



GOOD PRACTICE GUIDE:

Data Integrity – Manufacturing Records



Documents
Document Sharing Hub

GOOD PRACTICE GUIDE:

Data Integrity – Manufacturing Records

Disclaimer:

The *ISPE GAMP® RDI Good Practice Guide: Data Integrity – Manufacturing Records* provides practical and pragmatic advice to help ensure data integrity throughout the manufacturing business process. This Guide is solely created and 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.

© Copyright ISPE 2019. All rights reserved.

All rights reserved. No part of this document may be reproduced or copied in any form or by any means – graphic, electronic, or mechanical, including photocopying, taping, or information storage and retrieval systems – without written permission of ISPE.

All trademarks used are acknowledged.

ISBN 978-1-946964-20-5

Preface

The importance of achieving and maintaining data integrity cannot be overstated, as data integrity is the essential foundation to safeguarding product quality and patient safety. There are new regulatory citations for data integrity violations nearly every month for pharmaceutical companies worldwide. The diversity and complexity of manufacturing data compounds the data integrity issue. Links between the physical process and electronic sensors, conducting certain activities using paper records as part of a hybrid system, and employing multiple suppliers and systems within the manufacturing process all contribute to the challenges of maintaining the integrity of manufacturing records.

Data integrity relies on:

- Having the right technical controls for computerized systems used in support of the business process to prevent unauthorized changes to manufacturing data and to preserve the GMP content and meaning of that data as representative of the product quality
- Creating a culture that promotes and preserves data integrity in support of patient safety and product quality
- Implementing a review process for meaningful oversight of data regardless of the systems used

This Guide provides practical guidance for data integrity across the manufacturing business process for pharmaceutical companies at all levels of data integrity maturity. It is positioned under the *ISPE GAMP® Guide: Records and Data Integrity* and is aligned with the principles of *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 (Syneos Health, 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:

John Andrews	Andrews Consulting Enterprises, Ltd.	United Kingdom
Karen Ashworth*	Karen Ashworth Consulting, Ltd.	United Kingdom
Reetu Chopra	Teva Pharmaceutical Industries, Ltd.	United Kingdom
Kristina Dimitriadis	Werum IT Solutions GmbH	Germany
Brian Frederiksen	NNE	Denmark
Željko Granoša	Pliva Croatia, Ltd.	Croatia
Paul Irving	Northern Life Sciences Ltd.	United Kingdom
Hilary Mills-Baker	Emerson Automation Solutions	United Kingdom
Steve Murray	Werum IT Solutions	USA
Gregory Ruklic	Independent Consultant	USA
Catherine Smillie	Sharpe Solutions & Compliance Ltd.	United Kingdom
David Stokes	Convalido Consulting Limited	United Kingdom
Rajesh Thempadiyil	Dr. Reddy's Laboratories, Ltd.	India
Anders Vidstrup	NNIT A/S	Denmark
Christian Wölbeling	Werum IT Solutions GmbH	Germany

*Karen Ashworth significantly contributed to the quality of the final guide by restructuring and editing the guide after leading the team creating the process control systems content.

Contributors and Team Reviewers

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

Carsten Bierans	Werum IT Solutions GmbH	Germany
Daniel Brandl	Rockwell Automation	USA
Peter Falcon	Ferring International Center SA	Switzerland
Cristiano Behringer Ferrari	Nordika Do Brasil Consultoria Ltda	Brazil
Louis Hendriks	Global Value Web	Netherlands
Mark Wager	Eli Lilly and Company	USA

Subject Matter Expert Input and Review

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

Stuart Jones	MSD	Ireland
--------------	-----	---------

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
-----------	--	-----------

Special Thanks

The Leads would like to give particular thanks to Chris Reid (Integrity Solutions Ltd.), Arthur “Randy” Perez (Novartis – retired) 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

AVA Documents
Document Sharing Hub

Table of Contents

1	Introduction	7
1.1	Background.....	7
1.2	Purpose.....	7
1.3	Scope.....	8
1.4	How to Use This Guide	8
1.5	Terminology.....	10
2	Data and Records	17
2.1	What Is Special about Manufacturing Systems?.....	17
2.2	What Is Regulated Data?	19
3	Data Flows and Data Lifecycle.....	23
3.1	Mapping Data Flows	23
3.2	Typical GMP Data Flows.....	23
3.3	Data Lifecycle	27
4	Risk Management Approaches.....	41
4.1	Initial Data Integrity Risk Assessment and Identifying the Impact of Poor Data Integrity Controls	41
4.2	Identifying Data Flows with Impact on Patient Safety, Product Quality, and Data Integrity	41
4.3	Performing Functional Risk Assessments and Identifying Controls.....	41
4.4	Implement and Verify Appropriate Controls	42
4.5	Reviewing Residual Risks and Monitoring Controls	42
5	Detecting Data Integrity Issues	43
5.1	Data Review Requirements	43
5.2	Critical Thinking	52
6	Preventing Data Integrity Issues.....	57
6.1	Data Integrity Culture	57
6.2	Supplier Management Considerations.....	57
6.3	Essential Technical Features	61
6.4	Policies and Procedures	64
6.5	Access Considerations	66
6.6	Critical Validation Activities	70
7	Quick Wins	71
7.1	Quick Wins for Existing Systems	71
7.2	Quick Wins for New Projects	75
7.3	Quick Wins for Governance.....	76
8	Appendix 1 – Process Control System Risk Assessment Example	79

9	Appendix 2 – Risks to Data Integrity Associated with Specific Data Types.....	81
9.1	Automatically Captured Process Data	82
9.2	Simple Calculated Values	83
9.3	Manually Entered Process Data	85
9.4	Process Data Originating in Another System.....	86
9.5	Alarms and Events.....	88
9.6	Exception Reports.....	89
9.7	Images and Image Recognition	90
9.8	Complex/Statistical Transformations.....	92
9.9	Outputs Used Manually by the Operator.....	93
9.10	Production Definition Management – Master Recipes.....	94
9.11	Production Resource Management – Material Control.....	95
9.12	Production Resource Management – Equipment Control.....	96
9.13	Production Resource Management – Personnel Control.....	97
9.14	Production Dispatching.....	98
9.15	Production Execution – Control Recipes and Workflows.....	99
9.16	Production Data Collection and Production Tracking.....	100
9.17	Production Performance Analysis	100
9.18	Manufacturing Production Record Review, Release, and Disposition.....	101
10	Appendix 3 – Data Integrity Maturity Factors	103
11	Appendix 4 – Essential Technical Features	113
12	Appendix 5 – Policies and Procedures	121
13	Appendix 6 – Example Segregation of Duties.....	125
14	Appendix 7 – Validation Activities	129
15	Appendix 8 – Data Integrity Citations by Regulators.....	137
16	Appendix 9 – References	145
17	Appendix 10 – Glossary	149
17.1	Abbreviations and Acronyms	149
17.2	Definitions	151

1 Introduction

1.1 Background

Manufacturing systems, with their link to the physical process, have an obvious impact on patient safety and product quality, and have seen a notable increase in data integrity citations¹ over recent years [1, 2]. Manufacturing systems² have different inherent challenges compared to other systems, for instance, laboratory systems, such as:

- Process Control Systems (PCS) typically have a limited user interface that lacks the sophisticated levels of granular access controls present in higher level systems or other organizational areas.
- The sheer volume of acquired data from the myriad of sensors throughout the process may only be manageable through trending and exception reporting processes.
- A manufacturing process may involve packaged systems and integrated equipment sourced from many different suppliers, data from all of which may need to be transferred into a Manufacturing Execution System (MES) in a coherent format for incorporation into the electronic production record.

This Guide provides practical guidance for data integrity across the manufacturing business process, with particular focus on PCS, MES, and the interfaces and relationship between them.

This Guide uses the framework of the *ISPE GAMP® Guide: Records and Data Integrity* [3], a data lifecycle approach to the development and use of manufacturing IT systems such as MES and PCS.

1.2 Purpose

This Guide focuses on practical and pragmatic advice to help ensure data integrity throughout the manufacturing business process as well as providing data integrity considerations for systems at specific levels within the ANSI/ISA-95 hierarchy [4]. (See Section 1.5.1 for an explanation of the levels.)

- Levels 1 and 2: PCS
- Level 3: MES

This Guide provides a manufacturing business process framework to:

- Facilitate interpretation of regulatory requirements and expectations based on industry best practices
- Define the data integrity approach for planning and project activities
- Apply risk management to all aspects of the data lifecycle
- Identify roles and responsibilities to manage the data within the data lifecycle
- Integrate existing and new systems
- Scale activities and documentation appropriately

¹ Example citations categorizing the specific data integrity concerns are given in Appendix 8.

² The term “manufacturing systems” in this Guide refers to any type of manufacturing system, whether it is a PCS, MES, or other system used in the manufacturing process that creates regulated data.

- Leverage existing documentation and knowledge
- Develop an appropriate data strategy for manufacturing IT systems

1.3 Scope

The scope of this Guide is data integrity for MES and PCS, i.e., ANSI/ISA-95 Levels 1-3, used in a GMP environment [4].

ISPE GAMP® Good Practice Guide: A Risk-Based Approach to GxP Process Control Systems [5] and *ISPE GAMP® Good Practice Guide: Manufacturing Execution Systems – A Strategic and Program Management Approach* [6] provide detailed guidance on validating and maintaining PCS and MES in a compliant state throughout their operational life; therefore, these topics are not covered within this *ISPE GAMP® RDI Good Practice Guide: Data Integrity – Manufacturing Records*.

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) [7]
- FDA Guidance for Industry: Data Integrity and Compliance with Drug CGMP Questions and Answers (December 2018) [8]
- MHRA “GXP” Data Integrity Guidance and Definitions – Version 1 (March 2018) [9]
- EMA Guidance on good manufacturing practice and good distribution practice: Questions and answers, Data integrity (August 2016) [10]
- PIC/S Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments draft guidance (PI041-1 Draft 3)³ (November 2018) [11]
- China FDA Draft Guidance on Drug Data Management Practice⁴ [12]
- Health Canada GUI-0001 Good manufacturing practices guide for drug products (February 2018) [13]
- TGA Data Management and Data Integrity (DMDI) (April 2017) [14]

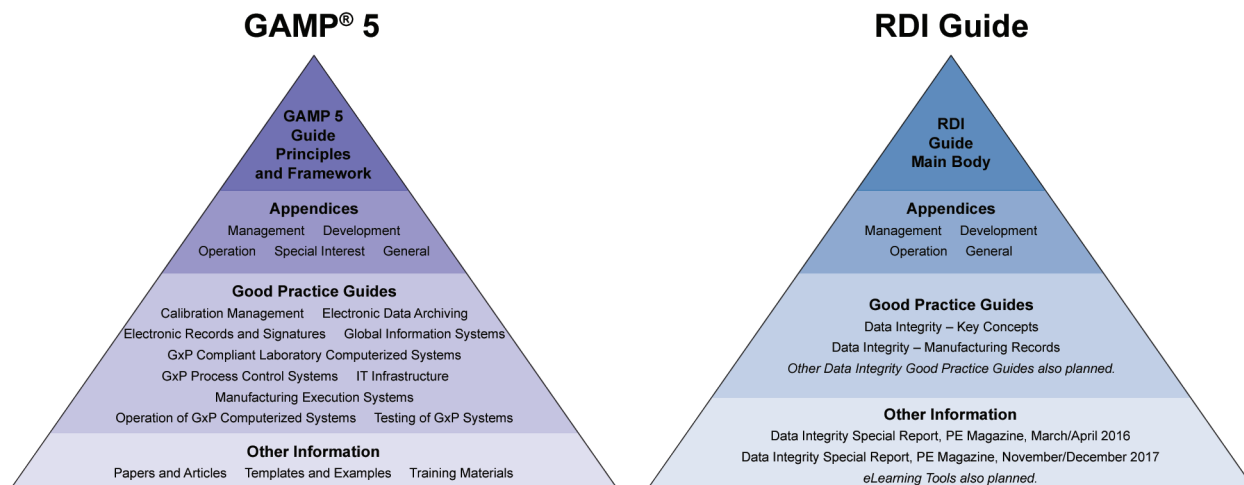
1.4 How to Use This Guide

This Good Practice Guide is positioned under the *ISPE GAMP® Guide: Records and Data Integrity* (RDI) [3] and is intended to provide additional and specific detail on data integrity, in the same way as *ISPE GAMP® 5* [15] has Good Practice Guides that provide additional and specific detail on computerized systems validation.

³ **Note:** PI041-1 [11] is a draft guidance at the time of the *ISPE GAMP® RDI Good Practice Guide: Data Integrity – Manufacturing Records* publication.

⁴ **Note:** Guidance on Drug Data Management Practice [12] is a draft guidance at the time of the *ISPE GAMP® RDI Good Practice Guide: Data Integrity – Manufacturing Records* publication.

Figure 1.1: Structure Basis of This Guide



The main body of this Guide addresses the requirements for data integrity in a GMP environment, with sections providing guidance on:

- Regulated records
- Data flows and data lifecycle
- Risk management approaches
- Detecting data integrity issues
- Preventing data integrity issues

The Guide also contains a chapter on “quick wins” (Chapter 7), which suggests actions that can create considerable improvement in the integrity of manufacturing system data with only modest resources. This section may be useful when prioritizing actions to improve data integrity; it should not be seen as a full list of the required activities.

