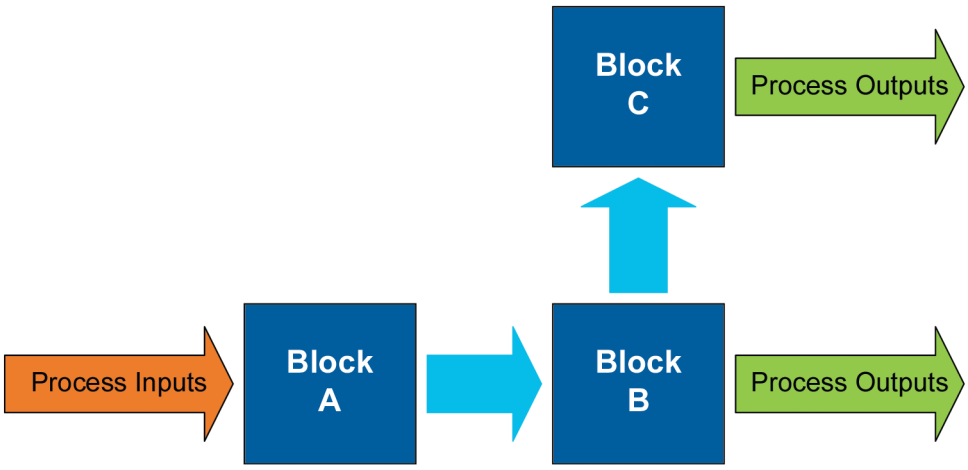
Figure 3.6: Business Process Representation

At the simplest level, a business process can be represented as a set of process inputs that are subjected to some process to become process outputs.

3.2.3 Breaking Down the Process

The next stage in understanding is to break the process down into smaller blocks.

Figure 3.7: Detailed Process



A block can involve a manual operation or a computerized system, and the process may pass from physical actions to data-only blocks, such that the process output may not always exit from the last block in the process.

[2] Such as Business Process Model and Notation, or Unified Modelling Language as used in Appendix 6 of this Guide.

Table 3.4: Block Definition Examples by Example Function

Function	Process Inputs	Block A	Block B	Block C	Process Outputs
Manufacturing	Raw materials	Weighing and mixing station	Sterilization and filling line	Collating run data from Process Control System (PCS) into EBR system as batch record	Finished goods from Block B, batch record from Block C
QC Laboratory	Samples from production	Sample preparation	Chromatographic analysis	Collating multiple analytical and physiochemical results to a Laboratory Information Management System (LIMS) system	CQAs for C of A from Block C
Clinical Study	Subject accepted into study, with Subject ID	First visit, interview on medical history and demography	Electrocardiogram, blood sample taken, data queries	Statistical Analysis System	eCRF data from Block C

3.2.3.1 Manufacturing Example

In the manufacturing area, the process inputs are the raw materials, which are delivered to the weighing and mixing station. Here, trained operators weigh out the active ingredients, then manually transfer them to the mixer, and add a specified quantity of Water for Injection (WFI) from the WFI loop. The operator records the material ID, lot number, and weight onto a paper form.

The mixer is manually started and continues for a specified time period, after which the aqueous solution is automatically pumped to the sterilization and filling line.

Sterilization is achieved using a combination of pre and final filters, with the final filter being a 0.2µm sterilizing grade filter. After the final filter, the solution is passed directly into the filling line where it is dispensed into heat-sterilized glass ampoules under aseptic conditions.

All of the Critical Process Parameters (CPPs) generated during the manufacturing process are recorded into an Electronic Batch Record (EBR) by the EBR system. The process outputs here are both physical (the injectable finished product in ampoules) and data (the batch record).

3.2.3.2 QC Laboratory Example

In the QC laboratory, samples from the production area are the process inputs; these are received and recorded into the sample register. An analyst prepares the samples for chromatographic analysis according to the governing SOP. The analyst manually records the sample ID, weight, pH, reagents used, and the vial ID into their paper laboratory notebook or controlled worksheets.

Multiple samples are collected into a sample set. The sample set information (sample ID, sample weight, etc.) is manually entered into the Chromatography Data System (CDS) and an instrument method is selected. The CDS initiates the analysis using the parameters from the instrument method and acquires the data from the chromatography instruments. The data is then processed with the CDS, with the final results passing to the Laboratory Information Management System (LIMS) where it is collated with results from the other analyses and physiochemical tests (the Critical Quality Attributes (CQA)) to produce the Certificate of Analysis (C of A).

The physical samples are destroyed after analysis so the process output from the QC laboratory process is simply data in the form of the C of A.

3.2.3.3 eCRF Example

At the clinical trial site, the process inputs are typically interviews with, and direct observations of, the study subjects, but also study subject data from the Electronic Health Record (EHR) system, from diagnostic systems (e.g., electrocardiograms or radiology systems), and from clinical laboratory results. The study nurse or Principal Investigator (PI) enters the data either directly or transcribes it from the EHR into the eCRF.

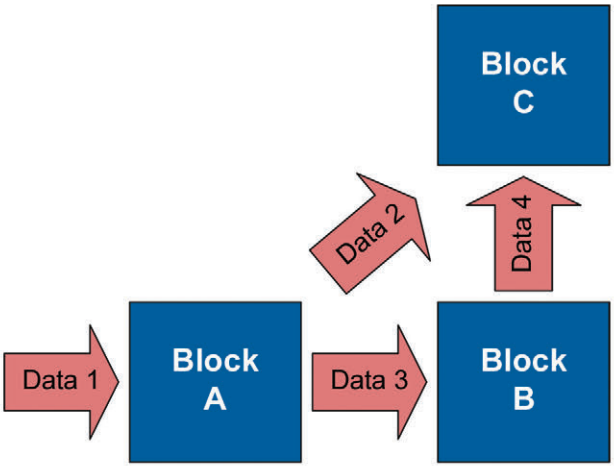
During data entry, edit checks are triggered by the eCRF requiring immediate correction of apparently implausible data (e.g., body temperature of 30°C for a living human). Once the data is submitted to the database, the data is typically reviewed by the sponsor's data management, who may raise queries back to the clinical site, which may lead to the correction of data by the study nurse or PI through a query process. After the last visit, each CRF is reviewed and signed off by the PI.

Sponsor oversight is performed through monitoring visits and source data verification, and audits performed by the sponsor.

The output of the process is the data in the eCRF system, which may be certified copies, but also any paper originals, and the original data in the EHR, diagnostic systems, or clinical laboratory results, which are the source data.

3.2.4 Data Flows within the Process

Figure 3.8: Data Flows within the Process



Having broken the high-level business flow into smaller blocks, it is now possible to determine the data flows between these blocks.

Table 3.5: Data Flows for Manufacturing Process

GxP Impact	Data 1	Data 2	Data 3	Data 4
High – Impacts GxP	• Material ID	• API weight • WFI volume • Batch Number • Alarms and Events	• Batch Number • Product type	• CPPs e.g., differential pressures • Alarms and Events
Medium – Supports GxP	• Material Lot Number	• Recipe used • Operator ID • Balance ID • Flow Meter ID	N/A	• Parameters and Setpoints • Operator ID
Note: Each organization must determine for themselves which records are High, Medium, or Low GxP impact. The example is for illustrative purposes only and should not be taken as a recommendation.				

Table 3.6: Data Flows for QC Laboratory Process

GxP Impact	Data 1	Data 2	Data 3	Data 4
High – Impacts GxP	• Sample ID	• Sample ID • Vial ID	• Sample ID • Sample weight • Vial ID	• Components • Amounts • Impurities
Medium – Supports GxP	• Sample • Register	• Analyst ID	• Reagents used • pH meter and balance used	• Methods • Audit trails • Analyst ID
Note: Each organization must determine for themselves which records are High, Medium, or Low GxP impact. The example is for illustrative purposes only and should not be taken as a recommendation.				

Table 3.7: Data Flows for eCRF Process

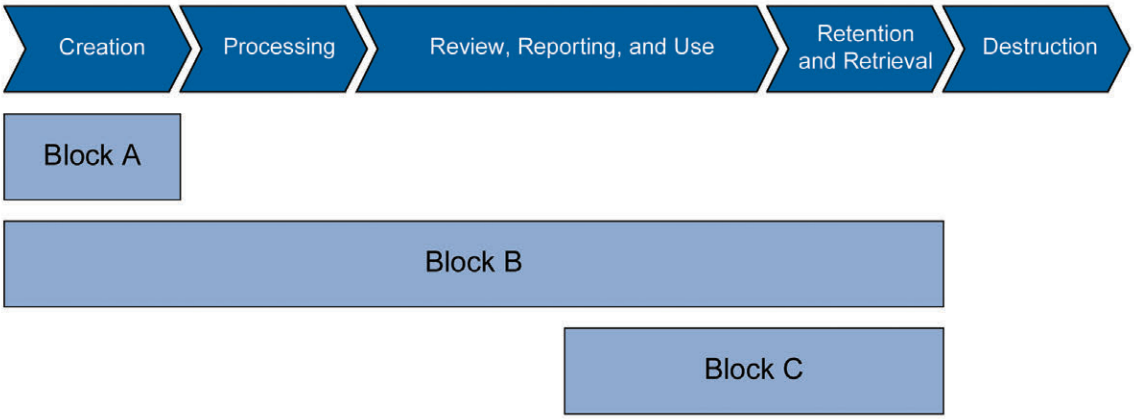
GxP Impact	Data 1	Data 2	Data 3	Data 4
High – Impacts GxP	• Subject ID	• Subject ID • Medical history • Demographic data	• Subject ID	• Updated clinical observations • Electrocardiogram data
Medium – Supports GxP	• Study ID	• Operator ID • EHR metadata on medical history	• Data on data sources • Queries	• Audit trails • Operator ID
Note: Each organization must determine for themselves which records are High, Medium, or Low GxP impact. The example is for illustrative purposes only and should not be taken as a recommendation.				

As the level of detail in the business process is broken down into smaller blocks (activities, tasks, or individual systems), it becomes clear which are manual and which are automated processes, as well as how the data is passed through the process and which data has the highest GxP impact. When assessing the integrity risks to the data, consider if the data can be changed or reprocessed within this activity.

3.2.5 *Mapping the Data Life Cycle*

Understanding the data flow provides the understanding of the data to map the blocks against the data life cycle.

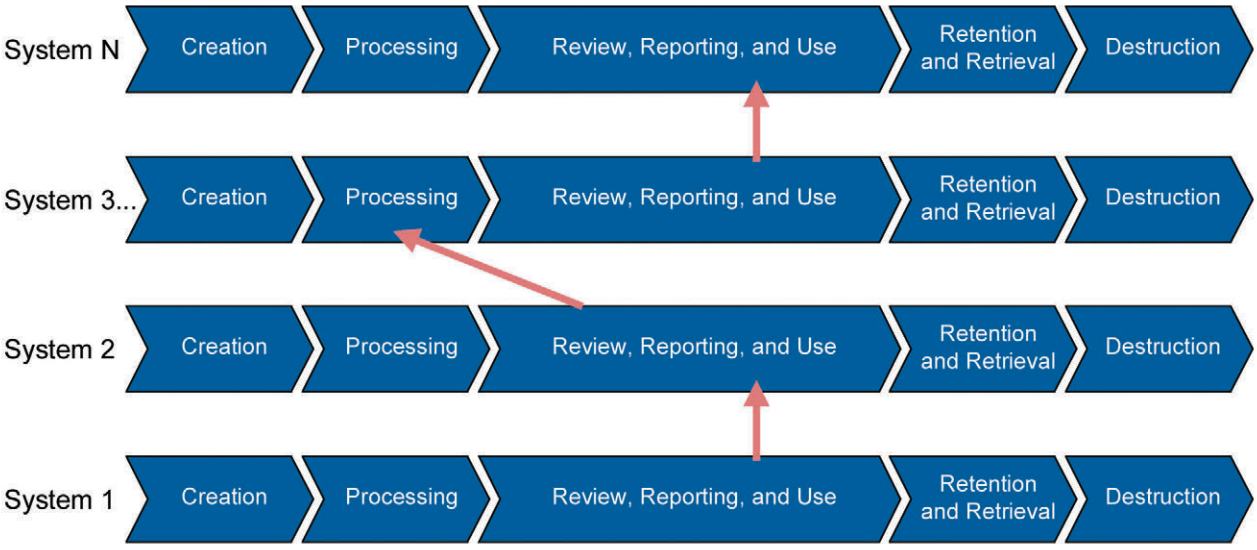
Figure 3.9: Subprocess Blocks within the Data Life Cycle



Data integrity issues in the early stages of the data life cycle, in a low-level process block, will pass up the chain of activities (systems) and compromise the integrity in the higher-level systems, such as LIMS, Enterprise Resource Planning (ERP), and the statistical analysis system.

Additional issues can occur when records or data are migrated, duplicated, or transferred from one computerized system to another, such as an extract of records exported to another application for analysis, review, trending etc., which can result in a loss of metadata and audit trails. It is therefore important to define (and document) across the data life cycle which records and data are considered the primary record and will directly support GxP-decision making.

Figure 3.10: Data with GxP Impact Passing up a Process Hierarchy



3.2.6 *Process Risk Assessment/Data Integrity Risks*

A (business) process risk assessment is a non-system-specific high-level assessment of the business process and data flow. It is aimed at identifying key process-level risks to patient safety, product quality, integrity of regulated records, and identifying the essential controls to manage those risks. Typically at this stage, no assumptions are made about the exact nature or functionality of any computerized systems supporting the process.

In this example the overall process, the blocks within each process, and the data flows through the process have been defined, allowing the process risk assessment to focus on risks specific to the systems and planned implementation.

Computerized systems generally do not reduce the severity of the harm of a risk within the business process (e.g., the severity of harm of losing GxP data is unchanged by the presence or absence of a computerized system); however, when appropriately implemented and operated a computerized system may reduce the probability of occurrence of the risk (e.g., reduce the likelihood of losing data) or increase detection of failure (e.g., audit trail entry recording the deletion of the data).

Table 3.8: Process Activities and Process-level Data Integrity Risks

Functions	Block A	Block B	Block C
Manufacturing	Manual Weighing • Accidental or deliberate recording of inaccurate values • Incorrect entry of mixing time • Incorrect materials used	Process Control System • Incorrect recipe created or selected • Sensor failures • Filter failure • Unauthorized operator intervention	EBR System • Data incorrectly or not received from lower systems • Unauthorized changes within EBR
QC Laboratory	Manual Sample Preparation • Accidental or deliberate recording of inaccurate values • Use of incorrect or expired reagents • Failure to follow SOP • Samples accidentally or deliberately switched with other samples or standards • Physical influence on the balance pan e.g., finger, off center weigh boat	CDS • Incorrect method created or selected • Sample set incorrect-ly defined • Instrument or column failures • Processing method unsuitable or manipu-lated • Reprocessing to re-place failed results • Non-reporting of failures	LIMS • Data incorrectly or not received from lower systems • Unauthorized changes within LIMS
eCRF	eCRF System • Incorrect study subject selected • Transcription errors from paper or from EHR	EHR/EDC System • Entry errors when responding to queries • Inadequate review by PI • Data entry or recording errors in EHR system, or in Electronic Data Capture (EDC) system, or laboratory (see QC laboratory example) • For EDC, copying of result files to save money on EDC runs	Statistical Analysis System • Incorrect transformations to SDTM or ADaM data standards • Inadequate changes of randomization list or of inclusion/exclusion lists • Hardcoding of test outcomes or estimator values in study-specific computational code

The process-level data integrity risks here include inappropriate human intervention, inadequate technical controls within computerized systems, and failures in the transmission of data between process activities whether by manual or electronic transfer.

Data integrity risks also include physical factors that can affect the accuracy of results, such as a sensor failure in a process control system, or an inadequately-maintained and uncalibrated instrument in a chromatography system.

3.2.6.1 Behavioral Controls

Behavioral controls are essential where there is a manual operation or manual intervention on a computerized system, as the manual intervention potentially carries a high risk to data integrity. Where such manual interventions are necessary and must be permitted, it is essential that a quality culture supporting data integrity is strongly promoted and supported through the organization as part of the holistic approach required to achieve data integrity.

Manual operations are likely to be covered by an SOP detailing how and when the interaction shall be completed and recorded. It is the quality culture within the organization, however, that determines if the SOP is followed carefully and conscientiously or whether the task and associated documentation are completed in the most convenient way possible.

For the eCRF example, the data integrity culture of the clinical site is an important factor to measure and consider during site selection.

Section 2.3.2 of this Guide discusses the behaviors that support ALCOA+ requirements and Section 2.1 examines the wider data integrity culture.

3.2.6.2 Procedural Controls

Section 2.2.2 of the *ISPE GAMP® Guide: Records and Data Integrity* [1] discusses the overall paradigm for procedural controls, with an emphasis on process-level and workflow-based procedures.

It is recommended that there are company or site-wide SOPs governing (but not limited to) the following activities with potential impact on data integrity and supporting general GxP compliance as part of the regulated company's Quality Management System (QMS):

- Training
- Support Services
- Performance Monitoring
- Repair Activity
- Periodic Review
- Change Control and Configuration Management
- Handover
- Security
- Incident Management
- Disaster Recovery

The *ISPE GAMP® Good Practice Guide: A Risk-Based Approach to Operation of GxP Computerized Systems* [29] provides detailed guidance on the essential content for such SOPs.

Both manual and computerized system processes require governing SOP(s) to define best practices that should be followed at all times.

Where computerized systems are used, SOPs are needed for:

- Routine use and data review and audit trail review, if required
- System Administration
- Access Management and Access Review
- Backup and Restore

For manual processes, SOPs or other procedures must clearly describe the scope and order of activities to be performed and provide instructions on how the tasks should be completed, documented, and reviewed.

In the QC laboratory example used in this Chapter, specific SOPs also need to cover:

- Integration (how to identify when the automated integration has failed to sufficiently resolve peaks or peak tails etc., and when and how should manual integration be used)
- Review of data electronically, including audit trails (how to detect data integrity issues)

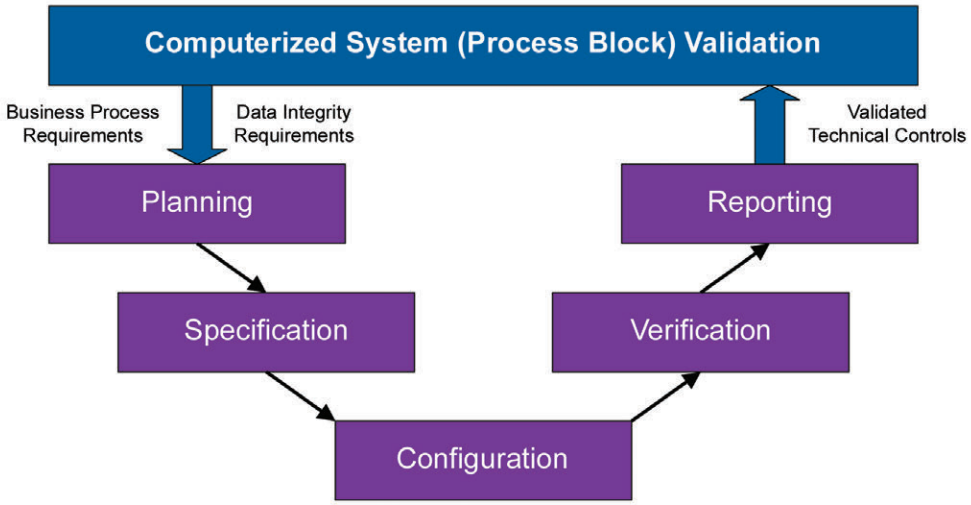
Personnel must be trained on the content and importance of all SOPs that they are to follow.

In the eCRF example, the clinical site personnel will not work according to sponsor SOPs, but according to training materials provided by the sponsor during the site initiation visit. In a hospital, the personnel entering data into the EHR or working with the diagnostic systems or in the laboratory will follow hospital procedures. Often, personnel are not trained during site initiation visits, but only the staff directly involved with the CRF and eCRF. Source data verification and monitoring activities need to consider these departments, and the monitors need to be trained so that they can challenge computerized systems and detect associated data integrity issues.

3.2.6.3 Technical Controls

Technical controls are achieved by the use of computerized systems in support of the business process. The understanding of process-level data integrity risks across the business process (e.g., from Table 3.8) can drive the initial definition of the essential requirements for the system. The computerized system implementation project then translates those requirements into technical controls during specification or configuration activities to manage the identified risks.

Figure 3.11: Business and Data Integrity Requirements driving System Specification and Validation



A computerized system implementation is only as good as the requirements defining what the system must do to support the business process, and how it must protect data integrity, patient safety, and product quality. The business process supported by the system will be subject to regulatory requirements that must be included in the requirements planning. Appendix 6 contains a detailed discussion on requirements planning.

A documented and detailed functional risk assessment may be performed to determine the extent and rigor of the verification activities during the project phase. *ISPE GAMP® 5, Appendix M3* [2] contains a widely-used methodology suitable for this purpose. The focus of the verification activities will be to confirm that the system as designed, implemented, and configured has provided effective controls to mitigate, minimize, or eliminate the risks to data integrity and/or to facilitate detection of data manipulation (typically through reports/queries).

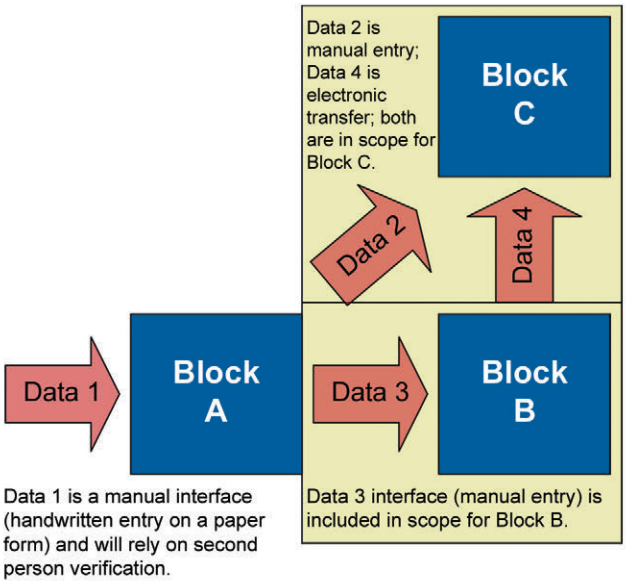
General technical controls are discussed in the *ISPE GAMP® Guide: Records and Data Integrity, Appendix D3* [1].

3.2.6.4 *Interfaces*

Interfaces are covered in detail in Section 4.4.

When planning the validation activities, clearly define which interfaces are included in the scope of each system's validation and confirm that all automated interfaces have been validated, and that all manual interfaces have robust procedural controls and, where necessary, second person verification to reduce the data integrity risk around this manual touch point.

Figure 3.12: Defining Interfaces within Validation Scope

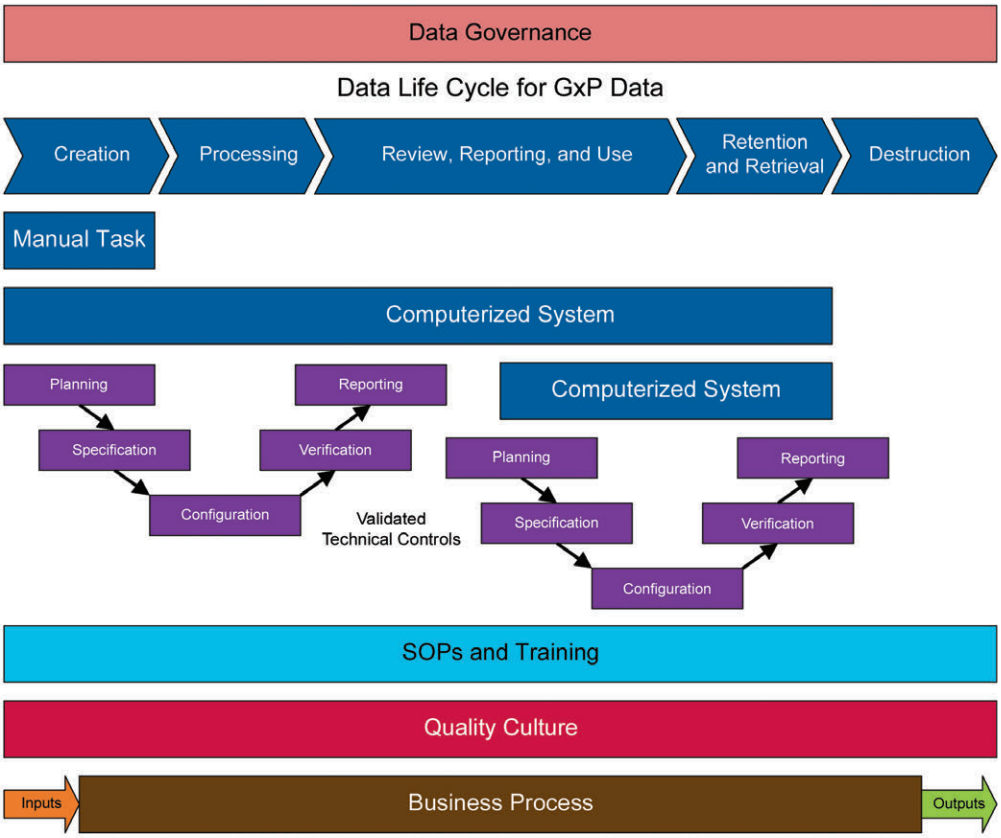


3.2.7 *Aligning Data and System Life Cycles*

Data integrity requires a holistic approach, combining technical, procedural, and behavioral controls to support data integrity throughout the business process. Within that data life cycle, there may be any number of manual tasks and computerized systems supporting the business process.

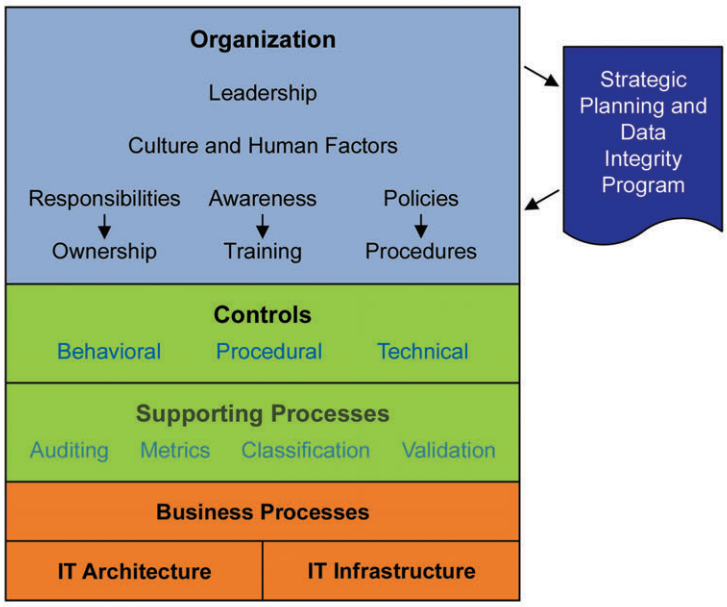
Figure 3.13 shows the alignment needed between the multiple systems, manual or computerized, that impact critical data during its life cycle.

Figure 3.13: Alignment of Data and System Life Cycles for a Single Data Life Cycle



It must be recognized that above the individual data and system life cycles there is an overarching data governance requirement, providing the essential holistic approach to safeguarding data integrity throughout the organization and across all data life cycles. This is represented in Figure 3.14, originally Figure 3.2 in the *ISPE GAMP® Guide: Records and Data Integrity* [1].

Figure 3.14: Data Governance Framework Essential for All Data Life Cycles [1]



4 Risk Management Approaches

4.1 Focus of Risk Management

The focus of any risk management approach is to identify potential risks to patient safety, product quality, and data integrity, and then implement controls to reduce, eliminate, or mitigate those risks or to increase detection. This is reflected in the Principles of Quality Management as defined in ICH Q9 [30]:

> *"Two primary principles of quality risk management are:*
>
> - *The evaluation of the risk to quality should be based on scientific knowledge and ultimately link to the protection of the patient; and*
> - *The level of effort, formality and documentation of the quality risk management process should be commensurate with the level of risk."*

The actions taken to identify and safeguard against risks to data integrity, patient safety, and product quality should inherently reduce compliance risks as the aim of any regulation is itself to reduce risks to data integrity, patient safety, and product quality; however, reduction of compliance risk should not be the primary driver in the risk management approach. It should be noted that patient safety is also intrinsically impacted by the integrity and quality of the data on which a regulatory decision is based.

This Section of the Guide is structured around the logical progression of a computerized system implementation project:

- Supplier selection and management
- GxP Computerized System
- System interfaces
- Access controls

Throughout all of these activities, the potential risks to patient safety, product quality, and data integrity should be evaluated and addressed through the requirements planning process. (Further detail on requirements is contained within Appendix 6.)

4.2 Supplier and Third-Party Management

4.2.1 Introduction

ISPE GAMP® 5 [2] describes the activities and responsibilities expected of the supplier in the provision of products and services to a regulated company, but not the specific considerations required for data integrity.

With the increase in the use of electronic computerized systems, it is no longer necessary for all people contributing to a body of work to be co-located. The ability for work to be performed at multiple sites has enabled regulated companies to increase their use of third parties.

This Chapter makes a distinction between third parties (e.g., system integrators or service providers) and suppliers that deliver a product to a regulated user as there are some differences between the required activities. A third party may:

- Be an external entity or an internal group of the regulated company (e.g., internal IT departments)
- Manage data and/or systems
- Act as a product developer if they provide the services through applications or systems developed and owned by them such as SaaS solutions

This Chapter describes the technical and procedural measures that need to be in place to ensure the integrity of the data under third party control and the controls necessary to ensure the integrity of products provided by a supplier.

Regardless of the use of third parties, the responsibility for compliance remains with the regulated company.

4.2.2 Third Party Quality/Technical Agreements

In the case of an internal third party (same legal entity as the regulated company), a Quality Contract/Technical Agreement is needed, but not a legal contract.

If the third party is a separate legal entity, a legal Quality Contract/Technical Agreement must be in place that defines the obligations of the regulated company and the third party with respect to data integrity.

Written Quality Contracts/Technical Agreements between the regulated company and the third parties are a requirement from some regulatory agencies [17, 31] and an expectation from others; whatever the specific driver, they are needed to clearly establish the duties of each party as good business practice. The requirement for a Quality Contract/Technical Agreement between the entities is critical to ensure that any work provided by a third party meets the requirements of the regulated company including requirements to ensure the integrity of the data.

The third party may itself not be governed by any healthcare regulation, and thus may be unaware of, nor be able to comply with, such legislation. Additionally, the third party may be subject to other regulations or licensing bodies.

Even if the third party is not subject to healthcare regulations when they enter into a Quality Contract/Technical Agreement with a regulated company, they must be aware of the requirements of the environment in which they are working. Where the third party is providing a non-regulated service or activity, they are not required to follow the principles and guidelines of GxP, but rather the good practices and frameworks (e.g., ISO [32], IT Infrastructure Library (ITIL®) [33], or even *ISPE GAMP® 5* [2] supplier good practices) that apply to their activities/sector, legal requirements, and requirements as stipulated in agreed contracts.

Where the third party is undertaking regulated activities on behalf of the regulated company (e.g., contract manufacturing or testing, or contract research organization), the regulated company should ensure the Quality Contract/Technical Agreement stipulates, as a minimum, that:

- The principles and guidelines of the appropriate GxP are followed by the third party
- The third party must submit to inspections from competent authorities and to immediately inform the regulated company on receipt of a notice of inspection when the scope includes any products or services undertaken on behalf of the regulated company
- The third party must keep the regulated company updated about the inspection progress and findings throughout the duration of the inspection
- The third party shall update the regulated company of any inspectional findings on products or services not

related to the regulated company that may potentially impact the products or services undertaken on behalf of the regulated company

- In the event of an inspection of the regulated company, the third party must provide support as required, including access to original data

If the third party is found to be negligent in their procedures or practices, any regulated company using the third party is potentially vulnerable to the same deficiencies and must determine the risk to their data or products.

The level of detail within a Quality Contract/Technical Agreement should be commensurate with the risk and type of service provided by the third party. The Quality Contract/Technical Agreement defines the roles and responsibilities as they impact the services, and specifically data integrity.

A key element of data integrity is accountability, that is, it must be clear who performed what, when, and how, so it is necessary for the third party to ensure accountability in all activities through their QMS.

The restrictions on systems and data access should be defined in the Quality Contract/Technical Agreement for both companies, including sufficient granularity to ensure the integrity of the data.

Separate and additional to the underpinning contracts discussed above would be the Service Level Agreement (SLA) and/or Operational Level Agreement (OLA). SLAs and OLAs are discussed in detail within the ITIL® [33].

4.2.3 Assessment of Third Parties

The regulations state that the QMS of the regulated company include the control and review of any outsourced activities.

With regard to data integrity, the quality system of the regulated company should stipulate not only how third parties are managed, but how aspects of data integrity are addressed by the third party.

Third-party assessments must include data integrity elements. The third party must establish controls similar to the controls within the regulated company since the third party performs activities on behalf of the regulated company.

The assessment conducted by the regulated company determines the maturity and competence of the third party to successfully carry out the required activities prior to establishing the contract. The extent and rigor of the assessment is based upon the risk to patient safety, product quality, and data used to make regulated decisions from the work being contracted.

The assessment should determine if the third party:

- Has adequate premises, equipment, and competent personnel with the appropriate level of knowledge and experience

- Uses qualified infrastructure, qualified analytical instruments and production equipment, and appropriately validated applications (In the case of a service provider, ITIL® [33] approaches may be applied as an alternative to infrastructure qualification.)

- Maintains and follows an established QMS that adequately addresses the necessary regulatory requirements including the essential data integrity controls

- For third parties undertaking regulated activities on behalf of the regulated company, the third party has completed a maturity assessment per *ISPE GAMP® Guide: Records and Data Integrity, Appendix M2* [1], and that assessment shows that data integrity is adequately supported within the third-party organization.

Suppliers experienced with the regulated life sciences industry are more likely to have the appropriate quality practices to meet the needs of regulated organizations.

Often the regulated company may need to partner with the third party to train them or to help them establish the necessary controls, and the third party may benefit from applying processes and procedures analogous to those in place at the regulated company.

In cases where the third party hosts data, it is necessary to detail in the Quality Contract/Technical Agreement who owns the data, and how the owner will have access to the data during its life cycle. For example, an analytical laboratory at a hospital may perform analyses in connection with a clinical study. Data may be kept in paper format or in electronic format, and it is not unusual to find that the electronic repositories are owned by different legal entities, and that the paper archives are not under the control of the hospital.

A sponsor of a clinical trial should ascertain the locations of the data and the ownership for these locations. It is important to ensure data access (and control over the data) through procedural and contractual controls as data loss or the inability to access data can impact lives.

When performing assessments of hosting services, it is essential to determine if the third party hosts the data or if they have subcontracted the hosting to an Infrastructure as a Service (IaaS) provider. If they have subcontracted the hosting, the regulated company must have assurance that these services have been evaluated and that there are appropriate quality and technical controls in place including security, backup, and disaster recovery.

The regulatory expectation is that the regulated company monitors and reviews the performance of the third party. Periodic assessment of the third party should be allowed within the Quality Contract/ Technical Agreement. The frequency and rigor of the assessment should be based upon the risks associated with the contracted activities or services.

The regulated company is responsible for reviewing and assessing the records and results related to the outsourced activities. If the data is transferred from the third party to the regulated company, the integrity and security of the data must be ensured during transfer. The physical and logical security (including encryption) for the transfer of confidential or private information should be defined in the contract and based on existing documented procedures.

