MHRA GMP Data Integrity Definitions and Guidance for Industry March 2015

Introduction:

Data integrity is fundamental in a pharmaceutical quality system which ensures that medicines are of the required quality. This document provides MHRA guidance on GMP data integrity expectations for the pharmaceutical industry. This guidance is intended to complement existing EU GMP relating to active substances and dosage forms, and should be read in conjunction with national medicines legislation and the GMP standards published in Eudralex volume 4.

The data governance system should be integral to the pharmaceutical quality system described in EU GMP chapter 1. The effort and resource assigned to data governance should be commensurate with the risk to product quality, and should also be balanced with other quality assurance resource demands. As such, manufacturers and analytical laboratories are not expected to implement a forensic approach to data checking on a routine basis, but instead design and operate a system which provides an acceptable state of control based on the data integrity risk, and which is fully documented with supporting rationale.

Data integrity requirements apply equally to manual (paper) and electronic data. Manufacturers and analytical laboratories should be aware that reverting from automated / computerised to manual / paper-based systems will not in itself remove the need for data integrity controls. This may also constitute a failure to comply with Article 23 of Directive 2001/83/EC, which requires an authorisation holder to take account of scientific and technical progress and enable the medicinal product to be manufactured and checked by means of generally accepted scientific methods.

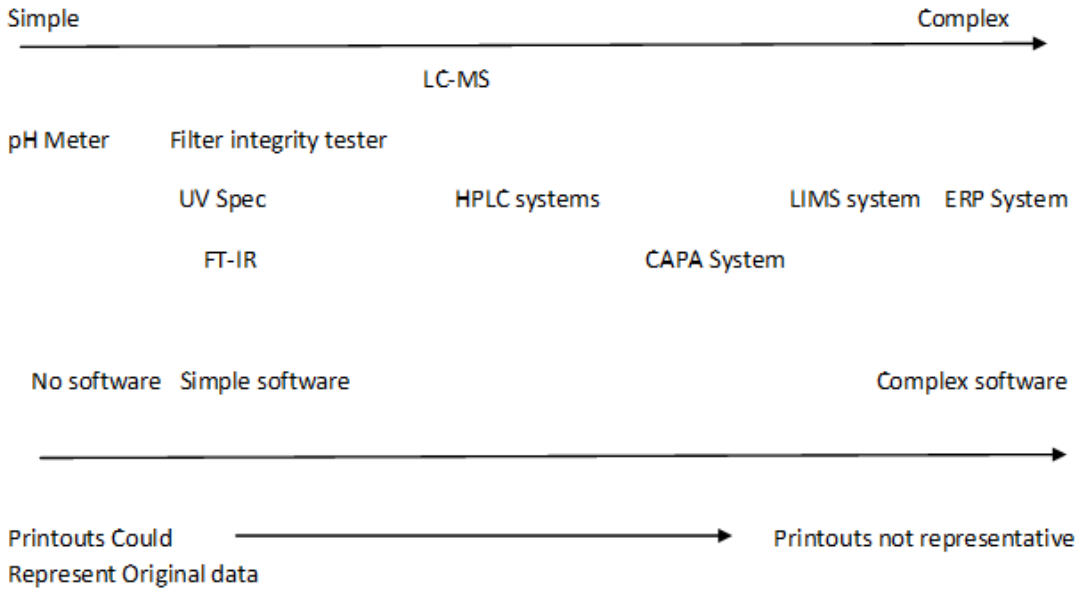
Throughout this guidance, associated definitions are shown as hyperlinks.

Establishing data criticality and inherent integrity risk:

In addition to an overarching data governance system, which should include relevant policies and staff training in the importance of data integrity, consideration should be given to the organisational (e.g. procedures) and technical (e.g. computer system access) controls applied to different areas of the quality system. The degree of effort and resource applied to the organisational and technical control of data lifecycle elements should be commensurate with its criticality in terms of impact to product quality attributes.