The extensive appendices in this Guide provide system-specific examples of:

- Risk assessments
- Risks relating to data types and usage
- Maturity factors
- Technical features
- Policies and procedures
- Segregation of duties
- Critical validation activities to support data integrity

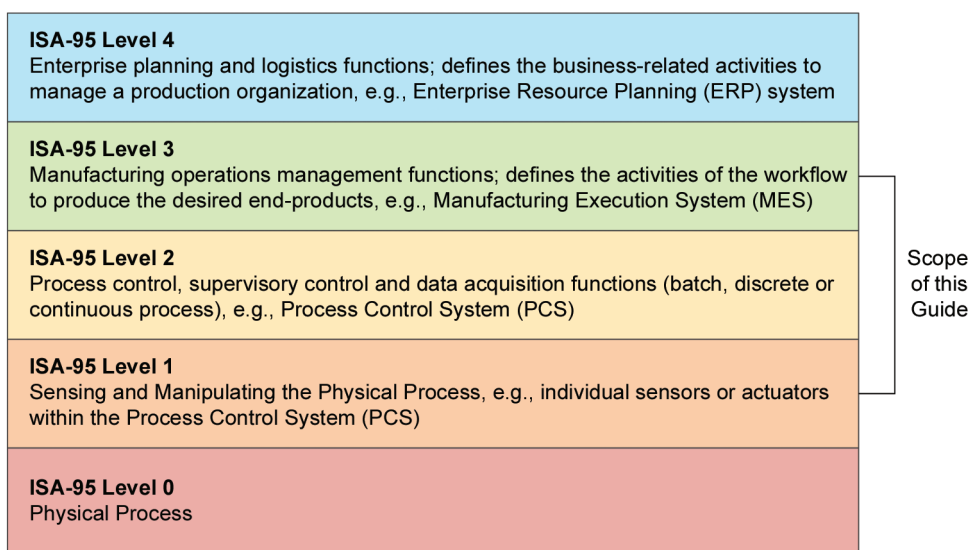
1.5 Terminology

The term “manufacturing systems” in this Guide refers to any type of manufacturing system, whether it is PCS, MES, or other system used in the manufacturing process that creates regulated data.

1.5.1 ANSI/ISA-95 Terminology

Within the ANSI/ISA-95 standard [4], levels of functional hierarchy are defined as shown in Figure 1.2:

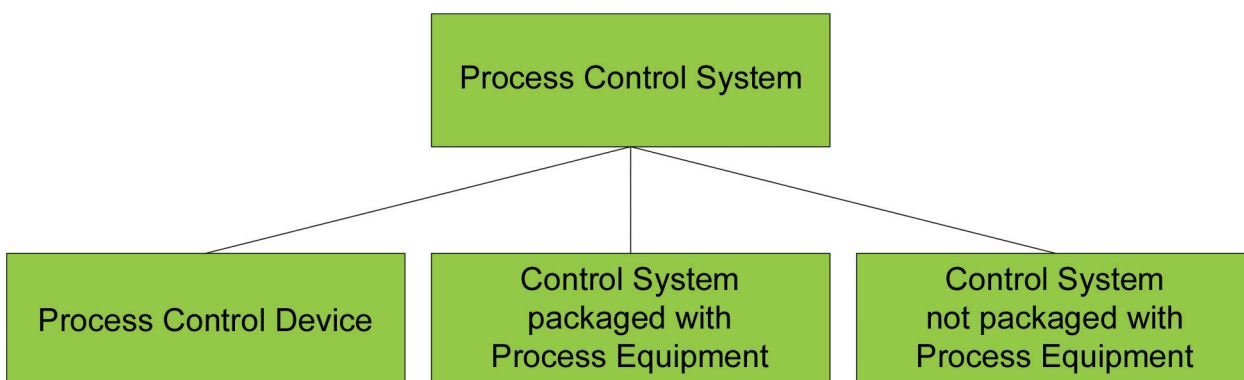
Figure 1.2: ANSI/ISA-95 Functional Hierarchy Levels [4]



1.5.2 Process Control System Terminology

The *ISPE GAMP® Good Practice Guide: A Risk-Based Approach to GxP Process Control Systems* [5] defines the types of PCS as shown in Figure 1.3.

Figure 1.3: Types of Process Control Systems [5]



Process Control Device [5]

A single device from a single supplier comprises an entire standalone PCS (e.g., single loop controllers)

Control System Packaged with Process Equipment [5]

Process control elements are embedded within process equipment into a “packaged” product from a supplier where the supplier Quality Management System (QMS) provides a level of assurance of the system in effect as though it were a commodity (may comprise multiple instruments, Programmable Logic Controllers (PLCs), and SCADA/HMI components).

Control System Not Packaged with Process Equipment [5]

PCS elements are developed independently in order to deliver a specific business process (may involve integrating multiple process control devices or packaged systems).

1.5.3 Manufacturing Execution System Terminology

ISPE GAMP® Good Practice Guide: Manufacturing Execution Systems – A Strategic and Program Management Approach describes the MES Domain as [6]:

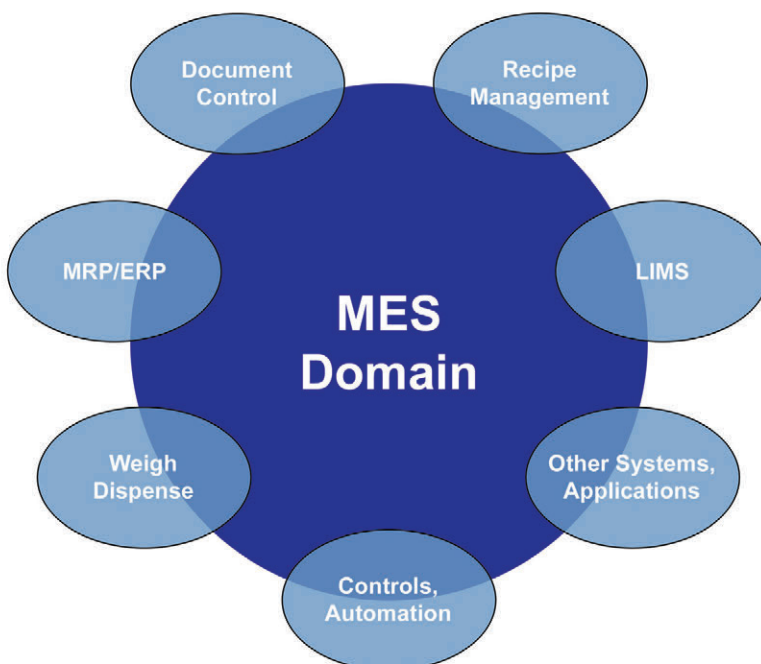
“The MES Domain is composed of computer systems, applications, and equipment functionality, along with any related data and information, which is used to produce desired intermediate results and final products.”

“An MES domain includes functionality relevant to the intended scope of manufacturing activities; this may incorporate only the necessary functionality from systems components or applications as necessary to support the domain defined by an organization.”

Within a manufacturing process, master and operational data may be spread among more than one computer system and/or across multiple ISA-95 layers [4]. Additionally, data may flow in between systems and layers within the data lifecycle. An example is shown in Figure 3.1.

To assist in defining, assessing, and managing this data, it is often useful to consider functional domains. A functional domain defines a logical boundary around functions and data that relate to a common business process, even where the functions and data may reside in separate systems in different layers of the ISA-95 model [6]. See Figure 1.4.

Figure 1.4: An Example View of the MES Domain [6]



1.5.4 Batch Record Terminology

This Guide uses the same terminology as the *ISPE GAMP® 5: A Risk-Based Approach to Compliant GxP Computerized Systems*, Appendix S2 [15] and the *ISPE GAMP® Good Practice Guide: Manufacturing Execution Systems – A Strategic and Program Management Approach* [6], which harmonized the terminology related to recipe-based systems and the ANSI/ISA-88 [16] Batch Control standard as shown in Table 1.1.

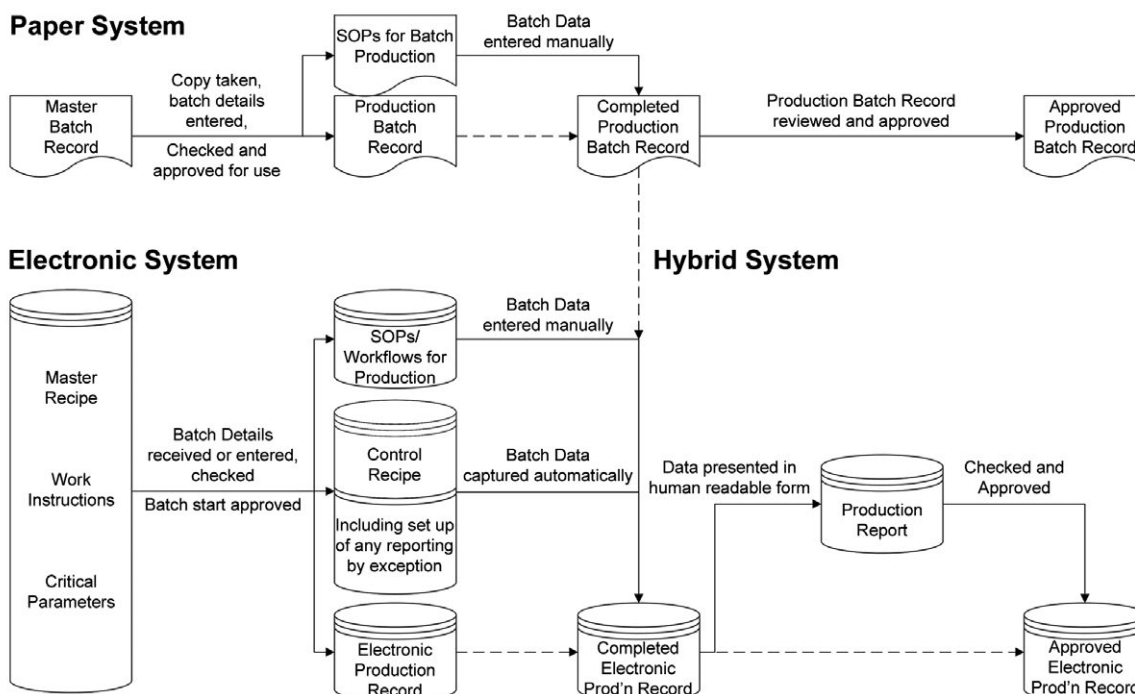
Table 1.1: Batch Record Terminology [6]

Paper System	Electronic System Equivalent Structure	Description
Master Batch Record (MBR)	Master Recipe	Contains product name or designation, recipe designation or version, formulas, equipment requirements or classes, sequence of activities, procedures, normalized bill of materials (quantity per unit volume to produce)
	Work Instructions (optional)	Additional detailed instructions – may include electronic SOPs or SOP references
	Critical Process Parameters (optional)	Required Process Parameters that are to be checked or monitored or are to be downloaded to other systems such as automation
Production Batch Record	Control Recipe	<ul style="list-style-type: none"> A Master Recipe dispatched or otherwise made available in manufacturing-related areas for execution Includes Master Recipe information with the addition of schedule, specific quantity to make, actual target bill of materials quantities, and other data for the batch and production instance
	Electronic Production Record	<ul style="list-style-type: none"> A store of data and information created by systems or entered by personnel during execution of Control Recipes May be located in one or more systems or databases Data may or may not be stored in human readable format
	Batch (Production) Report	Data and information in human readable format, presented either in electronic or paper format for activities, such as review, disposition, investigation, audit, and analysis

Records that contain definitions of required formulae, processes, and monitoring activities (whether these apply to a batch or continuous production) are referred to as “Master Recipes,” “Work Instructions,” and “Critical Process Parameters” (CPPs) in this Guide (see Figure 1.5).

Records that relate to a particular production instance are referred to as “Control Recipes,” “Electronic Production Records,” “Production Records” (to include for elements not stored electronically), and “Production Report” in this Guide (see Figure 1.5).

Figure 1.5: Production Record Terminology



1.5.5 Audit Trail Terminology

This Guide uses the same terminology as the *ISPE GAMP® Guide: Records and Data Integrity* [3] when referring to data audit trails as clearly distinguished from system technical logs and other event logs.

Data audit trails, as required by various regulations [7, 17, 18, 19], record operator actions that create, modify, or delete GMP records during normal operation, and should be clearly distinguished from other system and technical logs.

In a traditional paper-based system, if a user recognizes that a specific data entry is wrong, they strike out the wrong data in a way that it is still readable and put the correct value next to it with their initials, date, and in some cases, the reason.

In an electronic system, the data audit trail provides equivalent traceability when users create, modify, or delete GMP records and data as part of normal operations by recording the original and new values, the user, the time stamp, and in some cases, the reason.

In a paper-based manufacturing system, modifications made to the production batch record during normal operation are traced manually as indicated by the shaded items in Figure 1.6, while the master batch record and SOPs are likely controlled via a separate (document management) system. Note that changes to approval status are traceable within the record, and that destruction of the approved production batch record at the end of the retention period is also traceable through entries in an indexing system.