The regulated company should define within the Quality Contract/Technical Agreement if the third party can subcontract services. If subcontracting is allowed, it is important to ensure that the subsupplier or subcontracting party has been properly evaluated by the third party and has processes and procedures to ensure the integrity of the data.

In the same way that the regulated company is required to evaluate the third party prior to entrusting any work to them, the third party should not subcontract any work without evaluation of any subcontracted provider. Often contracts require approval by the regulated company prior to the use of a subcontracted service.

If the subcontracted activity is considered high risk or critical, the Quality Contract/Technical Agreement should include a clause that permits the regulated company to audit the outsourced activities performed by the mutually-agreed subcontractors. Where these subcontractors are involved in activities with significant impact on data integrity, the regulated company may be inclined to perform such an audit and not rely on the third parties' subcontractor assessment.

It is important that the third party not make changes to the contracted services outside the terms of the Quality Contract/Technical Agreement that may adversely affect the quality of the outsourced activities for the regulated company.

For example, if the contract states that the third party shall only use permanently employed staff for the contracted services, it is a breach of contract to use temporary staff. Using temporary staff even for activities such as manning an IT helpdesk is not uncommon, but this may have serious implications on data integrity, for example where that helpdesk manages user access accounts.

The regulated company should assess the potential impact the third party, or their product, will have on the overall data integrity maturity level of the regulated company itself, and fit with the regulated company's data governance approach.

4.2.4 QMS

It is the responsibility of the regulated company to assess the legality, suitability, and competence of the third party to successfully carry out the assigned activities prior to outsourcing the activities. The regulated company is also responsible for ensuring, by means of the contract, that the activities performed by the third party assure the integrity of the data.

The Quality Contract/Technical Agreement defines the procedures to be followed and related documentation. The regulated company may insist that their procedures be used, although often it is more appropriate for the third party to use their own procedures.

The assessment performed by the regulated company identifies gaps between the regulated company's QMS and that of the third party.

Where mitigation actions are implemented to address the identified gaps, care should be taken that these do not conflict with the quality system of the third party. If the service is an SaaS or IaaS, often the procedures from the third party are used because the regulated company does not have the expertise nor the appropriate procedures in place.

Within a company there may be different legal entities that provide services to each other. For example, it is not uncommon to find information management outsourced to a legally independent company, even though that company is physically located on or within the premises of the originating company.

The third party may be owned by the contract giver, such as a service group/department. Depending on the legal relationship, the level of control of the contract giver over the third party varies, which in turn impacts the type and content of the Quality Contract/Technical Agreements to ensure appropriate controls over data integrity.

4.2.5 Supplier's Quality System

When a regulated company obtains hardware or software from a supplier, it is the responsibility of the regulated company to ensure that the system is fit for their purpose in their environment before using it.

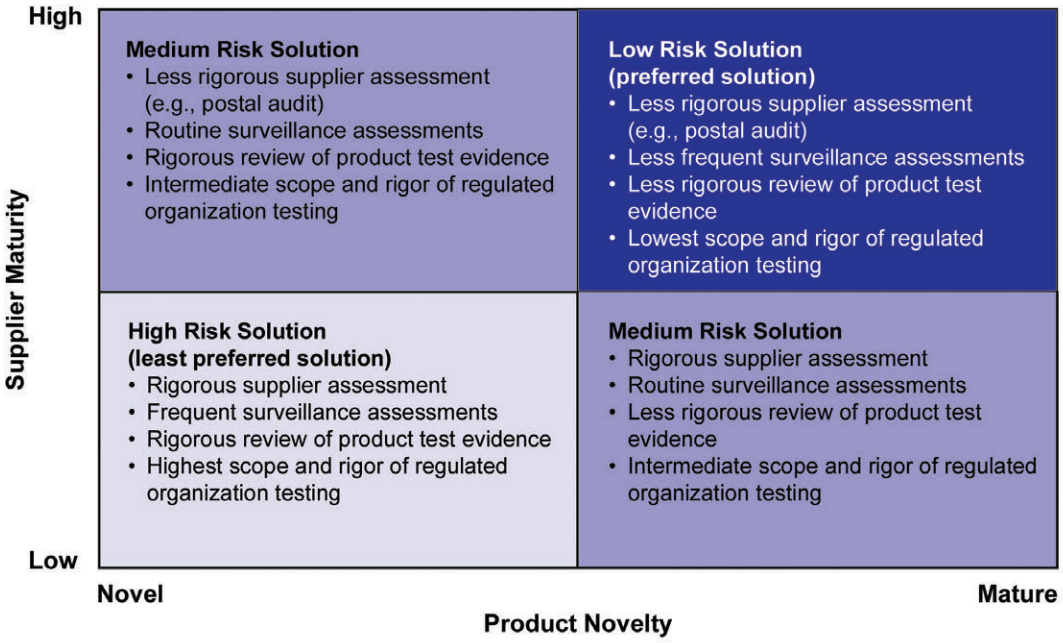
Similar to third parties, regulated companies must assess suppliers and their products to ensure that the product has the technical capability to ensure data integrity. The regulated company should, based upon a risk assessment, determine the extent of a supplier assessment and the scope of the review. The risk assessment should factor in the type of system, the intended use of the system by the regulated company, and the availability and findings of any independent assessments such as SSAE16 standards [34] and SOC 2 reports [35], and/or PCI attestation [36] and ISO 27001 certification [14].

The supplier assessment should evaluate the supplier against good practice activities as listed in *ISPE GAMP® 5, Table 7.1* [2].

As with third parties, suppliers are usually not governed by healthcare regulations and may be unfamiliar with the regulations governing the regulated company. Though it is assumed that the supplier will have a well-documented and implemented QMS, their procedures may not fully meet the expectations of the regulated company.

During the assessment, the supplier's procedures are evaluated as well as their knowledge of data integrity principles. The assessment should establish the relative maturity of the supplier and the product.

Figure 4.1: Supplier and Product Maturity [37]



It should be remembered that a supplier may provide multiple products, and each may be the subject of a different QMS and provide varying degrees of data integrity controls.

The regulated company evaluates the supplier's system against the regulated company's requirements including the data integrity controls for GxP-regulated computerized systems as listed in *ISPE GAMP® Guide: Records and Data Integrity, Appendix D1* [1].

Such an evaluation can ofte be done by reviewing the technical product documentation to determine if the software product or computerized system will support the GxP technical requirements and can be implemented within the business process of the regulated company. This evaluation should assess the level of testing performed by the supplier and determine if any can be leveraged by the regulated company.

The result of the supplier assessment may be a series of actions for one or both parties.

Activities for the supplier may include:

- Additions to the supplier's QMS to include GxP and data integrity expectations
- Training provided for employees involved in life cycle activities in relation to data integrity, regulatory expectations for security, audit trails, etc.
- Assessment of the supplier's maturity level

Activities for the regulated company may include:

- Risk assessment of any nonconformances to the regulated company's User Requirements Specification (URS), arising from a lack of technical controls in the supplier's system
- Revision of the Validation Plan to cover additional activities required based upon review of the supplier's documentation

- Additional testing by the regulated company to address data integrity requirements
- Additional testing by the regulated company to ensure the system is fit for purpose and supports the integrity of the data

4.2.6 Problem Management by Software Suppliers

The software supplier should have a clearly defined problem management process that includes a customer notification process in the event of a critical defect in the software. The software defects also need to be documented, resolved, and tracked to completion.

The defect tracking system should allow regulated companies to report defects, and uniquely identify and track them to completion. Traceability is important to ensure that issues are resolved in a timely fashion. The supplier needs to determine the root cause of problems to prevent recurrence.

If the software supplier does not conduct comprehensive software testing, then the regulated company must make sure that it is covered in-house. These controls can be anything, including completely validating the software in-house if the software supplier's quality systems are inadequate. Even if the supplier's quality system is excellent, user acceptance testing needs to be done in-house because the electronic environment and data volumes are significantly larger in a pharmaceutical company than at most test environments at the software vendor.

4.2.7 Free and Open-Source Software

Where a regulated company opts to use an open-source or freeware software product, there is unlikely to be a contract with the provider, and therefore the regulated company must exercise due diligence in verifying all aspects of the open-source software product's data integrity controls.

Additionally, products supplied as open source are provided "as is" with no ability to request bug fixes or updates. It may not be possible for the regulated company to gain knowledge of the provider's QMS and software development life cycle. (See *ISPE GAMP® 5, Appendix M2* [2] for further information about vendor assessment.)

It must be remembered that the ultimate responsibility for the use of the software rests with the regulated company.

4.3 GxP Computerized Systems

4.3.1 Introduction

The gap assessment performed on existing systems (see Section 2.4) aims to identify systems with unacceptable levels of risk to data integrity, patient safety, and product quality caused or increased by a lack of technical control or inadequate life cycle implementation. Typical examples are:

- System has been in use for an extended operational life and hardware and software can no longer be maintained or upgraded
- Software in use is not at or close to the currently available version
- System relies on outdated hardware and infrastructure
- Changes to the system have accumulated over time (e.g., uncontrolled software changes, changes of use)
- Limited investment and support for the system (e.g., insufficient licenses and/or no training for new staff)
- System requires combined electronic and manual record management practices to meet business needs and regulatory expectations

- System is operated under outdated practices, procedures, and functions
- System has been preserved for tasks or data that cannot be migrated
- System was purchased and implemented without due regard for data integrity controls, may have initially been believed to have no GxP impact, or both
- System was supporting a process not regulated by GxP but by change of product use now supports a process that falls under GxP

It is recommended that a documented review of all new, planned, or existing systems managing regulated records be performed to assess technical controls and procedural requirements for managing data integrity. Key areas to assess are contained in *ISPE GAMP® Guide: Records and Data Integrity, Appendix D1* [1]. Where a lack of technical control associated with increased risks to data integrity, patient safety, and product quality are identified, the remediation approaches in this Section may be appropriate.

Regulated companies must implement a robust process to address the requirements of records and data integrity. This process must involve an effective combination of suitable strategies and management practices to replace and/or remediate systems in order to protect data integrity, patient safety, and product quality in line with the quality risk management paradigm in ICH Q9 [30].

While reducing risk to data integrity, product quality, and patient safety is the primary driver for replacing systems, every opportunity should be taken to include tangible business benefits, such as efficiency improvements and waste reduction, obtainable from implementing new, compliant, and validated systems to allow the use of technical controls supporting fully electronic records and electronic signatures.

4.3.2 Drivers for Change

4.3.2.1 Technology Advances

The technologies supporting the regulated life sciences industries have leveraged and utilized significant advances within:

- Measurement and control devices
- Enterprise information management
- Software operating platforms and security
- Development and integration of applications/interfaces
- Service models including cloud (SaaS, PaaS, IaaS) and virtualization
- Data storage and analytics
- Blockchain and similar evolving technologies

These advances have driven the need for increased computer processing, capability, hardware, integration, and capacity. New computerized systems can accommodate this need within the concept and project stages of the system life cycle; however for older systems, addressing this need requires upgrades, changes, and replacement within the operational phase of the system life cycle.

4.3.2.2 Regulatory Progression

The regulatory expectation for integrity of data is not new, however, recent years have seen an increase in enforcement actions based on data integrity issues.

Regulatory authorities' expectations toward systems have also required that where technical safeguards cannot be met, a suitable upgrade (or replacement) must be considered:

MHRA 'GXP' Data Integrity Guidance and Definitions March 2018 [5]:

> *§6.13 "Where systems do not meet the audit trail and individual user account expectations, demonstrated progress should be available to address these shortcomings. This should either be through add-on software that provides these additional functions or by an upgrade to a compliant system. Where remediation has not been identified or subsequently implemented in a timely manner a deficiency may be cited."*

This expectation to upgrade the system is a direct consequence of **Article 23 of the EU Directive on Medicinal Products 2001/83/EC [38]:**

> *"1. After a marketing authorisation has been granted, the marketing authorisation holder shall, in respect of the methods of manufacture and control provided for in Article 8(3)(d) and (h), take account of scientific and technical progress and introduce any changes that may be required to enable the medicinal product to be manufactured and checked by means of generally accepted scientific methods."*

While not subject to the EU Directive, Australia has also indicated an expectation that manufacturers update systems, as stated in the **Australian Government Department of Health Therapeutic Goods Administration – Data Management and Data Integrity Policy [10]:**

> *"Where manufacturers identify gaps or vulnerabilities in the current controls, it is expected that you develop and document appropriate corrective and preventative actions for resolution. This may include the need to update the quality management system, equipment and software in cases where the in-use systems are unable to meet current expectations for Data Management and Data Integrity."*

The definition and scope of data integrity expectations is also expanding as evidenced by the number of regulations and guidance documents issued by various competent authorities. For example, the MHRA have followed their March 2015 GMP-only [39] guidance with a 2018 document applicable to all GxPs [5].

Health Canada is the first regulatory body to have updated their GMP specifically to add data integrity focus [9]:

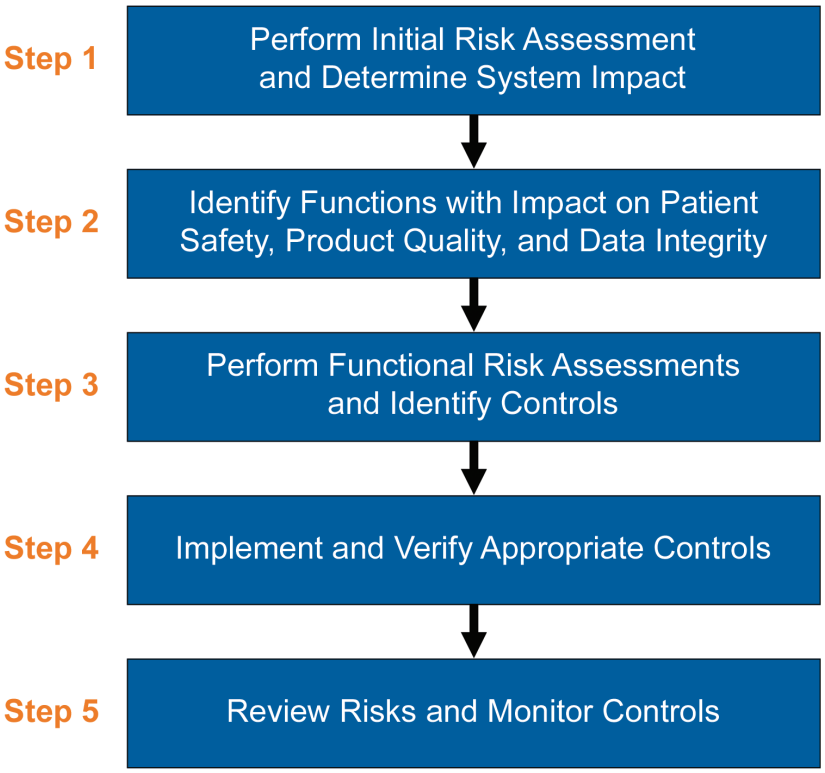
> *"Records must be reliable, complete, consistent and accurate.*
>
> *You must establish a data governance system to ensure controls are in place to prevent and detect data integrity issues throughout the product lifecycle. This includes:*
>
> - *having policies and standard operating procedures that clearly indicate management's expectations for how data should be acquired, modified, reviewed and stored*
> - *validating and maintaining equipment and associated computer systems*
> - *checking the preventative measures put in place periodically to verify their implementation and effectiveness*
>
> *These are standard principles under a pharmaceutical quality system, regardless of the media used (e.g. paper records or electronic records)."*

4.3.3 Risk Assessment of Computerized Systems and Data

4.3.3.1 Risk-based Approach to Mitigation and Control

The gap assessment within the corporate data integrity program, Section 2.4, may have identified systems with inadequate technical features and controls to support data integrity. The next step is to apply quality risk management to both the system and the records that it creates, processes, reports, transmits, or stores.

Figure 4.2: Quality Risk Management Process [2]



The risk management process should:

- Identify the system impact, based on an understanding of the business processes, process risk assessments, user requirements, regulatory requirements and known functional areas. System impact should include identification and use of GxP data within the system.
- Identify the functions that impact patient safety, product quality, and integrity of data supporting regulated decisions
- Perform the functional risk assessment and identify controls (behavioral, procedural, or technical) that can be instituted
- Implement and verify the control measures
- Review the risks and routinely monitor the controls (periodic reviews, data audits, etc.)

4.3.3.2 Computerized Systems Risk

Performing a risk assessment for a new computerized system is a well-established process integrated as part of the system life cycle approach. (Refer to *ISPE GAMP® 5, Chapter 5* [2] and the *ISPE GAMP® Guide: Records and Data Integrity, Appendix D1* [1].) This risk-based approach may be used to identify and quantify data integrity risks and define requirements (controls).

For existing systems, the risk assessment process can follow a similar approach but must also consider the existing business process (workflow) and the current data life cycle.

The *ISPE GAMP® Guide: Records and Data Integrity* contains additional guidance for identifying data integrity risks in *Chapter 5* and *Appendix D1* [1], and assessing the system use and technical and procedural controls against the verification questions contained in *Appendix D1* of the same Guide [1].

Control and corrective actions determined for areas of weakness require a suitable implementation approach that is technical, procedural, training, replacement, etc. The corrective activities should be incorporated into the QMS or Corrective and Preventive Action (CAPA) system.

4.3.3.3 Records Risk

The risk assessment of records and data must be based upon clear business process understanding, data life cycle and data flow knowledge, and awareness of the GxP decisions that will be made based on the data. The risk assessments should be periodically reviewed. Key risks that can compromise data integrity include:

- Poorly designed or implemented systems and poorly defined practices
- Incorrect data management across data life cycle stages
- Undefined data status (primary source, supporting, copy)
- Inappropriate data safeguards and controls

Managing data across the data life cycle stages requires identification of the record and data types supporting GxP decisions, and the impact of that data. Examples of record and data types include:

- Identification (e.g., material, label, reference)
- Original data
- CPP/CQA (e.g., results, in-process measurements)
- User account levels and access rights
- Event data (e.g., activity, user entry, training)
- Metadata (e.g., user IDs, audit trail entries)

For each record type it is important to define the primary source upon which the GxP process and decisions are made. Refer to Chapter 3 for further discussion on record and data types.

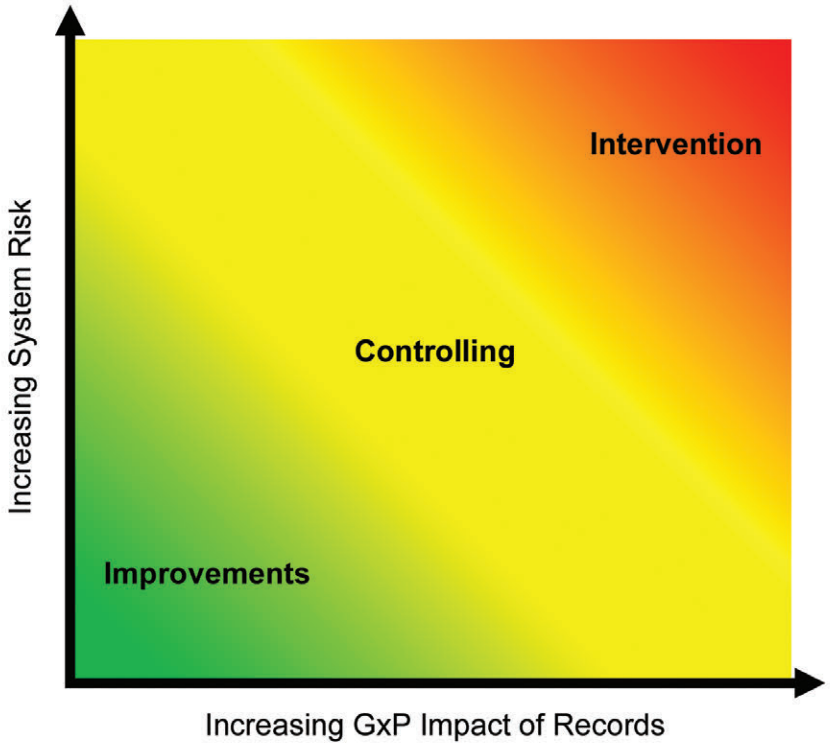
Evaluation of record and data integrity risks should focus on the impact to patient safety, product quality, data supporting regulated decisions, and the business process, considering risks from people, process, and technology on the system, records, and data.

4.3.4 Systems Remediation

4.3.4.1 Risk Management

There are likely to be multiple systems requiring remediation, some of which may have similar levels of risk, or may be dealing with records of similar GxP impact. Each organization must decide whether the remediation priority is determined by the risks associated with the computerized system, the GxP impact of the record, or some weighted combination of the two.

Figure 4.3: Example of a Risk-Based Continuum for Remediation



Implementation of remediation actions should be commensurate with the impact of the system and the risk to data integrity, product quality, patient safety, and regulated decisions. It is important to ensure the correct system stakeholder involvement with the risk management actions; the roles in Table 4.1 are defined in the Roles and Responsibilities in Section 2.2.

Table 4.1: Examples of Remediation Actions and Management Involvement for Systems

Remediation Actions	Improvement	Controlling	Intervention
	• Improvement plan specific to system • Actions and activities defined as part of QMS and/or Operational Excellence • Routine auditing of system and improved practices	• Additional quality oversight and controls relating to daily use • Increased review and checks relating to daily use • CAPA to improve data integrity with specific metrics and measures in place to assess the effectiveness of the actions	• Prohibit further use of system • Change or modification to the system to ensure a "step change" increase in data integrity robustness • Identify an alternative workflow or process that can quickly replace the system.
Management Level Involved	• Process Owner • System Owner	• Site quality function • Process Owner • System Owner	• Senior Management • Site quality function • Process Owner • System Owner
Reporting Measures Needed	As part of QMS and/or Operational Excellence (improvement)	As part of QMS and/or system-specific action plan	Quality council review system-specific action plan as part of site risk profile
Escalation and Vigilance Required	Routine level, internal audit, and progress review of activities	Escalated level, frequent review of action plan, assessment of use and processed product	Highest escalation level, immediate actions relating to use and processed product, and mandatory reporting to relevant health authorities

When focusing on the risk management of computerized systems, it is important to also retain visibility of the overall business process. Section 3.2.6 discusses the behavioral, procedural, and technical controls that can be used to remediate risks.

4.3.4.2 Develop a Detailed Remediation Plan

A detailed remediation plan is the starting point to effective implementation of the necessary data integrity mitigation actions and controls. The plan details the actions to be taken, responsible personnel, and the dates for action completion.

It is important to remember that the remediation plan can only offer a means to reduce or eliminate data integrity risks for data or records created following the completion of the remediation, but will not remediate integrity failures in previously created data or records. Any dependencies between associated actions should also be considered.

Where a decision to replace, remediate with third-party software, or upgrade a system has been taken, any associated remediation activities from the gap assessment of the original system should be reconsidered and re-evaluated.

The execution of the data integrity remediation plan for a site/department/area could be a significant amount of work; therefore, additional resources (budget and personnel) to achieve the plan in an acceptable time period may have to be escalated to senior management for approval.

Once the cost of remediation is determined, it may be necessary to review the plan and adjust priorities based on the expense and effort required against the level of risk involved with the ongoing use of the system.

This is not to say that financial investment is a justification to not replace a system; rather that the order in which systems are replaced should be based on the level of risk (to data integrity and the ultimate impact on patient safety, product quality, or a regulatory decision) and the overall department or site risk. Some examples of such considerations are:

- Replacing multiple workstation systems with a single networked solution provides remediation for all of those workstations in one project, whereas upgrading each one individually is cost prohibitive.

- Investing in a third-party solution that reduces the data integrity risk priority from high to medium across a wide range of outdated standalone systems may gain more overall risk reduction than investing the same amount of money to replace one or two systems to reduce them from high to low.

4.3.5 Examples of Risk Reduction

4.3.5.1 Risk Reduction and Remediation

Remediation of identified data integrity risks may require a combination of multiple mitigating actions, such as:

- Implement a procedural control or workaround to strengthen a weak technical control (this may only be acceptable as an interim measure for a limited period of time until the system is upgraded or replaced)

- Create a well-defined workflow and review process

- Change the system use, configuration, or architecture to improve data security

- Leverage other systems upstream or downstream within the data life cycle to reduce the requirements for the existing system, such as using a dedicated archiving system for record retention rather than relying on maintaining the record securely in the original system indefinitely

- Upgrade the existing system to the latest version if the new version possesses all or most of the technical controls lacking in the existing system (discussed below)

- Implement a third-party solution to provide the missing technical features (such as individual logon or audit trail), or to capture the data into a secure system immediately after creation (Note: this is only a partial mitigation as the data remains at risk during creation)

- Replace the existing system with a system that possesses most of the required technical controls (discussed below)

If most risks are, or can be reduced to, an acceptable level as defined by the organization, the system can be retained for ongoing use, subject to the implementation of any necessary remedial actions (technical and/or procedural). If there are unacceptable risks that cannot be mitigated, then upgrade or replacement of the noncompliant system is required.

Case studies are included regarding the management of data integrity risks with a potential impact on product quality, patient safety, or regulated decisions for different system types:

- Appendix 11 – Case Study: DBA and Security Controls for an ERP System in a Medical Device Manufacturing Environment

- Appendix 12 – Case Study: Laboratory Computerized System

- Appendix 13 – Case Study: Uncontrolled Spreadsheet

- Appendix 14 – Case Study: Process Control System
- Appendix 15 – Case Study: Business Application System

4.3.5.2 Quick Wins

All GxP computerized systems and associated records and data integrity practices should operate robustly with consistently low levels of data integrity risks. Setting best practices across the company and systems is an important step toward building a sustained compliance status and ensuring product quality and patient safety, and facilitating regulatory decisions based on data whose integrity is assured.

As part of a corporate governance program, the following approaches should be considered company-wide:

- Personnel are trained on good data integrity practices and the principles of ALCOA+ to assure the data integrity risks within each area are well understood and controlled.
- Awareness of primary records and the data life cycle for all data used for GxP decisions, including knowledge of where and how all data is created and stored (paper or electronic, archives, etc.), and how it fits within the data life cycle.
- Robust security access controls for all computerized systems are established.
- Modifications to time and date stamp are not permitted.
- Setting of granular privileges is used to restrict specific actions to a set of specified users.
- User privileges/profiles are correctly defined and implemented for the level of responsibility and actions performed, based on the principles that:
  - Each user has the least privileges needed to do their daily job.
  - No one with a direct interest in the content of the data is able to delete data.
- Backup and retrieval tests are verified, traceable, and applicable across the software versions used, and periodically tested.
- Archive and restore tests are verified, traceable, and applicable across the software versions used, and periodically tested.
- Where there is full Electronic Records and Electronic Signature (ERES) functionality available, but not utilized in an installed system, a project is implemented to move from paper records to fully electronic. If this ERES functionality has not been validated, validate the ERES functionality before the move is completed.
- A procedure is established for signing paper printouts where electronic signatures are not available. The procedure should describe the link between the electronic and signed paper and confirm the requirement to maintain the signed paper record (or a verified true copy of the signed paper record) as well as the original electronic record throughout the retention period.
- Controlled templates or forms are created to provide consistency in manual recording of data as part of the hybrid system, providing sufficient space for manual data entries and clearly describing what data should be recorded, when and where. The form includes space to record any second person verification of manual data entries and crosschecks against original source data where required based upon risk.
- Methods, templates, and user-defined equations are verified and brought under formal change control.

- Changes to master data are controlled according to GxP impact and implemented under change control.
- Disaster recovery and business continuity plans are documented and tested and periodically verified.

If these controls can be achieved via existing SOPs that have to be reviewed, or through new ones, SOPs should be created or updated in a timely fashion, verified, and issued for use rapidly to get the controls in place, including training the users on the SOPs.

4.3.5.3 Hybrid Situations

The *ISPE GAMP® Guide: Records and Data Integrity* [1] defines a hybrid situation as a situation where paper and electronic record and signature components coexist. This is typically where data is held electronically but a summary record is printed for the purposes of affixing a handwritten signature in lieu of electronic signatures. The data integrity risks with this situation arise from ineffective signature record linking leading to uncertainty around which records have been signed, and the ability to change an electronic record without a corresponding change to the approved status by handwritten signature.

WHO [3] identifies a hybrid approach as non-preferred and encourages the implementation of electronic signatures (see Appendix 5 for regulatory definitions).

Recent common misusage and misconception has used hybrid to describe existing computerized systems that lack technical control being partially remediated by the use of paper-based processes, such as a manually-recorded paper log in place of independent, computer-generated audit trails on electronic data. A manually-recorded paper log lacks the independence of the computer-generated audit trail and is unlikely to contain entries documenting deliberate data manipulation. Using a paper log in a hybrid situation introduces significant data integrity risks and is not recommended.

This is further confirmed in the MHRA 'GXP' Data Integrity Guidance and Definitions, §6.13 [5], which specifically names software upgrades and add-on software as expected mitigations.

4.3.5.4 Investing in New or Alternative Systems

Upgrade, Replace and/or Supplement

GxP computerized systems may have known data integrity gaps or may not have had a previously performed data integrity assessment. Systems implemented with a superficial review of data integrity elements may have had little or no verification of technical and/or procedural controls intended to protect a system's data.

For critical GxP systems that cannot be remediated into a compliant state by either technical or procedural means, the likely resolution for serious data integrity gaps is system decommissioning and replacement with a new system.

Considerations for New Systems

Considerations when planning for a new system include:

- Alignment and fit with the Data Integrity by Design paradigm as applied to the overall business process and/or production process
- Application of the quality risk management approach
- Review, revision, and optimization of current business workflows (processes) and data flows as these were influenced by what was feasible in the existing system and may not be possible, desirable, or optimal in a new system. Blindly using existing workflows often results in unnecessary customizations and inefficiencies.

- Standardization of processes (i.e., using the same process for the same type of work throughout the company). This also facilitates standardization of applications as discussed in the next bullet point.

- A Business Plan, quantifying the improvements in efficiencies and/or reduction in waste and time expected from the new system as part of justifying the Return on Investment

- It is essential to involve Data Stewards, Data Owners, and Quality personnel in this workflow review and optimization activity. Data life cycles, URSs, and SOPs for supported business processes must be updated accordingly.

- Standardization of applications, platforms, technologies, and vendors. When considering the implementation of new systems, solutions with the following characteristics are preferred:

  - Applications currently in use in the company

  - Applications possessing interface capability to other systems and equipment in use in the company

  - Applications with the potential to mitigate data integrity risks across multiple systems as a supplement to those systems

  - Platforms and technologies (e.g., database technology, operating systems, virtualization software or hardware components) that are the preferred choice within the company (or at least well known)

  - Known vendor with a strong history in the industry and excellent support services

Requirements and Design Review

The general approach when upgrading, replacing, and/or supplementing systems is no different from the approach described in *ISPE GAMP® 5* [2] and in the *ISPE GAMP® Guide: Records and Data Integrity* [1]; hence the following text is a short summary of the process.

The URS is generated based on the intended use of the system, the business process, the workflow, the data life cycle that it is intended to support, and the data integrity controls required. A detailed discussion of requirements planning is contained in Appendix 6 of this Guide and the *ISPE GAMP® Guide: Records and Data Integrity, Appendix D1* [1].

A design review process is utilized to evaluate which vendor offerings best meet the user requirements, with emphasis not only on the considerations listed above but also on the robustness of data integrity controls integral to the system.

Supplier Considerations

The supplier assessment process is fully documented within *ISPE GAMP® 5, Appendix M2* [2]. The data integrity considerations to be investigated during the conduct of a supplier assessment are discussed within Section 4.2.3. As a minimum, consideration is given to the:

- Robustness of the supplier's quality system as determined during the supplier assessment

- Capability of the supplier's product to provide the essential technical controls for data integrity, providing full mitigation for the computerized system shortfalls that have motivated this upgrade, replacement, or supplement

For service providers, and dependent on the service being provided, consideration additionally is given to:

- Knowledge levels and training relating to GxP, RDI, and ALCOA+

- Review of supplier quality systems and procedures to ensure incorporation of RDI and ALCOA+ expectations
- Review of supplier service contracts and practices to ensure incorporation of RDI and ALCOA+ requirements

4.3.5.5 Future Proofing

Companies should aim at future proofing their GxP computerized systems and electronic records from a quality risk management perspective as well as from a business perspective.

From a quality risk management perspective, future proofing ensures that:

- Risks to patient safety, product quality, and data supporting regulated decisions are monitored and the controls around those risks remain effective throughout the operational life of the system
- Complete data ensuring content, meaning, and integrity of electronic records (including metadata and audit trails) are preserved throughout the retention period
- Electronic records remain accessible in a suitable format (static vs, dynamic, as discussed in more detail in Section 3.1 Data Definitions and Requirements) and with the relevant capabilities (e.g., to search or reprocess) throughout the retention period. For considerations on this, refer to *ISPE GAMP® Guide: Records and Data Integrity, Appendix O1* [1]
- Systems are compliant with current regulations and remain compliant within the foreseeable future (e.g., in line with evolving regulatory thinking and any anticipated changes in regulatory requirements)

From a business perspective, future proofing ensures, but is not limited to:

- Securing continuous alignment with business needs
- Providing capacity and flexibility for future expansion and change of intended use (e.g., expansion from the QC laboratory into the Development laboratory)
- Ensuring that systems in use (including operating systems, middleware, hardware, and other system components) are supported by the vendor. This encompasses evaluating plans for upgrades or replacement of aging components and technologies.
- Consolidating systems, platforms, and vendors
- Standardizing processes, applications, platforms, and technologies
- Future proofing is covered in data life cycle descriptions and plans/strategies addressing specific systems, platforms, technologies, etc. as appropriate.

4.3.6 Preserving Existing Data

This Section briefly considers data migration from systems and is not intended to be a complete guide. For more information, refer to *ISPE GAMP® Guide: Records and Data Integrity, Appendix O1* [1].

4.3.6.1 Characteristics for Records

It is important to note that some of the key aspects to be addressed when migrating electronic GxP records in general (e.g., audit trail entries, metadata, and integrity of data) are by nature lacking or insufficient in older systems, and are being substantially upgraded or replaced as part of the data integrity quality risk management process.

Typical characteristics in relation to records in such systems requiring upgrade or replacement include:

- Audit trail is lacking or insufficient (e.g., audit trail may be present, but the ability to identify individuals who have created, modified, or deleted critical data is lacking due to use of shared logins)
- Integrity of data has not been properly assured at the time of creation. Migrating and protecting data of questionable integrity will not rectify existing integrity issues, but the data is at least secured ongoing from the point of migration.
- Supporting data and/or metadata is not adequate (e.g., previous versions of analytical methods are not retained in the system)
- Data is not stored in a way that supports proper records management
- A large number of the records are semi-active or inactive in the context of the records life cycle (i.e., regular access to the records is no longer required)
- Records may be stored in a proprietary data format (i.e., with no or limited possibility of being read outside of the native application)

4.3.6.2 Initial Planning

Any replacement or upgrade of a system results in the need to manage the ongoing preservation of any regulated data from the system for the full retention period.

When replacing systems, attention must be paid to the migration of records and data from the original system into the new system to ensure the maintenance of data integrity with no consequential loss or corruption of data and associated metadata.

If migration of data and records to the new system is not necessary or feasible, then the complete records and data from the original system must be securely retained or archived as complete records in a legible and retrievable format for the required retention period.

The key to planning data migration from a system is to conduct a risk assessment on the future need for this data (e.g., the need for the ability to view, search, and reprocess data) considering the business process and the regulatory expectations (e.g., static vs. dynamic data), technical possibilities, and associated cost.

MHRA 'GXP' Data Integrity Guidance and Definitions – Version 1, March 2018 [5]:

"6.8. Data transfer / migration

Data transfer is the process of transferring data between different data storage types, formats, or computerised systems.

Data migration is the process of moving stored data from one durable storage location to another. This may include changing the format of data, but not the content or meaning…

Data transfer should be validated. The data should not be altered during or after it is transferred to the worksheet or other application. There should be an audit trail for this process."