Data may be generated by (i) a paper-based record of a manual observation, or (ii) in terms of equipment, a spectrum of simple machines through to complex highly configurable computerised systems. The inherent risks to data integrity may differ depending upon the degree to which data (or the system generating or using the data) can be configured, and therefore potentially manipulated (see figure 1).

Figure 1: Diagram to illustrate the spectrum of simple machine (left) to complex computerised system (right), and relevance of printouts as 'original data'



(diagram acknowledgement: Green Mountain QA LLC)

With reference to figure 1 above, simple systems (such as pH meters and balances) may only require calibration, whereas complex systems require 'validation for intended purpose'. Validation effort increases from left to right in the diagram above. However, it is common for companies to overlook systems of apparent lower complexity. Within these systems it may be possible to manipulate data or repeat testing to achieve a desired outcome with limited opportunity of detection (e.g. stand-alone systems with a user configurable output such as FT-IR, UV spectrophotometers).

Designing systems to assure data quality and integrity

Systems should be designed in a way that encourages compliance with the principles of data integrity. Examples include:

- Access to clocks for recording timed events
- Accessibility of batch records at locations where activities take place so that ad hoc data recording and later transcription to official records is not necessary
- Control over blank paper templates for data recording
- User access rights which prevent (or audit trail) data amendments
- Automated data capture or printers attached to equipment such as balances
- Proximity of printers to relevant activities
- Access to sampling points (e.g. for water systems)
- Access to raw data for staff performing data checking activities.

The use of scribes to record activity on behalf of another operator should be considered ‘exceptional’, and only take place where:

- The act of recording places the product or activity at risk e.g. documenting line interventions by sterile operators.
- To accommodate cultural or staff literacy / language limitations, for instance where an activity is performed by an operator, but witnessed and recorded by a Supervisor or Officer.

In both situations, the supervisory recording must be contemporaneous with the task being performed, and must identify both the person performing the observed task and the person completing the record. The person performing the observed task should countersign the record wherever possible, although it is accepted that this countersigning step will be retrospective. The process for supervisory (scribe) documentation completion should be described in an approved procedure, which should also specify the activities to which the process applies.

Term	Definition	Expectation / guidance (where relevant)
Data	Information derived or obtained from raw data (e.g. a reported analytical result)	Data must be: A - attributable to the person generating the data L – legible and permanent C – contemporaneous O – original record (or 'true copy') A - accurate
Raw data	Original records and documentation, retained in the format in which they were originally generated (i.e. paper or electronic), or as a 'true copy'. Raw data must be contemporaneously and accurately recorded by permanent means. In the case of basic electronic equipment which does not store electronic data, or provides only a printed data output (e.g. balance or pH meter), the printout constitutes the raw data.	Raw data must: • Be legible and accessible throughout the data lifecycle. • Permit the full reconstruction of the activities resulting in the generation of the data

In the following definitions, the term 'data' includes raw data.

Metadata:	Metadata is data that describe the attributes of other data, and provide context and meaning. Typically, these are data that describe the structure, data elements, inter-relationships and other characteristics of data. It also permits data to be attributable to an individual.	Example: data (bold text) **3.5** and metadata, giving context and meaning, (italic text) are: *sodium chloride batch 1234,* ***3.5****mg. J Smith 01/07/14* Metadata forms an integral part of the original record. Without metadata, the data has no meaning.
Data Integrity	The extent to which all data are complete, consistent and accurate throughout the data lifecycle.	Data integrity arrangements must ensure that the accuracy, completeness, content and meaning of data is retained throughout the data lifecycle.
Data governance	The sum total of arrangements to ensure that data, irrespective of the format in which it is generated, is recorded, processed, retained and used to ensure a complete, consistent and accurate record throughout the data lifecycle	Data governance should address data ownership throughout the lifecycle, and consider the design, operation and monitoring of processes / systems in order to comply with the principles of data integrity including control over intentional and unintentional changes to information. Data Governance systems should include staff training in the importance of data integrity principles and the creation of a working environment that encourages an open reporting culture for errors, omissions and aberrant results. Senior management is responsible for the implementation of systems and procedures to minimise the potential risk to data integrity, and for identifying the residual risk, using the principles of ICH Q9. Contract Givers should perform a similar review as part of their vendor assurance programme