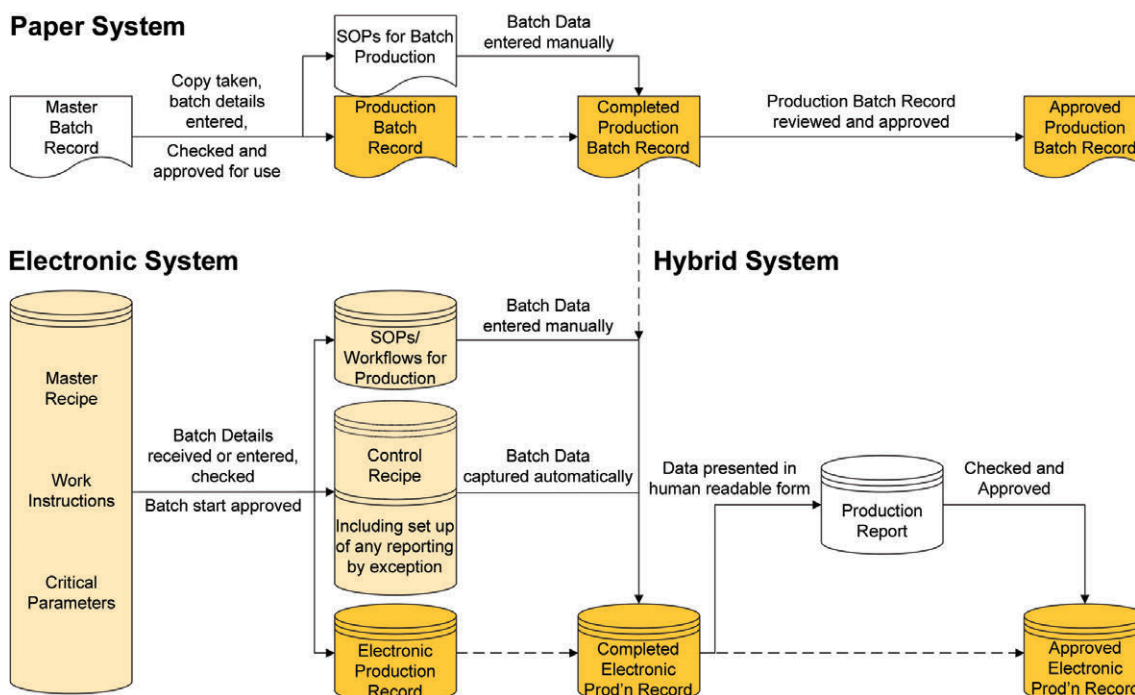
In the equivalent electronic manufacturing system, the data audit trail captures modifications that are part of normal operations. It also tracks activities that change the status of a record (e.g., applying an electronic approval signature) and to deletion of the production record at the end of the retention period.

However, electronic systems need to control and consider data audit trail requirements for the following situations:

- Many MES allow for the master recipe and associated work instructions and CPPs to be controlled with their own lifecycle (versioning, review, and approval workflows, etc.)

- Workflows for production are derived from master work instructions and are likely to be read-only (the equivalent of SOPs). A production process that deviates from the workflow is captured as an exception in the production record. If workflows are supported by lower level workflow configurations guiding an operator through individual processes, they are likely controlled through the configuration management.
- Control recipes include a number of elements:
 - The “recipe” itself derived from the master recipe (typically read-only with no manual inputs; deviations are captured as exceptions in the production record)
 - Lower level configurations that automatically control unit operations, phases, and reporting by exception, usually controlled under configuration management
 - There may be recipe set points and/or settings involved in reporting by exception that need to be manually transferred to a lower level PCS. Where this is the case, consideration should be given to how this process is captured. Thresholds used in reporting by exception can present a particularly high risk to data integrity as the results presented in the production report are relied upon for GMP decision making. There is a strong argument that changes to these values should be included within the data audit trail so that they can be reviewed as part of the release process.
- The validated production report is typically a read-only human readable presentation of data from the electronic production record with no manual input.⁵
- A documented risk assessment should be used as the basis for determining audit trail requirements and controls. See Chapter 4 for additional detail on the risk management process.

Figure 1.6: Data Audit Trail Example for Production Records⁶



⁵ In a hybrid system, manual interventions may be captured onto supplemental paper records.

⁶ An audit trail in a paper-based system may be provided in part by document management and change control processes.

It is recognized that manufacturing systems use many different terms including “audit trail” and “event log”, and that they may save any or all of the following within their storage locations:

- Data that is part of a GMP record in its own right (e.g., capture of an operator data entry or confirmation along with operator details and time stamp)
- Activities that affect GMP records (e.g., a correction to a value previously entered into the electronic production record, along with operator details, explanation, and time stamp)
- Technical system logs of non-GMP activities (e.g., login activity or a change to a non-GMP setting – note these may also be captured with operator details, explanation, and time stamp)

It is important to identify which items from such storage locations are part of the GMP record or are part of the data audit trail, as these will need to be presented within the production report during the review and release process.

2 Data and Records

2.1 What is Special about Manufacturing Systems?

The *ISPE GAMP® Good Practice Guide: A Risk-Based Approach to GxP Process Control Systems* [5], *Manufacturing Execution Systems – A Strategic and Program Management Approach* [6], and *A Risk-Based Approach to Testing of GxP Systems* [20] identified a number of characteristics that are special to manufacturing systems; see Table 2.1. These “special” attributes provide a useful framework for looking at data integrity in these systems.

Table 2.1: Characteristics Specific to Manufacturing Systems and their Impact on Data Integrity

Special Attribute of a Manufacturing System	Impact on Data Integrity
<p>Interfaces to a physical process...</p> <p>...being controlled in real time</p> <p>[5, Section 3.2]</p>	<ul style="list-style-type: none"> Sensors generate enormous amounts of data that generally has hysteresis applied/is sampled/is filtered prior to being stored as part of a required record. This processing is typically carried out in a predetermined and validated way without manual intervention. Decisions about what data needs to be stored (e.g., sample rates, hysteresis, filtering, different sample regimes when an excursion condition is present, requirements for recording maximum excursion point) should be based on a documented risk assessment. <ul style="list-style-type: none"> Note that not all data sampled for control purposes (which can be at a relatively high frequency in the order of milliseconds or seconds) needs to be considered for data storage. See the discussion and example in Section 3.3.1. Data from sensors can introduce a degree of uncertainty in the measurements (e.g., due to sensor capabilities, calibration errors, noise) that can be mitigated through appropriate specification setting, good installation practices, calibration, and filtering or smoothing of signals to ensure that required tolerances are met. The process model (see Figure 3.1) may also process sampled data prior to being stored as part of a required record. Sampled data that is essential to the reconstruction of the process should be kept. <ul style="list-style-type: none"> When the required record contains all data necessary for the reconstruction of the process, the decision whether to keep the sampled data may be based on a risk assessment. This assessment should include the risk to patient safety, product quality, as well as data integrity factors, such as the degree to which the processing is predetermined, validated, and without manual intervention. It is important to clearly differentiate between sampled data that comes directly from the process for control purposes, and measured/processed data retained for quality purposes. The process model may supply a summary of calculated data into required records (e.g., an overall pass/fail for a processing step or a statement that there have been no excursion alarms within the period of the batch). Retention of sampled data needs to be subject to the same considerations as the bullet point above. See Section 3.3.2.2 for additional discussion on summary data and alarms. Maintaining the synchronicity of data can be extremely important⁷, as well as maintaining clock synchronicity in different parts of the system and in different systems, and preventing access to the time stamp. Networking of computerized systems, synchronized to the same reference time source helps ensure that time stamps are correct across the data generated and stored.

⁷ The time stamps of recorded values need to be included with the “meaning” of that data and transferred correctly.

Table 2.1: Characteristics Specific to Manufacturing Systems and their Impact on Data Integrity (continued)

Special Attribute of a Manufacturing System	Impact on Data Integrity
<p>Interfaces to a physical process...</p> <p>...being controlled in real time</p> <p>[5, Section 3.2]</p> <p>(continued)</p>	<ul style="list-style-type: none"> When quality control testing is part of the release process, product release may take place sometime after the end of batch processing. It is important that signatures and data audit trails generated during the release process are associated with the correct production record. The data or records may be recorded within the manufacturing system or within a separate (computerized or paper-based) system. The real-time nature of control combined with the likelihood of failures of sensors, measuring equipment, control equipment, and actuators lead to the use of redundancy. This affects how primary records are identified and kept current.
<p>Multiple components and multiple suppliers...</p> <p>...supported by multidisciplinary teams</p> <p>[5, Section 3.1]</p>	<ul style="list-style-type: none"> There tend to be many interfaces within and between manufacturing systems. These interfaces create challenges to ensuring that the “meaning” of data (including aspects such as synchronicity, data quality, and data audit trail of manual actions) is transferred alongside the data. Hybrid systems without the capability to electronically record/transfer all data and metadata to other systems or production records may require printouts as part of the record to ensure complete data. Some suppliers may not have an appreciation of the importance of complete and accurate records, that is, data integrity. Supplier management becomes more important when the supplier maintains required records for the sponsor company. In addition to required records, manufacturing systems often provide large amounts of data for other purposes, such as maintenance, system monitoring, and process equipment monitoring. It is important to ensure that data that forms part of required records (i.e., regulated data) is adequately managed and protected to maintain data integrity. Risk management may help guide the amount of rigor applied to those records. Segregation of duties can be complex, for example, denying the operational team access to change validated recipe parameters, while allowing them access to parameters required to operate the system and equipment. Organizations should avoid conflicts of interest by careful design of the way data flows through and between manufacturing systems. When many groups are involved in system design and operation, accommodating the different uses of the same data while maintaining the integrity of the data used for regulated purposes can be challenging (e.g., data used by QA to support product release, by production operations to calculate equipment utilization, and by the maintenance department to plan routine maintenance).
<p>Interfaces to and degree of embedding with process equipment...</p> <p>... which may also be required to control a multipurpose plant</p> <p>[5, Section 3.3])</p>	<ul style="list-style-type: none"> Embedded control systems that are delivered as a standard part of packaged systems may lack flexibility in the format of data and metadata offered for interfaces. Metadata may need to be reapplied as data is transferred to the MES to become part of the production record, and this will need to be covered in the validation of the interface. There may be restricted facilities for data integrity controls relating to access control, logging, and audit trailing, etc., available from embedded control systems delivered as a standard part of a particular item of process equipment. This could be overcome by sending data to an historian and slaving control from a PCS with more functionality. Process equipment may be multipurpose; with different data integrity control requirements depending on its purpose, or with different owners for data from different processes or different time periods. Non batch-related use may be captured in an electronic logbook system to enable electronic review of data processes.

Table 2.1: Characteristics Specific to Manufacturing Systems and their Impact on Data Integrity (continued)

Special Attribute of a Manufacturing System	Impact on Data Integrity
<p>Systems which cross many functional domains...</p> <p>... Functional domains which cross many systems...</p> <p>.... Often with systems and functions installed in a phased manner</p> <p>[6, Section 1.6.2 and Appendix 3])</p> <p>[5, Section 3.3]</p>	<ul style="list-style-type: none"> • Systems may include data sets with differing degrees of GMP impact, making it important to identify data that is required for regulated purposes and ensure that appropriate controls are applied. • Data sets may be passed across multiple interfaces between systems resulting in challenges to maintaining consistency and synchronicity, and to ensuring data is complete and always accompanied by the relevant metadata. • It may be necessary to assess risks to data integrity resulting from intermediate stages of implementation where some phases of the final data flow are present. For example, a processing vessel where the control system has been upgraded to pass data directly to an historian within the MES, but the dispensing system has not yet been upgraded, so records are still printed to paper copies.