This concept of data migration involving a change to the format of the data is further reinforced in the OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring Number 17 Advisory Document of the Working Group on Good Laboratory Practice Application of GLP Principles to Computerised Systems, §67 [23]:

> *"Conversion of data to a different format should be considered as data migration (e.g. from a proprietary data format to PDF"*

Data transfer may occur as part of archiving, where the data is moved out of the original system storage once the inactive phase of the records management life cycle is reached or when regulation requires. (Records life cycle is discussed in the *ISPE GAMP® Guide: Records and Data Integrity, Appendix M7* [1]).

Data transfer and migration may be required several times for electronic records with a long retention period due to upgrade or replacement of the system managing the records.

The scope of data to be migrated must be established. Not all data will need migrating, for example data that is at the end of the retention period can be destroyed under a controlled process. See *ISPE GAMP® Guide: Records and Data Integrity, Appendix O1* [1] for further information on evaluating this and what to consider.

4.3.6.3 Transfer and Migration Strategies

Data transfer and migration can encompass many strategies, depending on the technological options available, such as:

- **Directly transferring the complete data in its existing format into a newer version of the system (the upgraded system)**

The system needs to have the capability to read data created in older versions of the system, and to convert the data into the newer format used in the upgraded system.

A level of verification, based on risk, is needed to confirm that all data (including metadata and audit trails) has been copied across, and that data integrity is not affected by the conversion to the newer format. An application where the vendor includes such backwards/forwards compatibility can significantly reduce the effort required for, and the risks involved with, data migration.

Consideration needs to be given to any changes in the processing algorithms used in the newer version of the system compared to the older version; this could cause reprocessed data to give a different result in the newer version of the system compared to data processed in the older version.

- **Transferring the data in its existing format into a secure offline archive**

The data may have questionable integrity from its life in the original system; however, the secure offline archive will at least prevent further modification or deletion.

Consideration needs to be given to how to retrieve and read the data (including metadata) during the retention period. Typically, this requires preserving a working copy of the original system; options for this are discussed in the point below.

Complete data in the secure, offline archive may not be easily found and retrieved if the archive indexes are not well constructed and maintained.

- **Transferring the data in its existing format into a long term, secure data repository**

Where data is in a proprietary format not readable outside of the original application, at least one instance of the original application needs to be preserved (possibly as a virtual machine to avoid potential issues with obsolete computer hardware and operating systems) so that the complete data can be restored to that original application to satisfy the regulatory data retention requirement and the "Legible" requirement within ALCOA+.

Consideration needs to be given to how the data can be identified and queried within the repository such that it is readily retrievable; a repository that can extract metadata and searchable tags from the data is one solution to this requirement.

- **Migrating (transforming) the complete data from its original format to a neutral format for subsequent import into the new system**

Each conversion brings risks to the integrity of the data, and extensive verification is needed to ensure that data is not lost or corrupted during each conversion. With this approach it is unlikely that all of the data can be migrated, so a risk assessment is needed to determine the impact of loss of data, metadata, and supporting data, and the ability to interact with, or reprocess the data.

The Allotrope Foundation [40] is a current collaborative effort to create a cross-vendor format for the acquisition, exchange, and management of laboratory data.

- **Migrating (transforming) the data into a static format readable through the long-term retention period**

This is typically achieved either by converting the data into a static electronic format such as PDF, or by printing the data to a paper printout. Where the record was originally in an inherently static format, this strategy is low risk and simple to implement, and provides a true, accurate, and complete copy of the original data. However, the regulators have expressly stated that a static record cannot be considered a true, accurate, and complete copy of dynamic data where the user needs to interact with the record content. (See Appendix 5 for further details.)

A laboratory example is chromatography data, which cannot be reprocessed from a static printout. In the clinical realm, an example is the eCRF database archived as a PDF, which can no longer be searched electronically for specific study events or criteria.

The value of maintaining a dynamic record may decrease with time and needs to be considered during this transformation. The likelihood of needing data in its dynamic format may decrease over time. For example, chromatography data up to one year old may need to be reprocessed as part of the annual product review if there are any concerns about data integrity, however after a period of ten years it would be extremely unusual for there to be a need to reprocess the data.

A risk assessment should be performed before converting data into a static format, to evaluate the likelihood of the dynamic format being needed in the future and the extent of regulatory risk in not maintaining the data dynamically.

4.3.6.4 Verifying Data Transfer and Migration

Data transfer, as a simple copying of data from one media or system to another with no conversion, requires a level of verification to confirm that the complete data was successfully transferred with no corruption (often achieved using a checksum). This is discussed in detail in Section 4.4.

Data migration, as mentioned in the Section above, creates more complex data integrity risks, especially when elements of data cleansing and conversion are involved. The *ISPE GAMP® Guide: Records and Data Integrity, Appendix O1* [1] discusses the special considerations around data integrity for data migration, and the *ISPE GAMP® Good Practice Guide: A Risk-Based Approach to Testing of GxP Systems, Appendix T9* [37] details how to plan and verify data migration processes.

4.3.7 Ongoing Monitoring

4.3.7.1 Track and Review Progress

Good monitoring and reporting procedures must be in place to ensure completion according to the agreed schedule as well as escalation of issues to the program team if necessary. Metrics include the execution of effectiveness checks against the measures implemented to remediate the data integrity shortfalls.

Routine internal audits of GxP computerized systems must form part of the company quality management approach to ensure the environment, behaviors, and practices performed by the system stakeholders remain compliant and appropriate to ensure data integrity.

4.3.7.2 Maintaining a System

To maintain data integrity and the validated state, computerized system repairs and upgrades must be appropriately handled through Operational Change and Configuration Management and Repair Activity processes (refer to *ISPE GAMP® Good Practice Guide: A Risk-Based Approach to Operation of GxP Computerized Systems* [29]). Additional care is needed to ensure that data (including metadata) are not compromised by the system repair and upgrade activities.

4.4 System Interfaces

4.4.1 Introduction

The current networked world requires consideration of a multitude of system interfaces. Data flows through different types of computerized systems, starting with the creation or first entry of data, continuing through the transfer from one system to another, and ending with the destruction of the data transferred.

In this context, data integrity is an important topic from the beginning; a challenge that increases as systems are added and uses of the transferred data become more varied.

This Section focuses on data integrity issues affected or caused by interfaces between computerized systems that are in the sphere of influence of a regulated company.

The following proposals, issues, and recommendations may also be applicable, but not be intended, for inbound and outbound interfaces, for example, from or to business partners or authorities, which should be the subject of a validation project following *ISPE GAMP® 5* principles [2].

The same principles described here for data interfaces should be considered when data is migrated from one system to another, such as the migration or transfer of data into a new or replacement system. One major difference is that during data migration, the process is only used once.

The key concept of “interfaces” is characterized by a generic approach for all types of interfaces:

- End-to-end process approach from the “sending” process step in the source system to the “receiving” process step in the target system
- Risk-based approach that supports data integrity analysis by focusing on typical issues, data with GxP impact, and risks related to interfaces
- Placing emphasis on the roles and responsibilities for the validation of interfaces and the need for a common understanding around the use of an interface by all parties involved
- Identifying the need for a thorough interface specification, incorporating all aspects of the interface, as a solid foundation to find and mitigate data integrity risks.

Performing a risk-based validation for every critical interface, based on the knowledge of the interface context, the coupled processes, and the typical risks, along with a coordinated specification as the foundation, should prevent most data integrity issues related to interfaces.

Whether a system is a new implementation or a system under remediation, validation is essential to achieve reliable interfaces, ensuring that the integrity of data is maintained during the passage from the source system to the successful receipt and use at the target system.

4.4.2 Manual Interfaces

Manual interfaces involve either direct data input by an operator, or manual transfer of files from one location to another. Manual interfaces require strong procedural controls, no conflict of interest, and documented verification that the complete and correct data were successfully transferred. Wherever possible, manual interfaces should be replaced with automated interfaces.

4.4.3 Purpose of Interfaces

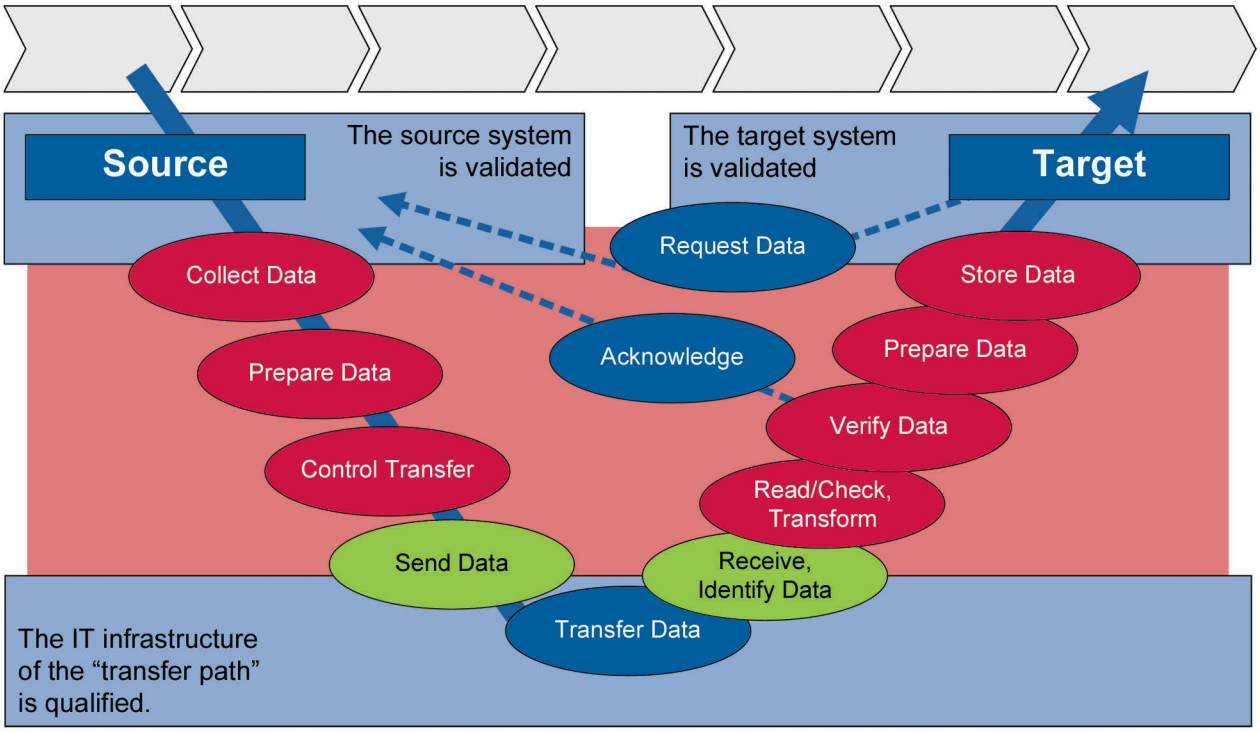
Ideally, all data within a regulated environment is transferred without modification between validated systems using a validated interface on qualified IT infrastructure. It is relatively simple to check whether the source or target systems are validated and if the IT infrastructure was qualified, but it can be more difficult to detect if data was transformed, modified, deleted or corrupted before, during, or after the transfer process.

Most interfaces between computerized systems have more functionality than just transporting data. Figure 4.4 provides an overview of typical functions performed by an interface.

In order to facilitate data integrity analysis:

- Every function of the new interface should be specified and documented before implementation, reflecting the needs of the supported business processes.
- If the interface exists, the embedded logic and intended use of the interface functionality needs to be thoroughly understood.

Figure 4.4: Typical Tasks of an Interface



4.4.4 Understanding the Context of the Interface

It is important to include interfaces within the scope of the validation effort either as part of the source system, the target system, or as a separate validation effort.

A structured validation process helps to verify that all requirements are fulfilled; in the same way a well-structured process is needed to define the requirements for the interface specification.

In order to assess and prevent data integrity issues across networked systems, it is essential to understand the context of the interface(s).

An interface consists of at least four parts:

1. **Connected Systems**

   The sending system (source system) delivers the outbound data, and the target system receives the inbound transferred data.

   In some IT environments there may be additional intermediate systems along the transfer route that collect data from different sources, store the data temporarily, then sort, mark, and forward the data to the target systems.

   There should be written system descriptions for all systems involved as required in EU GMP Annex 11 [17].

2. **Data**

   Data to be transferred, including supporting data (e.g., related records and change history related to the transferred data) and metadata (e.g., audit trail entries and contextual data).

   The data field mapping of the source and target data storage should be available (e.g., the database structure, tables, field lists, import requirements of the target system, or description of export files).

3. **Processes**

   The business process/GxP process includes the audit trail entries/change history related to the transferred data across systems, or two or more processes linked by the interface.

   A process description and process flow chart are helpful to allocate the process steps that trigger sending data, and, if applicable, show the process step on the target side where the data is received and verified.

4. **IT Infrastructure**

   The underlying IT infrastructure components are the parts that transport the data and operate the connection between source and target systems: network, communication server, router, temporary data storage (buffer).

4.4.5 Allocating Responsibilities

It is inherent in the nature of interfaces that there are multiple parties involved.

- Process Owner of the source system, of the target system and, if applicable, of intermediate systems

- Data Owner of the data to be transferred and, if the interface crosses organizational boundaries, the new Data Owner on the receiving side (Data Owner may be the Process Owner for the source or target system)

- Process Owner of the process that provides data for transfer and triggers sending, and the Process Owner who is responsible for accepting, storing, and processing the received data
- System Owner of IT infrastructure components that are part of the transfer route, and, if defined, System Owner of the IT service processes, e.g., network services, backup and restore
- Suppliers/Developers for the interface itself, or for connecting elements to source or target systems

If these roles and responsibilities are not clearly defined there is a risk that some important requirements and changes will not be addressed, tested, or validated.

4.4.6 Further Information

Appendix 7 contains further information on creating an interface requirements specification and identifying data integrity risks associated with interfaces.

4.5 Access Controls

4.5.1 Introduction

Computerized systems in a GxP environment require controls, management, and documented processes to ensure that data integrity is maintained. There are many areas of concern to organizations, including:

- Classifying systems and data
- Controlling access at user/administrator/supplier levels
- Segregating duties between individuals and functions
- Historical and ongoing auditing

Systems and databases may contain a large array of data; therefore, organizations should classify their systems and the data contained within them; see Section 2.4 for more information on data classification.

Organizations should have defined and accountable data/asset owners who understand the value of the data, and thus the level of protection required and a clear framework by which to apply controls. While each organization will have a different framework, it is important to ensure that it is agreed, and documented. It should be understood that the value of the data, and consequently the design of the control framework, changes over time.

Organizations are generally aware of the need to manage system user level access with appropriate controls and processes. Privileged access should be treated with the same, if not more, rigor and thoroughness.

Privileged access is a term given to define enhanced permissions to perform administrative tasks that require additional system/application visibility, such as issue resolution activities and system modifications/development, but also could include actions, for instance legitimate transactional data corrections under appropriate approvals and change documentation.

Privileged access to systems/databases should be kept to a minimum and regularly reviewed to ensure that data deletion or unauthorized/malicious modification does not occur. It is important to note that segregation of duties between interested parties in relation to data owners prevents a conflict of interest.

Organizations should also confirm that an appropriate audit strategy is in place to monitor:

- Use of privileged access accounts
- Internal and external failed access attempts
- Activities undertaken upon gaining entry
- Traceability back to Data Definition Language (DDL, defines data structures within a database) and Data Manipulation Language (DML, manipulates data within a database) statements
- Other potentially malicious activities, for example, out of hours requests for access

This audit strategy should be risk-based, repeatable, and periodically reviewed to ensure that it remains viable and effective in a changeable environment and that data integrity is not compromised. Appendix 19 contains guidance on auditing access controls.

Individuals with privileged access have the technical means to access and modify data, bypassing the user interface and often without any audit trail of their activities; therefore, it is advisable to consider implementing additional controls and safeguards around privileged access. Modern database engines offer granular access controls and database-level audit trail functionality, which can be used to extend data integrity controls to the database layer.

Privileged access to databases is one component of overall database security that should include controls for physical access and remote access to the server(s) and data warehouse(s), and controls for access to the application using the database and a means to review such access, such as those described within 21 CFR Part 11 [41] among other regulations.

Privileged database access is an especially important component of overall system security as those with such access can not only perform a large array of tasks but can do so without affecting the audit trail(s) owing to the level of system insight required for the role.

It is imperative that those with privileged access are not only trained on the importance of ensuring that access is controlled in relation to patient safety, product quality, and data generated to support regulated decisions, but also to have a clear understanding of the implications and consequences of participating in any unauthorized activities or knowingly allow them to occur.

Additional consideration should be given to outsourced privileged accounts where personnel may have insufficient understanding of the consequences of changes to data.

4.5.2 Segregation of Duties

In the regulated life sciences industries, it is necessary to have segregation of duties. There are checks and balances required by regulations, which in this case come in the form of a Quality Assurance or an Authorized Person role.

The same principle holds true when a computerized system is implemented to automate a business process. In general terms, it is necessary to separate tasks to assure different people have the responsibility to perform different parts of the activity, known as Segregation or Separation of Duties (SOD).

There should be a role for those who create data and a role for those who approve or release data (i.e., the quality specialists). The SOD is clear and ensures that there are different functional groups responsible for the creation and the approval of data or records.

When implementing and using computerized systems, assurance of SOD at the operating system, application, and/or database level is needed.

As discussed, users with database privileged access roles have the ability to change the way the system works, the ability to turn off the audit trail, or even the ability to modify or delete data created in the system, and often without traceability. As such, controls are needed to ensure that a separate person or distinct function performs these privileged activities.

This person should not have any responsibility, accountability, or direct interest in the data/records created and maintained in the computerized system and should be detached from the business process they support. This can be done by an Engineering or IT representative with no direct interest in the data/records in the system.

All actions performed by individuals with privileged access must be authorized either by operating procedures and/or appropriate change controls before the activity takes place, with clear traceability attributing the actions to an individual, along with the dates and times the actions took place.

4.5.3 Access Control and Review

Managing system access can be likened to managing access to a home. The home owners (system administrators) have full control over both internal and external access to the property and the ability to grant or deny access to others. Family members (super/power users) have full knowledge of the house and can visit and leave as they please; however, they may not have access to all areas within the property. Family friends (system users) have far less access but are able to visit the property without prior authorization; while tradesmen (vendor users) must schedule a visit and be supervised by the home owners for the duration of the visit.

Access to computerized systems should only be granted in situations where individuals are adequately trained to perform the activity and have a legitimate business reason to do so. A record of the request for access or additional access should be documented in accordance with SOPs.

Revocation of user access should occur in a timely manner in accordance with SOPs, and care must be taken to ensure that no residual access remains, for example, the application access was removed but not the server/host access.

For restricted and sensitive data, often read access needs to be as strictly controlled as write access; for example, incorrect granting of read access could lead to the unblinding of randomized clinical study data.

4.5.4 Further Information

For further information around these topics, see the following Appendices in this Guide:

- Appendix 9 – Security Controls
- Appendix 10 – Case Study: DBA and Security Controls for an RTSM System in a GCP Environment
- Appendix 11 – Case Study: DBA and Security Controls for an ERP System in a Medical Device Manufacturing Environment
- Appendix 19 – Auditing Access Controls

5 Critical Thinking

5.1 Auditing

5.1.1 Introduction

Sound scientific practice requires reliable data that is generated and maintained with integrity. This is not a new concept but a fundamental principle in critical decision-making processes.

In recent years significant problems with data integrity have been found in the pharmaceutical, biotechnology, and medical device industries worldwide, and the heightened awareness of these issues has changed the approach used by regulators when conducting inspections.

For example, the US FDA now dedicates a significant portion of their Pre-Approval Inspections to data integrity auditing [42]:

- *"Objective 1: Readiness for Commercial Manufacturing*
- *Objective 2: Conformance to Application*
- *Objective 3: Data Integrity Audit"*

Similarly, there has been an increased focus by the regulators on systems supporting GCP-regulated processes. As early as 1995, the World Health Organization published WHO Technical Report Series, No. 850, Annex 3: Guidelines for good clinical practice (GCP) for trials on pharmaceutical products [43], which included validation and other system aspects.

More recently, ICH released the Integrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6(R2), November 2016 [44], which brings additional focus on computer system validation and data integrity.

The implementation of technical and procedural solutions to meet regulatory requirements throughout the business process proves to be challenging for organizations that do not have expertise in incorporating data integrity into their daily activities. Maintaining data integrity requires an understanding of how the data is handled throughout the data life cycle; once the integrity is lost it cannot be restored.

This Chapter emphasizes the importance of data integrity in critical decision making, and describes how to apply that understanding to auditing for data integrity.

This Chapter is structured as a tool for use in the auditing process to determine if the basic requirements of data integrity are in place. It is not intended to be a comprehensive checklist for conducting data integrity audits, but more of a guidance of what to review to identify gaps in data handling processes, especially at interfaces between systems.

Detailed review of complex workflows from multiple sources is a difficult task. It is different in each organization and might vary between different areas of the same company. Auditing all the different variations and data flows requires strong critical thinking skills.

The complexity associated with finding data integrity breaches can lead to many things that may look out of line but in reality, are not important or not important enough, to take significant action. An effective auditor must understand the business process and the technical controls provided/implemented in the system. Having adequate technical knowledge, along with experience in knowing how the technology is used, in conjunction with any paper records in the process, will help in understanding the importance and impact of key elements and the associated risks.

5.1.2 Data Integrity Audits Compared to Regular Auditing

QMS audits typically focus on verifying that reports or results have been generated in accordance with the governing SOP and may include cross checking paper and electronic records for transcription errors and consistency. This type of audit usually does not require review of core system information or any detailed knowledge of how electronic systems work and does not typically involve additional extensive data checks.

Data integrity auditing verifies the authenticity of electronic data through detailed examination of the data within the electronic systems, and then cross-referencing to an array of other sources, including data audit trail entries and paper-based records. This necessitates a deeper knowledge of how electronic data is handled, and how complex systems work and interact to establish the validity of the information being examined. It extends beyond the traditional data review process for a deeper dive into the verification process in order to establish the integrity of the data.

In pharmaceutical research and manufacturing, data integrity auditing often involves the use of electronic or paper data with cross-referencing and tracing across complex workflows to establish the facts or evidence of the integrity of the data used for critical decision making.

5.1.3 Leveraging a Risk-Based Approach

The audit process may be tailored to match the risk profile of the data and the business process.

- Clearly identify the risks and apply an appropriate amount of effort to identify the important gaps so mitigation or remediation can be implemented.

- Choose the systems or business processes that have the highest potential impact or risk to the product quality, patient safety, or data on which a regulated decision is made. For example, this may be clinical trial data submitted in support of an NDA, or QC test data used to make the critical decision to release a batch to the market.

- Include applications with ongoing issues or concerns as reported by the users within the business area. Select a few key items and follow them into the systems, paying special attention to points where data is transferred between systems.

5.1.4 Data Governance within the QMS

The organization's overarching QMS can significantly impact data integrity if it does not include the correct quality attributes, establish and maintain a state of control, and prioritize continuous improvement for its data integrity assets.

5.1.4.1 Data Integrity Policy

The organization's Data Integrity Policy should cover the data integrity approach across the company. It should include details about the data integrity training required for the organization, which should address methods for detection and prevention of data integrity issues.

5.1.4.2 Documentation and Data Standard

There should be a Documentation and Data Standard that details data handling. It should cover paper and electronic forms of data if both are used. The level of detail should be aligned with the practices and may include additional detailed data integrity procedures at the implementation level.

Gathering some quick understanding of how data is captured and reviewed will give a general idea of the overall quality culture of the organization. Check how the data and access to the data are handled to ensure it is stored, protected, complete, and access restricted appropriately.

5.1.4.3 Deviation Process

Deviations should have been raised when there has been a departure from an approved standard, when something unexpected has happened, or when the standard did not adequately cover the process. If deviations are rated for impact, make sure the assessment considers the risk to patient safety, product quality, or the integrity of data supporting regulated decisions.

Review associated root cause investigations, to ensure there are enough details to get to the root cause. Search for repeat deviations as a sign that root cause is not being addressed. Check for holistic reviews so if a problem is found in one system or area it is also fixed in other areas.

Review all the incident reports that result from not following a procedure or from having any kind of unexpected event. This may help determine if there is an alternate means to record deviations. If an organization has no deviations recorded, it may represent problems across several layers of the organization including an overall ineffective quality system.

5.1.5 Data Life Cycle

Establish the data life cycle for GxP data and walk through the business process, following the data flow from creation to destruction.

Analyzing complex data flows requires critical thinking combined with an understanding of the data transfers between systems. Leveraging experience and knowledge about how and where processes can go wrong in a highly complex workflow and adapting to follow the many paths the data can take is more effective than following a checklist.

The auditor will greatly benefit by watching people conduct their daily activities, verifying that their practices match the approved procedures (SOP, method, etc.). This provides an opportunity to clarify points during the observation rather than relying on an explanation from a company representative who may not be fully knowledgeable of the intricacies of the process.

5.1.5.1 Creation

The business process and the associated data life cycle may involve one or more computerized systems as well as paper-based recording. Manual recording of data increases the risk of accidental or deliberate errors in the recording process, and therefore represents a risk to the integrity of the data.

If paper is used to record original data that is later transcribed into an electronic system, review the procedure that ensures the integrity of the manually-entered data, such as second person verification of the transcription, confirming that the process is well documented and followed.

5.1.5.2 Processing

Automated processing via a predetermined, validated method, such as the conversion of an analog signal to a pressure reading by means of fixed calibration factors, offers fewer risks than more complex processing operations that are potentially vulnerable to manual intervention. The auditor should assess and understand:

- The complexity of the processing operation (e.g., simple summation or trending vs. complex integration)
- The possibility, extent, and occurrence of manual intervention in the processing (e.g., who can impact the processing? How much can the human intervention impact the result? How often is manual intervention happening?)
- The criteria governing when manual intervention is appropriate (e.g., is there a clear SOP defining “good vs. bad,” and when intervention should occur?)

- The depth of documentation of the manual intervention actions taken

- The rigor of data review resulting from manual intervention (e.g., does the reviewer investigate all the previous versions of the results to understand the effect of the intervention, and assess whether the intervention met the SOP criteria?)

Excessive use of manual intervention can indicate a lack of robustness in the process and may result in deliberate data manipulation.

5.1.5.3 Review, Reporting, and Use

Review

Section 3.1.4.1 and Appendix 5 discuss in detail the principles of static and dynamic data. As part of the auditing process, verify that data review for dynamic data is not solely based on the review of a printed summary report.

Printed reports suffer two main limitations:

1. The user may be able to determine which subset of data they chose to include (and therefore exclude data with or without scientific justification).

2. In the case of chromatograms and spectra, a paper printout does not allow scaling of the graphical elements to examine peaks and baselines in additional detail.

Data review and approval may involve a hybrid situation as defined in Section 4.3.5.3, where the record is held electronically but a summary report is printed for the purposes of bearing a handwritten signature. In this situation, key considerations are:

- Presence of linkage between the electronic record and the paper signature, to evaluate the traceability between them

- Controls in place to prevent further editing of the electronic record after the paper-based approval has been completed

- Review of the electronic record conducted prior to the approval signature being executed, rather than relying only on the printed summary report if the printed report is not complete and accurate with regards to dynamic data

Reporting

Out of Specification Process

Out of Specification (OOS) results should have a rigorous process for investigation, including well-documented conclusions and outcomes. If the OOS results impact a batch of materials or a series of results, a deviation should have been issued and any other batches or results impacted by the failure should have been included in the investigation.

OOS results have a notorious history of being discarded for retesting, which creates a testing into compliance situation. This is why the audit trail should be always enabled, reviewed, and no inappropriate access allowed so that discarded results or retesting can be detected.

Failing results cannot be discarded but can be invalidated with sound scientific justification, which are then classed as Invalidated OOS results. Within the FDA draft Guidance for Industry, Submission of Quality Metrics Data, the metric Invalidated OOS Rate is included as an indicator of the operation of the laboratory [45]. Invalidated OOS are also discussed in Appendix 12.

Metrics and Trending Process

Metrics are a key aspect of monitoring and trending as part of a continuous improvement process and for management review. They help ensure that a state of control is established and maintained across the organization. Escalation of serious issues to management allows resources to be allocated to address the issues. Reviews and actions taken as a result should be documented and available for request along with the governing procedure.

Deviations, change controls, and OOS results among other things, are primary items to be monitored and trended. For example, a sudden increase in deviation numbers or OOS results could be an indication that something is deteriorating in an area. Appropriate actions should be taken to stop the deterioration and prevent it from happening again (CAPA), all of which should be documented for review. Be aware of metrics driving the wrong behaviors, such as punitive actions for having a certain number of deviations or for obtaining failing results. Analytics are further discussed in Section 5.2.

Use

The original data from the computerized systems may generate results that are either directly uploaded or summarized and then uploaded to another system. All data transfers must be well controlled to maintain the integrity of the data. In addition, there may be varying degrees of manual operations involved in any of the data flows, which complicates the processes and controls.

For example, QC laboratory data may be captured in a LIMS that generates the reported results. The results may then be transferred to an enterprise-level system like SAP, where they are combined with other data from the manufacturing execution system to collectively form the data set necessary to conduct the batch release.

Retention and Retrieval

For GxP data, either the original record or a true copy must have been maintained. A full discussion of original records, true copies, and static vs. dynamic formats is found in Section 3.1 and Appendix 5.

If paper is used in the process, confirm it is kept in a paper archive for long-term storage. Paper is more difficult to handle and store in a suitable fashion, so it may get neglected in the overall process. It is also more likely to get lost because it is a manual process and rigorous procedures need to be in place to ensure proper handling.

Verify that paper is not being substituted for the retention of original electronic (dynamic) records.

The metadata and supporting data (see Section 3.1) must be moved to the long-term archival storage location along with the associated original data. The metadata should be moved carefully to avoid separating it from its associated original data or risk breaking the bond between them, which may permanently compromise the integrity of the data.

Archival storage of data should be long term for the retention period of the record. Data acquired with many systems are in a proprietary format and require the original native application to read the data and metadata. Over a long retention period, it can become increasingly difficult to maintain a functional copy of the native application, and data migration or system virtualization strategies may become necessary. This is discussed in more detail in Section 4.3.6.

Verify that data can still be located, retrieved, and read electronically.

Toxicology and clinical data have extraordinarily long retention times (e.g., 30 years) to keep the original data in human-readable form. Additionally, during corporate acquisitions, there needs to be a process in place ensuring that both the data and the ability to open records are not lost in the transition.

Destruction

Destruction of the data at the end of the retention period must be governed by a formal process and logged. It is rarely included in the scope of a data integrity audit as the primary focus is on data in the preceding stages of the data life cycle.

5.1.6 Reviewing in the Reverse Direction

Working backwards from selected data with GxP impact in a final report or submission is useful because it provides a review of the entire chain of data used to generate that report. Reviewing data from the final report back to first creation is counterintuitive and helps to avoid assumptions about data flows and traceability. Selected critical examples can be traced back through the entire data workflow to detect gaps along the way. This review should include a check of how the data is saved and protected as well as checking human touch points and interfaces, as these are particularly high risk to data integrity.

5.1.6.1 Identify the Critical Path

It is important to understand the critical path so the auditing efforts focus on the highest-risk aspects of the workflow. Following the path is facilitated by starting at the end of the data chain using selected review examples from data with direct GxP impact in final reports or submissions.

5.1.6.2 Final Reports or Submissions

Reviewing from the end of the data chain backwards is not the normal direction of review, so it can reveal problems that might have been overlooked as part of the standard forward-review process.

- Obtain a copy of a report that was approved, sent to the client or the FDA, or however it was presented in the final form. This should be a report used to make critical decisions to support the business.

- Choose a few examples of data or have the business area choose a few critical examples to trace (e.g., data used to release a batch of material).

Note: Only a few key examples are needed for the initial review. If problems are found, additional reviews should be performed to confirm the observation.

- Trace the data back to the source of the original data, checking everything and every system along the way.

- Walk to the actual systems and have the operator do the trace. Do not rely solely on printouts for review as it is too easy to hide things with printouts.

- Pay particular attention to interfaces between systems, especially when there was human intervention.

- Review the audit trail and confirm that the organization is reviewing it for data with GxP impact changes.

5.1.7 Computerized Systems

5.1.7.1 Complexity

Over time, the systems used to generate and manage data have become increasingly complex. Ensuring the necessary elements in a system are configured properly is difficult if the details of that system and how it interfaces with other systems are not well understood. A clear understanding of the business process and how the computerized system supports this process is an important aspect of auditing for data integrity.

Care must be taken to review configuration settings to identify consequences to data integrity arising from a poor choice of privileges and system controls. Whether the integrity of the data is compromised through neglect or deliberate manipulation, the outcome may be the same; it would be considered a wrongful act and the integrity of the drug application [data] may be compromised [46].

The sheer volume of data can also be a problem. Some systems may capture operator entries that impact GxP data in the same audit trail or log as less critical or routine system activities. As the audit trail grows, a row by row review process becomes ineffective and the likelihood of spotting an error among thousands of entries is nearly impossible.

A focused review, where the entries impacting GxP can be segregated and sorted based on the data integrity risk associated with the particular activity, provides a more effective detection of issues including those that would not be found without the audit trail review. This can form part of a validated exception reporting process where the system functionality permits this [5].

Laboratory informatics (e.g., CDSs or e-LN) are both configurable and complex, and potential data integrity issues can hide within that complexity.

5.1.7.2 Standalone Systems

Standalone systems have no way to automatically transfer the data and metadata to long-term archival storage, so human intervention is required with checks to make sure all data collected is moved in a controlled fashion. The systems cannot be automatically updated for security defects or virus protection. Access controls are applied locally and configuration may vary system to system.

With so many manual requirements, preventing conflicts of interest is also required. That means many more procedural controls and documented verification processes need to be used on standalone systems to ensure the integrity of the data is maintained, and the audit approach may need to examine the effectiveness of those manual controls.

5.1.7.3 Networked Systems

As long as the networked systems are configured and used appropriately, the problems mentioned in the standalone systems above can be avoided. Automated and validated transfer of data, user group access control, automatic updates, and far fewer procedural controls require considerably less effort to maintain the integrity of the data. Verify the system validation reflects the intended use, and examine any software upgrades and network system expansion (e.g. additional clients) installed since the validation project ended. Verify that operational change control and configuration management have been applied since the system went live to maintain the validated state.

5.1.7.4 Access Management

Investigate who has access to what system functionality. Obtain the organizational chart and the access or user list for the system and confirm that there are no conflicts of interest, especially for elevated access privileges. This should be done for both electronic and paper data.

5.1.7.5 Conflicts of Interest

- Verify that no one in the business area has administrative or delete access to the original data or audit trails.

There may be exceptional cases, such as for standalone systems with small user groups where segregation of duties may be accomplished with multiple roles assigned to an individual. In this situation, an additional review is recommended to ensure that the appropriate role has been used in all instances. For example, the elevated privilege role has not been used for low-level daily tasks.

- Verify that no one can approve their own work.

- Verify if anyone with elevated privileges has left the area and still has access.
- Verify the appropriateness of the various access levels and the people assigned the roles.

5.1.7.6 Audit Trail Review

- Verify that the audit trails are always on and cannot be shut off by anyone in the business area.
- Verify that the audit trail is collecting "who did what and when" (and "why" when changes are made).
- Verify that unique user IDs are in use with no shared accounts.
- Verify that the business process for audit trail review clearly specifies what in the audit trail is critical to review. Then choose some examples of those critical items to review and look for documented evidence that the business is conducting their own reviews on those items.
- Verify time stamps to assure the data is collected and recorded contemporaneously; reusing data or capturing the last perfect experiment would have all time stamps happening within a very short (or impossible) period of time.