2.2 What is Regulated Data?

In assessing the risks associated with data, it is important to understand how the data is used. Data may be:

- **Regulated:** used for a regulated decision or to support a regulated process, that is, data as required by, or in support of, the predicate rules – what we have to keep
- **Operational:** non-regulated data used for business process decisions such as performance analysis and maintenance scheduling – what we want to keep
- **Unnecessary:** data not needed due to either the circumstances of its creation (e.g., during a non-regulated activity or process) or because that data does not provide additional context, metadata, or meaning for the activity or process – what we do not need

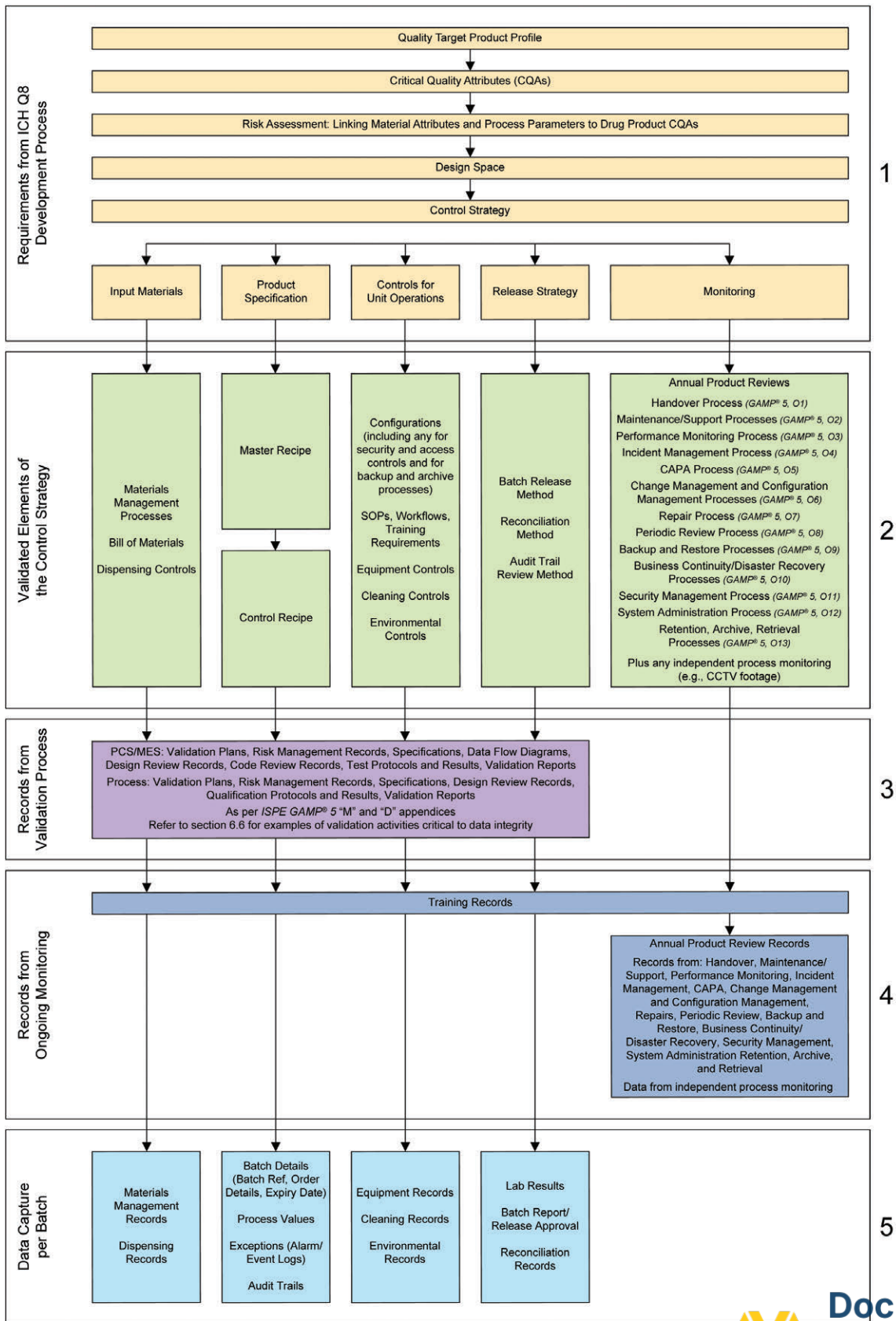
Within regulated data, it is recommended to consider the data risk⁸ and the data criticality⁹ in order to determine the appropriateness of the data integrity controls applied.

Regulated data within manufacturing systems is required to demonstrate the quality of the end product and which, in a well understood process, should be traceable back to the Critical Quality Attributes (CQAs) of the drug product and the CPPs that support these [21]. Figure 2.1 illustrates this. The identification of CQAs and CPPs and the derivation of an appropriate control strategy is the subject of ICH Q8(R2) [21].

⁸ Data risk: opportunity for data alteration and deletion, and likelihood of detection/visibility of changes by the manufacturer's routine review processes [9].

⁹ Data criticality: impact to decision making and product quality [9].

Figure 2.1: Regulated Data in Manufacturing Systems



The records demonstrating that the required control strategy is in place and effective can be split into four groups:

- Records that form part of the control strategy itself (shown as Box 2 in Figure 2.1)
- Records demonstrating successful initial validation of the entire control strategy (shown as Box 3 in Figure 2.1)
- Records from ongoing monitoring demonstrating that the validated state is being maintained (shown as Box 4 in Figure 2.1)
- Records captured for every batch demonstrating that the control strategy kept the CPPs within the required limits and hence delivered good quality product (shown as Box 5 in Figure 2.1)

The records that prove successful validation of the entire control strategy are essential to assure product quality. These are documents or configurations that should be managed through established systems. (Refer to *ISPE GAMP® 5* [15] and to the *ISPE GAMP® Good Practice Guide: A Risk-Based Approach to GxP Process Control Systems* [5], and *ISPE GAMP® Good Practice Guide: Manufacturing Execution Systems – A Strategic and Program Management Approach* [6] for additional details.)

Validation activities to demonstrate that data integrity controls operate as intended are addressed in Section 6.6. The ongoing maintenance of the validated state should be managed through established systems. (Refer to the *ISPE GAMP® Good Practice Guide: A Risk-Based Approach to Operation of GxP Computerized Systems* [22] for further information.) Additional review activities to ensure that data integrity is maintained are discussed in Section 5.1.

The remainder of this Guide mainly focuses on the data captured for every batch demonstrating that the control strategy kept the CPPs within the required limits.

3 Data Flows and Data Lifecycle

3.1 Mapping Data Flows

The mapping of data flows is a prerequisite to understanding the potential threats to data integrity. The output from this activity will be used to define user requirements and to develop data integrity risk management. (Refer to Chapter 4 for information on risk management approaches.)

Within the manufacturing execution layer, data flows can be extremely complex, necessitating a full process mapping exercise to identify the data flows required to support each business process.

When creating process maps, the business analysis team should interview the current operational staff. The subsequent process steps and system workflow mapping sessions should include data specialists from the business to ensure they understand the data challenges and exchanges through multiple platforms within the future system architecture. Where data flows result in duplication of data sets, it is important to document where the primary records reside.

The process mapping exercise should define the use of the data within the business processes. Data flow diagrams can then be created to map data flows into, through, and out of individual manufacturing systems. Performing this mapping along with the definition of the process will ensure that all data requirements are fully understood and integrated into the system design.

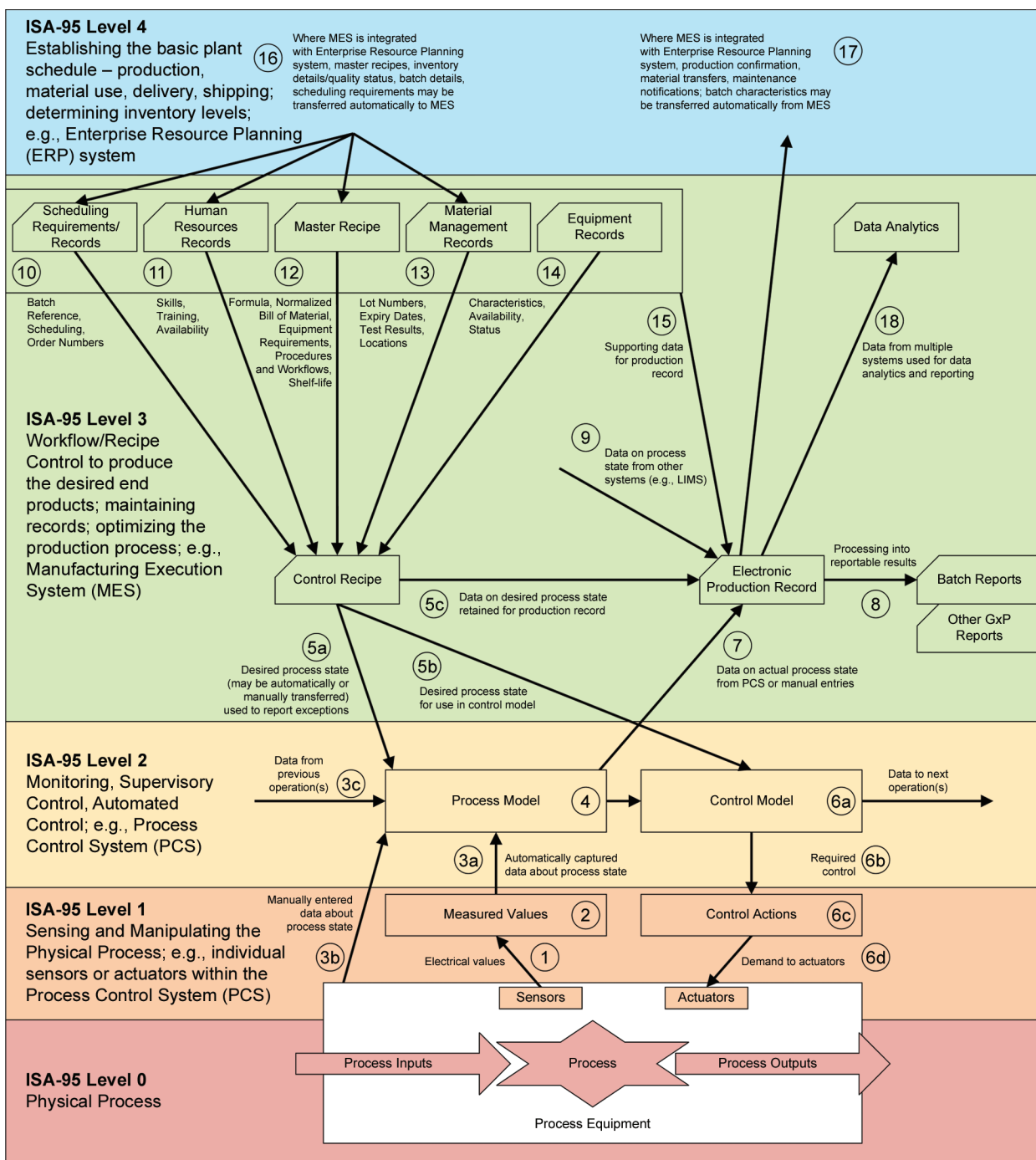
Establishment of the data flows will also confirm that links are correctly established between Level 4, Level 3, Level 2, and reporting systems. In addition, determining the data requirements can help support master data management processes within an organization. This may lead to potential efficiency gains with regard to selecting standard data models and transfer processes.

3.2 Typical GMP Data Flows

Generically, GMP data flows within manufacturing systems can be summarized as in Figure 3.1. Note that many modern PCS include elements of Level 3 (recipe management and data historians), and that many different systems are often part of each MES functional domain (see Figure 1.4); therefore, the focus should be on the data flows themselves rather than the exact system within which they reside.

For example, GMP materials can be defined in one system (ERP), released for use in another (LIMS), and consumed in another (PCS or MES). The risk to data integrity within the materials management functional domain is determined by looking across all of these systems.

Figure 3.1: GMP Data Flows in Manufacturing Systems



This Guide is concerned with Levels 1 to 3 as shown in Figure 3.1, and does not replicate areas common to other computerized systems (e.g., infrastructure and network concerns; what happens to records transferred outside of the system; or how and where records should be stored long term).

Typical data flows within manufacturing systems (identified by the numbers given in Figure 3.1) are as shown in Table 3.1.

Table 3.1: List of Typical Data Flows in Manufacturing Systems

Reference (See Figure 3.1)	Generic Manufacturing System Data Flow
1	<p>Physical quantity converted to electrical signal by sensors and transmitters and passed into system.</p> <p>These sensors provide measured values that are related to attributes of the process inputs, the process itself, or the process outputs.</p> <p>For example, weigh cells provide ingredient quantities, thermocouples measure the temperature, and near IR sensors detect the moisture content of the mixture being passed to tableting.</p>
2	Electrical signal processed into measured value (e.g., applying range, linearization, and calibration settings).
3	<p>Passing measured value(s) into process model (e.g., via a communications link):</p> <ol style="list-style-type: none"> Values collected automatically by sensors Process values entered manually by an operator (e.g., ingredient weights) or data recorded by a scribe¹⁰ witnessing an operation (e.g., in aseptic processing). Data passed from previous operations to next operations (e.g., previous clean in place result, or details of continuous liquid addition rates to allow for corrections to pH after a time delay during which a particular reaction takes place)
4	<p>A process model maps these measured values onto those attributes of the process or equipment to be controlled.</p> <p>(This might be a very simple model, such as:</p> <ul style="list-style-type: none"> The measured value from the thermocouple is representative of the temperature of liquid in the reactor. The signal from the speed sensor is proportional to the speed of the conveyor carrying medical device test strips through the dryer. <p>It can also be a complex multivariate or statistical model that relates what is measurable to the variables of interest through complex calculations or statistical models, and anything in between.)</p>
5	Passing in data about desired process state to create records that summarize whether the process is in the desired state (e.g., deviation alarms, pass/fail entries on reports).
6	<p>Deciding on the required control and passing signals to actuators to perform the required activity:</p> <ol style="list-style-type: none"> A control model compares the actual process or equipment state to the desired state and decides what control actions are needed to create or maintain the desired conditions. (This can be a very simple model such as: “The temperature is too low so turn the heater on.” or “The speed is too high so reduce the output to the motor.” It can also be achieved through very complex multivariate control algorithms.) The control model passes this required action to the elements of the system that control individual actuators, or in some cases makes a request for manual intervention, such as addition of an ingredient or removal of a failed item. The required position and state for individual actuators is determined and turned into an electrical signal. The actuators make the required adjustment to the process.

¹⁰ WHO TRS 996 Appendix 1 contains the requirements for compliant use of scribes where needed [7].

Table 3.1: List of Typical Data Flows in Manufacturing Systems (continued)

Reference (See Figure 3.1)	Generic Manufacturing System Data Flow
7	<p>Data from the process control layer is transferred to the electronic production record for use in production reports or other records (e.g., maintenance record). This may include data directly collected from interfaced instruments, such as weigh scales and bar code readers, as well as data acquired and transmitted by the process control layer.</p> <p>This record incorporates the batch details, material flows, the manufacturing process flow, equipment flow, verification of the data captured, and operator activities. The data captured allows confirmation of the right materials with valid status, the correct quantities of the material, correct calculations, equipment with the right status, etc.¹¹</p>
8	Data is processed into reportable results (e.g., generation of final production report) and used in GMP activities such as the review and release of product based on the reported CPPs and exceptions.
9	Data from other systems (e.g., LIMS) is gathered by the manufacturing system.
10	<p>Scheduling requirements created within the manufacturing system or read from a Level 4 system (e.g., ERP) are used to generate control recipes and identify data for a specific batch.</p> <p>The scheduling functional domain is typically the source of supporting data that permits traceability from batch data in the electronic production record to order references and shipment details in Level 4 systems.</p>
11	<p>Human resources records created within the manufacturing system or read from a Level 4 system (e.g., ERP) with details about operator skills, training, and availability are used in determining where a specific batch will be made and by whom.</p> <p>The human resources functional domain is typically the source of supporting data regarding the competence of operations staff involved in the production of a batch.</p>
12	<p>Master recipes created within the manufacturing system or read from a Level 4 system (e.g., ERP) and used to provide information on material characteristics and formulae, procedures and workflows, required equipment characteristics, etc., needed to generate a control recipe for a specific batch.</p> <p>Where the master recipe for a product is generated, reviewed, and approved in the manufacturing system, version control of the master recipe and control over which version is current are provided from within the recipe management functional domain.</p> <p>The recipe management functional domain creates the control recipe for a specific batch. This provides data on the desired process state that is transferred both to the electronic production record and to the process control layer. This allows the process control layer to generate appropriate control responses and flag exceptions or deviations for operator action; it allows the processing of batch data into reportable results that clearly identify exceptions or deviations.</p>
13	<p>Material management records created within the manufacturing system or read from a Level 4 system (e.g., ERP) are used to determine which ingredients should be sourced from which location as part of the control recipe for a specific batch. These records are likely to include material source and quality data, material usage, material inventory, and detail of rejected materials.</p> <p>The materials management functional domain is typically the source of supporting data regarding the suitability of the materials used in the manufacture of a specific batch.</p>

¹¹ Where data from multiple systems at the process control level are passed to the same electronic production record, a data references dictionary may be needed to ensure that values from all systems are mapped to a consistent format in the MES level.

Table 3.1: List of Typical Data Flows in Manufacturing Systems (continued)

Reference (See Figure 3.1)	Generic Manufacturing System Data Flow
14	<p>Equipment and asset management records normally reside at the manufacturing execution level – the availability and status information may be passed automatically from the process control level or may be manually entered. These records are likely to include equipment specification details (e.g., vessel type, vessel capacity), equipment calibration status, and equipment cleaning records.</p> <p>The equipment management functional domain is typically the source of supporting data regarding the suitability of the equipment used in the manufacture of a specific batch.</p>
15	<p>Depending on the degree of integration of the various MES functional domains, supporting data for a particular batch may be transferred to form part of a separate electronic production record, or may be retained within the original domain from where it is queried as data is processed into reportable results, such as a production report.</p> <p>Data may also be processed within the MES to confirm that the required workflows have been followed and to generate exceptions if not automatically recorded (e.g., because a manual entry was made in place of a barcode scan; a manual correction was made to a status or data point or target value; a manual intervention occurred during processing).</p>
16	<p>Where MES is integrated with ERP systems, master recipes, inventory details and quality status, batch details, scheduling requirements, etc., may be transferred automatically to the MES via a standard or customized interface.</p>
17	<p>Where MES is integrated with ERP systems, production confirmation, material transfers, maintenance notifications, and batch characteristics may be transferred automatically from MES for use in other functions.</p>
18	<p>Data from multiple systems is transferred for use in data analytics to allow analysis of, for example, yield and quality metrics against raw materials used.¹²</p> <p>Many outputs from the data analytics domain may be production management or maintenance related (operational data) rather than regulated data, but others, such as those used in periodic or annual product reviews, have GMP impact.</p>

Additional details on the types of data that make up these flows and the resulting risks to data integrity are included in Appendices 1 and 2.

3.3 Data Lifecycle

Some example activities within the data lifecycle associated with a manufacturing system production record are shown in Figure 3.2. Different types of data, their risks, and retention approaches are discussed in Appendix 2.

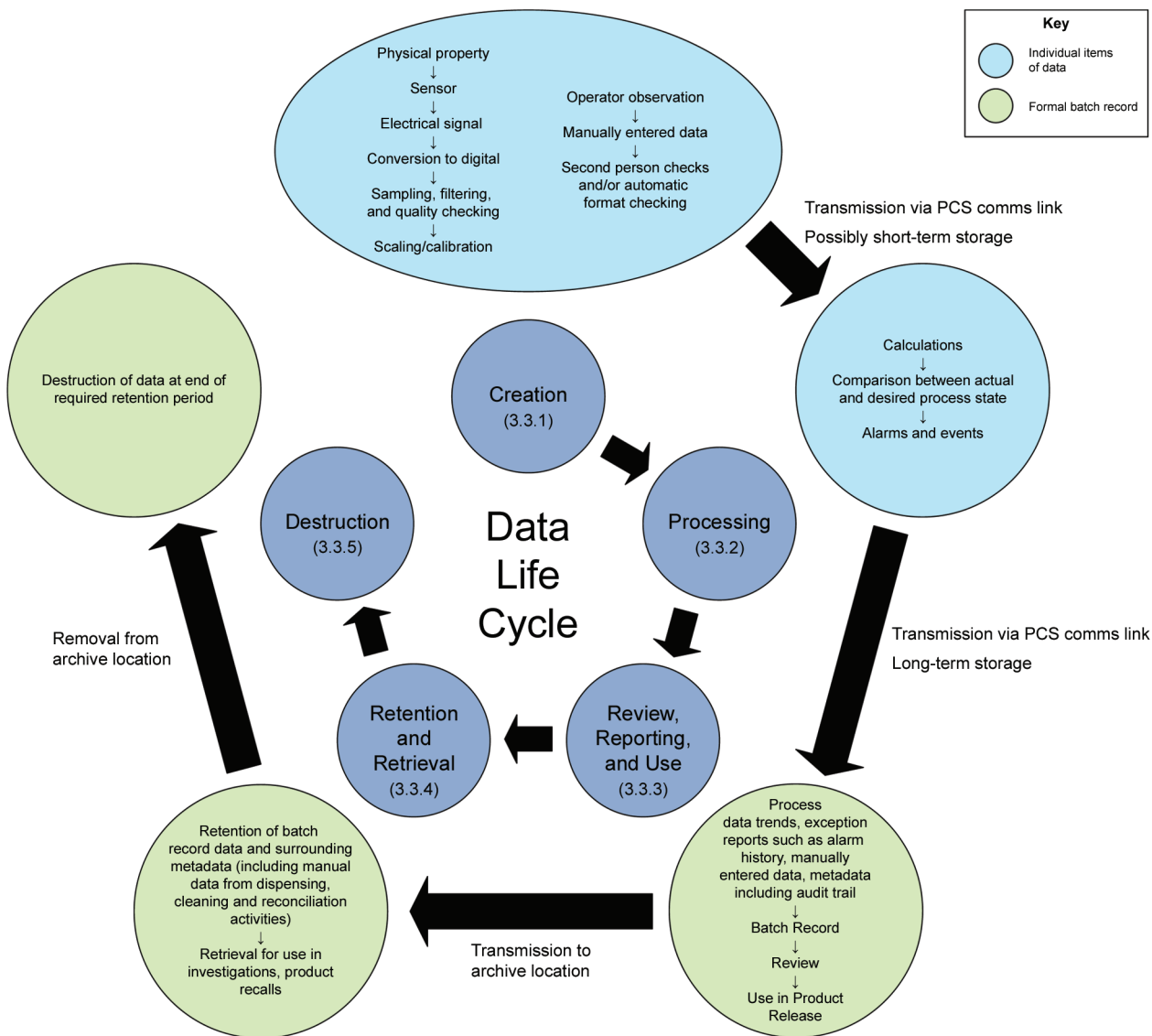
GMP regulations and guidance allow for flexible and risk-based strategies for record and data retention. Firms should implement meaningful and effective strategies to manage their data based on their process and product understanding, and knowledge of technologies and business models, and based on a justified and documented risk assessment.

¹² Since data sets from multiple systems are passed to a relational database, a data reference dictionary is needed to ensure that values from all systems are mapped to a uniform format in the MES level and that all data is labeled with the appropriate site, facility, material name, and batch number metadata.

Such retention approaches should ensure that:

- Critical GMP content and meaning are preserved
- Applicable specific predicate requirements are met
- The regulated company's ability to meet statutory obligations is maintained

Figure 3.2: Example Data Lifecycle Activities of a Manufacturing System Production Record



The following points are important to note:

- Record definition and linkage

The required regulatory records need to be defined and linked back to the supporting product CQAs and the CPPs. This allows the rigor of controls around the data in these records to be based upon risk to product quality, patient safety, and the integrity of GMP data.

- GMP usage of records

Individual items of data collected by a manufacturing system may or may not form part of these required records. (For example, some data not part of the production record may be important to retain for process performance analysis.) Some data may be used for multiple purposes, not all of which are regulated activities. Understanding the flow of data through the system to the point of use allows the rigor of controls around the data to be based upon the risk to data integrity.

- Changes in ownership of data

Ownership of the data may change during its lifecycle. A copy of the same data may also be owned by the plant maintenance manager when used for maintenance or calibration purposes.

This point can be particularly important in contract manufacturing situations, where the data owner may be someone in the customer's organization and the data steward may be someone in the contract manufacturing organization. This is further complicated when a system contains data for multiple customers with multiple data owners.

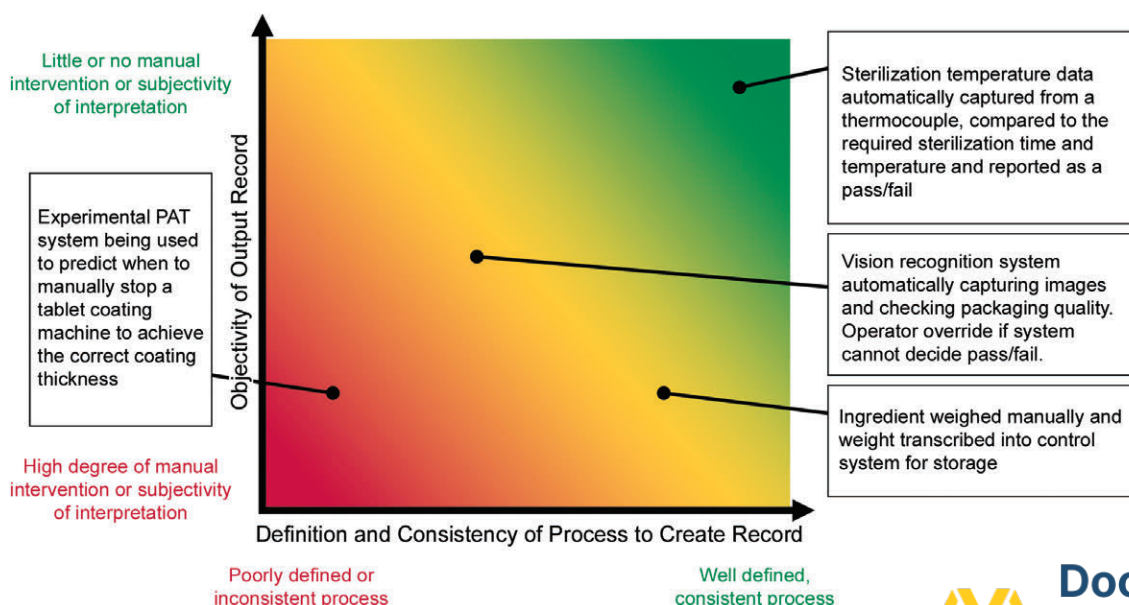
- Degree of manual intervention and subjectivity

The risk to the integrity of a set of data as it undergoes a particular process depends on how well defined and consistent that process is, and on how much manual intervention and subjectivity of judgement is involved in creating the output of the process [11].

Where a process is well defined ("we know exactly how to do this") and consistent ("if we do it like this, we always end up with the correct result"), has no manual intervention ("it all happens automatically") and an objective output ("we all agree on the result"), data integrity controls can be achieved by validating the system and maintaining it in a validated state. Figure 3.3 illustrates this concept. It is discussed in more detail in Appendix 2.