5.1.7.7 Transfers and Touch Points

Transferring data between systems, especially when human touch points are involved, may introduce data integrity issues. If the transfers are not automated and validated, there is a significant risk of human error for either deliberate or negligent integrity problems.

Manual transcription of data or manual file transfers require strong procedural controls, without user conflict of interest, and necessitate documented verification that the complete data was successfully transferred.

As an example, if second person verification is implemented for microbiological plates, that verification must be contemporaneous because the plates are discarded after use, so later readings of the plates are not feasible. Plus, they may require additional verification when the data is transcribed into an electronic system.

How the plates are handled before and after use also has an impact on the integrity of the data obtained from the plates, so those processes should be robust. Knowledge of the business process is critical to understand the necessary data integrity controls.

5.1.7.8 System Interfaces

Automated data migration between systems can be a source of problems if the interfaces are not validated (either separately or as part of the source or target system validation) to ensure they are suitable for intended use. Even if validated initially, if the systems are modified and the interfaces are not tested to ensure that the validity was not impacted, then the transfers may not work as intended. Data can be lost or corrupted if the interfaces are not working properly.

Verify that acceptable processes are in place and followed that validate and maintain the data transfers between systems. Request change controls and review the impact assessment for the interfaces when major changes were made to the systems. Check the testing documentation to see if the process was well documented.

Best practice for interfaces is covered in Section 4.4.

5.1.7.9 Change Control Process

A change control process should be used when changing approved processes or systems. It should include:

- Detailed description of what is being changed
- Assessments that cover the impact and risk of the change to ensure it does not have unintended consequences
- Review
- Approval
- Follow-up to ensure the change was implemented without issue

The follow-up should also detail anything unexpected and the impact of those issues:

- Review the incidents and change logs for the system, as that will give an indication of the system stability and control
    - When reviewing change controls, remember that deviations might be concealed within the change control system to reduce the number of deviations.
- Check entries in the audit trail for changes to configuration settings and cross check against the change control request
    - How are configuration settings tested and recorded?
    - How often are they confirmed?
    - Who has access to make changes to the configuration?

It may be valuable to determine the proportion of emergency changes performed against the total number of change requests. If the number of emergency changes is disproportionately high, it may indicate system changes done to hide testing into compliance.

5.1.8 Comprehensive Data Integrity Auditing Examples

Comprehensive data integrity auditing requires knowledge of the complexity and risk profiles for the system or workflow under review. Tailor the audit scope to what is planned for review, with the understanding that adaptation may be required that could change the scope during the audit.

Examples are provided in the Appendices of this Guide as follows:

- Appendix 16 – Reviewing Laboratory Systems
- Appendix 17 – Reviewing IT Systems
- Appendix 18 – Reviewing Supporting Data
- Appendix 19 – Auditing Access Controls

5.1.9 Evaluating the Audit Findings

Evidence of data manipulation is an obvious critical deficiency; however, there are other findings that may indicate potential for data integrity issues, such as:

- The organization is not saving the original data.
- The audit trails are not turned on all the time.
- There is no process for review of the complete data including audit trails.
- Inappropriate access to the data or inappropriate privileges allowing inappropriate behaviors.
- Anything that would make the validation of the software questionable:
  - A significant number of incidents or bug fixes indicating the software may not have been robustly tested.
  - Critical defects are discovered with no action taken.
  - The patch for a critical defect requires that it be run several times before it takes effect. Control over validated software should include reliability when it comes to applying a critical patch.
  - Failing to maintain objective evidence of testing could indicate that there was a switch to all informal testing and was not well controlled.
- A breakdown in communication such as not being able to get answers or replies to inquiries.
- The vendor or supplier has a new owner. The new company may not have the same quality philosophy or decide that quality is too expensive to maintain, so it will be abandoning those practices.
- A high staff turnover rate drives a lack of consistency and quality. This could also be a sign of severe cost cutting measures that can adversely affect quality.

It may be helpful to consider regulatory guidance on classification of audit findings when evaluating findings from a data integrity audit, and these are reproduced for convenience in Appendix 20.

5.2 Use of Analytics to Detect Data Integrity Issues

5.2.1 Introduction

Data gathered during essential process control and monitoring activities offers a means to identify opportunities for continual improvement based on analytics and trending. Such analytics can also provide a mechanism to identify patterns and abnormalities within the data that may represent data integrity issues. Regulators recognize this and have made it clear in enforcement findings that managers and quality personnel need to oversee their processes, review the data, and look for aberrant results [47, 48, 49].

While it is difficult to detect all potential data integrity issues, there are a number of queries that can be employed to uncover aberrant results meriting further scrutiny by a qualified person. The reports from these queries increase the likelihood of exposing integrity issues and have the added benefit of providing insights into routine behaviors of the organization.

A robust data integrity governance approach employs analytics and measures their effectiveness, knowing that the reports generated require an investment of time for the investigation of the issues found. The organization may also leverage published enforcement actions and internal assessments for inspiration to create new reports and refine existing ones to determine if data integrity issues exist.

Though many of the examples in this Section are specific for laboratories, they can be used as ideas to create reports for other types of systems.

5.2.2 What is Aberrant Data?

ISPE GAMP® Guide: Records and Data Integrity [1] defines an Atypical/Aberrant/Anomalous Result as:

> *"Results that are still within specification but are unexpected, questionable, irregular, deviant or abnormal. Examples would be chromatograms that show unexpected peaks, unexpected results for stability test point, etc."*

Within the context of data integrity, the term "aberrant" is used whenever the integrity of data is suspect and the data may have been falsified or manipulated.

Which definition to apply depends on the situation. For example, a person can create a data value that appears to be similar to the population and merge this fictitious value into the population. While this false value appears to be similar to the population, it is not natural or right – it is a false value, hidden among true ones.

This is in contrast to a value that is outside the normal population (a statistical outlier) different from the population, but is nevertheless a legitimate value.

While tools to detect and manage statistical outliers have existed for decades, detecting values that lack integrity has only recently been identified as a necessary part of data review and oversight.

A primary complicating factor for integrity detection is that falsified data does not immediately appear to be different than the population, thereby excluding common statistical approaches. This is not to say that falsified data cannot be detected with statistical approaches, it can [50]; however, approaches for detecting data falsification are more complex than those that detect, for example, population outliers in a Shewhart chart.

Appendix 21 contains a detailed discussion on using analytics and tools to identify aberrant data.

5.2.3 Reality of Data Integrity Metrics

Before a company implements a program that includes data integrity metrics, it is important to recognize the limitations of such an effort. For many data integrity query reports, one report record does not equal one confirmed issue. Rather, every record requires an investigation to understand the situation before a valid determination can be reached. This necessity of human confirmation limits the number of query reports that an organization can review effectively and should be considered when creating a suite of data integrity queries.

As a part of implementing a data integrity program, it is recommended that the investigation time and the number of confirmed issues be tracked to assess the effectiveness of each report in the suite. Data integrity reports will not have equal value to an organization; reports that appear to be ineffective should be removed, so that more effective monitoring (or technical controls) can replace them.

Manually-prepared reports are problematic and have little utility in a robust data integrity program. Due to the effort involved in collecting data and creating the report, they are useful only for exceptional situations. Additionally, manual reports have the possibility of data manipulation (data changed or excluded) if the report creator(s) has a stake in the report outcome (e.g., executives use the report for performance evaluations). In practice, routine data integrity reports are created from electronic data.

Due to the technical limitations of some standalone systems, there are data integrity issues that cannot be reported. For example, many systems may only transmit data when initiated by the user. With these devices, users can analyze or enter data multiple times before electing to transmit a value. The receiving system has no record of these additional values.

Data integrity monitoring may change some bad behaviors by simply moving them to new forms of bad behavior. For example, users who are questioned about a value named "test" found by a query looking for suspicious names, might simply find a new term, one which is not routinely searched for in a report.

In other words, monitoring can force manipulators to become more creative in their practices. This type of behavior, the "moving target", is difficult to detect without the cooperation of all people engaged in the business and is one more example where a mature quality culture provides benefits beyond its cost.

5.2.4 Data Integrity Metrics Sources

In addition to the report ideas presented later in this Chapter, there are other sources of metrics.

1. Regulatory enforcement observations are a rich source of ideas.

This citation is from an FDA Warning Letter (4) [51]:

> *"…the analyst at your firm altered the file name in the spectrophotometer containing the sample identification information for* ***(redacted)*** *API lot #* ***(redacted)****, tested on April 2, 2014, to support the release of two previously manufactured lots…"*

This leads a company to ask *"If that event happened in our firm, how would we detect it?"* If copied/renamed files have a new Modified Date, but the same Creation Date as the original file, could a query be designed to report files with duplicate creation dates?

This single question turns observations into knowledge bases for integrity reports, and established data integrity programs can use observations in reverse to assess their coverage: *"Do we have a means of detecting this observation?"*

2. CAPA/deviation events can be reviewed for issues that merit data integrity reports.

For example, a vendor defect might leave a data table open to manipulation; until the defect can be technically prevented (if possible), a report can look for improper manipulation. Trigger events might include the lack of a valid reason entry, an improper/non-validated user ID, an out of sequence time stamp, or any known mechanism an unauthorized person might use to manipulate a system.

3. Life cycle status reviews can be a source of useful report ideas, both for data integrity and data governance.

For example, the "normal" path of the business process might result in all results with either a status of "Approved" or "Withdrawn", environmental samples with a status of "Complete", all active and inert materials with statuses of "Approved", the final batch "Approved", and any CAPA items related to the batch as "Closed" or "Monitoring".

Start by analyzing the statuses if someone in the organization does something with the potential to invalidate the released material, e.g., a purchased material is discovered to fail the related substance specification for a specific country, or a test result is withdrawn and reprocessed. These changes will trigger status changes in one/many related systems.

Next, answer the questions *"If someone makes a change after a material is released to market, can we detect it?"* and *"If someone creates samples related to a batch, and then creates new samples that are used for batch release, can we detect the set of samples that were abandoned?".*

This type of detection can create many combinations of scenarios, but careful analysis can yield algorithms that detect potential issues. This type of reporting becomes more critical as multiple sites and interacting systems are added to the manufacturing process.

5.2.5 Governance vs. Operations Reports

Operations reports for data integrity should target the data reviewer as the primary consumer. They should provide details that permit a rapid review of data in the most efficient manner.

Ideally, these reports give the reviewer the search options such as manufacturing site, product number, material ID, date range, and status, to create small reports that expedite data review and release. The majority of data integrity reports fall into the operations category of reports.

Additionally, a well-designed and validated exception reporting process can be used.

It is a regulatory expectation to thoroughly review all GxP data collected during the execution of regulated activities. Data review should be performed in the process of the manufactured product quality decision. Data for review is collated by batch reporting functions from production records per approved report formats. From the EU regulation [27]:

> *"Where a validated process is continuously monitored and controlled, then automatically generated reports may be limited to compliance summaries and exception/out-of-specification (OOS) data reports."*

Following this principle, ISPE has adopted the term Review by Exception as follows [52]:

> *"The approach whereby manufacturing and quality data generated in validated Manufacturing Operations is screened to present or report only critical process exceptions required by approvers for review and disposition of intermediates and products; requirement for human review/approval of completed Electronic Production records only when a production parameter is out of specifications or there is some other discrepancy or critical exception condition."*

The Quality unit may adopt the Review by Exception practice following a thorough risk assessment to capture all of the batch review requirements under existing GxP regulations.

Governance reports should monitor the use of operations reports at all sites with the goal of ensuring that tools are effectively deployed and used to monitor business processes for aberrant results.

Moreover, governance is the logical place to collect metrics on the number of report records that are investigated, the amount of time consumed in aberrant result investigations, and the number of confirmed integrity issues identified and resolved. The ultimate purpose of this data collection is the evaluation of query reports for efficiently finding issues in practice.

In addition to metrics on the use and performance of reports at each site, governance should evaluate other sources of data integrity issues, such as unplanned issue discoveries, CAPA events, internal assessments, regulatory inspections of sites, and findings from other sites in the same industry to identify the need for new queries and retirement of existing queries that no longer provide detection value.

Another outcome of governance is the identification of issues that should be converted from detection queries to technical controls that prevent occurrence at the outset of the process, as technical controls are superior to detection queries.

Figure 5.1: Prevention and Detection

Reliable  Inconsistent	Prevent	Technical	Controls configured into system, to prevent data integrity issues
		Human	Procedures to address gaps in technical controls, or creation of manual data (data outside of electronic system)
	Detect	Technical	Queries to highlight suspicious/undesired patterns or situations in electronic data
		Human	Data review to identify issues not detectable by queries, or manual data (data outside electronic system)

5.2.6 Report Ideas

This Section provides a list of ideas for both operations and governance reports. The operations reports (Table 5.1) are to be used by personnel involved with operational activities, while governance reports (Table 5.2) are to be used periodically by supervision or senior management to assess the update and effectiveness of data integrity metrics in the organization.

Table 5.1: Data Integrity Operations

Title	Use/Purpose	Notes
Audit Trail Reasons	Listing of user name, reason for change, and date for all audit trail entries over a specific date range.	Permits quick review of audit trail reason codes. Detects situations where users add improper or overly cryptic comments that would not be defensible. If a system does not force an audit trail entry, blank fields can be detected as well.
Short Run (Analytical Laboratory)	For any system that permits a sample run (multiple samples in a group), including chromatography. Selection criteria would have a threshold that is considered “short”, for example, two samples or less.	Detects testing into compliance, where a sample is reassayed until a desired result is obtained.
System Backup	List of system, backup date, success indicator. For a specific date range, and sorted by system, then date.	Prevention of data loss is a fundamental data integrity objective. This is a critical report to review every few days.
Manual Integration Detail (Analytical Laboratory)	List of method, number of samples manually integrated, total number of samples. Sort by method for a specified date range. Applicable to chromatography systems.	Provides insight on which methods require manual intervention. Lowering the number of manual integrations will improve data integrity and efficiency – a key metric for any chromatography initiative.

Table 5.1: Data Integrity Operations (continued)

Title	Use/Purpose	Notes
Reprocessed Runs (Analytical Laboratory)	List of run name, date, and user. For any sample runs that were reprocessed (re-calculated). Could add reason, if collected by the system. For a specific date range.	Processing a sample run multiple times impacts efficiency and potentially data integrity. Any reworking should be scrutinized, as data manipulation is a possible reason for the change.
Aborted Runs (Analytical Laboratory)	List of run name, date, and user. For any system that permits a sample run (multiple samples in a group), where the user can stop the run any time prior to completion.	Aborting a run is one way to stop the creation of unwanted data (orphaned data). Large numbers of aborted runs also indicate other issues: anytime data is not used, there is inefficient use of resources.
Manual Override	List of method, process, date, and user. Include reason for override if system collects it. For a specific date range. Could also filter by user or method/ process.	Manual overrides indicate an abnormal event. This could be legitimate (such as resampling/retesting per a protocol) or suspicious (changing value to meet a specification).
Suspicious Samples (Analytical Laboratory)	List of samples assayed in system: run name, date, sample name, and user.	Look for names such as “demo”, “test”, “trial” or others, that might indicate bad behaviors, such as retesting until a desired value is obtained.
Method Performance (Analytical Laboratory)	List of site, analyst, method, start time, end time, and total time. Could also summarize performance for a time period (e.g., average time, number of executions).	Report has obvious performance uses, but the integrity goal is primarily looking for performance that is “too good to be true” and where shortcuts may be taken. In some countries, the analyst name should be anonymous to avoid disclosure of personal information.
Data Review (Analytical Laboratory)	List of method, review start time, review end time, and total time. Could also summarize performance for a time period (e.g., average time, number of executions).	Similar use to Method Performance report, for review/release of any method or process. Looking for performance that is “too good to be true,” indicating the possibility of shortcuts in review.
Duplicated File	List of time stamps, file #1, file creator, file #2, file creator.	Look in directories for files with duplicate creation dates. This might indicate the copying of a source file that is renamed and presented as a new sample result. This report is only applicable for systems that capture data in file-based structure.

Table 5.2: Data Integrity Governance

Title	Use/Purpose	Notes
Report Summary	List by site, business unit, each data integrity operations report, with number of report executions for a specific date range. Optional: add user ID of person executing the report to the list.	Intent is to learn if parts of the business operation are using data integrity reports available to them. Low numbers from reports such as System Backup, Audit Trail Reasons, and Short Run might indicate insufficient attention to data integrity as a routine part of process performance.
Chromatography Methods, Manual Integration (Analytical Laboratory)	List method, total number of injection peaks, number of manually integrated peaks, and percent manual integration. For a specific time interval.	This metric provides an opportunity for method improvement by finding methods with large numbers of manual integrations and applying technical expertise to convert them to automated integration, resulting in integrity and efficiency improvements.
Status Out of Sync	List batch, system out of sync, status in question, and date of status change. Search for all batches where expiry date is today or future date.	Primary use of this metric is detecting batches where test result status and batch status do not match, based on the normal process. For example, an environmental test for an approved batch should be “Approved/Released” and not “In Process”.
Right First Time	Depending on area, a list of either batches (manufacturing) or tests (QC laboratory) that were approved more than once, for a given date range.	In addition to an efficiency issue, this is also an indirect indicator of issues with data review, as an issue was discovered after a test or a product was initially released.
Access Changes	List of user, system, previous role, new role, and date of change. For a given date range.	Detects situations where users are granted elevated access, then returned to lower level in a short time period as this is not detected by “point in time” reports from systems.
Access Summary	List of site, system, role, and number of people with that role. This report is the access status on a specific date.	Permits a review for number of people in every system with enhanced access roles.
Multiple Roles	List of system, user ID, role for each person with more than one security role in a system. Sort report by system and user so duplicates are in successive rows of report for review.	This is indicated in situations where a single user has multiple roles. It permits review that assesses the potential for conflicting actions.

6 Appendix 1 – Data Integrity Gemba Checklist in the Laboratory

This Appendix provides examples for Gemba walks in a laboratory environment [11]. The table contains guidance for the leader as well as coaching questions that can be asked during the walk.

Table 6.1: Example of a Gemba Walk in a Laboratory Environment [11]

GEMBA Guide (A) Leader Self-Ask Questions	GEMBA Guide (B) Leader Coaching Questions
1. What is the process? Look for: Steps that add value, flow between steps, standardization of tasks	**1. What is the standard?** Hopefully it will be clear at a visual glance. Is a data integrity checklist used? Are laboratory leaders visible? Are your laboratory metrics visible? Is data integrity discussed in the laboratory?
2. What is normal/abnormal? Look for: Standard work, expected state, variation to the expected state	**2. How do we develop a standard?** Used where a standard is ambiguous or lacking. Are your data integrity reviews visual? How do you know if data integrity checks are done consistently?
3. What is working well? Look for: Standards being followed, ideas being generated, lessons shared	**3. How clear is the standard to those doing the work?** Reveals the depth to which standards have been used. Do you discuss laboratory metrics? Do you discuss OOS and data integrity CAPA items?
4. What is not being followed? Look for: Checklists not populated, equipment in poor condition, poor housekeeping, variation to standard work	**4. How clear is the standard to those not doing the work?** Leaders should require that they can quickly assess the status of safety, quality, and on-time output.
5. What is broken? Look for: Equipment requiring repair, safety hazards, status of line clearance controls	**5. How well are we performing against the standard?** The variation in responses can reveal a lot about how well people understand their standards. Who administers your laboratory systems? Are your standards controlled? Are personnel trained in data integrity?
6. What is not understood? Look for: Variation to standard, poorly constructed procedures, understanding of team priorities	**6. Why are we not performing to the standard?** This is a golden opportunity for a leader to practice the five why questioning. Do laboratory technicians understand their data integrity responsibilities/ procedures?
7. What is creating waste? Look for: Any forms of waste – transport, inventory, motion, waiting, overproduction, over processing, defects	**7. What can we do to improve the current condition?** This question can be used as a catch-all in any situation, any condition, any Gemba. How do your laboratory employees provide feedback on the data integrity program and opportunities for improvement?

Table 6.1: Example of a Gemba Walk in a Laboratory Environment [11] (continued)

GEMBA Guide (A) Leader Self-Ask Questions	GEMBA Guide (B) Leader Coaching Questions
8. What is creating strain? Look for: Poor workstation design, inadequate environmental and/or ergonomic design factors, overburdening of activities	**8. How can we make the abnormal condition more immediately visual?** Often the reason problems persist is because they go undetected. Do you make your data integrity audit findings visible? Do you share resolutions?
9. What is creating unevenness? Look for: Uneven production schedules, variation in staffing levels, process interruptions	**9. Why do you think I asked you these questions?** Uneven schedules and staffing may cause unintended consequences. As an analyst, do you have time to provide your data integrity checks? Does your sample run rate (volume/products) allow for proper data integrity review or audit?
10. What is not visible enough? Look for: Signals to problems, performance indicators, management presence, communication of team priorities, standards	**10. What other questions would you have liked me to have asked?** The main use of this question is for the leader's learning. It is also helpful for the team's assessment and learning of the data integrity review within the laboratory.

7 Appendix 2 – IMPACT Tool Applied to Data Integrity

This Appendix demonstrates the application of Braksick's IMPACT tool [12] in a data integrity context.

Table 7.1: Example of IMPACT Tool [12] Applied to Data Integrity

I	**Identify Goal**	Increase organization's knowledge and understanding of data integrity through development of a training and audit program for data integrity.
M	**Select the Measure To Deliver Goal**	1. Standard number of audits for data integrity within core functional areas 2. Increase number of personnel who have completed data integrity training course to 100% 3. Develop data integrity champions for the different functions and areas 4. Analyze data integrity observations for trends
P	**Pinpoint the Behaviors**	1. Encourage a speak-up culture where data integrity concerns, issues, or suggestions are shared in a timely manner in a neutral, constructive forum. 2. Promote and coach for enhanced attention to detail where data integrity is everyone's job through leadership and vision, as well as Gemba walks 3. Leaders engaged in data integrity audits, reviews, and metrics and actions established and communicated.
AC	**Activate the Consequences**	1. Cross-functional reviews and continuous improvement discussions as part of functional area KPI reviews. 2. Leaders privately acknowledge individuals for raising data integrity concerns. 3. Governance established for data integrity program.
T	**Transfer Knowledge and Skills To Sustain Change**	1. Lessons learned are documented and shared with wider workforce. 2. Establish Data Integrity Community of Practice to develop new skills and facilitate questions and answers and between different areas.
	LQIs/LBIs	**Leading** 1. Data integrity audits scheduled/planned 2. Examples of predictive analytics 3. Percentage of data integrity champions identified **Trended Lagging** 1. Percentage complete on data integrity training 2. Percentage of data integrity related internal audit observations

8 Appendix 3 – Corporate Data Integrity Program Case Study

This Appendix is provided as a case study to demonstrate how an organization can improve their data governance through practical measures.

8.1 Background

One company had active data integrity initiatives underway for some time but saw an opportunity to put more structure and governance in place to ensure greater visibility, prioritization, and coordination across the organization (GxP).

To enhance its data integrity program, the firm expanded its initiatives to include global and cross-functional activities as well as targeted site and functional activities throughout the organization. This included the paper and electronic data processes used in the functional areas of Manufacturing (GMP), QC Laboratories (GMP), and Clinical (GCP).

The company captured the key elements of its data integrity program in a global document approved by senior management from both Quality and Medical/Regulatory/Safety. The document describes the objectives needed to achieve the necessary identification, prioritization, and governance of its data integrity activities or action plan. The approach taken at both a corporate and functional level was outlined in this document to appropriately manage and ensure data integrity.

8.2 Program Objectives

The company defined the objectives of its data integrity program as follows:

- Ensure appropriate controls and processes are in place to assure data integrity
- Expand the initial data integrity program approach beyond commercial manufacturing
- Coordinate data integrity activities across multiple areas to ensure consistency and to enhance visibility of progress and outcomes
- Implement targeted improvements and program elements in sites/functional areas
- Enhance the existing QMS to improve or clarify data integrity expectations
- Educate, promote awareness, and share learning across the sites/functional areas related to data integrity
- Ensure effective overall governance of the company's data integrity initiatives
- Monitor results from escalations, reviews, investigations, site/function self-inspections and/or audit outcomes to ensure completion of corrective actions, and to detect and investigate trends
- Ensure the alignment of key third parties (such as contract manufacturing organizations, contract laboratory organizations, contract research organizations, and external partners) with data integrity expectations commensurate with risk

8.3 Governance

Because the company is a large pharmaceutical firm, it needed to develop a governance structure that would ensure both the appropriate integration of activities across the individual sites/functional areas, as well as provide visibility of the program and progress toward its action plan to Senior Quality Management. The company accomplished this by forming a Data Integrity Executive Committee and a Data Integrity Integration Committee, in addition to local site/functional committees.

The Data Integrity Executive Committee is comprised of Senior Quality Leadership from Corporate Quality Systems, Laboratories and Auditing, Development, Manufacturing, and Medical/Regulatory/Safety. This committee is responsible for:

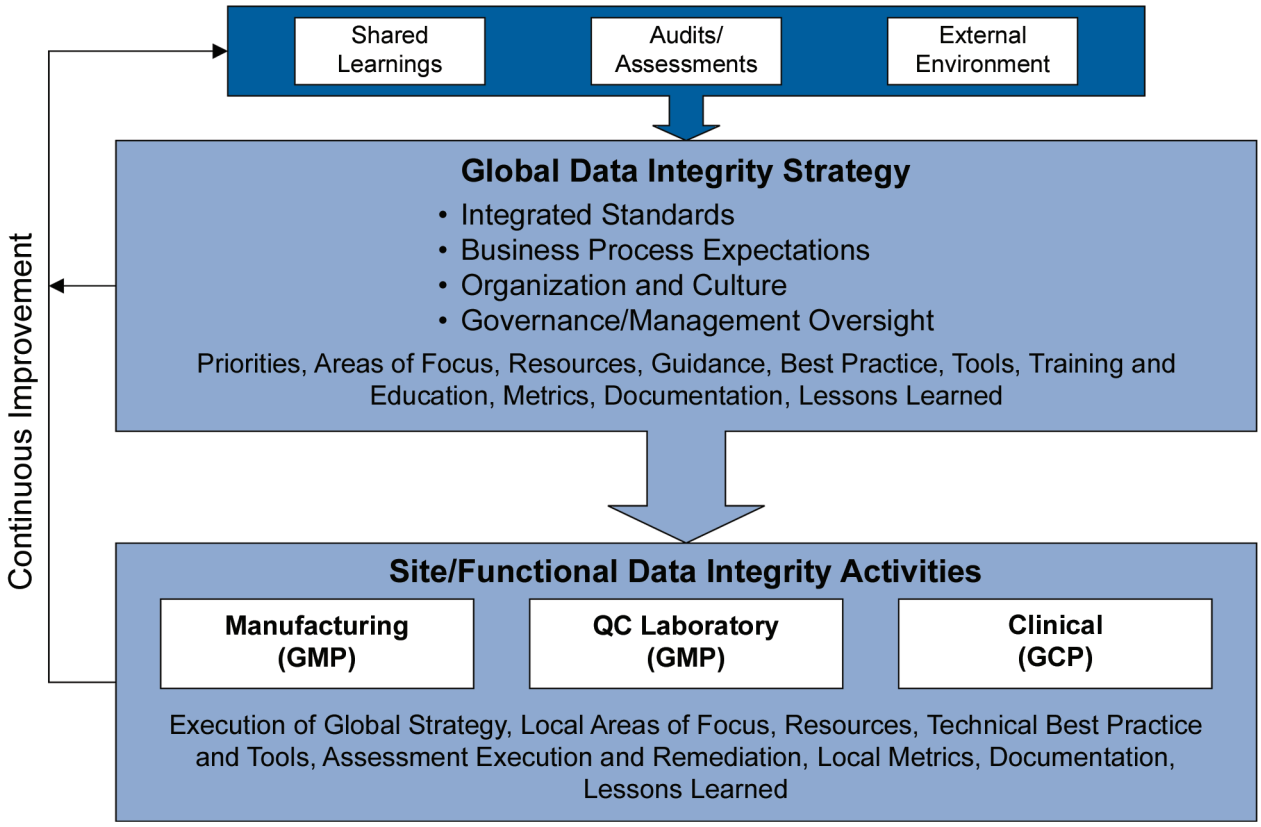
- Providing sponsorship for the execution of the program in their area
- Setting priorities and areas of focus for the program
- Ensuring that adequate resources are in place to support the program
- Reviewing key management control metrics
- Overall data integrity performance

The Data Integrity Integration Committee is comprised of management and associates from the same areas represented on the Data Integrity Executive Committee. This committee is responsible for:

- Maintenance of the program
- Communication of the program and expectations to practitioners
- Ensuring translation of the program into executable action plans
- Seeking and compiling input and opportunities from relevant site/functional areas
- Ownership of appropriate action plans and oversight of plan execution
- Identification and collection of overall data integrity performance measures

Figure 8.1 depicts the interactions of the global and site/functional areas in the program's scope.

Figure 8.1: Interactions within the Program's Scope



8.4 Program Action Plan

A detailed data integrity action plan was created to capture and provide visibility to the key activities needed to advance the program's objectives. The Data Integrity Integration Committee provides periodic updates to the Data Integrity Executive Committee throughout the year. The action plan is updated as needed, but at a minimum annually.

8.5 Conclusion

The more structured approach from the data integrity action plan successfully improved the data integrity mindset of employees at all levels in the company, and advanced the company as a whole further along the timescale of cultural excellence maturity.

9 Appendix 4 – Culture and Continuous Improvement Capability Road Map

This Appendix provides an example of a Culture and Continuous Improvement Capability Road Map [11] that could be applied as part of an organizational improvement paradigm.

Table 9.1: Culture and Continuous Improvement Capability Road Map [11]

Overall Process	Plan	Do/Assess	Check/Act	Monitor	Communicate
		Identify and Analyze	**Decide and Formulate**	**Review/Control**	
Identify and discuss concerns, issues, risks, and improvement opportunities with supervisor/ manager Provide input and/or propose improvement opportunities solutions as experts in the areas they are responsible for Engage in improvement opportunities Be aware of your actions' effects and potential consequences on output quality	Work with line management to align goals and develop a plan for your specific area	Identify risk, issue, or opportunity Access root cause, impact, and potential solutions	Decide to act or escalate the identified risk, issue, or opportunity using improvement methodologies Propose or participate in the solution	Provide feedback to supervisors on effectiveness of the solution KPIs LBIs Controls	Know when to communicate to your supervisor and peers Know what information to document for the record or future reference: knowledge management/ lessons learned Document event and solutions, decisions and rationale appropriately

Table 9.2: Learning Road Map [11]

Behavioral Requirements	Training				
	Technical Skills	Soft Skills	CI Program Specific (as needed)	Advanced (optional)	Support
System accountability matrix	Continuous improvement essentials		1. PDCA 2. DMAIC and others, as appropriate	1. Local training material for CI processes 2. Root cause analysis and Risk Management 3. Decision-making	1. Readily available job aids and tools • Templates/checklists/ examples • Case studies 2. Consultation provided by leaders, managers, and subject matter experts on specific identified opportunities/ improvements 3. Opportunities to practice • Real-time quality event identification and communication • Regularly-scheduled Gemba walks 4. Coaching • Quality event identification and communication • Practical application of continuous improvement 5. Knowledge sharing • Best practices • Lessons learned • Documentation and communication 6. Reinforcement and recognition • Expectations for proactive improvement are in place and managed as part of performance management • Proactive improvement discussions integrated into activities (e.g., daily, weekly, quarterly)

10 Appendix 5 – Regulatory Definitions of Data Terminology

This Appendix contains the original regulatory definitions that were compared in Section 3.1.

Table 10.1: GMP Regulations Using Raw Data, Complete Data, and Complete Information

Source	Reference	Content
US FDA GMP	21 CFR Part 211.194 (a) [26]	Laboratory records shall include complete data derived from all tests necessary to assure compliance with established specifications and standards, including examinations and assays
US FDA GMP	21 CFR Part 211.188 [26]	Batch production and control records shall be prepared for each batch of drug product produced and shall include complete information relating to the production and control of each batch.
EU GMP	EudraLEX Vol. 4 Chapter 4 [27]	**Records:** Provide evidence of various actions taken to demonstrate compliance with instructions, e.g. activities, events, investigations, and in the case of manufactured batches a history of each batch of product, including its distribution. Records include the raw data which is used to generate other records. For electronic records regulated users should define which data are to be used as raw data. At least, all data on which quality decisions are based should be defined as raw data
WHO	TRS No. 996 Annex 5 [3]	Data means all original records and true copies of original records, including source data and metadata and all subsequent transformations and reports of these data, which are generated or recorded at the time of the GXP activity and allow full and complete reconstruction and evaluation of the GXP activity. Data should be accurately recorded by permanent means at the time of the activity. Data may be contained in paper records (such as worksheets and logbooks), electronic records and audit trails, photographs, microfilm or microfiche, audio- or video-files or any other media whereby information related to GXP activities is recorded.
PIC/S	PI 041-1 (Draft 2) [7]	**Complete:** All information that would be critical to recreating an event is important when trying to understand the event. The level of detail required for an information set to be considered complete would depend on the criticality of the information…A complete record of data generated electronically includes relevant metadata.
UK MHRA	'GXP' Data Integrity Guidance and Definitions [5]	**6.2. Raw data (synonymous with "source data" which is defined in ICH GCP)** Raw data is defined as the original record (data) which can be described as the first-capture of information, whether recorded on paper or electronically. Information that is originally captured in a dynamic state should remain available in that state. Raw data must permit full reconstruction of the activities. Where this has been captured in a dynamic state and generated electronically, paper copies cannot be considered as 'raw data'…. In all definitions, the term 'data' includes raw data.

Table 10.2: GLP Definitions of Raw Data

Source	Reference	Content
US FDA GLP	21 CFR Part 58.3(k) [21]	Raw data means any laboratory worksheets, records, memoranda, notes, or exact copies thereof, that are the result of original observations and activities of a nonclinical laboratory study and are necessary for the reconstruction and evaluation of the report of that study.
US FDA GLP 2016 Proposed definition (not yet in force)	21 CFR Part 58.3 [21]	Raw data means all original nonclinical laboratory study records and documentation or exact copies that maintain the original intent and meaning and are made according to the person's certified copy procedures. Raw data includes any laboratory worksheets, correspondence, notes, and other documentation (regardless of capture medium) that are the result of original observations and activities of a nonclinical laboratory study and are necessary for the reconstruction and evaluation of the report of that study. Raw data also includes the signed and dated pathology report.
OECD	GLP No 1 [22]	**Section 2.3 item 7.** Raw data means all original test facility records and documentation, or verified copies thereof, which are the result of the original observations and activities in a study. Raw data also may include, for example, photographs, microfilm or microfiche copies, computer readable media, dictated observations, recorded data from automated instruments, or any other data storage medium that has been recognised as capable of providing secure storage of information for a time period as stated in section 10, below. (Section 10 not reproduced here.)
OECD	GLP No 17 [23]	**Data (raw data):** Data (raw data) may be defined as measurable or descriptive attribute of a physical entity, process or event. The GLP Principles define raw data as all laboratory records and documentation, including data directly entered into a computer through an automatic instrument interface, which are the results of primary observations and activities in a study and which are necessary for the reconstruction and evaluation of the report of that study. **Data (derived data):** Derived data depend on raw data and can be reconstructed from raw data (e.g., final concentrations as calculated by a spreadsheet relying on raw data, result tables as summarized by a LIMS, etc.).

Table 10.3: GCP Definitions of Source Data

Source	Reference	Content
EMA/CHMP/ ICH Tripartite GCP	Guideline for good clinical practice E6(R2) [24]	**Source Data:** All information in original records and certified copies of original records of clinical findings, observations, or other activities in a clinical trial necessary for the reconstruction and evaluation of the trial. Source data are contained in source documents (original records or certified copies). **8.1 Addendum** The sponsor should ensure that the investigator has control of and continuous access to the CRF data reported to the sponsor. The sponsor should not have exclusive control of those data. The investigator/institution should have control of all essential documents and records generated by the investigator/institution before, during, and after the trial.
US FDA Guidance for Industry	Electronic Source Data in Clinical Investigations [25]	Source data includes all information in original records and certified copies of original records of clinical findings, observations, or other activities in a clinical investigation used for reconstructing and evaluating the investigation. Access to source data is critical to the review and inspections of clinical investigations.

Table 10.4: Definition of Static and Dynamic Data Types

Source	Reference	Content
WHO	TRS No. 996 Annex 5 [3]	**Dynamic record format.** Records in dynamic format, such as electronic records, that allow for an interactive relationship between the user and the record content. For example, electronic records in database formats allow the user to track, trend and query data; chromatography records maintained as electronic records allow the user (with proper access permissions) to reprocess the data and expand the baseline to view the integration more clearly. **Static record format.** A static record format, such as a paper or pdf record, is one that is fixed and allows little or no interaction between the user and the record content. For example, once printed or converted to static pdfs, chromatography records lose the capability of being reprocessed or enabling more detailed viewing of baselines.
UK MHRA	'GXP' Data Integrity Guidance and Definitions [5]	A static record format, such as a paper or electronic record, is one that is fixed and allows little or no interaction between the user and the record content. For example, once printed or converted to static electronic format chromatography records lose the capability of being reprocessed or enabling more detailed viewing of baselines. Records in dynamic format, such as electronic records, allow an interactive relationship between the user and the record content. For example, electronic records in database formats allow the user to track, trend and query data; chromatography records maintained as electronic records allow the user or reviewer (with appropriate access permissions) to reprocess the data and expand the baseline to view the integration more clearly. Where it is not practical or feasibly possible to retain the original copy of source data, (e.g. MRI scans, where the source machine is not under the study sponsor's control and the operator can only provide summary statistics) the risks and mitigation should be documented.
PIC/S	PI 041-1 (Draft 2) [7]	Many electronic records are important to retain in their dynamic (electronic) format, to enable interaction with the data. Data must be retained in a dynamic form where this is critical to its integrity or later verification.
US FDA GMP	Data Integrity and Compliance with CGMP Guidance for Industry [4]	For the purposes of this guidance, static is used to indicate a fixed-data document such as a paper record or an electronic image, and dynamic means that the record format allows interaction between the user and the record content. For example, a dynamic chromatographic record may allow the user to change the baseline and reprocess chromatographic data so that the resulting peaks may appear smaller or larger. It also may allow the user to modify formulas or entries in a spreadsheet used to compute test results or other information such as calculated yield.

Table 10.5: Definitions and Requirements for Original Records and True Copies

Source	Reference	Content
UK MHRA	'GXP' Data Integrity Guidance and Definitions [5]	**6.11.1 Original record** The first or source capture of data or information e.g. original paper record of manual observation or electronic raw data file from a computerised system, and all subsequent data required to fully reconstruct the conduct of the GXP activity. Original records can be Static or Dynamic. **6.11.2 True copy** A copy (irrespective of the type of media used) of the original record that has been verified (i.e. by a dated signature or by generation through a validated process) to have the same information, including data that describe the context, content, and structure, as the original. A true copy may be stored in a different electronic file format to the original record if required, but must retain the metadata and audit trail required to ensure that the full meaning of the data are kept and its history may be reconstructed. Original records and true copies must preserve the integrity of the record. True copies of original records may be retained in place of the original record (e.g. scan of a paper record), if a documented system is in place to verify and record the integrity of the copy. Organisations should consider any risk associated with the destruction of original records. It should be possible to create a true copy of electronic data, including relevant metadata, for the purposes of review, backup and archival. Accurate and complete copies for certification of the copy should include the meaning of the data (e.g. date formats, context, layout, electronic signatures and authorisations) and the full GXP audit trail. Consideration should be given to the dynamic functionality of a 'true copy' throughout the retention period (see 'archive'). Data must be retained in a dynamic form where this is critical to its integrity or later verification. If the computerised system cannot be maintained e.g., if it is no longer supported, then records should be archived according to a documented archiving strategy prior to decommissioning the computerised system. It is conceivable for some data generated by electronic means to be retained in an acceptable paper or electronic format, where it can be justified that a static record maintains the integrity of the original data. However, the data retention process must be shown to include verified copies of all raw data, metadata, relevant audit trail and result files, any variable software/system configuration settings specific to each record, and all data processing runs (including methods and audit trails) necessary for reconstruction of a given raw data set. It would also require a documented means to verify that the printed records were an accurate representation. To enable a GXP compliant record this approach is likely to be demanding in its administration.
US FDA GMP	21 CFR Part 211.68(b) [26]	Hard copy or alternative systems, such as duplicates, tapes, or microfilm, designed to assure that backup data are exact and complete and that it is secure from alteration, inadvertent erasures, or loss shall be maintained.

Table 10.5: Definitions and Requirements for Original Records and True Copies (continued)

Source	Reference	Content
US FDA GMP	21 CFR Part 211.180(d) [26]	Records required under this part may be retained either as original records or as true copies such as photocopies, microfilm, microfiche, or other accurate reproductions of the original records. Where reduction techniques, such as microfilming, are used, suitable reader and photocopying equipment shall be readily available.
US FDA GMP	Data Integrity and Compliance with CGMP Guidance for Industry (draft) [4]	Electronic copies can be used as true copies of paper or electronic records, provided the copies preserve the content and meaning of the original data, which includes associated metadata and the static or dynamic nature of the original records. True copies of dynamic electronic records may be made and maintained in the format of the original records or in a compatible format, provided that the content and meaning of the original records are preserved and that a suitable reader and copying equipment (for example, software and hardware, including media readers) are readily available.
PIC/S	PI 041-1 (Draft 2) [7]	The original record can be described as the first-capture of information, whether recorded on paper (static) or electronically (usually dynamic, depending on the complexity of the system). Information that is originally captured in a dynamic state should remain available in that state.
EMA/CHMP/ ICH Tripartite GCP	Guideline for good clinical practice E6(R2) [24]	**Source Documents:** Original documents, data, and records (e.g. hospital records, clinical and office charts, laboratory notes, memoranda, subjects' diaries or evaluation checklists, pharmacy dispensing records, recorded data from automated instruments, copies or transcriptions certified after verification as being accurate copies, microfiches, photographic negatives, microfilm or magnetic media, x-rays, subject files, and records kept at the pharmacy, at the laboratories and at medico-technical departments involved in the clinical trial). **Certified Copy:** A copy (irrespective of the type of media used) of the original record that has been verified (i.e., by a dated signature or by generation through a validated process) to have the same information, including data that describe the context, content, and structure, as the original. When a copy is used to replace an original document (e.g., source documents, CRF), the copy should fulfill the requirements for certified copies.
EMA	EMA Guideline on GCP compliance in relation to trial master file [53]	A certified copy is a paper or electronic copy of the original record that has been verified (e.g. by a dated signature) or has been generated through a validated process to produce a copy having the exact content and meaning of the original.

Table 10.6: Definitions around Hybrid Systems

Source	Reference	Content
UK MHRA	'GXP' Data Integrity Guidance and Definitions [5]	**4.3 Hybrid** Where hybrid systems are used, it should be clearly documented what constitutes the whole data set and all records that are defined by the data set should be reviewed and retained. Hybrid systems should be designed to ensure they meet the desired objective.
WHO	TRS No. 996 Annex 5 [3]	The use of hybrid systems is discouraged, but where legacy systems are awaiting replacement, mitigating controls should be in place… A hybrid approach might exceptionally be used to sign electronic records when the system lacks features for electronic signatures, provided adequate security can be *maintained*… Replacement of hybrid systems should be a priority.
EU GMP	Chapter 4 [27]	**Generation and Control of Documentation** 4.1 All types of document should be defined and adhered to. The requirements apply equally to all forms of document media types. Complex systems need to be understood, well documented, validated, and adequate controls should be in place. Many documents (instructions and/or records) may exist in hybrid forms, i.e. some elements as electronic and others as paper based.

Table 10.7: Audit Trail Definitions and Requirement

Source	Reference	Content
UK MHRA	'GXP' Data Integrity Guidance and Definitions [5]	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 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.
EU GMP	Annex 11 [17]	**9. Audit Trails** Consideration should be given, based on a risk assessment, to building into the system the creation of a record of all GMP-relevant changes and deletions (a system generated "audit trail"). For change or deletion of GMP-relevant data the reason should be documented. Audit trails need to be available and convertible to a generally intelligible form and regularly reviewed.
US FDA GMP	Data Integrity and Compliance with CGMP Guidance for Industry (draft) [4]	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. An audit trail is a chronology of the "who, what, when, and why" of a record. For example, the audit trail for a high performance liquid chromatography (HPLC) run could include the user name, date/time of the run, the integration parameters used, and details of a reprocessing, if any, including change justification for the reprocessing. 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.
OECD	GLP No 17 [23]	**3.4. Audit trails** 80. An audit trail provides documentary evidence of activities that have affected the content or meaning of a record at a specific time point. Audit trails need to be available and convertible to a human readable form. Depending on the system, log files may be considered (or may be considered in addition, to an audit trailing system) to meet this requirement. Any change to electronic records must not obscure the original entry and be time and date stamped and traceable to the person who made the change. 81. Audit trail for a computerised system should be enabled, appropriately configured and reflect the roles and responsibilities of study personnel. The ability to make modifications to the audit trail settings should be restricted to authorised personnel. Any personnel involved in a study (e.g. study directors, heads of analytical departments, analysts, etc.) should not be authorised to change audit trail settings.

11 Appendix 6 – Requirements Planning

11.1 Introduction

The Computerized System Life Cycle denotes a framework of human and technical activities to plan, specify, configure or code, verify, and report on computerized system validation. The objective of the computerized system life cycle is to ensure that data is complete, consistent, and accurate during all phases of the data life cycle (capture, processing, review, reporting and use, retention and retrieval, and destruction).

The activities within any life cycle are inextricably linked: a gap or lack of rigor in one activity will detract from the success of subsequent activities, whereby latent data integrity vulnerabilities are built into the computerized system. It is important to get a solid foundation.

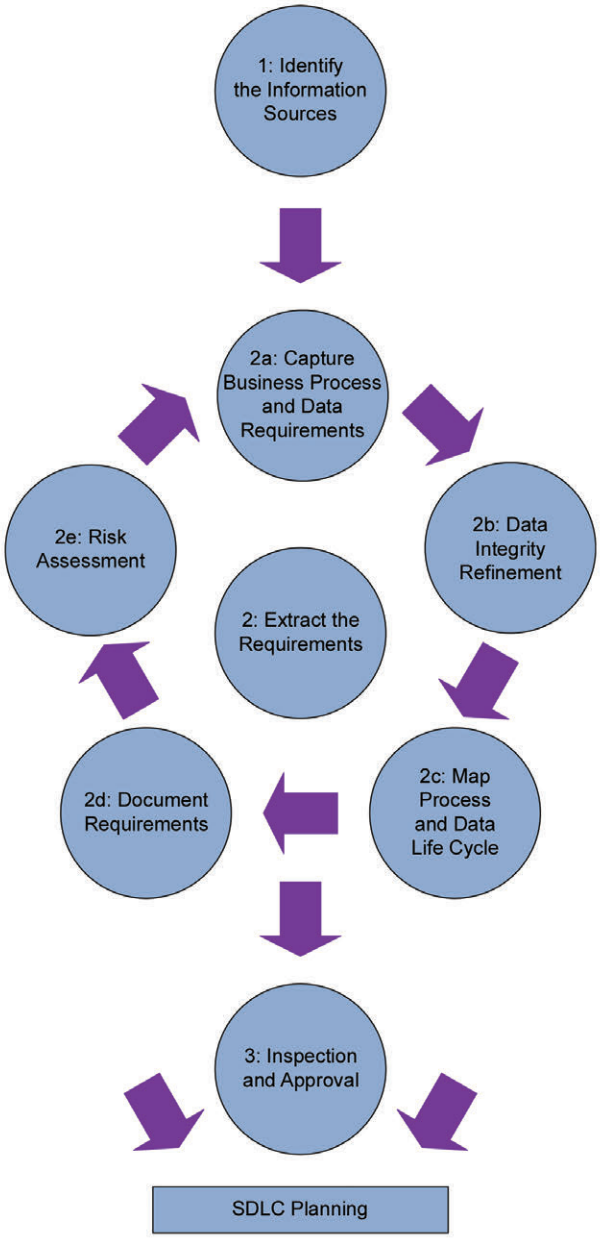
The focus of this Appendix is the requirements activity within the Project phase of the computerized system life cycle. This Appendix aims to increase awareness of the importance of the requirements definition and ways to approach it in order to reduce the risk of designing and building data integrity vulnerabilities into the system. Suggestions of tools are provided to help in this regard.

Data integrity requires a holistic approach across all computerized systems supporting the data life cycle, and it is important to ensure that there is a robust definition of the required processes and data so that the design and implementation activities have a solid foundation from which to work.

The expected computerized system life cycle activities (e.g. specification and verification approaches), derived first from the process risk assessment, and then from an FMEA-type approach at the system level, should be included in quality agreements and contracts with suppliers.

An outcome of the requirements gathering process can be the definition of what documentation needs to be created, the associated content of deliverables, and access to data, audit trails, data ownership, and location.

Figure 11.1: Requirement and Planning Steps
Used with permission from Empowerment Quality Engineering Ltd., www.empowermentge.com.



11.2 Requirements

Requirements are a major source of errors. Research by Wiegers [54], found that 41-48% of errors originated in the requirements phase, and that the typical root cause was false assumptions (what was produced vs. what was expected).

When a requirement is poorly defined, its functionality is built on a false assumption. This issue is compounded as subsequent functionalities are invariably impacted (as a result of control flows and interconnectivity of functionality). Other functions may leverage the outcome of the flawed requirement implementation and become infected by a ripple effect.

Identifying a false assumption late in the life cycle (e.g., at the verification stage) can result in data integrity failure. Fixing this issue may reveal other issues as a result of ripple effects. Pressures to fix these issues late in the life cycle are constrained by time and cost issues, with subsequent reduction in quality; all of which increases data integrity risk.

The creation of requirements normally concentrates on capturing the business process for development into a computerized system; however, data integrity demands additional attention to quality, data, interfaces, architecture, design, implementation, testing, and operational needs.

11.3 Requirements Analysis

The requirements phase is where the needs of the computerized system are identified, such as:

- Business process to be automated
- Types of users who will operate and support the computerized system
- Data to be created, ingested, processed, transferred, and stored
- Speed of interactions
- Interfaces between modules and systems
- Storage capabilities
- Training
- Security
- Compliance

An omission or incomplete capture of a requirement can have serious connotations to the data integrity capabilities of the computerized system.

Requirements analysis is the *"process of studying user needs to arrive at a definition of a system, hardware, or software requirements (IEEE)"* [55] for the proposed computerized system: the process and associated data life cycle.

Requirements analysis is the activity of identifying, documenting, studying, and qualifying the risk for the computerized system needs. It is an iterative process as requirements can be identified at any time during the system life cycle. The details become more granular as more information is elicited.

The following steps provide guidance on conducting the requirements definition phase of a computerized system with a focus on data integrity. These steps can be performed sequentially or in parallel.

Step 1: Identify and Document Your Sources of Information

This step should include as many stakeholders as possible, (e.g., business owners, SMEs, quality, developers, IT, testers, and suppliers). Common source of the "needs" of a computerized system include:

- Workshops: business process walk-through sessions
- Interviews with stakeholders, SMEs and end users
- Observations of existing process

- Tailored questionnaires
- Information gathering: periodicals and industry trends
- SOPs, work instructions, and templates
- Help and user guides
- Audit observations
- CAPA lists (Tool)
- Defect lists (Tool)
- Regulations (e.g., 21 CFR Part 11 [41]) and guides
- Industry standards
- Trade shows, periodicals, industry groups (e.g., ISPE)
- FDA warning letters

Document all sources of information and their owners during the requirements process for future reference.

Step 2: Extract the Information

This step usually involves a top-down approach based on the product and process understanding (*ISPE GAMP® 5, Appendix D1* [2]), but there may be instances where a bottom-up approach is used when access to an SME provides granular details.

High-level activities:

- Focus on establishing the business processes.
  - Iteratively decompose the generic business process into more granular, distinct processes and functions
  - Capture and classify the processes, data, and associated interactions
  - Identify the data life cycle
  - Assess the overall process risk
- Investigate potential data integrity vulnerabilities

Step 2a: Capture the Process and Data Requirements

Requirements gathering involves an act of posing questions to multiple information resources. The starting point is to discover what the current system does and what the proposed system needs to do. The “5 W’s +” (Who, What, Why, Where, When, How), is a useful and simple technique to capture the requirements.

Process-Focused Questioning

- **Who:** Who or what subprocess is involved with or controls the action to be performed? Who is the process owner? What are their roles and what are their permissions?

- **What:** What is the specific action to be performed? What constraints are applicable when automating this process (e.g., regulations)? What initiates the action? What concludes the action? What is the importance of this process? What happens after the action? What is the data involved with the action: input and output? (See Data-Focused Questioning, below). What does the process do with the data? What are the known problems with the current manual process? How can these be prevented (via automatic or manual checks and controls)? What is the business process manual reconciliation routine, such as record counts, range checks, time sequence checks, to be automated?

- **Why:** Why is this process required? Can it be automated? What is the rationale and the importance of this process? What is the importance of the data that is ingested, processed, transferred, and stored?

- **Where:** Where does this activity occur? Where can it occur? Is there a system that interacts with the process, either providing input or receiving output?

- **When:** When does this process occur? Does it occur within a sequence of processes or in parallel? Can it occur in isolation? Is it dependent upon any previous action, state, or data?

- **How:** How is the activity performed? Is it efficient? What are the activities within the process? How is the process executed? What are the rules that govern the process? Where are these defined? How are errors in the business process made known? What is the impact of a process error on the data? How is a business process error corrected?

Document any assumptions made for clarification and risk assessment. Document the activity owner and the source of the information. Capture any data identified during this step.

Data-Focused Questioning

For each data item identified above, identify its data life cycle (*"generation and recording through processing (including analysis transformation or migration), use, data retention, archive/retrieval and destruction"* [5]) and constraints for the creation of a data dictionary. For example, some questions to ask in relation to each process elicited above:

- **How:** How critical is the data? Will it be used for quality, finance, pharmacovigilance, regulatory submissions, etc.? Is the data item with GxP impact dependent upon another data item? Is the data item required for the system audit trail?

- **Who:** Who reviews and approves the data item? Who owns the data item (Attributable)? What are their roles? What are they permitted to do? What are they not permitted to do? Who are users of the process (roles and responsibilities)? How will these be replicated within the computerized system (e.g., operator account permissions, data base account permissions, interface account permissions)? Who has access to the audit trail?

- **What:** What is the data item (e.g., numbers, text, images)? What is the minimum and maximum size of the data item? What happens to the data when it has been used or processed? What is the data range applicable to the data item? What type of information can the data item contain? Is the data dependent on other data? Is the data used in reports or transferred to other systems? What is the meaning of the data and is this captured? Is all of the metadata captured such as size, origin, range, and so forth?

- **Why:** Why is the data item used? Who can create, ingest, process, transfer, and store the data item?

- **Where:** Where does the data originate from? Where will it be stored? Where will it be transferred to?

- **When:** When is the data item created, ingested, processed, transferred, and egested/stored during the process? What is the sequence of access and control on the data? What are the roles and permissions of those involved? When is the data item reviewed and approved?

- **How:** How is the data item ingested and created? How is it processed? How is it egested to another process or to storage? Will this be a manual or automatic process? How will the data be used in line with other data? Are these of the same data type, for example, a whole number and a decimal number?

- **ALCOA+:** Is the data uniquely identified and does it conform to the definitions of ALCOA+:

  - Attributable
  - Legible
  - Contemporaneous
  - Original
  - Accurate
  - Complete
  - Consistent
  - Enduring
  - Available

What checks could be applied to the system to ensure ALCOA+ at each data integration? For example, only a specific role or process can write data to a storage location during a specific sequence of events. Document any assumptions made for clarification and risk assessment. Document the data owner and the source of the information. Capture any process details identified during this activity.

Non-Functional Requirement Questioning

The previous examples center largely on the functionality of the system; however, non-functional requirements also need to be considered. Of all of the information sources listed in Step 1, is it is recommended to engage technical SMEs (e.g., IT administrators, database administrators) to assist in this activity. Some items to consider when eliciting non-functional requirements include:

- **Performance:** A discussion centered on the high volume use of the computerized system may reveal the need for requirements associated with processor capacity or physical memory consumption. This discussion may result in dedicated design considerations or the use of a specific programming language that supports robust process memory management.

- **Data Replication:** A discussion regarding the risk of data corruption when it is at rest, e.g., a system deletion or overwrite may result in the requirement for a referential database management system. This discussion may then evolve into the requirement for data replication.

- **Architecture:** A discussion regarding the integrity of the computerized system may raise concerns regarding denial of service, altered system footprints, opened system ports (reflecting a security breech), or services password management.

The discussion may result in new requirements, for example, self-diagnosis (i.e., system tracing and event logging capabilities), changing vendor default passwords, shutting down unnecessary operating system services, applying enhanced technical considerations (e.g., technical defenses such as common password blacklists) in order to reduce the burden solely on end users for password, or enhanced system monitoring (e.g., monitoring failed password attempts).

- **Supplier Contract:** A discussion of internal or external supplier capability or resources involved in deployment, use, and maintenance may result in requirements centered on regulatory compliance training, data integrity training, delivery of life cycle documentation, and increased collaboration between supplier and validation.

Step 2b: Data Integrity Refinement

Negative testing "aims to demonstrate that the software does not work" as intended [56]. Apply this mindset to data integrity vulnerabilities. Review the information and ask: how would the proposed requirement not work? It should not cause data integrity exploits from, for example:

- Ineffective software implementation
- Inadequate software testing
- The potential for operator key stroke mistakes (87 rather than 78)
- Operator deceit (backdating test results)
- IT failing to act on a hardware alarm that results in data corruption

For each requirement or at each stage in the business process and the data lifecycle, consider potential data integrity issues and subsequent measures to counter them. Consider potential origins of data integrity faults across functional and non-functional requirements as well as associated life cycle processes in order to increase focus, for example:

- **Human Level:** Consider what was poor with the previous project. What issues were found during production? Was there an audit with a set of recommendations?
    - **Project Management:** Failure to allocate time for effective reviews during requirements, design, and implementation, or to allocate time for defect fixing cycles during testing
    - **Test Management:** Insufficient prioritization of high-risk areas, limited focus on data integrity test techniques, and lack of visibility of design outputs for test design all lead to ineffective testing that only focuses on the positive aspect of the system
- **Technical Level:**
    - **Functional Level:** Insufficient software and hardware design flows that ignore unfamiliar areas of functionality needs, or ineffective SOPs that fail to provide adequate instruction for backing up, restoring, and verifying data repositories
    - **Non-Functional Level:** Hardware material fatigue, defective operating system patches, power outages, or processor bottlenecks

Again, the approach is to ask questions when reviewing documented resources, interviewing SMEs, or conducting workshops. The maxim "there is no such thing as a stupid question" applies during this approach as a question can spark off a discussion thread in a previously undocumented area.

For example, how can the system prevent unintended changes to the data when it is at rest?

- How should the data item be stored? For example, refer to (Process and Data) Data flow model showing that data is shared by more than one process in Step 2c. Where should the data item be stored: in a flat file, or a hierarchical database, or a relational database, or an object-oriented database? What will the size of the data store be? How many data items will it contain and what is the maximum size of each data item?

- How big will the data store grow over time? What would happen to the processing data if the capacity of the data store is reached and the process data cannot be written to the store? What happens if connectivity to the data store is broken? Can the data store be moved (e.g., backup or archived)? Will the data ever need to be restored? Will any of these actions need to be logged anywhere for self-diagnosis (a topic for design and implementation)?

- What would the risk of the integrity of the data be, if any of the above questions resulted in a negative scenario? Would processed data be lost? Would stored data be overwritten?

- How can the risks be mitigated? Are manual checks required via SOPs or could automated controls be implemented?

Capture all unknowns, uncertainties, and assumptions (e.g., possible dependencies to other requirements) as these may be important in the data integrity aspect of the requirements process.

Tool Example: Checklists

Checklists help stimulate discussion. The following is a checklist with a specific focus on data integrity that builds upon that guidance. This list is not definitive and is provided as an example.

Table 11.1: Data Integrity Checklist Tool Example (in addition to *ISPE GAMP® 5, Appendix D1* [2])

Area	Example
Data	• Review the business data and control flow to be automated • Have the following been identified within the business process? - Data collection - Data ingestion into the system - Data processing (data control) - Data transfer (between functions or systems) - Data storage • Have the constraints on the data been identified? What can the data not do? (e.g., the data cannot be deleted; the data must not be amended by certain roles; the data must be encrypted in transit and at rest (storage)). How will contemporaneous saving of the data at the end of a step and before the next step be enforced? • Has the data model been defined by using ALCOA+ and are there any gaps in the ALCOA+ model? Is the metadata defined? • Define the metadata: - For example, what is the current size of the data item? - What will be the future size of the data item? Consider all of the data sizes of the system to determine the system storage capacity now and in the future. - What is the type of the data item? (e.g., is it a numeric value, floating point value (to three decimal places), alphabetical values, hexadecimal value, Arabic character set?) - Output format for display and printing purposes - Default value - Validation rules for checking allowable data upon ingestion - Derivation formula (if it is a calculated value) - Data owner - Data ranges (minimum values, maximum values) - Future data ranges (facilitates the future proofing of the system design for easier change control) • Include a requirement that the system will perform validation checks on manually-entered data or data received via an interface from another system. • What are the dependencies of the data to other data? What data must be in existence before this data can exist in the processing (referential integrity)?

Table 11.1: Data Integrity Checklist Tool Example (in addition to *ISPE GAMP® 5, Appendix D1* [2]) (continued)

Area	Example
Data (continued)	• Are there any hardware considerations when determining the control and data flow of the business process, e.g., disk storage size, or connectivity between subsystems? • What data will be presented to end users and how it will be presented? Are there risks associated with report generation and how may this impact data presentation, e.g., US letter page size vs. UK A4 page size? • How will the data be backed up and archived? • Will data need to be migrated from an existing system? Has the existing data been mapped to the new data dictionary? Have inconsistencies between the two data dictionaries been identified for further analysis?
Functions	• Will the computerized system indicate that something may go wrong via warning threshold functionality, and identify when something goes wrong by providing alarms? Will it help provide the capability to troubleshoot where a problem occurred, e.g., trace logging to identify when and where a data item was incorrectly changed, or system logging to indicate changes to the computerized system by an administrator account? • Will the computerized system indicate that something may occur that could have a negative impact on data integrity, or instantly indicate that there is a fault in the system? • Will the system clock be locked to all system users? Will it be synchronized to a centralized server clock?
Regulatory Needs	• Are the regulations (e.g., Annex 11 [17], 21 CFR Part 11 [41]) defined as specific requirements? • Will the system support the forensic review of data? • Are system administrator rights assigned to individuals with a direct interest in the data? • If data storage is to be contracted to a third party, do contracts specify the ownership of the data, the geographical location, and that archived data should be locked? • Will the audit trail be locked at the source?
Security	Security is highly coupled to data integrity: a strong data integrity-controlled system will also yield a more secure system. Consider system availability considerations, such as what happens if a subsystem is not accessible (e.g., denial of service) or if patient data is suddenly available to a public site and the range of mitigation actions that need to be applied. For example: • Will logical security be designed and built into the system (i.e., specific design and implementation requirements)? Will the system be deployed in a secure environment? Has a risk assessment been completed for this environment? • What security controls can be built in to enhance data integrity considerations (e.g., the adoption of design techniques such as state transition charts, defensive programming techniques, robust coding standards, or the use of a referential database instead of spreadsheets)? • What are the controls for administration of the system? What can the system administrator do? Do they have access to the data store? If so, how can this be controlled? Is the role of system administrator distinct and removed from the roles for operational use? If not, will there be a policy of role privilege escalation to denote when the administrator account is accessed? • Will the system topology be reviewed for security gaps? For instance, what operating system or firmware accounts are required to install and operate the system? What system ports and operating system services are required? What accounts, port numbers, operating services, etc. are not required for system operation? When will these be removed, closed, or shutdown?

Table 11.1: Data Integrity Checklist Tool Example (in addition to *ISPE GAMP® 5, Appendix D1* [2]) (continued)

Area	Example
Security (continued)	• What is the management policy for the creation, maintenance, and retirement of user accounts and associated technical controls to reduce the risk of compromise? Will the system detect and record all attempts that fail identification, authentication, or authorization constraints? Will user accounts be deactivated rather than deleted when a user leaves the organization? • What is the authentication, and encryption mechanism for data exchange between systems? Will the system prevent corruption of data during transit between interfaces? Will the system discover if data was corrupted during transit and have the intelligence to facilitate retransfer of data? • Will the data storage be in a secured location? Who has access to the location? What controls are required to ensure data storage security? Will the system/data location protect its hardware components from damage or theft? • What code policies will be used to implement the design? • What activities will be required to ensure that unauthorized malicious programs do not infect the system, either during implementation or in operational use? • Will the system need to be subjected to vulnerability assessments and penetration tests? • Will the system restrict the ingestion of data to predetermined types that are not susceptible to viruses?
Usability	• Consider how the screen layout of the Graphical User Interface (GUI) facilitates ease of data creation, processing, and review. How will the audit trail be presented for ease of use? Given that human interaction with a computerized system is a major source of data integrity issues, how will the use of the GUI design limit human actions with the system (e.g., reducing three key strokes to two)? • Will the system provide meaningful system feedback? For example, providing understandable error messages rather than obscure developer statements (e.g., "incorrect file format, please use either .xls or .csv" and not "#1548668: invalid pointer")? • When considering the usability of a computerized system, focus on the issues caused by the manual business process by checking the CAPA system or defect logs. Seek to consider these problems as technical controls. Focus on early error detection and feedback (e.g., introduce various levels of data checks upon entry into the system).
Non-functional Requirements	• What are the activities required to support the availability of the computerized system, such as high availability of fault tolerance, scalability, and performance? How will such aspects of the computerized system impact the integrity of the data? For example: - How easy will the system scale for future use? How many users should the system cope with now and how many should the system cope with in two years? - What happens when system performance limits are reached? For example, what is the impact to the data if system memory resources are maximized? Will in-memory queued data be dropped or overwritten if access to the data store or remote system is denied? What should happen at the performance extremes? Should the system slow down gracefully and seek to deal with all data processing or will the system simply run out of resources (resulting in a system crash and shut down)? How will this scenario affect the data being processed at that juncture? - What is the system availability? What happens if the system is not available for use (e.g., slow response to remote systems or broken resources)? How will this impact data processing of other aspects of the business? What will the SLA instruct the operations help desk to do to ensure rapid recovery after a failure event? - How will the support personnel know that there is a potential problem with the system? What are the monitoring needs of the system? Will the system support error recovery (e.g., fail over to "hot" standby system)? What will the backup and restore process involve?

Table 11.1: Data Integrity Checklist Tool Example (in addition to *ISPE GAMP® 5, Appendix D1* [2]) (continued)

Area	Example
Non-functional Requirements (continued)	- What are the data integrity concerns for system installation? Are there any checks for the configuration of the deployed system, default system accounts, security checks (e.g., hardened servers where unnecessary server accounts are removed)? Are these checks automated? - How will mitigation actions to such issues be verified?
Interfaces	• How can the end user influence data integrity? Can the number of interactions to the system be reduced to limit the number of key strokes and thus reduce the risk of error? • Can automated interfaces between systems reduce the need for users to manually transfer data between systems and thus facilitate data amendments? For example, add standalone systems to the network in order to utilize common security constraints, access constraints, and network time servers. • What is required to prevent data integrity issues from occurring between systems? For example, implement encryption of data messaging, use of automated error detection to ensure that the data was not corrupted in transit, the use of message handshaking and acknowledgements to ensure that the correct communication point is established and that the data has been received as expected.
Enduring	• Is it known how reliable and fault tolerant the system will need to be against failure? • Are mechanisms in place to detect/restart individual components that may fail? • Is it necessary for any kind of alternative system for business continuity to be provided? • Are data backup and recovery procedures required? • If yes, will they be manual or automated? • If no, what would happen if the data store crashed? • Is it necessary to log transactions or processes that have been initiated?
Performance	• Are the speed and accuracy requirements of any data processing known? • If yes, are they achievable? Are they measurable? • How will the speed and accuracy be accepted and tested? • What is the maximum number of events that are expected to occur at the same time? • Has the user's perceived performance (actual time for them to perform a task) been considered? • Have transaction times been taken into account for access and response to data stores and interfaced systems?
Requirements Assumptions	• Capture all assumptions (e.g., dependencies to other requirements) with the business process for reducing the scope of human error further within the computerized system project. • Test requirements • Consider the testability of all requirements. Is there any constraint on how the system can be validated? Can validation perform data integrity specific tests (e.g., performance tests, penetration tests, negative tests)? Have all negative scenarios been considered for testing? Are they economically viable to test?

Step 2c: Map the Process and Data Life Cycle for Further Analysis

Capture the evolving list of requirements in a spreadsheet, word document, or a requirements management tool. Invariably, the use of the 5W's+ approach will generate a lot of information that can become subject to change as more knowledge is attained.

At some stage in the gathering process, use of schematic notations will prove beneficial to capture the information in an easy to understand manner. Business Process Modeling is a schematic approach to facilitate a process for analysis, understanding, review, and agreement. There are many techniques available, e.g., Structured System Analysis and Design Method (SSADM), Unified Modeling Language (UML), or a customized approach that can be used to identify and structure this information. Some examples include:

- (Process) Use Cases help identify the actors of the process and associated actions of the controlling process.

  The following example demonstrates a file approval process. The diagram informs the reader of the actions performed (e.g., logon, select a file, and electronic signature). The diagram hints at the associated data required, such as login data (user ID, password), file status value, file (containing multiple data values), approval status, and so forth.

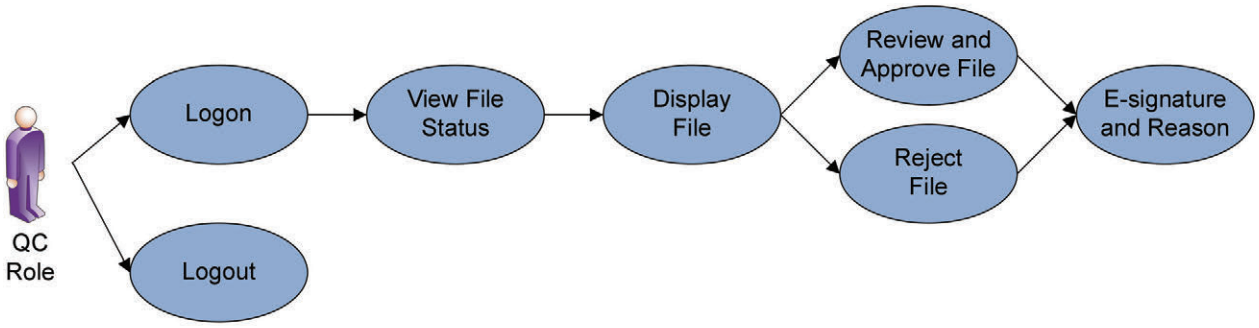
  Given that the diagram captures the process at a certain level, there are still questions to answer such as: What is the file? What data does it contain? What does an approved file look like? A new iteration of the 5W's+ will begin allowing a further decomposition of information to be captured.