Where any of these criteria are not met, or a possibility of repeated runs or excluded data exists, there is likely (depending on the criticality of the data to patient safety and product quality) a need for additional risk mitigation activities, such as audit trailing and reviewing operator activities, or retaining raw values.

Figure 3.3: Factors Influencing the Risk to Data Integrity



3.3.1 Data Lifecycle – Creation

A key consideration in a manufacturing system is data capture and what data (and associated metadata) must be retained. The environment in which records are created in a manufacturing system has the following characteristics:

- The majority of data is process values captured automatically by physical sensors.
- The definition of the desired process state (as defined in the master recipe for the process) may occur through:
 - Fixed settings (verified during original system validation)
 - Recipe values downloaded automatically
 - Manually entered values
- Some data relating to the actual process state is also entered manually, particularly relating to manual interactions with the process such as ingredient additions, or semiautomatically (e.g., via barcode scanners).
- Batch identification information may be entered manually, semiautomatically, or through an automatic transfer from an ERP system.

For automatically captured data, often the initially acquired values (considered as the initial digital data prior to any post acquisition scaling, linearization, or further processing) are not considered original data and do not need to be retained. This is because there is usually no manual processing of manufacturing data equivalent to that performed on laboratory data which potentially requires the acquired data to be “reprocessed” in the case of an Out of Specification (OOS) result or other investigation. This relies on the systems being sufficiently secure to prevent unauthorized access to the initially acquired data or the configuration of scaling, linearization, etc., parameters. (This is usually the case as such software is often stored and executed in inaccessible firmware.)

Data Creation Example

The temperature in a particular vessel is a CPP. A temperature exceeding 36°C for more than 30 s can cause product degradation that could cause patient harm.

The temperature is initially captured in the input/output subsystem of the PCS as an electrical value every 20 ms. This is filtered to remove noise, linearized and ranged, and any calibration adjustment added, then passed into the process control subsystem every 100 ms. (PCS typically acquire values at a rate in ms to allow control algorithms, safety interlocks, etc., to function correctly).

The dynamics of the process are such that the vessel temperature changes 3°C per minute under maximum heating or cooling conditions.

During process development, consideration should be given to:

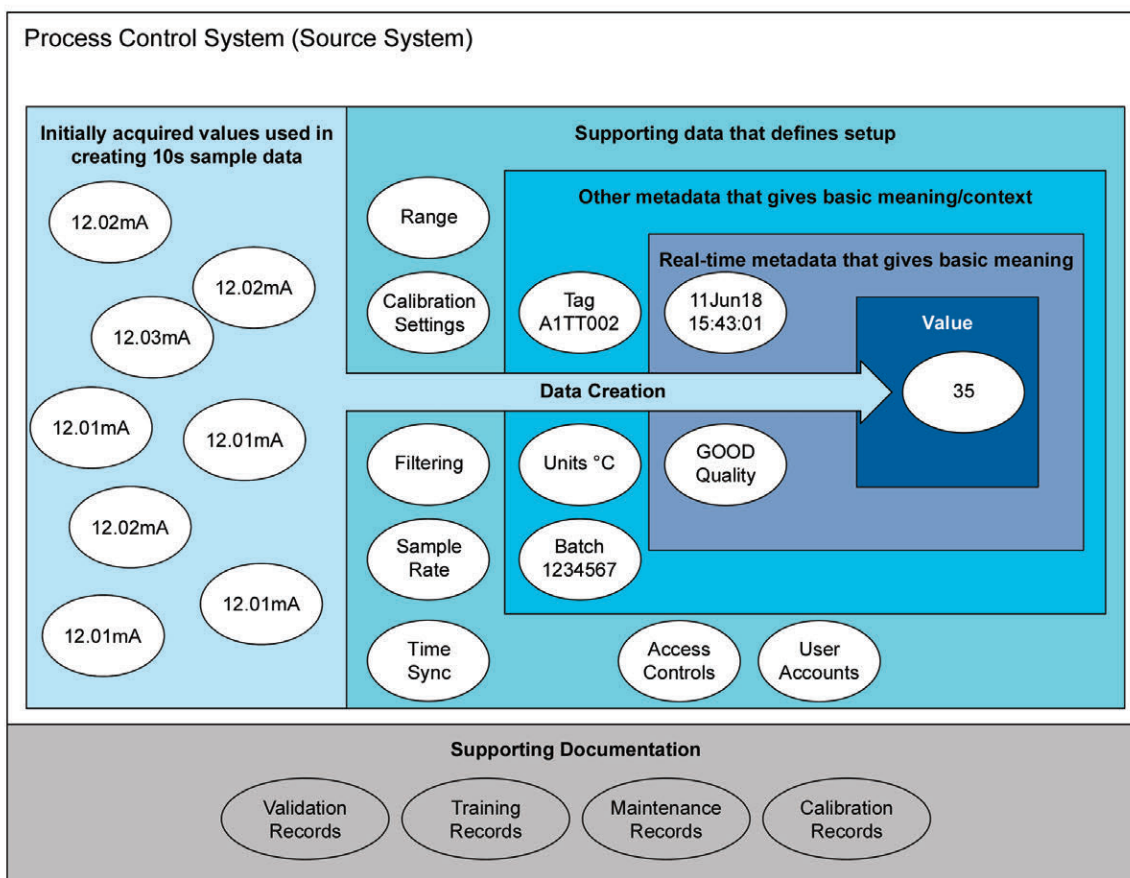
- The criticality to patient safety: *this signal impacts product quality – we need to record it frequently enough to be sure that it is within specification*
- The process dynamics: *if we sample every 10 s, the worst-case excursion we would not know about is a 10 s period where the temperature moved from 36°C to 36.25°C and returned to 36°C*
- The degree of manual intervention involved: *there is none – this is a completely automatic process and the parameters such as range, linearization, and calibration adjustment are suitably locked down*

The conclusion is that a 10 s sample rate for the stored data is suitable. The 10 s samples are committed to long-term storage and are treated as the “original” data for the temperature value.

Note that these 10 s values are supported by various items of metadata (shown in Figure 3.4):

- Basic meaning data, such as quality status and time stamp (items that change in real time and are normally stored alongside the value) as well as units, tag number, batch reference number (items that may be stored alongside the value or stored in a way that allows them to be cross referenced and retrieved along with the value).
- Other supporting data contained in system configurations such as range, sample rates, filtering, calibration settings, etc. (These are not normally modified during normal operation and are typically controlled via change control and configuration management procedures.)

Figure 3.4: Example of Data Creation



There are, however, cases where acquired values may need to be considered and retained, for example:

- Where there is insufficient security to prevent unauthorized access to the initially acquired data and the configuration of scaling, linearization, etc. parameters

In these cases, the initially acquired data may be stored directly to the historian to demonstrate that the data in the PCS has not been accidentally changed or deliberately altered to produce falsified data values.

- Where acquired data is subject to manual processing

While not usually the case for process equipment (highly automated with minimal or no manual intervention), it may be necessary if there is a requirement to manually process analytical data acquired directly from process instruments (e.g., manual intervention on an NIR spectrum used for Process Analytical Technology (PAT) on drying endpoints in granulation).

- To demonstrate during an investigation that the calculated value has been correctly derived from the data

Whereas standard algorithms such as those used in PID controllers can be considered mature, for novel control or filtering algorithms, it could be useful to retain the initially acquired data to demonstrate the robustness of the control software in the case of an OOS investigation.

For manually entered data, the initially acquired values together with metadata are considered part of the original record and must be retained [9].

3.3.2 Data Lifecycle – Processing

3.3.2.1 Transmission of Data

Where ever possible, manual transcription of data should be eliminated and be replaced with a validated interface providing automated transmission of data and associated and relevant metadata [9].

The environment in which data is migrated/transmitted in a manufacturing system has the following characteristics:

- Many PCS receive data from other systems either through standard interfaces (e.g., Fieldbus, TCP/IP) or custom interfaces (e.g., custom protocol convertors to allow analytical instruments to communicate with the PCS or MES).
- Data may be transmitted with its associated metadata, or the metadata may need to be reapplied in the receiving system, necessitating validation that this has been correctly configured.
- Data and metadata may undergo translation to local language; in this case, validation activities will be needed to ensure that:
 - The original meaning is correctly maintained
 - Translated expressions are not affected by truncation
 - Time and date formats are correct and data from different time zones correctly handled
 - Conversion of units of measurement (e.g., kPa to bar) is correct

This is likely to affect recipes, workflows, labels, and other system data outputs, as well as process data and production reports.

- It may or may not be technically possible for the data audit trail to be transmitted alongside the associated data.

Where possible, it is useful to have data audit trail information accompany the data for ease of review and to ensure that it is not separated from the record to which it applies.

Where the data audit trail continues to reside in the source system, it is important to determine how to include this information in review activities, how to achieve traceability between data audit trail entries and final records, and how to protect and archive the data audit trail.

- Specific risks to data integrity include corruption of data due to high ambient electrical noise or electromagnetic interference found in manufacturing areas, or security issues associated with the use of wireless networks, for example, in warehouses.

It is useful to specifically highlight data transfer (which is essentially a copying function from one networked device to another) as part of the data flow with a high risk to data integrity.

Qualification of such infrastructure and validation of the end to end transmission of the data should be sufficient to mitigate these risks. Challenge testing may be applicable, perhaps by testing the maximum reliable range of network devices, or testing data transmission under conditions of maximum electrical load when data and power cables cannot be separated.

- Many MES allow for interfaces to other systems that produce complex data sets (e.g., LIMS data if the final review of batch release is performed in the MES). The data is frequently transmitted as a copy of the summary results with the primary record (including the complete original data set) remaining on the originating system. In some cases, it is possible to transmit the entire record for storage as part of the electronic production record. Process mapping and data flows should identify where the data is used and for what purpose.
- Because a functional domain within a MES can be split across many systems, with data transferred between them, these systems can be particularly vulnerable to network outages, and infrastructure may need to be designed to minimize the risk to interruptions in real-time operations.

Business continuity procedures should be in place for managing data integrity during unexpected interruptions and subsequent recovery (e.g., a network outage may require data in a production recipe operation to be updated for new material status not transferred during the outage) and for managing data integrity during planned infrastructure downtime.

- Interfaces over which data is transmitted often offer the ability to report their status via alarms, and may have functionality to allow manual or semiautomatic recovery of missing data.

Where GMP data is transmitted, these functions should be used, and procedures in place both for responding to data transmission alarms and managing the recovery of missing data and its integrity.

- GMP variable data such as batch identification, material specification, status and availability, and recipe information may be downloaded to the manufacturing system from an external system (e.g., ERP).

Data such as the creation of inventory from process operations may also be interfaced to an ERP. These interfaces are often based on ready-made adapters, although sometimes they must be custom developed.

- Data may be transcribed manually from one system to another for use in a subsequent manufacturing process. (An example is transcribing weight and lot number information from a separate ingredient preparation station into the PCS.)
- Data may be output from a manufacturing system for inclusion with the product (e.g., data printed directly on the product or packaging, or accompanying literature, such as batch, lot or serial number; product name, formulation or code; date of manufacture; expiry date), or to prompt operator actions such as removal of reject product.

In this case the data “transmission” may be via an automated process (e.g., communication link to a labeling machine) or a manual process (e.g., manual removal of rejected product and its subsequent reconciliation).

- Existing records, for example, master recipes and equipment records, may be migrated for use in new manufacturing systems. Migration is a form of transmission in which the data is not only sent to a receiving system, but also changes format to suit a new system. This migration process may involve elements of processing in order to extract, cleanse, transform or enrich the data before it is loaded into the new system.

- Data within manufacturing systems is not typically encrypted. It does not include confidential personal data, and business confidentiality is normally addressed via system and infrastructure security.

However, where encryption is required, consider the following questions:

- What data requires encryption?
- When is encryption required, e.g., at rest, during processing, both?
- Is encryption of data required when transferring to and from other systems?
- How will the encryption be managed?

Data Transmission Example

The temperature in a particular vessel is a CPP. Temperature data created as in the Data Creation Example in Section 3.3.1 needs to be transmitted from a PCS to an MES.

In order for the 10 s sample values of the temperature to retain their meaning and context in the target system, there are some elements of metadata that must either be transmitted alongside the value or be reapplied at the target system. There are other elements of supporting data that may be transmitted to the target system or may remain in the source system. See Figure 3.5 and Table 3.2 for details on the different elements of metadata.

Figure 3.5: Example of Data Transmission

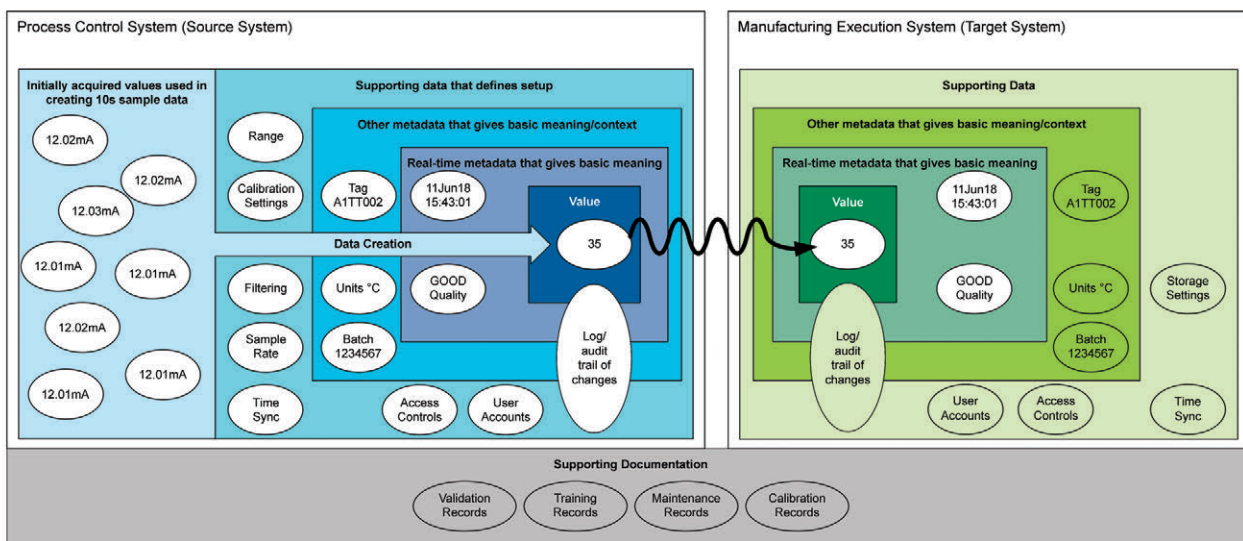


Table 3.2: Example of Data Transmission

Type of Metadata or Supporting Data	Needs to be Retained?	Required to Give Basic Meaning and Context?	Typical Decision on Transmission Between Systems	Typical Review Points for Data
Initially acquired values used in creating the 10 s sample data	No	No	As discussed in Section 3.3.1, it is the 10s samples of temperature value that are committed for storage and treated as the original record. The initially acquired values do not need to be retained and do not form part of the metadata that must be transmitted.	Only as part of confirming suitable filtering etc. during initial validation. Initially acquired values are not stored so are not available for review.
Real-time metadata that gives basic meaning	Yes	Yes	The time stamp and quality status of the temperature values are recorded in real time and need to be transmitted alongside the value. (The target system has no means to deduce or reapply these unless they are also transferred.)	Alongside data as part of product release
Other metadata that gives basic meaning or context	Yes	Yes	Metadata such as sensor identification tags, units, batch reference are required to give meaning and context to the value but may be either transmitted alongside the value or reapplied at the target system. Where metadata is transmitted, correct transmission should be validated. Where metadata is reapplied, the process of reapplication should be validated, with particular attention paid to any translation into local language or different units of measurement.	Metadata that is entered per batch or on a regular basis is reviewed as part of product release. Fixed settings are reviewed as part of initial validation, then control of these challenged (e.g., by sample checks, reviewing changes, checking change logs for modifications) during periodic review.
Supporting data	Yes	No	Most supporting data relating to system setup is retained in the source system and never transmitted to the target system. The target system will have its own similar set of supporting data relating to the setup of functions such as access control, time synchronization, and data storage.	Fixed settings reviewed as part of initial validation, then control of these challenged (e.g., by sample checks, reviewing changes, checking change logs for modifications) during periodic review.
Log/audit trail (see Section 1.5.5 for additional discussion of technical system logs versus data audit trails)	Only for items considered part of data audit trail	Only for items considered part of data audit trail	It may or may not be technically possible for logs and data audit trails to be transmitted alongside the associated data. Where possible, it can be useful to have data audit trail information for GMP data accompany the data for ease of review and to ensure that it is not separated from the record to which it applies. Where the data audit trail continues to reside in the source system, it is important to determine how to include this information in review activities, how to achieve traceability between data audit trail entries and final records, and how to protect and archive the data audit trail.	Data audit trail relating to activities that potentially impact GMP data values is reviewed as part of product release Technical system logs relating to activities done under change control, configuration management, or calibration management processes are reviewed as part of periodic review.

Section 4.4 of the *ISPE GAMP® RDI Good Practice Guide: Data Integrity – Key Concepts* [23] discusses interfaces with respect to data integrity risks and appropriate mitigation.

3.3.2.2 Processing Data

The environment in which data is processed in a manufacturing system has the following characteristics:

- The processing of data to create the record is often a predetermined, validated method with few manual interventions.
- Depending upon its use, data processing may be simple (e.g., display of a scaled and linearized signal on an LED display) or complex (a custom multivariate control (MVC) algorithm, vision systems using artificial intelligence).

Simple, standard processing (e.g., using a commercial off-the-shelf product such as a chart recorder) may require little in the way of specific data lifecycle definition and documentation. It is however, important to assure that data from the chart recorder is captured into a secure database for storage.

- Because most physical processes cannot be efficiently controlled by operators constantly monitoring values, PCS generally automate the comparison of the actual process state to the desired process state, and identifying problems through alarms.

Where alarms are relied upon to make GMP decisions, the alarm occurrence and clear messages, together with operator acknowledgements, are often defined to be part of the required record.

Where operators can change the settings of such alarms, a risk to data integrity is introduced that may be mitigated through a combination of controls, such as:

- Limiting access to change the settings (procedural access controls may be implemented as a temporary quick win, but long-term mitigation requires technical control)
- Including in the data audit trail changes that might impact alarms used to flag exceptions related to GMP decision making
- Including the alarm setting and/or the audit trail information in the final production report for confirmation during product release

It is likely that GMP alarms are prioritized differently to system and other process alarms for ease of alarm management. It can be helpful if this separation is also provided in the way alarm history information is presented for review.

- Use of PAT, Statistical Process Control (SPC) or MVC can involve processing complex data sets.

Although the processing is normally automated, the underlying process models may have been built semiautomatically by integrating data from individual statistical experiments. Some systems may also contain adaptive or self-tuning elements. In these cases, more acquired data may need to be retained (compared to non-adaptive models) for additional verification of the self-tuning elements, and additional controls considered.

- As with control functions, data processing functions can be implemented using standard, configured, or custom-coded software.

Where a data processing function is implemented in standard software by a mature supplier¹³, the regulated company should seek to leverage work completed by the supplier in designing and testing the function to ensure it meets specifications, and where applicable, regulations, and in documenting the method of compliance against each clause of those specifications/regulations. Regulated companies can then focus testing efforts on data integrity risks related to the specific business process and custom elements within the system, or systems not provided by mature suppliers.

- Outputs such as labels may be generated that, through a manual or automatic process, become affixed or integral to the finished product.

3.3.3 Data Lifecycle – Review, Reporting, and Use

The environment in which data is reviewed, reported, and used in a manufacturing system has the following characteristics.

3.3.3.1 Use by Operations Staff in Real Time

- Views of the stored data may be created and presented to users as part of a GMP-decision making process during operation; for example, information presented to operators about failures that is used to manually remove and/or reconcile rejected items.
- A decision as to whether displayed values need to be stored to support an investigation of operator actions should be based on risk. Where operators make GMP decisions based on data displayed but not stored long term (e.g., an analogue pressure gauge), a contemporaneous manual record of the reading should be made and the need for a second person verification determined. Long term, replacing such manual readings with electronic readings and automated interfaces would reduce data integrity risk.

3.3.3.2 Use for Product Release and Subsequent Regulated Activities

While many different reports can be created from a manufacturing system, the production report is the essential human readable output from the system. The production report is ideally generated through an approval workflow, that is, as each record from an individual operation is approved, it is collated into the production record, from which the production report is derived. Some equipment may lack the ability for an electronic signature workflow, for example, autoclaves. In such cases, alternative workflows should be evaluated based on risk.

- Views of the stored data are often created as part of product release processes. Refer to Section 5.1 for additional detail on the data and metadata required to be presented for review as part of the release process.
- In most large modern manufacturing systems, data to create these views is stored in a relational database from which it is extracted to form the production report. This may require reports to be configured or custom coded to provide the required views of the stored data on a time period or batch basis.
- In some simple systems, reports may be captured from run time values during a process without the source data being separately stored. The production report thus ends up as the only record of the data that cannot be processed to provide alternative views, filters, or searches.

This is not a problem for product release activities as the production report should contain the required information for that purpose. Companies should consider in their risk-assessment process whether the data needs to be filtered or searched as part of other regulated activities, and appropriate mechanisms put in place. (As an example, if records are retained only as flat file PDF reports and a product recall was required of all batches that used a particular active ingredient lot, could a standard search facility find specific material lot numbers within the report or would an index need to be created manually?)

¹³ See the ISPE GAMP® Good Practice Guide: A Risk-Based Approach to Testing of GxP Systems, Section 3.5, regarding requirements testing for a preferred supplier [20].

- Product sign-off may or may not be captured via the manufacturing system itself (or may be captured for multiple PCS within a single MES).
- Process control data may be in the form of a printout and may include:
 - Trends on chart recorders
 - Alarm log printouts
 - Printouts from simple instruments such as weigh scales

The data integrity lifecycle should include all data that is output to printout, with consideration as to:

- The integrity of copies of data held locally for printing (e.g., in a chart recorder memory, in a “temporary” print file, in the memory of the instrument)
 - Whether the printout is used for regulatory purposes such as product release or product recall (if not, it should ideally be identified as “not to be used for regulatory purposes”, e.g., printed in a header/footer, as a watermark, or in prominent text)
 - Whether the original electronic data or the printout is subsequently considered the primary data for regulatory purposes
 - Whether the data is static or dynamic¹⁴, whether there is any loss of GMP content or meaning in producing the printout, and whether any additional risk is introduced if the quality unit reviews the printout rather than the original electronic data
 - Who is allowed to generate a printout?
 - The location and security of printout output devices
 - Validation to ensure that the data included in the printout is an accurate representation of the GMP content and meaning of the electronic record. If the data is static, then the printout may be a true copy of the electronic record.
- In a hybrid situation, the production report may have attachments such as printouts from equipment not interfaced (e.g., a filter integrity tester), and specimens of packaging materials. These attachments need to be signed and dated by the person generating the record along with traceability information such as sample ID or batch number and need to be reviewed as part of the overall approval workflow [11].
 - Some very simple electronic systems, such as balances, pH meters or simple processing equipment that does not store data, generate directly-printed paper records. These types of systems and records provide limited opportunity to influence the presentation of data by (re-)processing, or changing electronic date and time stamps. In these circumstances, the original record should be signed and dated by the person generating it and information to ensure traceability, such as sample ID, batch number, etc. added onto the record. These original records should be attached to batch processing or testing records.

3.3.3.3 Use for Process and Equipment Control

A great deal of data (and derived and/or calculated data) is used to automatically control processes or equipment to ensure that CPPs and other controlled parameters remain within specification.

¹⁴ See the *ISPE GAMP® RDI Good Practice Guide: Data Integrity – Key Concepts* [23] for an explanation of static and dynamic data.

Much of this information is short-term data that does not need to be stored (e.g., the proportional and integral components calculated as part of the output of a PID controller) as long as the software is appropriately validated and can be relied on to perform as intended.

In some cases, it may be necessary to store a subset of this data in order to support any subsequent investigations. For example, where MVC is used, it may be necessary to store all of the measured process parameters in order to demonstrate that the process was under control at all times, even though the control action is based only on the output of the MVC algorithm. (This is because there may be multiple combinations of the input parameters that create the same output condition.) Appropriate sampling and storage periods should be considered in these cases.

3.3.3.4 Use for Maintenance

Manufacturing system data may also be used for maintenance purposes. Some of this will have regulatory significance (e.g., data linked to or included in calibration records or maintenance logs, or data used to provide a rationale for preventive maintenance frequency), while other maintenance data may only have operational value.

3.3.3.5 Use for Performance Analysis

Other manufacturing system data (or copies of data) may be used solely for performance analysis, such as equipment utilization, or efficiency tracking. Unless data is considered a CQA or CPP, this data does not have regulatory significance.

3.3.4 Data Lifecycle – Retention and Retrieval

Once manufacturing system data is archived, data integrity concerns are often similar to those for all other system types. Some items specific to manufacturing systems include:

- The production report needs to remain available throughout the required retention period. In some systems this is achieved by archiving as flat file production reports, while in others it is accomplished by archiving the electronic production record and maintaining the ability to create a production report from it.

When making decisions about how regulated data should be retained, the degree of interaction with the data required for future GMP activities needs to be determined. Where the ability to search, filter, and trend is required, data needs to be retained in a form that permits these activities.

- Large data sets such as trend data sets, alarm history, technical system logs, and data audit trails, may be retained as evidence of operational compliance, and used to support the investigation of incidents or OOS results. Retention should be based on risk and regulations.
- Simple instruments and controllers are likely to store data in volatile memory and may need uninterruptable power supplies.
- Some simple devices may not connect to a network so data for archiving may be migrated via portable media. The physical and logical security of such devices needs to be risk assessed. This is to ensure that data cannot be accidentally or deliberately overwritten by connecting the storage to an unauthorized device, or lost if it cannot be stored, for example, if someone removes the storage media.
- Because of the real-time nature of process control and its potentially high impact on patient safety and product quality, PCS often incorporate elements of redundancy in order to prevent loss of data. Where this is the case, duplicate copies of data will need to be defined and controlled (including questions such as: which elements of the data lifecycle are duplicated; what is treated as the primary copy; what implications are there for backup and recovery).