  Thinking negatively is also beneficial. For example, what could happen if a user tries to change the PC time clock or turn off the audit trail?

Tool Example: Use Case (UML)

The following use case displays the QC role reviewing and approving a file.

Figure 11.2: Use Case for QC of a File
Used with permission from Empowerment Quality Engineering Ltd., www.empowermentge.com.

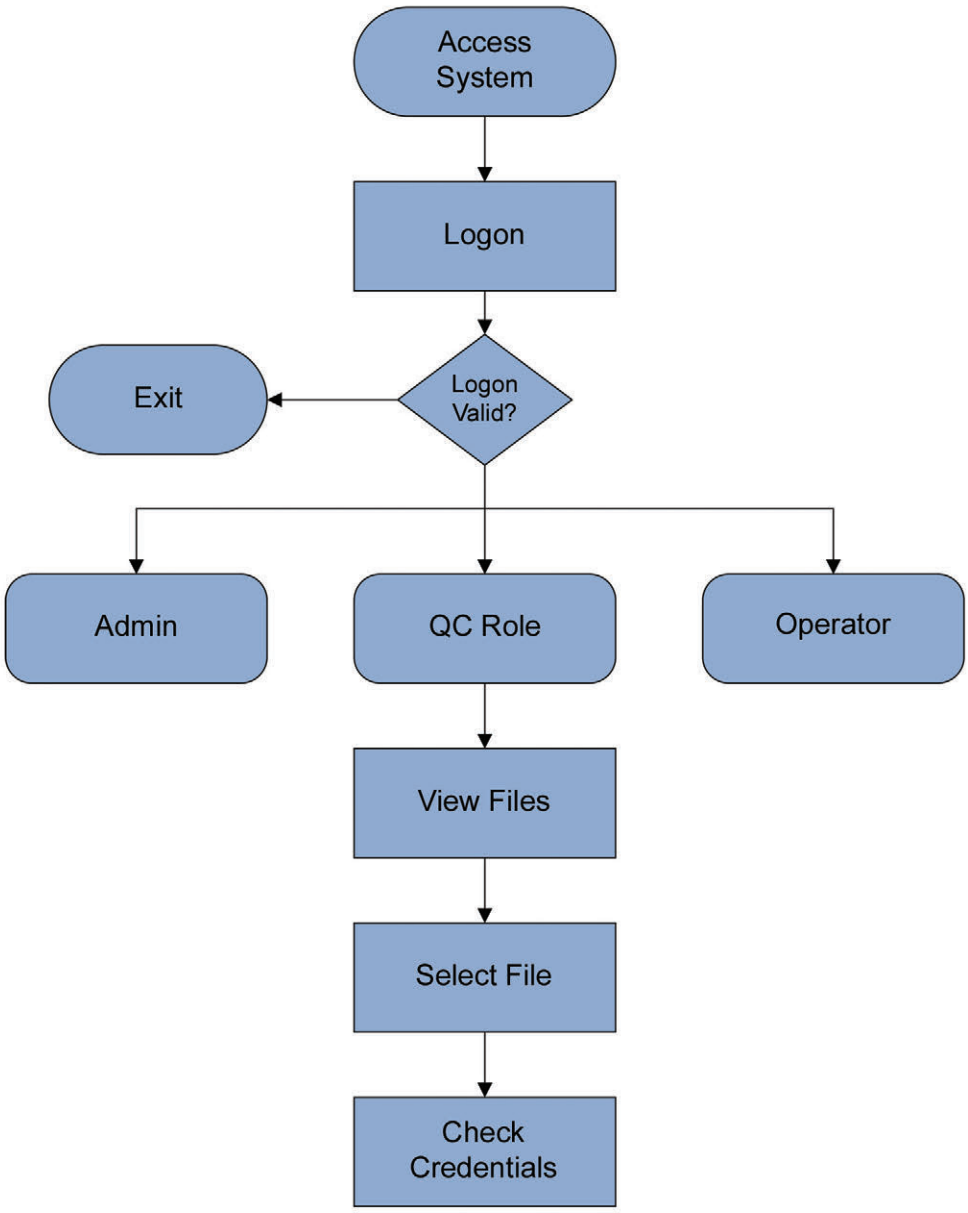


- A (Process) Flow Diagram identifies the control of flow of a process that affects each entity and the sequence of when events can occur and cannot occur. In the following example extract, the action of check credentials cannot occur before the action of view files.

Tool Example: Flow Diagram

The Flow Diagram highlights the logon and activities of the three roles (for brevity only an aspect of the QC role is displayed).

Figure 11.3: Example Flow Diagram Extract for QC Role
Used with permission from Empowerment Quality Engineering Ltd., www.empowermentge.com.



- (Data) Data Modeling identifies the entities (things about which business needs to record data, e.g., employees, PCs, ingredients), attributes (property of an entity, e.g., name, make), and the relationships between entities (associations), as well as functional dependencies (e.g., employee name is functionally dependent on user ID). This approach is very effective in documenting the data items.

Tool Example: Data Modeling

The following data model provides an abstract of the Patient and Clinical data model, identifying the entities, attributes, and relationships.

Table 11.2: Patient and Clinical Data Modeling Example

• Entity	• Patient
• Attributes	• Patient ID, Patient name, Patient address, Patient telephone, Patient email, Patient age, Patient sex, etc.
• Entity	• Clinical Trial
• Attributes	• Trial ID, Trial sponsor, Trial protocol, etc.
• Relationship	• Patient enrolls into Clinical Trial

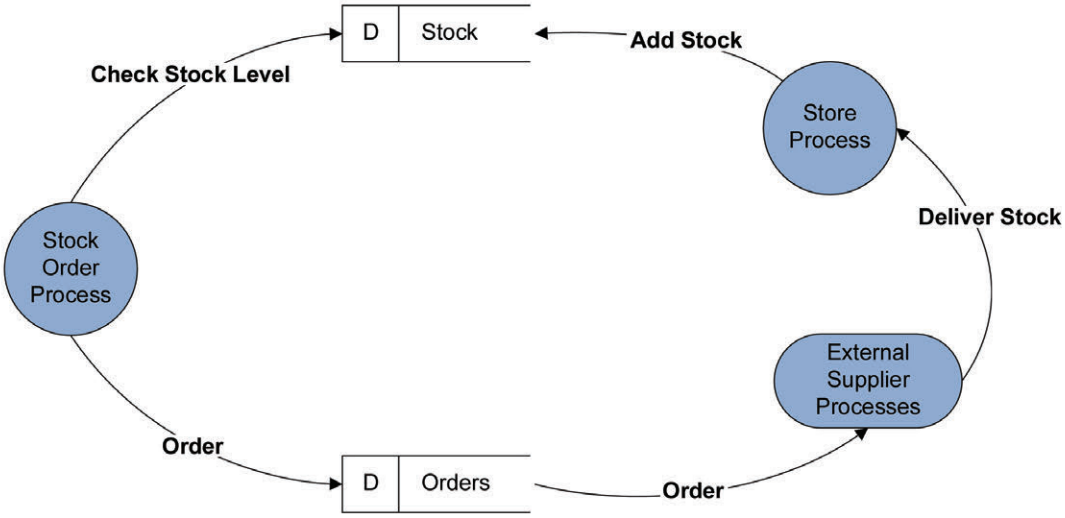
Further examples include SSADM, ALCOA+ model.

- A Data Flow Model captures how data moves around the business process and identifies processes that interact with the data, such as storage and external entities.This model helps visualize how the process can gain control of the data and how the data can be used. By walking through the various steps in a workflow, and thinking about how data can be inadvertently changed, potential data controls can be identified.

Tool Example: Data Flow Diagram (SSADM)

The following data flow shows that the stock data is shared by two processes: the Stock Order process reads from the Stock data repository and the Store Process writes data to the Stock data repository. The technical aspect of this interaction is not detailed at this stage, e.g., whether this is an internal interface or an external interface, whether the Stock data is stored in a flat file or database.

Figure 11.4: Example: Data Flow of Material Data Item between Processes
Used with permission from Empowerment Quality Engineering Ltd., www.empowermentge.com.



Some of the approaches listed above are design techniques; however, they can be very useful for identifying the process and data flow of the system that is to be computerized. They will also form input into the design stage for decomposition into technical design.

Step 2d: Document the Requirements

By iteratively using a range of information resources, the 5W's+, and schematic diagrams, a comprehensive understanding of the process flow and data flow of the business process is obtained, along with a list of uniquely identifiable requirements and assumptions.

Requirements must be documented and the data associated with each requirement identified. Documented requirements should contain a unique identifier, priority, and identification of any dependencies. Per *ISPE GAMP® 5, Appendix D1, Section 3.1.2* [2], requirements should be written as SMART (Specific, Measurable, Achievable, Realistic, Testable). *ISPE GAMP® 5, Appendix D1, Section 3.3* [2], provides guidance on how to document requirements.

Traceability

Requirements must be adequately traced to lower level life cycle artifacts. In practice, requirements are often not allocated to design elements or to their source documentation, making it difficult to identify the impact of changes to requirements, design, architecture, code, and test artifacts. This increases the risk of artifacts becoming out of step with each other, increases the burden of performing life cycle activities, and increases the risk of data integrity vulnerabilities.

It is important to ensure traceability among artifacts and to instigate this from the outset. For custom-developed systems, it is recommended to ensure that the requirements traceability approach per the validation approach is specified in the contract with the systems developer.

Tool Example: Mnemonics

The identification of life cycle artifacts can make it difficult to provide meaning and hamper understanding, for instance, in risk action lists, progress reports, defects, traceability, etc. This is particularly applicable to customized systems. The use of project mnemonics (designated within a high-level plan such as a project, quality, or validation plan) is a simple but powerful approach to ease readability and context.

It can be difficult to read through traceability matrices in order to understand whether all of the requirements have been captured without having to repeatedly cross check associated documentation. The goal is to ensure that requirements are uniquely identified and meaningful to the reader.

Given that a user requirement can generate one or more functional/system requirements, for example, functional/system requirements are a response by the supplier to the user requirements, what is needed is a decomposition of user requirements into more granular detail for transfer of the requirement (what the system will do) into the technical design (how a system will do it).

It is recommended to devise a mnemonic for requirements to facilitate traceability and relevance.

For example, the following user requirement involves the system administration; the mnemonic informs the user that it is a user requirement from the URS, it belongs to the security (defined in this example as SEC) group or requirements, and it is the first requirement (001) within that group:

- URS-SEC-001 | System Administration

The following requirements are functional requirements (Functional Requirements Specification) created by the supplier in response to the user requirements. They belong to the security functional group of requirements; they are derived from the first security requirement from the URS (SEC-001) and are the first and second functional requirements within the group (001 and 002):

- FRS-SEC-001-001 | System Administration User ID and password

- FRS-SEC-001-002 | System Administration create user groups and assign permissions

Another example is where HLD denotes High Level Design:

- HLD-SEC-UC-001 | System Administrator High Level Design Use Case

Step 2e: Perform the Risk Assessment

Requirements risk analysis is essential to the data integrity of a computerized system. Try to envisage data integrity failure scenarios as a result of something not going as planned, and then try to establish how the failure can be either prevented or contained.

Table 11.3: Failure-Mitigation Example

Failure	Mitigation
Hard disk material fatigue (e.g., spindle crash) results in data loss	Mirror (replicate) the data across several hard disks so if one hard disk fails, the data is still available.
Requirements are incomplete, missing, conflicting, or generate false assumptions resulting in data integrity failure	Perform robust requirement reviews to all requirements. Ensure the review meetings are held and assign specific roles to reviewers. Assign data role to one reviewer who will concentrate on the integrity of the data items within the review.

Applying the idea of assumptions during requirements gathering and analysis is a simple approach to reduce faults: what assumptions could be possibly derived from this requirement? How can the assumption(s) have ripple effects within the life cycle? If the assumption was built into the life cycle, as a fault, how soon could it materialize? What activity or test would reveal the fault? What checks could be applied at the requirements review or design review stage to ensure that an assumption was not made? How could the requirement be structured to eliminate the assumption and save on the effort?

In the above example, a mitigation strategy needs to be designed, built, and verified within the computerized system. The computerized system now has an additional requirement (and associated life cycle activities, e.g., installation, testing and IT considerations), that the business community may not have thought of, but is deemed necessary for the data integrity of the computerized system.

Activities to perform include (see *ISPE GAMP® 5, Chapter 5,* for more detail on risk [2]):

- Decompose the business process works and identify what additional activities the computerized system will be required to perform by using the 5W's+, iteratively elicit the high level process via a modeling approach, e.g., use case diagrams or flow charts.
- Identify the controlling functions and associated data. Document the requirements.
- Continuously refine and optimize the process and data flow.
- Identify vulnerabilities in the manual process and categorize the data with GxP impact as a result.
- Identify potential failures in the computerized system such as technical complicities and security (e.g., turning off the clock and audit trail).
- Define mitigation strategies to counter the failures, (e.g., enhanced administration access controls, locking the audit trail within the computerized system).

- Ensure that appropriate life cycle activities are assigned to counter any identified risks, (e.g., performance engineering, performance code reviews, performance testing).

- Define the verification of the mitigation strategy to be implemented in the computerized system.

- Review the requirements and risks.

- Repeat the process as often as required.

Tool Example: Failure Mode Effect Analysis (FMEA) [57]

FMEA is a proven tool when seeking to prevent data integrity issues and to ascertain the scope of life cycle effort to be applied for each requirement. The FMEA template can contain the requirement ID, the potential failure mode, the risk priority, and the mitigation strategy. In addition, a suite of verification activities can be identified to confirm the effectiveness of the mitigation strategy at various stages in the life cycle based upon the risk category.

FMEA can provide the following benefits:

- Identify potential faults associated with each requirement that may result in a data integrity fault

- Identify mitigation strategies to counter the fault (e.g., a new requirement and/or application, design constraints, or the creation of a new SOP)

- Depending on the priority, identify one or more verification activities from across the full verification and validation spectrum

- Identify the scope of activities to be applied during the life cycle

- Initiate the requirements traceability matrix, (FMEA can be taken further and act as the repository for the traceability matrix during progression through the life cycle)

- Initiate the test design and scoping process. The identification of applicable test activities/phases within the life cycle will indicate the types of tests to be designed. Furthermore, the act of thinking of tests will result in another walk through of the requirements flow from a positive and negative perspective and help to identify any additional false assumptions at the requirements stage.

Step 3: Perform Robust Requirement Inspection

Once the requirements have been elicited into a stable state, documented, and are ready for baselining, they must be inspected to identify and remove errors. The inspection is a very important process that should not be undervalued or underestimated.

Note: There is a difference between an inspection and a review.

- **Inspection:** The FDA defines the inspection as:

  "A manual testing technique in which program documents [specification, (requirements, design)…], are examined in a very formal and disciplined manner to discover errors, violations of standards and other problems. Checklists are a typical vehicle used in accomplishing this technique." [55]

- **Requirements Review:** The FDA defines the requirements review as:

  "A process or meeting during which the requirements for a system, hardware item, or software item are presented…for comment or approval." (IEEE) [55]

During a requirement inspection, the requirement is evaluated based on its structure, content, source of information, domain knowledge and compatibility, and the SMART definition [2]. The inspection meeting is used to capture the issues and recommendations.

Inspecting the requirements is recommended as it provides a greater error detection focus than a document review.

Recommendations for successful inspections:

- Nominate a chair to facilitate the inspection activity. The chair allocates the requirements document to the evaluators.
- Chair of the inspection meeting may allocate specific roles to evaluators, e.g., regulatory perspective, testability perspective, business perspective, security perspective, data integrity perspective, to ensure greater coverage.
- Evaluators inspect the requirements document prior to the inspection meeting, capturing errors, violations of regulations, and suggested recommendations. Refer to the Requirements Checklist, below.
- At the inspection meeting, the chair of the meeting walks through each requirement and asks for any responses from the evaluators. The evaluators provide their issues and recommendations. The meeting must not attempt to discuss solutions. The scribe of the meeting documents the responses and provides them to the requirements document author.
- The author will either implement the responses or liaise with the evaluator for further discussion and agreement. The updated requirements are then forwarded to the evaluator for sign off on the implementation of their responses.

Note: This approach can be used for generating and tracking metrics across the life cycle for trend analysis and QMS process improvement.

Although more intensive and time consuming, the inspection is a valuable tool in reducing the risk of technical and process data integrity issues from errors and false assumptions, for example, by:

- Checking that requirements are not processing the same data item in parallel whereby one process overwrites the result of another process. (Process A ingests data 1 for processing. Before it is finished, Process B ingests data 1 for processing and writes the result. Process A then finishes processing and overwrites Process B's result and removes it from the system).
- Checking that subjective language is not used whereby developers assume different connotations of a requirement.
- Checking that there are no gaps in the requirements or that no critical omissions exist that may pose a risk to the design and implementation of the requirements.
- Checking that there are no conflicting, incomplete, or ambiguous requirements.
- Checking that regulations are documented as requirements.
- Checking that dependencies or relationships between requirements are captured.
- Checking that requirements are uniquely identified, prioritized, and assigned an owner.

Tool Example: Requirements Issue Checklist

Create a checklist of typical requirement issues as an aide memoire for evaluators. Update the checklist after each requirement inspection to include new issue types. See Table 11.4 for an example.

Table 11.4: Requirements Issue Checklist

Issue Type	Example
Ambiguous	The requirement is open to interpretation, increasing the risk of false assumptions and data integrity vulnerability. For example: "The system will be easy to use" is open to interpretation by individuals; a developer's definition of "easy to use" will be different from a tester's, which will be different to an end user's.
Errors	There is a mistake in a requirement, e.g., against a business process, regulation, guidance, or policy. For example: The requirement states that the password length must be at least six characters whereas corporate policy is eight. Or "The maximum volume is 50 cl" when the correct value is 5 ml.
Incomplete	A requirement is omitted from the document or information is missing from the requirement (e.g., priority, owner, unique ID). For example: "The system will be fast." How fast? What is the minimum acceptable response time? What is the minimum acceptable response time during high usage?
Inconsistent	The requirement is not consistent with other requirements in the document. For example: "The disk space warning lights will be yellow and blink in the following fashion: Blink-Blink-Off; Blink-Blink-Off" vs. "The storage space will warn if there is a problem with capacity. The system administrators will see a display of blue lights flashing."
Irrelevance	The requirement is beyond the scope of the project. For example: "The system will play a tune upon completion of the workflow tasks."
Misplaced	The requirement is located under the wrong section. For example: A GUI requirement is located under the data storage requirements.
Redundant	The requirement is a duplicate, whereupon a change to one of the requirements may not be replicated and lead to anomalies. For example: Password credential details are replicated across multiple requirements.
Technical Constraints	In some instances, it is necessary to state technical constraints, for example, integrating with an existing system. However, resist the temptation to state "how" the system will be implemented rather than "what" the system should do; otherwise, constraints are placed on architects, designers, implementers, et al., which may impede the provision of a solution that improves the business process. For example: "The system must write to a relational database.", where upon a hierarchical or object-oriented database may be a better solution for the system implementation but is now discounted.
Testability	The requirement is not implementable due to technology constraints or operational constraints. For example: The data will be sent to the procurement system via a new underground dedicated coaxial wired communication channel to be provided by IT. The use of interfaces to a remote system that is not available for integration testing may result in the creation of test harnesses (software code written to drive tests into the system) and of test stubs (software code written to receive tests out of the system) to verify the interface connectivity and data transfer protocols.

12 Appendix 7 – Requirements Specification and Data Integrity Risks for Interfaces

This Appendix provides guidance on creating an interface requirements specification, and understanding the data integrity risks associated with interfaces.

12.1 Interface Requirements Specification

An appropriate system specification is a key element in the validation process for computerized systems. Similarly, the specification of an interface is the base document from which to identify possible data integrity risks and to define controls to prevent data integrity issues.

The project manager needs good coordination and communication skills to ensure that the necessary input is obtained from all involved parties, and may consider leveraging a well-structured specification template to collect the input and document the requirements from different views.

A sample structure of an interface requirements specification is shown in Table 12.1.

Table 12.1: Interface Requirements Specification

Interface Requirements	
1 General Sections	Description of: • Objectives, Purpose, Intended Use • Responsibilities, including a list of all stakeholders and involved parties • Restrictions and Design Rules to be adapted to requirements definition
2 Data Transfer from Source to Target	Source system *sends* the data to the target system which *receives* the data.
2.1 Process-related Data Requirements of Target System	Description of use cases for target system, including roles, process owner, and reference business process
2.2 Interface Requirements of Source System	• Required data for transfer • System Coupling: source system to target system • Trigger for data transfer and frequency of data transfer • Audit trail entries including time/date/time zone information for data transfer • Dependencies and conditions relating to the source system • Data extraction • Data processing • Data transformation • Error handling mechanism • Interface configuration variables • Data format for transfer

Table 12.1: Interface Requirements Specification (continued)

Interface Requirements	
2.3 Interface Requirements of Target System	• Data format for reception • Audit trail entries including time/date/time zone information for data transfer • Dependencies and conditions relating to the target system • Transformation of data received • Processing of data received • Error handling mechanism • Interface configuration variables
3 Data Transfer from Target to Source	Sending requests for data or acknowledgements or if the interfaces support bi-directional transfer, e.g., telegram-based communication process
3.1 Process-related Data Requirements of Source System	Description of use cases for source system, including actors and process owner, and reference business process
3.2 Interface Requirements of Target System	• Required data for transfer • System Coupling: source system to target system • Trigger for data transfer and frequency of data transfer • Dependencies and conditions relating to the target system • Data extraction • Data processing • Data transformation • Error handling mechanism • Interface configuration variables • Data format for transfer
3.3 Interface Requirements of Source System	• Data format for reception • Dependencies and conditions relating to the source system • Transformation of data sent • Processing of data sent • Error handling mechanism • Interface configuration variables

Additional information about tools for the effective development of requirements is contained in Appendix 6.

The purpose of a structure such as Table 12.1 is to regard all directions of data transfer separately, and to always describe both sides of the interface.

Topics that are not applicable are noted as such but not deleted. Every chapter should reference and identify the responsible party. All responsible and accountable parties have access to the complete specification to consider the corresponding requirements.

The designers and decision makers of an interface should document their statements and requirements for all elements of an interface such that the coordinated specification contains the essential elements to verify and support data integrity.

12.2 Typical Data Integrity Issues Related to Data Interfaces

There are some typical data integrity issues and challenges related to data interfaces derived from the requirements of ALCOA+.

Table 12.2 lists risks relating to interfaces but is not intended to be comprehensive. It is a starting point to assess the risks and data integrity issues and to consider controls or monitoring measurements corresponding to a specific interface.

Table 12.2: Risks Relating to Interfaces

Risk to Data Integrity	Possible Causes	Possible Mitigation Actions or Controls
Loss of data due to inadequate security features of the sending system or device; e.g., if a device does not store data long enough or is not able to resend after an error occurs	• Failure of interface components • Target system not available or not ready to receive data • Failures not recognized • Failure messages not escalated	• Implement a data buffer or intermediate data store • Implement a feedback procedure between source and target systems • Verify availability of target system before sending • Set up failure detection and handling procedures • Set up escalation procedures in case of detected failures
Loss of data due to inadequate security features of the receiving system or device; if the target system is not able to handle exceptions, e.g., to detect and handle erroneous or updated data/ records	• Erroneous data is ignored or deleted • Insufficient capacity of target system to process received data • Source is not authorized • Data from source not identified, is misinterpreted, or attached to wrong target data • Original data or records are overwritten by newer data • Record updates or updated data ignored or rejected	• Implement error detection and handling procedures • Introduce an intermediate system to facilitate the data transfer between source and target systems • Ensure that all likely errors have corresponding error handling routines in both the source and target systems • Conduct a risk analysis • Define events and implement rules for updating data
Loss of data due to inadequacy in IT infrastructure	• Insufficient network performance in terms of bandwidth, performance stability, and latency • Insufficient storage resources to buffer data between source and target	• Implement dedicated networks or connections for critical interfaces • Upgrade the IT network
Loss of data due to multiple interfaces clashing	• Multiple systems requesting data from the same source system at the same time • Timing of requests may result in some requests being lost or ignored	• Define all interfaces inbound to and outbound from each system • Implement prioritizing, synchronization, and timing delays to prevent clashing requests from different systems
Data/Records not identifiable (loss of meaning); e.g., this could occur if a target system is connected to multiple sources	• Identifiers of sources are not unique • Target system is not able to differentiate sources or data • Different sources are sending at the same time • Different devices send data without identifying attributes, e.g., sending values without units	• Evaluate the methods of data, record, and device identification • Implement unique identifiers, especially when adding new devices • Use temporary exclusive connections to specific sources (check-in/check-out) • Check the rules to detect source ID

Table 12.2: Risks Relating to Interfaces (continued)

Risk to Data Integrity	Possible Causes	Possible Mitigation Actions or Controls
Mismatching of data	• Identifiers used within the network are not unique • Date/time of source and target are not synchronized • Shift or backshift in data sequence after failure or interruption of production or quality processes • Identifying attributes are unknown on target side • Device changes are not detected by target system • Network type and design allows a change in record sequence after sending	• Monitoring of synchronous data exchange • Exception handling • Perform risk assessment and verify stop and restart procedures for interfaced components in production and quality processes • Define all known change events; specify and verify the routines to handle these changes
Data/Records are not complete	• Unintended filtering of records on source side caused by, e.g., missing release, missing data, manually selected records • No control of completeness on target side • Unintended aggregation of data from the same source • Target system is not designed to accept all data attributes such as audit trail data	• Comprehensive functional specification based on defined functional requirements • Start with a data integrity analysis to complete the data sets/record definition on the source side • Perform a risk assessment to find sufficient controls • Perform end-to-end verification from sending process to receiving process
Data/Records are not current	• Data received at target system is not the latest/current data	• Review timing of requests • Compare time stamp on data in originating system to time stamp in target system and flag if the data is older than “x” (dependent on data types and risk)
Unintended change to original data (change of meaning)	• Standard interface forces transformation of data, e.g., automatically transforms real number to integer, truncates • Interface uses unqualified conversion tools • Intermediate or target system replaces a time stamp with its own time stamp	• For standard interfaces, derive functional specification from supplier requirements or descriptions • Perform a risk-based verification • Identify all tools used for an interface, and assess their impact • Evaluate if time stamps are assigned correctly

It is important that system interfaces, whether manual or electronic, are controlled to ensure that the data transferred is complete and integrity has been maintained.

13 Appendix 8 – Example of a Four-Tier Classification System of a Life Science Company

This Appendix provides a four-tier classification level example.

Table 13.1: Example of a Life Science Company Four-Tier Data Classification Level

Classification Level	Description of Classification Level	Impact	Risk	Classification Mapping					Access Control	
				Discovery and R&D	Clinical Operation	Commercialization	Post-market	Support Functions	Target Audience	Granting Access or Sharing
Restricted	Extremely sensitive information that if compromised, is likely to have a direct negative impact to the organization's competitive advantage in the market, reputation, or compliance with applicable regulations. Overall impact to the company if the data were lost would exceed X Million in direct or indirect costs.	• Significant operation disruption • Loss of investor confidence • Loss of consumer confidence • Loss of key personnel • Monetary loss due to regulatory sanction or breach of contract • Damage to reputation, brand, and public image	• Availability • Integrity • Confidentiality	• PII/Personal data/ Personal information • Medicare HCIN, Social Security Numbers (SSN), etc. • Protected Health Information (PHI)	• PII/Personal data/Personal information • PHI	• PII/Personal data/Personal information • PHI	• PII/Personal data/ Personal information • Medicare HCIN, SSN, etc. • PHI	• System credentials (User ID and password) • Critical infrastructure (e.g., research facility blueprint, data center location)	Only approved individuals with a documented business need to know. For example, approved individuals on R&D teams.	Access shall be limited to authorized users with a legitimate business interest and a need to know. All access shall be approved by an appropriate data owner and tracked in a manner sufficient to be auditable. Before granting access to external third parties, contractual agreements that outline responsibilities for security of the data shall be approved by the Legal, Privacy, or Compliance Office.

Table 13.1: Example of a Life Science Company Four-Tier Data Classification Level (continued)

Classification Level	Description of Classification Level	Impact	Risk	Classification Mapping					Access Control	
				Discovery and R&D	Clinical Operation	Commercialization	Post-market	Support Functions	Target Audience	Granting Access or Sharing
Confidential	Sensitive information, that in conjunction with other data, could have a negative impact to the organization's competitive advantage in the market if compromised. Sensitive information protected by regulatory statutes.	• Significant operation disruption • Loss of investor confidence • Loss of consumer confidence • Loss of key personnel • Monetary loss due to regulatory sanction or breach of contract • Damage to reputation, brand, and public image	• Availability • Integrity • Confidentiality	• R&D programs • R&D knowledge • R&D work product • Laboratory data and test results • Intellectual properties	• Clinical trial data • Biostatistics • Treatment plan • Diagnostics • Monitoring data • Performance and patient outcomes	• Pricing details • Product information • Marketing intelligence	• Monitoring data • Drug evaluation • Compliance data • Documents under attorney-client privilege	• Merger and acquisition or divesture decision • Attorney work product • Employment information (e.g., compensation) • Merger and acquisition (prior to public release) • Treasury and financial transaction data	Only approved internal groups/ departments with a documented business need to know.	Access shall be limited to authorized users with a legitimate business interest and a need to know. All access shall be approved by an appropriate data owner and tracked in a manner sufficient to be auditable. Before granting access to external third parties, contractual agreements that outline responsibilities for security of the data shall be approved by the Legal, Privacy, or Compliance Office.
Business Use/ Internal	Information which is considered sensitive but is not likely to have a significant impact to the organization if compromised.	• Damage to competitive advantage • Loss of personnel confidence	• Availability • Integrity	• Internal collaboration	• Clinical program list • Patient satisfaction • Quality reporting	• Manufacturing data • Distribution data • Supply chain data	• Update or summary reports without PII or PHI	• Employee directory • Internal email and messages • Internal meeting content and minutes • Policies, procedures, and training materials	Only approved internal groups/ departments with a documented business need to know.	Reasonable methods shall be used to ensure internal data is accessed by or shared with authorized individuals or individuals with a legitimate need to know.
Public	Information which is intended or approved for public release.	• None	• Availability	• R&D strategy and pipeline • Public alliance and partnership	• Company press release of clinical trails	• FDA submission and approval • Public marketing material • Legal disclosure	• Legal disclosure	• Annual reports • Shareholders communication • Contact information • Company press release	All internal employee/ approved third parties/ public.	Public sharing must be approved by management or legal department

14 Appendix 9 – Security Controls

This Appendix outlines considerations for technical security controls around a computerized system.

14.1 Security Controls

When planning a computerized system, implementing risk-based network-level and account-level technical security controls is recommended to limit access to the database. These controls include:

- **Tiered system design:** Web server, application server, and database server need to run on separate physical or virtual machines, with each of the tiers on a separate network segment separated by firewalls. When the system needs to be accessible from the internet, the web server will run in a Demilitarized Zone (DMZ) and the rest of the servers will run on an internal network. In this way, if one of the servers is compromised, the rest can still run securely protected by firewall rules.

- **Database connection restrictions for service account(s):** Applications use service accounts in order to connect to the database. If not protected, such accounts can be used by individuals to initiate unauthorized connections. To prevent this, service accounts must have a very high degree of protection. They would typically have much longer passwords and would only be allowed to initiate connection to the database from the application server(s).

- **Separate administrative accounts:** It is best practice to create separate user accounts for administrative purposes. In this way a DBA would use one account to log on to the network and access email, shared folders, corporate applications, etc., and the second account, with a different password, to connect to the databases. If the “normal” account is compromised, the impact will be limited because without the “admin” (second) account, the attacker does not have access to the databases.

- **Database connection restriction for DBA accounts:** It is recommended to limit database connections for DBA accounts to a particular network or several networks.

  For restricted data this control can be further enhanced by implementing a “jump host” or a “jump network.” This requires DBAs first to connect to this “jump host” or “jump network” using, for instance, Secure Shell (SSH) or a Virtual Private Network (VPN), and only then will they be able to connect to the database. In this way even if the DBA password is compromised, the attacker is unable to connect to the database without access to the “jump host” or “jump network.”

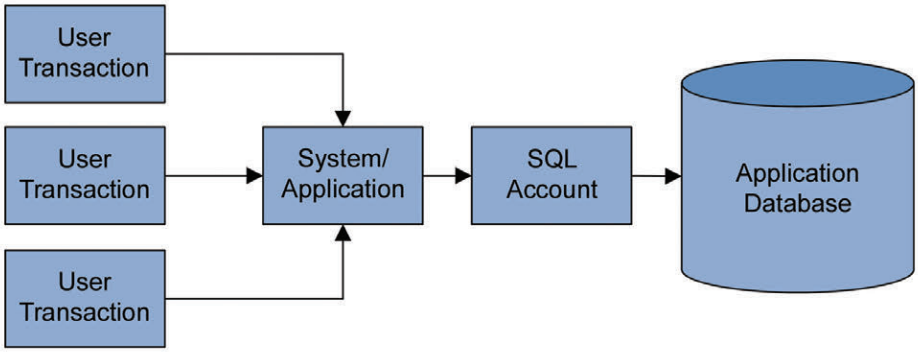
- **Database connection encryption:** This is recommended as an additional control for restricted and sensitive data. Encryption of data in transit between the application server and the database server is a good practice that prevents attackers from getting access to the data by listening to the network traffic.

- **Database encryption at rest:** This is recommended as an additional control for restricted and sensitive data. This control can be effective in cases when an attacker can gain access to either physical storage or to the operating system of the server on which database is running.

System access controls should be strictly managed, documented, and authorized. This includes unique usernames and passwords that expire in accordance with the organizational policy that adheres to 21 CFR Part 11 [41] and other regulatory requirements.

It should be noted that where applications are hosted, it is possible that individual user logons are imposed, but the application uses a generic account to access and store the data within the database/server, as shown in Figure 14.1. This is acceptable as long as there is traceability to the user and each transaction is attributable.

Figure 14.1 Database and Application access



Remote server access should be controlled in the same manner as the system/application account management; however, additional controls should be established to record who requires access as well as the rationale and duration of the required access.

An example is during system development where a third-party supplier requires access to the production server to promote modifications during Go-Live activities. With these controls in place it ensures that an audit log of user activities is recorded to facilitate an investigation should the quality of data be questioned.

Privileged database account access should be kept to a minimum as described initially by Jerome Saltzer in “The Principle of Least Privilege” (Saltzer and Schroeder [58]), and reiterated more recently in “The Common Sense Guide to Mitigating Insider Threats” (Collins, et al., 2016) [59].

Where there is only a single system administrator or database administrator account supported within the system (as is often the case with commercial databases), but the account needs to be used by multiple individuals, a password storage database (password vault) should be utilized. This practice allows individuals to use their own password (given that they have sufficient permissions) to log into the system. In this way, the administrator password is not revealed and the access is recorded within the audit log and attributable.

Within smaller organizations, procedural risk-based controls need to be in place so that requests for access are recorded and approved before access is granted. In addition, the password used for entry is changed afterwards, typically by QA or other independent group. The password is reset after each use and stored in a secure location, whether this is physically sealed in an envelope, or electronically in a password management tool.

14.2 Review of Controls

Review of the controls placed upon privileged database access is almost as important as the controls themselves. Where privileged access is required, it may be temporary or permanent depending upon the nature and complexity required for the access.

It is essential to avoid a situation where the individuals with privileged access to certain parts of the system(s) do not have their access removed when no longer required, leaving the system vulnerable to unauthorized modification. Regular review of the access privileges of users within the system/database via monitoring for dormant accounts and accounts where enhanced access was provided for a limited time but not removed provides an opportunity to prevent this situation.

15 Appendix 10 – Case Study: DBA and Security Controls for an RTSM System in a GCP Environment

This Appendix presents a case study on access and security controls within a clinical environment.

15.1 Background

A Randomization and Trial Supply Management System (RTSM) primarily deals with patient management, site management, and inventory management for investigational products. The RTSM also combines Interactive Voice/Web Response Technology with Electronic Patient-Reported Outcomes functionality.

The RTSM is integrated with other systems, such as a Clinical Trial Management System, which is a logistics system that automatically manages the delivery of investigational products when required.

Because the RTSM system includes ERES and is used for clinical trials worldwide, it falls under such regulations as GCP, 21 CFR part 11 [41], EU Annex 11 [17], and Japanese ER/ES [60].

The majority of the data in the RTSM system, including sensitive patient information, resides in the database. The data is classified as Restricted.

Issues with the integrity of the data in the RTSM system could result in the wrong drug/dosage assignment to the patients, drug shortages, incorrect assessments of the efficacy of investigational products, etc. These in turn would have a negative impact on patient safety and could cause (serious) adverse events and the failure of a clinical trial.

Although the RTSM is accessible from the internet, which adds additional IT security concerns, data theft is not a concern in this instance.

The regulated company decided to implement extended controls over the database privileged access for the RTSM database in order to ensure data integrity.

The following controls are implemented for the RTSM database, utilizing a risk-based approach.

15.2 Infrastructure Controls

The system is designed as a three-tier application with a load balancer in the DMZ, as well as web, application, and database servers for each tier on a separate network. The tiers are separated by firewalls, significantly reducing the risk of the database being compromised even if other tiers are penetrated.

Direct database connections using the application DBA account can only be established from the application server.

The administrative interface of the database server (SSH) and database connections using DBA accounts are only reachable from a special administrative network. DBAs have to log on to a VPN using two factor authentication before they can connect to the RTSM database. Access to this VPN is strictly controlled and is limited to the members of the IT infrastructure team.

Firewall, server, and database security are monitored in real time. Security events are analyzed by the SIEM (Security Information and Event Monitoring) solution and the IT security team is immediately alerted of any potential offenses.

15.3 Account Controls

DBAs use separate accounts to logon to the databases. These accounts are different from their network accounts used to logon to their laptops or workstations. They are only allowed to use their individual DBA accounts to logon to databases, thus establishing traceability of the actions to the individuals who performed them.

Use of shared accounts is strictly prohibited except for the cases when shared accounts are used for technical reasons. Passwords for shared accounts are stored in an automated password database that changes the password right after it was used.

Only individuals directly involved in maintenance of the RTSM database are granted DBA rights. There is a process in place for granting emergency temporary access to the database, which is time restricted and has to be revoked immediately after use.

There is an on and off boarding process in place that requires immediate access revocation when a DBA leaves the company or changes roles.

15.4 Segregation of Duties

There are segregation of duties rules in place to ensure that the members of the database management team cannot be involved in business functions related to the RTSM system, and that users in the functional area are not allowed to perform IT security, server administration, or network administration functions.

15.5 Periodic Reviews

There is a periodic database privileged account review process in place. Periodically an automated account report is produced on the RTSM database. This report shows all accounts defined on the database server, along with access rights defined for these accounts on different levels (sever/database/table, etc.). This report is reviewed and signed by the management of the database management team to ensure appropriate access.

In addition to the account review, a monthly database activity review process is performed. The database server sends its logs (including any SQL queries and commands run by the DBAs) to the SIEM solution. Neither the DBAs nor the operating system administrators have access to these logs.

The system automatically generates a monthly report of all the activities on the database (excluding activities generated by the application, which are captured by the application-level audit trail) and sends it to the database management team for review.

As there is no concern of data theft, SELECT statements are not reviewed, only statements and commands that change the data or the structure (INSERT, DELETE, UPDATE, ALTER, CREATE, etc.) are reviewed. The reviewer needs to verify that any action on the database was properly approved prior to execution, that is, there is an approved ticket in place for each case of data modification.

15.6 Internal Audit

The RTSM system is periodically audited by the internal audit function as a part of the annual audit program following a risk-based approach. Such audits include, among others, review of database management and security.

16 Appendix 11 – Case Study: DBA and Security Controls for an ERP System in a Medical Device Manufacturing Environment

This Appendix presents examples of access and security controls within a medical device manufacturing environment.

A manufacturing medical device company uses an ERP system/database built upon an SQL server. The ERP controls all manufacturing aspects: sales, manufacturing, purchasing, and finance with the following controls in place:

- Levels of data classification within the ERP system are defined and documented.
- System/server access is only provided upon documented successful completion of required training.
- Internal user access audits are undertaken annually to ensure that only trained individuals have access and that permission is granted only to the tasks required to be performed. Additionally, privileged access permissions are audited on a quarterly basis.
- Internal access to the servers hosting the application is only granted to trained individuals, each with a unique account and password that expires every 30 days.
- External access to the remote servers is provided to the support team from the supplier when a support ticket is raised. Upon resolution of the ticket, the password is reset and stored within a centralized password management application.
- Individual user logons are utilized for routine tasks, and accounts with increased access are only used to perform tasks that require the higher level of permissions. Additionally, the System Administrator account has been deliberately disabled due to the risk of automated password cracking tools, as such, guessing an unknown account name and password is exponentially more difficult than just cracking a password for the username System Administrator.
- Additionally, there is SOD in relation to permissions to ensure that no single individual can manipulate the system, for instance raising a purchase requisition and approving it.
- Passwords are required to be changed every 30 days with none of the 10 previous passwords accepted. Accounts are disabled promptly where individuals have left the business and/or the role.
- User accounts are not recycled for temporary employees.
- Data modifications within the system are categorized according to the previously agreed and documented risk assessment relation to the data/function, such as modification of vendor bank account details or correction of a system bug producing a “double pick” anomaly where the internet connection is momentarily interrupted.
- Data modifications within the system are controlled via an IT ticketing service to document the required change, which must be approved by an independent and documented department/area individual. Upon approval, the mechanism for making an amendment is documented and objective evidence is attached to each ticket.
- The system audit log of key fields is configured and reviewed on a regular basis with emphasis upon ensuring access permissions are not raised above the required level for the task(s), but also that legitimate changes have been documented and processed correctly. Additionally, the SQL database record logs are reviewed to ensure that only authorized and documented changes have been executed within the database.

- System development activities are documented as requirements, design specification(s), risk assessment(s), and operational/performance protocol(s) are executed before modifications are promoted from the testing environment(s) into the live/production database.

17 Appendix 12 – Case Study: Laboratory Computerized System

In a typical regulated laboratory with a diverse inventory of equipment and instrumentation, there may be many systems comprising older equipment lacking modern data integrity controls, each with an associated risk of data integrity failure. This Appendix provides a case study on identifying and remediating data integrity issues in a QC laboratory.

17.1 Typical Use Scenario

The instruments used in a QC laboratory must be qualified for operational verification against the supplier specification within the laboratory environment (after receipt and installation, and after major repairs). Then performance qualification is conducted, taking into account the actual conditions of the validated method for which the equipment will be used; the performance qualification should be periodically repeated [61] for ongoing confirmation of the instrument's performance for its intended use.

Equipment must be calibrated against standards on a regular schedule to ensure accurate and reproducible results. System suitability testing (chromatographic equipment) and point of use checks (e.g., balance check with a calibrated mass) are required daily to confirm that equipment is fit for purpose and to assure the accuracy of the results.

All equipment must have a logbook to record its use, maintenance, repair, and calibration activities. A calibration schedule is maintained and out of calibration instruments removed from use until repaired.

A QC laboratory in a manufacturing facility receives a set of samples for analysis. The samples are logged, and the relevant sample identity data captured.

To proceed further it is often necessary to divide the sample into several subsamples for a range of analytical testing. A single aliquot of sample may have one or more replicate determinations made that are averaged into a single reportable result, which is compared to the specification.

Each test is performed according to a validated method and compared to a pharmacopoeia or company specification. The same material may have to pass different sets of tests or specifications, depending on its intended market.

The methods followed are standard methods typically developed and validated in a separate method development laboratory, often associated with R&D rather than production. There must be a transfer protocol with defined acceptance criteria to establish the developed method in the QC laboratory.

Results that meet specification may be Out of Expectation (OOE) and/or Out of Trend (OOT). Understanding the cause of such results and implementing corrective actions can reduce the number of OOS results.

All OOS results must be investigated according to a formal process. OOS results where a sound scientific justification can be proved as grounds to exclude the result are classified as Invalidated OOS results. The invalidated OOS rate forms one of the reportable metrics required under the US FDA Quality Metrics Initiative [62].

Where the OOS or OOT result cannot be invalidated scientifically, a thorough investigation into probable root causes is conducted. All results – passing, invalidated OOS, OOS, OOT, OOE, including original data – must be retained for the retention period.

Laboratory analysts are professionally qualified and receive regular training in GxP-relevant topics (GMP, GLP, GCP) and GDocP, along with training in the use of equipment and the execution of experiments according to SOPs.

Training must be completed and evidence of competence proven and documented before the analyst can use a system. Refresher training is also carried out on a regular basis (often annually).

All analyst activities are recorded in the laboratory notebook or controlled worksheet (form), and where appropriate, a second person verification check is obtained. Calculations are frequently performed using locally developed and validated spreadsheets or entered into statistical packages.

To manage this complexity, many companies have invested in a LIMS.

Where a LIMS has not been implemented, the multiple results supporting, for example, a typical C of A, gives rise to a substantial amount of documentation containing a mixture of:

- Manual records
- Local printouts (from balances, autoclaves, etc.)
- Chart recorder output (paper charts or printouts)
- Printouts of PDF files, flat files
- Results in the form of proprietary format records that can only be read using the specific instrument and native software application

17.2 Records Risk Assessment and Controls Considerations

The process owner for a regulated laboratory (typically the Senior Analyst or Laboratory Manager) must ensure that:

- All laboratory processes have been identified and mapped into a coherent workflow
- Associated systems have undergone a risk assessment with respect to potential data integrity failures
- All regulatory records in the laboratory have been identified and formally designated as paper or electronic, or a hybrid situation (electronic records with handwritten signatures on a printed summary report)

A further complication in regulated laboratories is that acquired data from a system is subsequently processed to obtain a reportable result, and this result is used as initial data for the next step in the analytical or reporting process. Any unidentified or unreported data integrity failure in the lower levels will compromise the data integrity of all subsequent levels.

17.3 CDS Example

Table 17.1 contains an example of data integrity issues using paper laboratory notebooks and a local chromatography workstation (a standalone PC containing a CDS application) attached to a chromatograph for sample analysis. This is presented as an often-found scenario, mainly created by poor processes and controls; to stay realistic this example deliberately does not include every known data integrity failure associated with CDS but rather presents common bad practices.

Table 17.1: Typical Issues with CDS

Life Cycle Stage	Considerations and Issues
Data Creation	Sample identity and preparation data (weight, pH) are manually recorded into a paper laboratory notebook. This data is then transcribed manually into the CDS. The analyst logs into the CDS using the shared analyst account and manually selects the methods to be used with the sample set. CDS sends the instrument parameters (flow rate, pressure, etc.) to the chromatograph and receives the detector measurements. Final results (after processing) are manually transcribed into a spreadsheet to generate the C of A.
Data Processing	The data is extensively processed, initially using a processing method and then with manual integration and baseline adjustment. The processing method used in this step is slightly different on this workstation compared to the method stored in the next workstation on the bench.
Data Review, Reporting, and Use	The summary report is printed out showing the sample information, the chromatogram, and the peaks summary table. The reviewer does not go to the workstation in the lab to see if reprocessing was done to get the results, nor does the approver look for any additional injections (trial injections, duplicate injections) or unreported results. The result is approved based only on the summary report, containing a handwritten signature. There is no audit trail review, so the reviewer does not realize the analyst had turned off the audit trail for this data folder. The earlier results from the original processing method and first round of manual integration are OOS, but the analyst continued reprocessing to manipulate the results into specification. There was no reporting of the earlier OOS results.
Data Retention and Retrieval	The signed printout is carefully filed as the master record. The electronic data is left on the workstation, where backups are done annually because of the lengthy time it takes. Retrieving a particular result requires checking each of the five workstations to find which one has that data.
Data Destruction	Data is disposed of incidentally when a workstation is replaced due to obsolescence or component failure, or when data is deleted to create space on the workstation hard drive.

17.4 Remediation Plan

The prerequisites for good data integrity in any laboratory are a sound QMS supported by good document practices and good data integrity practices, supplemented by knowledgeable personnel operating in an environment free from pressure and opportunity to manipulate or misrepresent the results.

Each regulated laboratory should determine a strategic approach to transition their paper and hybrid situations to electronic records with electronic signatures within computerized systems along with full procedural and technical controls and supporting behaviors. Wherever possible, manual recording, processing, and transcription of data should be avoided; calculations should be moved away from manual spreadsheets and executed in validated systems such as LIMS and CDS. Replacing stand-alone laboratory instrument PCs with networked systems eliminates variation in methods and system configuration, and enforces central storage of data for ease of backup and retrieval.

18 Appendix 13 – Case Study: Uncontrolled Spreadsheet

This Appendix discusses a spreadsheet used in the QC laboratory; however, similar uncontrolled spreadsheets can be found in manufacturing, warehousing, and many other areas within an organization.

18.1 Scenario

A spreadsheet developed in Microsoft Excel is used to calculate trends in analytical results based on data from multiple sources:

- Data manually typed in, e.g., batch number, operator name, date
- Data copied and pasted from a standalone laboratory instrument
- Data from the CDS saved as a .csv file and opened in the spreadsheet application

The spreadsheet was created some years ago by a senior analyst who has since left the organization. Most of the analysts have copies of the unsecured spreadsheet on their PCs.

18.2 Records Risk Assessment and Controls Considerations

Often, the only way to uncover these forgotten spreadsheets is to walk through each stage of a business process, and at each stage determine:

- Where the data comes from
- How does it transfer to the next stage
- What happens in each stage

This is further explained in Section 3.2 Data and System Life Cycle Interrelationships.

18.3 Spreadsheet Example

Table 18.1 evaluates the typical use scenario for the uncontrolled spreadsheet.

Table 18.1: Typical issues with Uncontrolled Spreadsheets

Life Cycle Stage	Considerations and Issues
Data Creation	There is no logon to the spreadsheet application, so data can only be attributed based on the manually-typed operator name. Errors can be made when typing the batch number and other identifying data. The copied and pasted data may not include all of the data, and there is no way to check what data may have been excluded. The data, exported as a .csv file, has lost its traceability and linkage back to the original record.

Table 18.1: Typical issues with Uncontrolled Spreadsheets (continued)

Life Cycle Stage	Considerations and Issues
Data Processing	The calculations in the spreadsheet are not documented, nor have they been validated or protected from changes. Over time, the various copies of the spreadsheet have been "improved" by individual analysts so that each analyst uses a different version of the original spreadsheet. All data and calculations can be edited freely without a change audit trail allowing the analysts to eliminate outliers in the trended data. The data can be altered multiple times before a desirable result is achieved and the spreadsheet saved.
Data Review, Reporting, and Use	The trended data is emailed to the QC manager who reviews the charts in the spreadsheet and accepts them as evidence that the process continues to operate well within specification.
Data Retention and Retrieval	The spreadsheet file remains on the analyst's PC hard drive as a flat file, saved under a filename comprising his name and the date. To retrieve, all of the analysts would have to search their hard drives to find who trended that particular batch.
Data Destruction	When the analyst leaves the company, their PC is reformatted and reassigned to a new employee, so all the saved spreadsheets are deleted.

18.4 Remediation Plan

Possible mitigation actions are:

- Eliminate the spreadsheet by executing all calculations, statistical analysis, and trending within validated systems such as CDS, LIMS, Manufacturing Execution System (MES), or ERP
- Invest in a third party add-on that will provide the logon and audit trail technical controls absent in the spreadsheet application

(**Note:** that the spreadsheet will still not enforce saving data and all versions of results – "save" is still a manual operation.)

Spreadsheets are easy to develop, flexible in their application, and so prevalent that it may be impossible to eliminate them within an organization. A third party "add-on" alone is not enough to eliminate the data integrity risks, so additional technical and procedural controls are required, such as:

- Documenting the formulas and verifying the calculations and logic used within the spreadsheet, then creating a spreadsheet template using the validated functions
- Protecting the spreadsheet template so that only essential data entry cells can be edited
- Minimizing or eliminating manual data entry, and replacing it with validated data imports from the source applications
- Controlling distribution of the spreadsheet template, so that only the current version is available for use
- Implementing automated naming of spreadsheets generated from the template
- Ensuring that all spreadsheets generated are automatically stored in a secure location on a central server, which is automatically backed up nightly

19 Appendix 14 – Case Study: Process Control System

The case study in this Appendix is an example use situation for a production system representing typical bad practices that are commonly found.

19.1 Scenario

The blank batch manufacturing record is an issued copy of a master document containing spaces for recording the required CPP data, including operator identity and time of actions, which the operator writes in with a pen at the time of the operation.

Many machines have CPPs set at the beginning of the controlled process that do not change. The values of the CPPs and the date/time are data to be recorded.

Adjustment may be needed during the process. If the change is to a CPP, then this must be recorded as a change with a record of who, what, when, and why. Older systems often lack an audit trail to record a change and must rely on the operator manually recording the details of the change contemporaneously.

19.2 Records Risk Assessment and Controls Considerations

Machine data can be categorized as:

- Master data including machine settings, product settings, recipe instructions, user accounts, and user-type privileges
- Recorded data including alarms covering:
  - Date and time
  - Operator data including operator identity
  - Process data including quantities, weight, volume
  - Set points for control of operating parameters
  - Values of operating parameters
  - Processing times
  - Environmental conditions, relative humidity, pressure changes, room temperature

Table 19.1: Types of Data and Risks Associated with PCS

Type of Data	Risks	System Controls Shortfall
Master Data	Unauthorized access or changes to master data	No or limited access control to restrict functionality and system access Reliance on paper log entries in place of integrated audit trail of changes

Table 19.1: Types of Data and Risks Associated with PCS (continued)

Type of Data	Risks	System Controls Shortfall
Data Transfer	Manual recording and entry errors during transcription	No electronic interface to systems upstream or downstream of the process, or to any Supervisory Control and Data Acquisition (SCADA)/MES/data historian
Electronic Saved Data	Incomplete or insufficient storage capability	Limited data storage capacity First in, first out overwriting of stored data
Printed Paper Records	Completeness, accuracy of data contained in printed records	Systems may contain or use thermal printers so printouts are not durable There may be limited printed report options available
Data retention for all records	Not retained in a retrievable manner for the duration of the retention period	No worse or higher risks than any other paper-based records required to be retained

19.3 PCS Example

Table 19.2 contains an example of data integrity issues using a PCS with local operator panel. It is presented as a worst case, but still prevalent, scenario mainly created by poor processes and controls.

Table 19.2: Typical issues with a Noncompliant PCS

Life Cycle Stage	Considerations and Issues
Data Creation	CPPs are set on the machine by the operator and entered into the batch record. The CPPs are entered. The identity of the operator including date and time, and information about changes or incidents are reliant on the operator contemporaneously recording the data on the paper record. Some machines have an interface to an historian or SCADA system, which collects the data and produces a simple report that is printed out and put in the batch record. For most older systems, data is manually recorded and transcribed into either a paper record or a separate electronic system. Operators can overwrite sensor values in the PCS, or "optimize" the printouts to minimize CPP variations.
Data Processing	There is limited or no processing of data needed, but trending can only be done by manually typing batch values into a spreadsheet.
Data Review, Reporting, and Use	The completed batch record is reviewed by production management and by QA and is likely limited to reviewing the compiled loose paper records. Alarms and events are enunciated at the local control panel and should have been recorded by the operator into the paper checklist. There are no checks made by the reviewer on the alarm history.
Data Retention and Retrieval	Batch records are stored on paper for a defined period, typically shelf life plus one year.
Data Destruction	After the defined period the batch records can be destroyed. Note: batches used for clinical trials are retained for a longer time, typically the life time of the product.

19.4 Remediation Plan

Where possible, systems should be replaced with more integrated solutions where many of the login and audit trail functions can be provided by an MES.

In some instances, there may not be a direct replacement available that would provide all of the required technical controls.

While hybrid situations are not preferred (WHO TRS No. 996 Annex 5 [3]), they may be able to achieve an acceptable level of compliance when properly designed, implemented, managed, and monitored.

20 Appendix 15 – Case Study: Business Application System

This Appendix is a case study example use situation for an IT system representing typical bad practices commonly found.

20.1 Scenario

The required data is recorded electronically, audit trail functionality is enabled, but the electronic records are not reviewed; so, while the controls are available, the process is noncompliant.

In many cases, the IT system may not have electronic signature functionality and/or does not have the capability to search and filter audit trail entries for ease of reviewing (original data changes and/or deletions, as well as metadata changes and/or deletions).

Some IT systems lack sufficient technical controls, such as an IT system that uses one or more generic (shared) accounts with inadequate or no segregation of duty between operator and administrator.

Some IT systems may have been implemented for a non-GxP function but have either been adapted for GxP usage over time or may hold data supporting GxP records (for example, a call center interaction system). It may be possible to adapt the system configuration to support the required controls, but it is likely to take significant investment to upgrade or replace the system, and if the majority of the system is non-GxP, it may be difficult to justify such a financial outlay.

20.2 Records Risk Assessment and Controls Considerations

IT systems are often used across multiple groups, departments, and even sites, so a multidisciplinary team containing user representatives as well as IT administrators needs to be assembled to assess the system and controls.

Special consideration must be given to defining and understanding the business process or processes supported by the IT system, as the severity of the data integrity risk is affected by the GxP impact of the records under assessment.

20.3 IT Systems Example

Table 20.1 identifies data integrity issues arising from the use of an IT system resulting from poor practices and controls. It is a generic example only, and the specific use of the system affects the issues requiring remediation.

Table 20.1: Typical Issues with IT Systems

Life Cycle Stage	Considerations and Issues
Data Creation	Lack of access control Lack of personalized user accounts for operator, administrator, quality, etc. Date formats inappropriate (e.g., mm/dd/yy) and time zone offsets not recorded Interfaces between IT systems not validated commensurate with GxP risk
Data Processing	User-defined equations and queries are created ad hoc without documentation or validation
Data Review, Reporting, and Use	Lack of data review functionality and audit trail filtering Lack of electronic signature capability Lack of flagging or reporting of invalid and/or atypical data Inadequate report generation capability, therefore data is exported into spreadsheets for report creation
Data Retention and Retrieval	No record retention period defined and no formal archiving process Lack of technical capability to automate the deletion of data from the active database to guarantee data/record integrity
Data Destruction	Data disposal process not procedurally described, not documented, etc.

20.4 Remediation Plan

The remediation plan is determined by the use and GxP impact of the IT system:

- For an IT system solely supporting a GxP process, the focus should be on upgrading or replacing the system, with priority determined by the extent of the data integrity risk and the GxP criticality of the process.
- For a low GxP-impacting IT system containing GxP supporting data, and depending on the severity of the data integrity risk, improving the procedural controls and implementing a hybrid situation for signatures as detailed in Section 4.3.5.3 may be the best option.
- For an IT system containing both GxP and non-GxP data, several options may be considered:
  - Implement a system allowing different configurations for the GxP vs. non-GxP areas of the application
  - Treat the whole system as GxP critical and implement a remediation plan accordingly
  - Separate out the GxP and non-GxP processes and implement separate systems for each, with the appropriate controls for each

21 Appendix 16 – Reviewing Laboratory Systems

This Appendix gives detailed guidance on auditing laboratory systems. Audits or reviews should be conducted according to the policies and procedures of the regulated company. Example audit and review criteria are given below, but these are not prescriptive or exhaustive.

21.1 General Requirements

The laboratory system should be secure with restricted access to original data files, time clock, and audit trail functions. There should be unique user IDs in use for everyone. Shared accounts result in non-attributable data, a useless audit trail, and seriously compromised data integrity.

Conduct a computer system validation project for laboratory systems, including a list of requirements, design (configuration), and testing for the software and hardware, even for identical systems. Intelligently apply the risk-based approach.

Laboratory system validation involves documented qualification to assure the instruments are suitable for their intended use; however, the units are typically for one function, which do not require as comprehensive a validation effort as a complex IT system. Core validation elements should be performed: access controls, archival storage, reviews, qualification, configuration, documented suitability, change control, and maintenance records with adjustments using the risk-based approach.

Multiple identical systems can be covered under one validation package as long as the vendor qualification process indicates that each system is working as expected. Additionally, the system may be subjected to daily performance checks or periodic performance qualification to demonstrate the suitability of the system for its intended use.

21.2 Access Roster Review

Review the access roster at least annually for standard accounts and on a periodic basis for administrative access accounts. Ensure there is a policy or documented risk assessment that supports the choice of timing.

Look at the roles and responsibilities of everyone listed in the access roster and evaluate for conflicts of interest, especially in approver or release roles in the main part of the system.

- Compare the system access roster roles with the organization chart for the area and in Quality.
  - Verify that no one outside of Quality has a final approval role or data release role. Keep in mind that anyone with administrative privileges could also release data.
  - Verify that analysts cannot approve their own work.
  - Verify that access is disabled for anyone who has left the area.
- Determine if anyone in the business area has full administrative access to the system.
- Determine if anyone in the business area has full administrative access to the original data, either on the system or in the archive.

Look for trends in how access is handled to get a general idea of the business' state of control.

For example, a standard laboratory analyst left for another position in the company last year but still has access to five laboratory systems in their previous area. This may be five examples of inappropriate access, but if that access was well controlled, the likelihood of manipulating the data is low (analysts should never have delete permissions to the data). They need to be removed as part of a standard review process, but the risk dictates how often that review should be performed.

Conversely, if six analysts who left the area during the previous three years have access, a trend is indicated and a major problem with several processes has occurred. It also suggests scrutinizing administrator access for those systems to determine if the same neglect has been applied to administrator access.

21.3 Data and Transfers

There should be a clear understanding of what the original data is, and whether it is a manual recording into a laboratory notebook and/or controlled worksheet, or an electronic capture through a validated interface.

Original data for dynamic data will have been captured electronically and must be preserved as an electronic record for that record's retention requirement. It is easy to conceal issues with a printed paper copy; therefore, the electronic data needs to be retained not only for initial review by the organization but also by auditors and regulators.

Movement of data into the archive should be automated if possible. Tape is the backup for disaster recovery and is not considered long-term archival storage.

Set up the data archive for long-term archival storage (RAID arrays that are mirrored off site).

- Verify that members of the business area, or others with direct interest in the data, do not have the ability to delete the original data, either on the local drive or in the archive. This constitutes a conflict of interest, which is inappropriate access.

- Verify that the metadata is protected and stored (audit trails, method IDs, etc.); there must be a complete set of data with corresponding metadata to ensure data integrity.

- Verify that the organization can access the older data for the extent of the retention period.

    - If data from obsolete equipment is stored, can it still be opened in human-readable form?

    - If the old hardware was not stored, can it be opened in a virtual software environment? Keep in mind that virtual machines stop support of old software, for example, Windows 95 is no longer supported in the virtual software environment.

21.4 Data Processing

Data processing, or reprocessing and directions for performing manual integration, must be described by a procedure. Manual integrations should be rare and involve additional review (e.g., Quality).

Excessively reprocessing data or re-uploading data sets to IT systems can indicate a testing into compliance situation.

- Verify the manual integration process and some examples to ensure that the written process is followed.

- Determine if there is any evidence of only picking passing samples and discarding others. Ensure there is traceability in injection sequence numbers and timing – is everything accounted for, timing correct, and no orphan data?

- Verify the data review and approval timing: it should be a short time period, typically within one to two weeks.
- Verify that the organization is initiating deviations and conducting lab investigations.
    - Verify that root cause is found or if the organization relies on extensive retesting. Also, verify that the deviations are closed after the actions are completed.
    - Verify that CAPAs are applied across all stages of the process.
    - Verify that CAPAs cover physical and logical security.
- Review all of the manual integrations conducted for a method, looking for justification and a second person review for each one. Look for any with multiple manual integrations for a single injection; this is an indication of manipulating the data repeatedly until it passes.

21.5 Laboratory System Audit Trails

The audit trail should always be collecting data, recording who did what when, and logging why changes are made. The original entry should be retained in addition to the changed entry. No data should be obscured from review.

The audit trail should be on and reviewed as part of an approved process.

- Verify that data changes are part of a review process.
- For items that can be changed:
    - Review the audit trail for items such as proper reintegration, peak selection, proper data sets, and multiple re-uploads of data sets, etc.
    - Verify that “scientific judgment” is not used to rationalize changes. There should be clearly defined objective criteria for what is or is not acceptable in situations where data can be changed, and the changed data subjected to second person review. Anything out of the ordinary requires a formal investigation. Verify that CAPAs are completed and applied holistically.
- Review the time stamps for the data.
- Are there any gaps that would indicate the audit trail is being turned on and off (e.g., saving only the passing runs)?
- Do all the time stamps happen at the same or in an impossibly close period of time? Those conditions indicate that the organization is saving only the last perfect experiment and discarding the rest, which is not contemporaneous.

22 Appendix 17 – Reviewing IT Systems

This Appendix gives detailed guidance on auditing IT systems. Audits or reviews should be conducted according to the policies and procedures of the regulated company. Example audit and review criteria are given below, but these are not prescriptive or exhaustive.

22.1 IT System Overview

Computer system validation is a major component of ensuring data integrity along with providing evidence that the system is suitable for intended use. The system should be validated for its intended use (*ISPE GAMP® 5* [2], FDA [63], or other industry standard or practice). There should be a risk assessment that details the extent to which the system must be validated. The effort should drive the mitigation of any high or moderate risks down to a lower level that reduces the risk to the patient.

The main elements in a computer system validation that can be altered based on the impact assessment are typically the level of details in the requirements, design, and testing. The remaining core validation requirements are obligatory for most systems to ensure they are put into production use and then maintained in a validated state.

Predicate rule requirements must be met with defined quality requirements for electronic systems. The project must be a partnership between the business, QA, and IT using the risk-based approach for the area in which it is implemented. The business area needs to be involved in the testing and development of the system to ensure that it is validated for the intended use. Documented business acceptance testing provides proof that the system is suitable for its intended use.

Additionally, the business needs to use the system as intended or as it was validated. “Creative” use of a system may not be validated or tested and therefore may be unreliable.

22.2 User Access

There should be unique user IDs in use, and no shared accounts, as the data would not be attributable. The access roster should be reviewed periodically, with the frequency of review justified by a risk assessment and commensurate with system and data GxP criticality. Privileged access accounts may merit more frequent review than the standard accounts in the system; and manual or automated means should be used to ensure that user accounts that have not been used for an extensive period are disabled.

The system should be secure and access restricted, especially to the data files and audit trail. All of the data should be retained, including original entries when changes are made.

- Verify that unique user IDs are in use and users do not have multiple IDs where domains change due to global roles in different locations. Without this detail, the audit trail is worthless, electronic signatures are impossible, and the integrity of the data is seriously compromised.
- Verify roles and responsibilities of everyone listed in the access roster. Look for conflicts of interest especially in approver or release roles in the main part of the system. There should not be anyone that can approve their own work.
- Does anyone outside of Quality have a release role? Keep in mind that anyone that has administrative privileges could also release a batch.
- Does anyone in the business area have administrative access to the system? If so, audit trails can be turned on and off and data can be altered or deleted.

There is a database behind the user interface of the IT system where the data is stored and associations are made.

- Review the database administrator accounts for the system in which the data is stored for the main system.
- Verify if anyone in the business or anyone with a conflict of interest has access to the database.

The database usually does not have an audit trail; thus, any changes made are not recorded outside a documented change control process. Changes to the database require knowledge of the table structure and some unique skills but should not be neglected because inappropriate access can cause serious problems that cannot be detected by ordinary means.

- Verify how often the access roster is reviewed for database administrators. This is probably the highest-risk access to the system because of the lack of an audit trail combined with the ability to change anything in the system. Conduct the access review quarterly.
- Verify that database administrator access is removed when the person changes roles, or that the machine account is changed when someone leaves. The machine account is used by the database to allow access by the user interface to store and make associations with the data.

22.3 IT Audit Trails

In the main part of the IT system, the audit trail must always collect data. The audit trail should be on from system installation, and the data collected should be fully attributable, recording who did what when and why when changes are made. The business should have an SOP for audit trail review that clearly defines what should be reviewed in the system.

- Verify that the audit trail is collecting “who did what and when” (and “why” when changes are made).
- Verify that the audit trail review is part of an approved procedure.
- Verify what can be changed based on key criteria, and that the rest are locked down.
- Verify that the critical parameters are locked down. If they are not, who can change them? If the analysts or managers in the business can change them it is a conflict of interest.
- If nothing can be changed in the data (no reintegration, no updates) and the risk assessment clearly describes the process, then no audit trail review is needed. This is RARE.
- For items that can be changed, review the audit trail for elements such as proper reintegration, peak selection, proper data sets, and multiple re-uploads of data sets, etc.
    - Verify that “scientific judgment” is not used to rationalize the changes. There should be clearly defined objective criteria for what is or is not acceptable in situations where data can be changed, and these need to be reviewed (second person verification at a minimum).
    - Anything out of the ordinary requires a formal investigation. Verify that CAPAs are completed and applied holistically.
    - Review time stamps to assure data is collected and recorded contemporaneously – the last perfect experiment should not have all time stamps within a very short (or impossible) period of time.

The administrative activities within the main part of the IT system may or may not be audit trailed depending on how the system was developed. Additionally, even if the system provides for those activities to be audit trailed, the administrators may have the access to turn them on or off, so the perceived value is not there. The administrative role should be outside the business area to remove the conflict of interest and, as part of a formal change control process, requires oversight to make the changes.

- Verify that the change control process has appropriate oversight and that the changes are classified (e.g., major or minor) appropriately based on their impact to the system.

- Verify that emergency changes are implemented appropriately and not as a routine business practice. Emergency changes do not usually follow the same early review as routine system changes.

The database behind the user interface of the IT system may not have an audit trail or may not collect sufficient detail, so a formal change control process (with a clearly defined SOP) is usually required for the DBAs. This process should capture who did what, when, and why, plus assess the impact of the change. There should be oversight by the IT administrators in alignment with the business, and changes should be reviewed and approved. If SQL scripts are used to make changes to the database, then the scripts should be reviewed by a second person before use to make sure the scripts will not adversely affect the database. Additionally, a review of IT tickets related to back end data changes should be performed as these changes are often not subject to system audit trails.

22.4 IT System Validation

The system validation documentation should be robust and based on an industry best practice such as *ISPE GAMP® 5* [2] to validate the system for intended use. The requirements should be documented so that it is clear what elements the system needs and to ensure the rest of the validation effort is not misdirected or open for interpretation. The requirements design and testing should be traceable with no orphan elements such as a requirement with no testing.

- Verify validation elements based on the highest risk first.

- Verify that critical security elements are part of the requirements and trace to design (or configuration) and testing. Also, confirm that there is enough detail in each to understand what is required from the system. For example, depending on company policy, the requirements may include:

  - The system requires authentication with a unique user ID and a password before allowing access. There should also be a negative test that demonstrates that if an unauthorized individual tries to access the system, access is denied.

  - The system must require user passwords to be changed at least every 60 days.

  - The system must prevent users from reusing their current password.

  - The system must not allow passwords to be shorter than six characters.

  - The system must store passwords in an encrypted format.

  - After a stated number of consecutive unsuccessful authentication attempts, the user's account is disabled.