- This real-time nature also leads to business continuity requirements for rapid recovery of functionality that may be addressed via hot standby systems or virtualization. The implications of this for backup and restore of data should be considered, including how backed up data can be migrated or transferred to a standby or new virtual system. At the same time, the real-time nature of the systems may result in a disinclination to take a system out of use for calibration or testing of backup and restore.
- Unusual hardware platforms (meaning not those normally supported by an IT department) tend to be more commonplace in PCS than in other system types. This leads to issues over both regaining access during a disaster recovery process and retaining access to legacy data after a system is decommissioned. When sourcing a new system, if the same functionality is available in a system based on a standard hardware platform, this would be preferable. Where a system based on an unusual hardware platform is used, it is important to leverage vendor support for disaster recovery and legacy data access; this should be covered in a service level agreement.
- If the overall IT strategy states that the MES is the archiving system where regulated data will reside throughout the retention period, any attachments transferred electronically to the manufacturing execution layer will be included in the production report. If the archiving system for a particular data set is outside the system, then relevant attachments must be linked by ID, batch, or similar unique identification to allow traceability to and from the production report.

3.3.5 Data Lifecycle – Destruction

Where manufacturing system data is archived and retained outside of the system, considerations are similar to those for all other system types. Some items specific to information retained within a manufacturing system are:

- In most cases, short-term data that does not need to be retained is simply overwritten as part of the standard control processing.
- Instruments or controllers may have limited memory capacity and the oldest data may be overwritten by the latest data. This should be carefully assessed to ensure that devices have sufficient memory to retain data for the required period (e.g., to support any subsequent quality investigation) and should be reviewed as part of the equipment or system selection and specification.
- In some systems (particularly at the process control level), it is possible to configure the maximum period for which, for example, trend data is stored in the system, but this may not be possible for data such as alarm logs, which may have a finite capacity.
- Some systems (particularly at the process control level) may need manual intervention to remove files after they have been archived to avoid filling the available memory. Careful consideration needs to be given to the management of this process, how access to perform these processes will be restricted, and what (automatically or manually generated) records of the process need to be kept.

4 Risk Management Approaches

A complete assessment of the risks associated with the system requires careful consideration of data integrity and data quality risks, along with all other risk types (patient safety, product quality, safety, environmental, etc.). Following the model in Appendix M3 of ISPE GAMP® 5 [15], Risk Priority is determined by considering risk severity, probability, and detectability.

4.1 Initial Data Integrity Risk Assessment and Identifying the Impact of Poor Data Integrity Controls

The initial step of the risk management process is to understand system data in the context of the GMP process in order to assess the potential impact on patient safety, product quality, and the integrity of regulated records.

From the perspective of risks to data integrity in manufacturing systems, the risk severity should consider the worst-case impact if poor-quality (or no) data is acquired, processed, stored, or used. An example of a high-risk severity is data related to a Critical Quality Attribute where poor quality might kill or seriously injure a patient, such as data relating to the sterilization status of water for injection.

4.2 Identifying Data Flows with Impact on Patient Safety, Product Quality, and Data Integrity

Identifying individual data flows that may impact patient safety, product quality, and integrity of regulated records requires an understanding of the data lifecycle and of the way data flows through that lifecycle.

Managing the data integrity risk associated with a particular data flow starts by assessing how the data is used. For example, allowing an operator to change a CPP-related alarm limit could lead to a process excursion without the resulting alarm. Where alarms and events are reviewed as part of an exception reporting process, the failure to alarm could result in the excursion not being detected.

Knowledge of this impact allows the risk management process and the level of detail of data lifecycle modeling to be scaled appropriately.

4.3 Performing Functional Risk Assessments and Identifying Controls

Once the potential impact is understood, knowledge of the data flows within a system or functional domain allows potential threats to the integrity of the data to be identified and rated according to probability and detectability.

For example:

- The probability of a data integrity issue is lower for a validated data flow where values are captured from a sensor, stored, and reported with no manual intervention than it is for a value that is transcribed manually into the system.
- Some data integrity issues may be relatively easy to detect (e.g., missing or corrupted data). Other data integrity issues may be more difficult:
 - Data values within an acceptable range that have been deliberately falsified by copying data from a previous batch

- Repeated testing or aborted/failed runs or cycles run prior to the successful activity that are not reported in the production report. (Electronic review of the data in the system is required to detect this where the printed report from the system would not show it.)
- Values inadvertently overwritten by a subsequent batch before the data can be transferred for long-term storage
- Values from an instrument that has slowly slipped out of calibration

In addition, the closer a manufacturing system is to final product manufacturing or product release (e.g., secondary processing or packaging), the less likely data integrity issues such as a mix-up in batch ID will be detected prior to product release. Where few checks on data integrity are performed before batch release or where a batch is shipped under quarantine subject to subsequent review of the production record, the probability of detection is low.

It can be helpful to consider data flows by functional domain rather than just by system to ensure that the full impact of data flows between systems is captured, including those to and from systems not present in the production area.

The threats to data integrity can then be analyzed and appropriate controls identified, for example, modifying the system to replace manual transcription of weight with an automated data transfer; or the data quality is passed alongside the actual data values.

4.4 Implement and Verify Appropriate Controls

The controls then need to be implemented and verified as effective, for example, validate the automated data transfer to confirm the value is passed correctly, is not affected by noise, and has appropriate metadata such as units and tag number associated with it.

4.5 Reviewing Residual Risks and Monitoring Controls

Finally, the residual risks need to be reviewed and monitored. For example, has a new source of electromagnetic interference been introduced that affects the automated communications link?

These risks and the ability to effectively mitigate them can be assessed as part of the data lifecycle modeling described in Section 3.3.

An example of a detailed risk assessment for a CPP captured by a PCS and transferred to a production record in a MES is provided in Appendix 1.

5 Detecting Data Integrity Issues

5.1 Data Review Requirements

5.1.1 Define the Process for Data Review in a Procedure

The process for review should be described in a procedure [11] as set out in Section 4.4 of the *ISPE GAMP® Guide: Records and Data Integrity* [3]. Manufacturing data review is unlikely to be straight forward and the procedure may need to address:

- Which data belongs to which of the many systems involved in the production process
- The order in which data must be reviewed, for example, if it is stored across several systems and includes paper generated forms or reports
- How all passed and failed tests (e.g., filter integrity tests) or all successful, repeated, aborted runs or cycles (e.g., CIP) will be included with the production record for review to ensure full visibility to the reviewers
- An explanation of how to review the time synchronization across multiple systems
- An explanation on the use of trend data retrieval modes so that the right averaging or plot points are selected to ensure that meaningful data is obtained (e.g., selection of data on an hourly basis rather than a daily basis will lead to a very different type of review)
- Details on how the review is approved, especially if it involves several different departments, for example production and engineering
- Details on how errors or omissions detected are escalated, particularly if data issues are found relating to a system managed by a supplier
- Product release by exception and the implications for further data review

As far as possible, automated solutions should be used to collate items for data reviews, as this mitigates the risks associated with human correlations, interpretations, or manual extraction. For new manufacturing systems, this requirement should be a priority in the design.

On existing systems when there is a clearly identified requirement for frequent review of certain data items not currently automated, upgrading the systems should be considered so that review items are collated automatically and presented in a manner that makes the review easy. An example is the automatic extraction of alarm data that corresponds to each batch. There will be a cost to this but the long-term benefit is likely to be significant.

Data used in the automatic control of normal day-to-day processes or equipment may be short term and not stored, and therefore cannot be subject to review. The appropriate controls should have been verified during initial validation. Visual data only displayed instantaneously on a process control screen falls into this category. An example of how this is documented and handled is shown in Appendix 1, Table 8.1.

5.1.2 Types of Data to Review

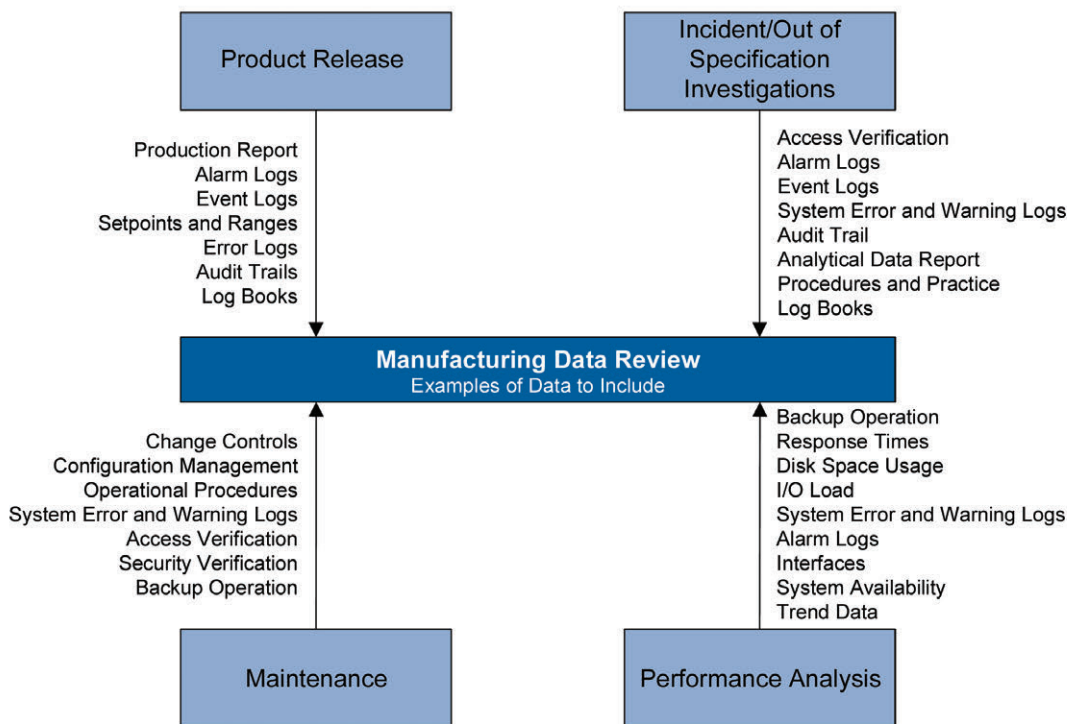
The data for review can be grouped as follows:

- Product release data
- Support of incident and OOS investigation data

- Maintenance data
- Performance analysis data

Figure 5.1 gives examples of items to review. What is reviewed, at what level, and at what frequency, is determined using a risk-based approach.

Figure 5.1: Data Review Considerations in Manufacturing Systems



5.1.2.1 Reviewing Data – Product Release

It is a regulatory expectation to review thoroughly all GMP data collected during the execution of manufacturing activities [24]. Data review should be performed as part of the product quality decision making. Data for review is collated by production reporting functions from production records using approved report formats. Under GMP regulations [24, 25]:

“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.”

ISPE has adopted the term “review by exception” to describe the application of exception reporting processes permitted by regulations [6]:

“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 ensure that all requirements under GMP regulations are captured. The exception reporting process and the configurations that underpin it need to be fully validated before use [11, 24].

The determination of product review and release requirements should consider what needs to be collated to confirm that:

- The manufacturing process was in line with the specification

This should focus on CQAs and CPPs, and with a recognition that in a multipurpose plant, critical data items may change with the change of product. (For instance, if a CPP during manufacture must stay in a specific range, a query could be developed to extract the maximum and minimum data, and set an exception indicator to flag within the production report when the range was not maintained.)

- No intended or unintended changes were made to the process

For example, where users must have access to change critical settings not automatically downloaded during batch setup, recording these parameters alongside the actual process values is essential. Review of data audit trails as part of product release is therefore especially important in this particular scenario where there is the possibility for users to misuse the access privileges to alter data in order to avoid an exception.

Where the PCS cannot download the critical settings from the MES or where the PCS lacks the security controls to prevent changes to critical settings, then reliance on audit trailing may be a compensating interim control and measures should be implemented where possible to improve the technical controls and eliminate the risk.

- The state of the equipment was appropriate

For example, providing evidence that the calibration status of instruments was good and that the equipment was cleaned before use.

- The input materials during manufacture were appropriate

For example, confirming that the materials used in manufacturing met specification, were within expiry date, and were added in the correct quantities.

This requires a good understanding of the data involved in the manufacturing process, including reference to the data flow diagram. (See Section 3.1)

The data reviewed for product release may be electronic (preferable), a printout, or a combination of both. An example is a production report that contains automatically generated data from a manufacturing system that is reviewed and approved within the manufacturing system, and includes a printout from a filter integrity tester.

If the best-case scenario has been achieved and everything to be reviewed during product release is collated into a well-designed, automatically generated production report, then the reviewer should easily be able to identify potential issues such as:

- Invalid or incorrect batch numbers
- Incorrect input materials
- Erroneous data entry or CPP values out of range
- Poor environmental conditions
- Incomplete or corrupt data
- Falsified data
- Actions by unauthorized personnel