  - The system must be capable of providing an access roster.

- Verify other requirements, design, and testing items critical to the system such as access restrictions to certain data, critical parameters, limits, release criteria, and other high-risk activities are covered by the system validation.

- Verify that the test execution for critical items contains objective evidence including actual results or outcomes, not just subjective assessments like pass or fail.

  Pass or fail can be attributed after the actual data is documented but when a review is conducted, the reviewer should be able to look at the data and come to the same pass or fail conclusion as the tester based on the data presented.

  For non-critical items it is acceptable to have pass or fail results if this is clearly defined in the test planning.

- Determine if critical calculations or parameters have been independently verified or tested, and that changes to them are restricted.

- Verify that periodic reviews are conducted for the system and that the reviews contain enough information to tell what happened during the past year. This is also an opportunity to review the changes and deviations associated with the system. Some key elements should be included:

  - Overall access roster review identifying accounts removed as a result of the review. The administrator accounts should also be identified as to when reviewed and any resulting changes made.

  - Change control analysis to determine if enough changes or significant changes were made that warrant regression testing.

  - Assessment of deviations including outcomes and CAPAs for the system. Make sure CAPAs are tracked to completion and applied across the system. Also, verify that deviations are not closed before the actions are completed.

  - Analysis of planned activities that are overdue to ensure critical items are not missed or are not implemented in a timely fashion.

  - For either the change controls or deviations, there can also be a review of classification for impact (e.g., major or minor) to ensure items are assessed correctly, or that items are not hidden in a parallel system (such as putting deviations into change controls).

  - Vendor management should be periodically reviewed to ensure that the information is current, routine audits and outcomes are completed, and the vendor relationship is maintained including notification of defects and tracking to completion. See more details about vendor management in Appendix 15.

IT systems are often put into production use in manufacturing areas with known defects. Generally, this is not important. The concern is, for example, if one or more of the defects affects a critical calculation or a status change in the system allowing the premature release of a batch, or release of a failing batch.

Computer system validation is designed to detect and repair, or implement compensating controls to prevent, issues from the known defects. Critical defects are detected, documented, and the fix is tracked to completion using a CAPA process.

22.5 IT System Data Flow

The data flow to and from the system should be defined. The data transfers and interfaces should be well controlled, automated, and validated if possible.

- Could an interface failure cause old data to be used as the basis for a regulated decision?
  - Who receives notice of interface failures?
  - How is the notice provided: passive or active?
- Verify human touch points in data transfers, as they are particularly risky.
  - Can the data be modified or deleted during the transfer?
  - Does the system track multiple uploads of data?
  - Is all the required data transferred with the right files?
  - Is all the required metadata moved along with the original data files?
- For IT systems, verify that the software development team uses a development environment for programming, a test environment for configuration and testing, and a production environment for the actual production usage.

  Keep in mind that the test environment may need a copy of the production data loaded (if available) to behave like the production environment during testing. Large data loads can affect the reliability of the system.

  Any data moved from the production environment to the test environment does not need to be fully verified aside from making sure it all moved; however, none of the data moved to the test environment should be moved back into the production environment.

  Process control systems may only have a production environment, so appropriate controls need to be strict to keep from having an adverse effect on what is being produced.

22.6 IT System Data Storage

The data in the IT system should be moved to long-term archival storage.

- Verify that archival storage uses an appropriate long-term storage media such as RAID arrays rather than the more short-term tape backup media.

  Additionally, where mirroring is used, the data should be mirrored off site to prevent loss in the event of a disaster such as a fire on site. If the mirror is in the same server room, then both copies could be lost if that room burns or is sprayed with water even if the fire does not reach that server.

- Tape backup is usually only for disaster recovery purposes. Typically, there is about a 30-day backup cycle, at which point the previous tapes are overwritten, so if a problem is not caught within 30 days, it can no longer be recovered.

  In addition, tapes, DVDs, etc., used for backup and archive should be stored within the environmental requirements stated by the manufacturer for archive use. If tapes are stored for a protracted period of time they must be periodically loaded and exercised or they may become brittle and unusable.

23 Appendix 18 – Reviewing Supporting Data

This Appendix gives detailed guidance on reviewing supporting data, for example, data outside of the GxP data flow, during a data integrity audit.

The following is not intended to be a comprehensive list of examples but it is presented to give an idea of how to compare data from different systems to detect problems or verify the integrity of data.

23.1 Time Card or Badge-in vs. Data or Batch Approval

Verify that personnel who performed specific functions (such as approved or released a record) were onsite or appropriately logged in at the time. This can be done by comparing the time card, badge-in, or logon information that clearly indicates when personnel are onsite or logged in, to product release, batch approval records, data collection or release, or other critical business process to confirm the date and time of the activity.

Verify that the electronic badge-in system is validated for intended use. This is a means of physical security to the facility and is designed to prevent access to unauthorized individuals.

Likewise, a remote approval or release requires that the person conducting the activity was logged in during the time the action took place. Keep in mind that physical activities like loading samples into an injector to start a run require a physical presence on site.

This review includes using the organization chart to verify appropriate access in a system, such as checking to see if anyone outside of Quality has a data release or batch approval role.

23.2 Maintenance Records vs. Data in Historian

Obtain copies of maintenance records for a piece of equipment and check maintenance and calibration dates. Then check for when the equipment was used in the data historian (or other system of usage documentation) to verify if the equipment was used when out of calibration, or when removed for service, or was in need of maintenance.

23.3 Batch Records vs. Component or Material Records

Review the batch record and note the equipment used during production. Verify that information against maintenance and calibration records to determine if the equipment was current and maintained. Also verify materials used in the batch record against the system used to track material handling and storage. Verify that the correct components were used and that the amounts agree with what was consumed or what remained.

Verify that approved suppliers were used for the incoming materials. This could include a trip to the warehouse and also a review of the material tracking system to assure it was validated for its intended use.

23.4 Concealing Things in a Parallel System: The Numbers Game

The deviation system is designed to document deviations and track CAPAs to completion. If there are many deviations, it could be an indication that there is a general lack of control, so a decision may be made to reduce the overall numbers. If time or money is not provided to improve the quality in the area, then the numbers may be reduced by other means, such as putting them in a parallel system and renaming them as change controls, trouble tickets, excursions, etc.

Detecting the parallel system is not easy because the behavior usually has been well rationalized. Go to the local business area and ask the personnel actually doing the work how problems are managed in broader terms than just how deviations are handled. The personnel managing the system may have a vested interest in protecting the parallel system that the local business area may not share.

Ask if there are problems in the area that are not critical enough to put in a deviation and how those are handled. Obtain a list with a description of those items if possible. If they have been scattered among different systems, examples may need to be compiled from each. Review the examples and see if there is a trend or they are critical enough to warrant further action.

23.5 Timing: Determining the Real Sequence of Events

Verify that the time stamps in a batch record match the use logs for the equipment. Verify that the operators were actually on site when they were recording data. Verify that the time and dates align with the expected batch production timing. The time and dates should be reasonable and not all at once or within an impossible timeframe, which would be an indication that the data was not recorded contemporaneously.

Check the data historian against the batch record to make sure the data was collected as the batch was being produced. Verify that the in-process laboratory results and final laboratory results were produced in alignment with the production records. Look for forward-processing decisions or release decisions that occurred before the laboratory results were available.

24 Appendix 19 – Auditing Access Controls

This Appendix contains guidance on auditing access controls with respect to data integrity.
Auditing of the access controls is recommended to ensure that the regulated company's security policy is sufficient to support data integrity. The most common security issues at the database level are external attacks, unsanctioned activities by authorized users, and mistakes.

Developing an appropriate audit strategy helps ensure that appropriate security measures are in place and improvements are identified, as well as facilitate a root cause analysis when an incident does occur.
The audit strategy should consider auditing the following:

- Privileged user access to determine who has accessed the database, when the access was obtained, how that access was obtained (i.e., from where it originated), and what data was accessed during the session.

- Failed access attempts, which can be indicative of attempts to gain unauthorized access. Monitoring and review of failed access attempts help determine the level and mechanism of unauthorized attempts to access the database and inform the organization of the level of threat to their data and the robustness of the remediation actions necessary.

- The activities performed when access was granted. This audit may be conducted at the statement, privilege, or object level, or may be a deep-dive audit, particularly in cases where a violation of data integrity is suspected or identified.

  The audit should review DDL and DML statements issued by ad hoc query tools. It should also seek to identify any vulnerabilities, such as default or weak passwords, denial of service vulnerabilities, unpatched buffer overflow, improper configuration, etc., and then review any attempts by users to exploit these vulnerabilities in the database.

- Suspicious activity to identify any unusual or abnormal access to sensitive data. Suspicious activity can be identified by looking for patterns that do not fit into the established baseline of use and so indicate misuse, such as attempting to obtain access at unusual hours, through unusual hosts, or hosts outside the trusted network.

- Account creation to ensure that all accounts with access at the database level were created through the correct channels and with the correct permissions, that is, that the account has not been created by a hacker trying to gain access to the database.

  The review should verify that all active accounts are still required and should retire any that are no longer needed.

- Changes and deviations from the baseline policy and configuration schema for the database. The audit should attempt to identify if any unapproved changes to the configuration, security settings, privileges, user accounts, authentication settings, or database structure were made, and by whom and when.

Auditing must be a methodical and repetitive process that should be reviewed periodically to ensure that the system sufficiently protects data integrity in an ever-changing IT environment.

Appendices 10 and 11 respectively contain case studies for a RTSM within a GCP environment, and an ERP system used with medical devices.

25 Appendix 20 – Regulatory Guidance Regarding Classification of Deficiencies

This Appendix provides quotations from regulatory guidances regarding classifying deficiencies relating to data integrity. These are provided to assist with evaluating audit findings.

TGA [10] provides a definition of a "critical" deficiency which includes data manipulation:

> *"A deficiency in a practice or process that has produced, or may result in, a significant risk of producing a product that is harmful to the user. Also occurs when it is observed that the manufacturer has engaged in fraud, misrepresentation or falsification of products or data."*

PIC/S draft guidance [7] offers an indicative list of classifications:

> *"Impact to product with risk to patient health: Critical deficiency:*
>
> - *Product failing to meet specification at release or within shelf life.*
> - *Reporting of a 'desired' result rather than an actual out of specification result when reporting of QC tests, critical product or process parameters.*
>
> *Impact to product with no risk to patient health: Major deficiency:*
>
> - *Data being miss-reported, e.g. original results 'in specification', but altered to give a more favourable trend.*
> - *Reporting of a 'desired' result rather than an actual out of specification result when reporting of data which does not relate to QC tests, critical product or process parameters.*
> - *Failures arising from poorly designed data capture systems (e.g. using scraps of paper to record info for later transcription).*
>
> *No impact to product; evidence of widespread failure: Major deficiency:*
>
> - *Bad practices and poorly designed systems which may result in opportunities for data integrity issues or loss of traceability across a number of functional areas (QA, production, QC etc.). Each in its own right has no direct impact to product quality.*
>
> *No impact to product; limited evidence of failure: Other deficiency:*
>
> - *Bad practice or poorly designed system which result in opportunities for data integrity issues or loss of traceability in a discrete area.*
> - *Limited failure in an otherwise acceptable system."*

26 Appendix 21 – Detecting Aberrant Results

This Appendix gives detailed guidance on assessing whether data is self-consistent, and using analytics and statistical sampling to detect aberrant results that may indicate data integrity issues. The definition and occurrence of aberrant results are discussed in Section 5.2.

Genuinely created data is valid even when OOS or OOT. More often than not, an aberrant result is completely within the acceptable limits and looks exactly as would be expected, sometimes even too exact.

26.1 Detection Methods

Fundamentally there are two ways to detect aberrant results:

- Take the initial measure twice or more, independently. Multiple measurements are a great way to prevent fraudulent data because falsifying data requires collusion between two parties, but it is very expensive (at least twice the work) and assumes that the process to take the measurement works as intended (i.e., the methodologies and calibration of the equipment are sound). Generally, complete double testing is impractical and by no means cost effective.

  Retesting is often performed on a random sample of the population to give a degree of confidence that the entire population is correct. Such a confidence interval is dependent on a number of factors including population size, distribution of false data, nature of the test being performed, etc. Even under the most conservative of sampling approaches false data can be missed.

- Compare a data set against itself to establish patterns and detect outlying data, generally referred to as data analytics. Data analytics can take many different forms and can range from being very simple (e.g., page numbering in a report) to very sophisticated and complex (e.g., machine learning algorithms).

While it would be impossible to list all the types of analytics that can be run (and data analytics are continually evolving and improving), there are some fundamental rules that a good analytics program should follow:

1. The analytic should be applied to the complete data set (not a sample). Effort should be taken to prove completeness.

2. Each analytic should have a clearly stated hypothesis or question to be answered.

3. Estimate a central tendency for the data (typically mean, median, and mode) to get an overall picture of what is going on with the data. Histograms, boxplots, and dotplots are very helpful in visualizing the dataset. The central tendency assists with selection of the appropriate analytical tools.

4. Develop the analytics around a set of principles, such as ALCOA+, that will help ensure multiple key facets of data integrity are scrutinized, and apply those principles consistently.

   One example would be Consistent, under ALCOA+: consistent records have a logical sequence; therefore, create a query to look for irrational event sequences and for unusual amounts of time (long or short) between event steps, which could be a sign of false data.

5. Establish the specific analytics but expect them to change as there are many elements that will continue to change over time, such as an increasing understanding of where and how to find issues.

Develop a “toolbox” of analytics but retain a flexible approach to encompass new situations and experiences. Understanding the motive and methods used by “bad actors” (people deliberately creating false data) allows the creation of analytics that are specifically designed to detect these methods. Once the “bad actors” become aware of the analytics, it likely that alternative methods will be adopted by them that the current analytics will not detect.

26.2 Assumptions

Often the analytic is run in a separate environment from which the source data is generated so it is important to understand any limitations or constraints arising from this. Data analytics often rely on unstated assumptions, such as assuming the data set is complete and unaltered, therefore consideration must be given to:

- Reliance on the underlying system generating the data: IT General Controls overview, Change Management, Logical Access, and IT Operations. Changes made outside of the system, for example in a manual preparation step before the assay, or changing the clock in the operating system on which the application runs, make it more difficult to detect false data.

- Extraction and Load techniques: Controls to ensure that the extracted dataset is complete (sum totals, record counts). Controls to ensure that data has remained unchanged since it was extracted. Controls to ensure the data is loaded completely and accurately into the analytic tool.

26.3 Grouping, Normalizing, and Profiling Data

Grouping, normalizing, and profiling data creates stratifications that can be used to establish normal trends.

Data represents real life observations, measurements, and or transactions. It is only natural that data collected at the same site, using a common technique, equipment, or system, and perhaps by the same person or persons, will have a greater consistency than data collected from random systems generated at random times around the world.

Grouping data provides a means to quickly compare trends between one group and the next, and to more easily identify and explore groups that might be of interest. The hierarchy for grouping should be established before running the analysis but it may also be necessary to create new groupings as the analysis progresses in response to perceived trends or similarities.

A typical grouping hierarchy for a pharmaceutical company may look like this (from top down):

Process/Procedure → Product → Site → User → Date/Time Range

Normalizing data can also become an important factor, especially if the analytic relies on means or standard deviations. Seasonal products are a common example, but also consider work schedules, natural events, etc.

Simple data profiles describe what data looks like in a given field. This should take into account maximum, minimum, mean, and standard deviation, but should also include descriptive characteristics such as precision (number of decimals), frequency of whole numbers, upper and lower allowable limits. In some cases, data may correlate to a natural occurring series (e.g., Fibonacci series, power series, etc.).

Complex data profiles seek to identify a correlation between two or more fields. If a strong correlation is found, the profiles can be the primary driver for the analytic. For example, if the frequency of whole number results increases as the shift time nears its upper limit, it may indicate that technicians are rushing results toward the end of their shift.

Data sets, particularly audit trails, typically contain sequential data. It often helps to understand the logical flow of the process used to generate the data, and to develop a hypothesis of a particular activity that should occur in sequence and about how long it should take to complete. It is also helpful to understand the maximum possible simultaneous results.

By mapping the process step to events recorded in the data set, and then calculating the time difference between those events, a dotplot can be used to determine if any data points or groupings fall out of order or are consistently outside of the expected range.

While automated queries and analytics are excellent for comparing large sets of data, the innate ability of humans to detect patterns should not be discounted. Graphical representations, data compiled into a chronological sequence, and normalized subsets of data can assist human pattern spotting.

False data is a subset of aberrant results, where the data has been deliberately falsified to achieve a result within specification. False data is concerning because it represents deeper issues around data integrity, and once found it is difficult to rely on the entire data set, even though the majority of it may be accurate. Evidence of false data will call into question the validity of all data generated or reviewed by the parties involved, and may need addressing through the organization's data integrity council.

27 Appendix 22 – References

1. *ISPE GAMP® Guide: Records and Data Integrity,* International Society for Pharmaceutical Engineering (ISPE), First Edition, March 2017, www.ispe.org.

2. *ISPE GAMP® 5: A Risk-Based Approach to Compliant GxP Computerized Systems,* International Society for Pharmaceutical Engineering (ISPE), Fifth Edition, February 2008, www.ispe.org.

3. WHO Technical Report Series, No. 996, WHO Expert Committee on Specifications for Pharmaceutical Preparations, Annex 5: Guidance on good data and record management practices, World Health Organization (WHO), 2016, apps.who.int/medicinedocs/documents/s22402en/s22402en.pdf.

4. FDA Draft Guidance for Industry: Data Integrity and Compliance with CGMP, April 2016, US Food and Drug Administration (FDA), www.fda.gov.

5. MHRA Guidance: 'GXP' Data Integrity Guidance and Definitions, Revision 1, March 2018, Medicines & Healthcare products Regulatory Agency (MHRA), www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency.

6. EMA (European Medicines Agency) Questions and Answers: Good Manufacturing Practice – Data Integrity, www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/q_and_a/q_and_a_detail_000027.jsp.

7. PIC/S Draft Guidance: PI 041-1 (Draft 2), Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments, August 2016, Pharmaceutical Inspection Co-operation Scheme (PIC/S), www.picscheme.org.

8. China Food and Drug Administration (CFDA) Draft Guidance on Drug Data Management Practice, Notice on RAPS (Regulatory Affairs Professionals Society), October 2016, https://www.raps.org/news-articles/news-articles/2016/10/asia-regulatory-roundup-cfda-releases-draft-guidance-on-data-management-(18-october-2016).

9. Health Canada GUI-0001: Good manufacturing practices guide for drug products, February 2018, www.canada.ca/en/health-canada/services/drugs-health-products/compliance-enforcement/good-manufacturing-practices/guidance-documents/gmp-guidelines-0001/document.html.

10. TGA Policy: Data Management and Data Integrity (DMDI), Australian Government Department of Health Therapeutic Goods Administration (TGA), April 2017, www.tga.gov.au/.

11. *ISPE Cultural Excellence Report,* International Society for Pharmaceutical Engineering (ISPE), Fifth Edition, April 2017, www.ispe.org.

12. Braksick, L. W., *Unlock Behavior, Unleash Profits: Developing Leadership Behavior That Drives Profitability in Your Organization,* Publisher: McGraw-Hill Education, 2007.

13. International Council for Harmonisation (ICH), ICH Harmonised Tripartite Guideline, *Pharmaceutical Quality System – Q10,* Step 4, 4 June 2008, www.ich.org.

14. ISO/IEC 27001, Information Security Management Systems, www.iso.org/isoiec-27001-information-security.html.

15. EudraLex The Rules Governing Medicinal Products in the European Union: Volume 4, Good Manufacturing Practice Medicinal Products for Human and Veterinary Use, Annex 11: Computerised Systems, 1991.

16. 21 CFR Part 11 – Electronic Records; Electronic Signatures – Scope and Application, 1997, Code of Federal Regulations, US Food and Drug Administration (FDA).

17. EudraLex The Rules Governing Medicinal Products in the European Union: Volume 4, Good Manufacturing Practice Medicinal Products for Human and Veterinary Use, Annex 11: Computerised Systems, June 2011, ec.europa.eu/health/documents/eudralex/vol-4_en.

18. PIC/S Guide: PE 009-14 (Annexes), Guide to Good Manufacturing Practice for Medicinal Products Annexes, July 2018, Pharmaceutical Inspection Co-operation Scheme (PIC/S), www.picscheme.org.

19. FDA Warning Letter 320-15-17, September 2015, US Food and Drug Administration (FDA), www.fda.gov.

20. EU NCR 14MPPP078, ISPE Thailand Data Integrity Day 18 October 2016, http://www.ispeth.org/web/documents/ALCOA_workshop_model_answers-all_groups.pdf.

21. 21 CFR Part 58 – Good Laboratory Practice for Nonclinical Laboratory Studies, Code of Federal Regulations, US Food and Drug Administration (FDA), www.fda.gov.

22. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, Number 1, OECD Principles of Good Laboratory Practice, January 1998, Organisation for Economic Co-operation and Development (OECD), www.oecd.org/.

23. OECD Series on Principles of Good Laboratory Practice and Compliance Monitoring, Number 17, Advisory Document of the Working Group on Good Laboratory Practice, Application of Principles of GLP to Computerised Systems, April 2016, Organisation for Economic Co-operation and Development (OECD), www.oecd.org/.

24. EMA/CHMP/ICH, EMA/CHMP/ICH/135/1995, Guideline for Good Clinical Practice E6(R2), Step 5, June 2017, www.ema.europa.eu/ema/.

25. FDA Guidance for Industry: Electronic Source Data in Clinical Investigations, September 2013, US Food and Drug Administration (FDA), www.fda.gov.

26. 21 CFR Part 211 – Current Good Manufacturing Practice for Finished Pharmaceuticals, Code of Federal Regulations, US Food and Drug Administration (FDA), www.fda.gov.

27. EudraLex The Rules Governing Medicinal Products in the European Union: Volume 4, Good Manufacturing Practice Medicinal Products for Human and Veterinary Use, Chapter 4: Documentation, June 2011, ec.europa.eu/health/documents/eudralex/vol-4_en.

28. ISO ISO/IEC/IEEE 24765:2017 Systems and software engineering – Vocabulary, September 2017, International Organization for Standardization (ISO), www.iso.org/home.html.

29. *ISPE GAMP® Good Practice Guide: A Risk-Based Approach to Operation of GxP Computerized Systems,* International Society for Pharmaceutical Engineering (ISPE), First Edition, January 2010, www.ispe.org.

30. International Council for Harmonisation (ICH), ICH Harmonised Tripartite Guideline, *Quality Risk Management – Q9,* Step 4, 9 November 2005, www.ich.org.

31. PIC/S Good Practices for Computerised Systems in Regulated "GXP" Environments, PI 011-3, September 2007, Pharmaceutical Inspection Co-operation Scheme (PIC/S), www.picscheme.org.

32. ISO, International Organization for Standardization, www.iso.org/home.html.

33. ITIL®, www.axelos.com/best-practice-solutions/itil.

34. SSAE16 standards, Statement on Standards for Attestation Engagements (SSAE) No. 16, Reporting on Controls at a Service Organization, April 2010, ssae16.com/SSAE16_overview.html

35. SOC 2 (Service and Organization Controls) audit reports, American Institute of Certified Public Accountants (AICPA), www.ssae-16.com/soc-2/.

36. PCI Attestation (Payment Card Industry), www.pcicomplianceguide.org.

37. *ISPE GAMP® Good Practice Guide: A Risk-Based Approach to Testing of GxP Systems,* International Society for Pharmaceutical Engineering (ISPE), Second Edition, December 2012, www.ispe.org.

38. EudraLex Volume 1 – Pharmaceutical legislation for medicinal products for human use, Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use, Article 26, ec.europa.eu/health/documents/eudralex/vol-1_en.

39. MHRA GMP Data Integrity Definitions and Guidance for Industry, Revision 1.1, March 2015, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/697053/Data_integrity_definitions_and_guidance_v2_Withdrawn.pdf.

40. Allotrope Foundation, www.allotrope.org.

41. 21 CFR Part 11 – Electronic Records; Electronic Signatures – Scope and Application, September 2018, Code of Federal Regulations, US Food and Drug Administration (FDA), www.fda.gov.

42. FDA Compliance Program Guidance Manual, Chapter 46 – New Drug Evaluation, "Pre-Approval Inspections," CPG 7346.832, May 2010, www.fda.gov.

43. WHO Technical Report Series, No. 850, Annex 3: Guidelines for good clinical practice (GCP) for trials on pharmaceutical products, World Health Organization (WHO), 1995, http://apps.who.int/medicinedocs/en/d/Jwhozip13e/.

44. International Council for Harmonisation (ICH), ICH Harmonised Guideline, *Integrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6(R2),* Step 4, 9 November 2016, www.ich.org.

45. FDA Draft Guidance for Industry: Submission of Quality Metrics Data, Revision 1, November 2016, US Food and Drug Administration (FDA), www.fda.gov.

46. FDA's Application Integrity Policy, June 2018, US Food and Drug Administration (FDA), www.fda.gov.

47. FDA Warning Letters, Sun Pharmaceutical 5/7/14, US Food and Drug Administration (FDA), www.fda.gov.

48. FDA Warning Letters, Canton Laboratories Pvt. Ltd. 2/27/14, US Food and Drug Administration (FDA), www.fda.gov.

49. FDA Warning Letters, Posh Chemicals 8/23/13, US Food and Drug Administration (FDA), www.fda.gov.

50. Carey, Bjorn, "Stanford researchers uncover patterns in how scientists lie about their data" 16 November 2015, news.stanford.edu/2015/11/16/fraud-science-papers-111615/.

51. FDA Warning Letters, Yunnan Hande Bio-Tech. Co. Ltd., April 2015, US Food and Drug Administration (FDA), www.fda.gov.

52. *ISPE GAMP® Good Practice Guide: Manufacturing Execution Systems – A Strategic and Program Management Approach,* International Society for Pharmaceutical Engineering (ISPE), First Edition, February 2010, www.ispe.org.

53. EMA, Guideline on GCP compliance in relation to trial master file (paper and/or electronic) for content, management, archiving, audit and inspection of clinical trials, EMA/15975/2016, March 2017, European Medicines Agency, www.ema.europa.eu/ema/.

54. Weigers, Karl. E., *More About Software Requirements: Thorny Issues and Practical Advice,* Microsoft Press, 2005.

55. FDA Inspection Guides, Glossary of Computer System Software Development Terminology, August 1995, US Food and Drug Administration (FDA), www.fda.gov.

56. BSC SIGIST, BS 7925-1 (Working Draft) Standard Glossary of Testing Terms, Ver. 6.3, British Computer Society Specialist Interest Group in Software Testing BSC SIGIST, testingstandards.co.uk/bs_7925-1_online.htm.

57. Mikulak, Raymond J., Robin McDermott, Michael Beauregard, *The Basics of FMEA,* Productivity Press; Second Edition, 2008.

58. Saltzer, Jerome H. and Schroeder, Michael D., "The Principle of Least Privilege" in "The Protection of information in computer systems," *Proceedings of the IEEE,* Vol. 63, No. 9, September 1975, pp. 1278-1308.

59. Collins, Matthew L., Michael C. Theis, Randall F. Trzeciak, Jeremy R. Strozer, Jason W. Clark, Daniel L. Costa, Tracy Cassidy, Michael J. Albrethsen, Andrew P. Moore, "Common Sense Guide to Mitigating Insider threats", CERT Insider Threat Center, CMU/SEI-2016-TR-015, Fifth Edition, 2016, Software Engineering Institute, Carnegie Mellon University, resources.sei.cmu.edu/library/asset-view.cfm?assetid=484738.

60. Japanese ER/ES, English Translation from Pharmaceuticals and Medical Devices Agency (PMDA), www.pmda.go.jp/files/000153231.pdf.

61. USP 41–NF 36, First Supplement, General Chapter <1058>, issued 2018, The United States Pharmacopeial Convention, www.usp.org.

62. FDA: "Request for Quality Metrics Guidance for Industry," July 2015, US Food and Drug Administration (FDA), www.fda.gov.

63. US Code of Federal Regulations (CFRs), Title 21 – Food and Drugs, https://www.ecfr.gov/cgi-bin/text-idx?SID=c39c6f19d8341dc2cca0000efa700849&mc=true&tpl=/ecfrbrowse/Title21/21cfrv1_02.tpl#0.

64. ISPE Glossary of Pharmaceutical and Biotechnology Terminology, www.ispe.org.

65. Techopedia, www.techopedia.com/.

28 Appendix 23 – Glossary

28.1 Abbreviations and Acronyms

ALCOA	Acceptable, Legible, Contemporaneous, Original, Accurate
ALCOA+	ALCOA, with the addition of Complete, Consistent, Enduring, Available
API	Active Pharmaceutical Ingredient
C of A	Certificate of Analysis
CAPA	Corrective and Preventive Action
CDS	Chromatography Data System
CFR	Code of Federal Regulations (US FDA)
CGMP	Current Good Manufacturing Practice
CHMP	Committee for Medicinal Products for Human Use
CI	Continuous Improvement
COTS	Commercial off the Shelf
CPP	Critical Process Parameter
CQA	Critical Quality Attribute
DBA	Database Administrator
DDL	Data Definition Language
DMAIC	Define, Measure, Analyze, Improve, and Control cycle
DMDI	Data Management and Data Integrity
DML	Data Manipulation Language
DMZ	Demilitarized Zone
EBR	Electronic Batch Record
eCRF	electronic Case Report Form
EDC	Electronic Data Capture
EHR	Electronic Health Record
e-LN	electronic Laboratory Notebook
EMA	European Medicines Agency (EU)
ERES	Electronic Records and Electronic Signature
ERP	Enterprise Resource Planning
FDA	Food and Drug Administration (US)
FMEA	Failure Mode Effect Analysis
GAMP®	Good Automated Manufacturing Practice
GCP	Good Clinical Practice
GDocP	Good Documentation Practice
GLP	Good Laboratory Practice
GMP	Good Manufacturing Practice
GUI	Graphical User Interface
GxP	Good "x" Practice
HCIN	Health Insurance Claim Number
IaaS	Infrastructure as a Service
ICH	International Council for Harmonisation

ID	Identification
ISO	International Organization for Standardization
IT	Information Technology
ITIL®	Information Technology Infrastructure Library
KPI	Key Performance Indicator
LBI	Leading Behavioral Indicator
LIMS	Laboratory Information Management System
LQI	Leading Quality Indicator
MES	Manufacturing Execution System
MHRA	Medicines and Healthcare products Regulatory Agency (UK)
MRI	Magnetic Resonance Imaging
NDA	New Drug Application
OECD	Organisation for Economic Co-operation and Development
OLA	Operational Level Agreement
OOE	Out of Expectation
OOS	Out of Specification
OOT	Out of Trend
PaaS	Platform as a Service
PC	Personal Computer
PCI	Payment Card Industry
PCS	Process Control System
PDCA	Plan-Do-Check-Act
PHI	Protected Health Information
PI	Principal Investigator
PIC/S	Pharmaceutical Inspection Co-operation Scheme
PII	Personally Identifiable Information
QA	Quality Assurance
QC	Quality Control
QMS	Quality Management System
R&D	Research and Development
RAID	Redundant Array of Independent Disks
RDI	Records and Data Integrity
RTSM	Randomization and Trial Supply Management
SaaS	Software as a Service
SAP	Enterprise software to manage business operations
SCADA	Supervisory Control and Data Acquisition
SIEM	Security Information and Event Monitoring
SLA	Service Level Agreement
SMART	Specific, Measurable, Achievable, Realistic, Testable
SME	Subject Matter Expert
SOC	Service and Organization Controls
SOD	Separation or Segregation of Duties
SOP	Standard Operating Procedure

SQL	Structured Query Language
SSADM	Structured System Analysis and Design Method
SSAE	Statement on Standards for Attestation Engagements
SSH	Secure Shell
SSN	Social Security Number
TGA	Therapeutic Goods Administration (Australia)
UML	Unified Modeling Language
URS	User Requirements Specification
VPN	Virtual Private Network
WFI	Water for Injection
WHO	World Health Organization

28.2 Definitions

Corrective and Preventive Action (CAPA) (ISPE Glossary [64])

A quality system defined by 21 CFR Part 820.100; the policies, procedures, and support systems that enable a firm to assure that exceptions are followed up with appropriate actions to correct the situation, and with continuous improvement tasks to prevent recurrence and eliminate the cause of potential nonconforming product and other quality problems.

Data Dictionary (IEEE/ISPE Glossary [64])

A collection of the names of all data items used in a software system, together with relevant properties of those items; e.g., length of data item, representation, etc.

Data Definition Language (DDL) (Techopedia [65])

A computer language used to create and modify the structure of database objects in a database. These database objects include views, schemas, tables, indexes, etc.

Data Manipulation Language (DML) (Techopedia [65])

A family of computer languages including commands permitting users to manipulate data in a database. This manipulation involves inserting data into database tables, retrieving existing data, deleting data from existing tables and modifying existing data. DML is mostly incorporated in SQL databases.

Data Integrity (MHRA, 2018 [5])

Data integrity is the degree to which data are complete, consistent, accurate, trustworthy, 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 appropriate quality and risk management systems, including adherence to sound scientific principles and good documentation practices.

Hybrid Situation

A situation where paper and electronic record and signature components coexist.

Metadata (MHRA, 2018 [5])

Metadata are data that describe the attributes of other data and provide context and meaning. Typically, these are data that describe the structure, data elements, inter-relationships and other characteristics of data, e.g. audit trails. Metadata also permit data to be attributable to an individual (or if automatically generated, to the original data source).

Metadata form an integral part of the original record. Without the context provided by metadata the data has no meaning.

Primary Record (MHRA, 2015 [39])

The record which takes primacy in cases where data that are collected and retained concurrently by more than one method fail to concur.

Redundant Array of Independent Discs (RAID) (ISPE Glossary [64])

Usually a server array that has redundancy built in so if one drive fails it can be replaced with no loss of data.

In automation, an array configuration and applications for multiple independent disk drives as if they were one large disk. RAID provides a method of accessing multiple disks as if they were one large disk. RAID is typically used for file servers, transaction of application servers, where data accessibility is critical, and where fault tolerance is required.

Raw Data (ISPE Glossary [64])

Any worksheets, records, memoranda, notes, or exact copies thereof that are the result of original observations, measurements, recordings, etc. of an activity, such as a study, operation, investigation, etc., and are necessary for the reconstruction and evaluation of the report of that activity. In the event that exact transcripts of raw data have been prepared (e.g., tapes which have been transcribed verbatim, dated, and verified accurate by signature), the exact copy or exact transcript may be substituted for the original source as raw data. Raw data may include photographs, microfilm or microfiche copies, computer printouts, magnetic media, dictated observations, and recorded data from automated instruments. As such, the raw data may exist in either hard/paper copy or electronic format.

Wrongful Act (FDA [46])

A wrongful act is any act that may subvert the integrity of the review process. A wrongful act includes, but is not limited to, submitting a fraudulent application, offering or promising a bribe or illegal gratuity, or making an untrue statement of material fact. A wrongful act also includes submitting data that are otherwise unreliable due to, for example, a pattern of errors whether caused by incompetence, negligence, or a practice such as inadequate standard operating procedures or a system-wide failure to ensure the integrity of data submissions. A wrongful act may be evidenced in a document, including informal documents such as correspondence or memoranda, or verbally, such as in telephone conversations or in one-on-one meetings. Regardless of the means, each suspected incident of a wrongful act should be reported and investigated to determine whether they raise significant questions regarding data integrity and reliability with respect to a regulated product.

600 N. Westshore Blvd., Suite 900, Tampa, Florida 33609 USA
Tel: +1-813-960-2105, Fax: +1-813-264-2816

www.ISPE.org