Data Lifecycle	All phases in the life of the data (including raw data) from initial generation and recording through processing (including transformation or migration), use, data retention, archive / retrieval and destruction.	The procedures for destruction of data should consider data criticality and legislative retention requirements. Archival arrangements should be in place for long term retention (in some cases, periods up to 30 years) for records such as batch documents, marketing authorisation application data, traceability data for human-derived starting materials (not an exhaustive list). Additionally, at least 2 years of data must be retrievable in a timely manner for the purposes of regulatory inspection.
Primary Record	The record which takes primacy in cases where data that are collected and retained concurrently by more than one method fail to concur.	In situations where the same information is recorded concurrently by more than one system, the data owner should define which system generates and retains the primary record, in case of discrepancy. The 'primary record' attribute should be defined in the quality system, and should not be changed on a case by case basis. Risk management principles should be used to ensure that the assigned 'primary record' provides the greatest accuracy, completeness, content and meaning. For instance, it is not appropriate for low-resolution or static (printed / manual) data to be designated as a primary record in preference to high resolution or dynamic (electronic) data. All data should be considered when performing a risk based investigation into data anomalies (e.g. out of specification results)

Original record / True Copy:	**Original record:** Data as the file or format in which it was originally generated, preserving the integrity (accuracy, completeness, content and meaning) of the record, e.g. original paper record of manual observation, or electronic raw data file from a computerised system **True Copy:** An exact verified copy of an original record. Data may be static (e.g. a 'fixed' record such as paper or pdf) or dynamic (e.g. an electronic record which the user / reviewer can interact with). Example 1: a group of still images (photographs – the static 'paper copy' example) may not provide the full content and meaning of the same event as a recorded moving image (video – the dynamic 'electronic record' example). Example 2: once printed or converted to static .pdfs, chromatography records lose the capability of being reprocessed and do not enable more detailed viewing of baselines or any hidden fields. By comparison, the same dynamic electronic records in database format provides the ability to track, trend, and query data, allowing the reviewer (with proper access permissions) to reprocess, view hidden fields, and expand the baseline to view the integration more clearly.	Original records and true copies must preserve the integrity (accuracy, completeness, content and meaning) of the record. Exact (true) copies of original records may be retained in place of the original record (e.g. scan of a paper record), provided that a documented system is in place to verify and record the integrity of the copy. It is conceivable for raw data generated by electronic means to be retained in an acceptable paper or pdf format, where it can be justified that a static record maintains the integrity of the original data. However, the data retention process must be shown to include verified copies of all raw data, metadata, relevant audit trail and result files, software / system configuration settings specific to each analytical run*, and all data processing runs (including methods and audit trails) necessary for reconstruction of a given raw data set. It would also require a documented means to verify that the printed records were an accurate representation. This approach is likely to be onerous in its administration to enable a GMP compliant record. Many electronic records are important to retain in their dynamic (electronic) format, to enable interaction with the data. Data must be retained in a dynamic form where this is critical to its integrity or later verification. This should be justified based on risk. * computerised system configuration settings should be defined, tested, 'locked' and protected from unauthorised access as part of computer system validation. Only those variable settings which relate to an analytical run would be considered as electronic raw data.

Computer system transactions:	A computer system transaction is a single operation or sequence of operations performed as a single logical 'unit of work'. The operation(s) that make up a transaction may not be saved as a permanent record on durable storage until the user commits the transaction through a deliberate act (e.g. pressing a save button), or until the system forces the saving of data. The metadata (i.e., user name, date, and time) is not captured in the system audit trail until the user commits the transaction. In Manufacturing Execution Systems (MES), an electronic signature is often required by the system in order for the record to be saved and become permanent.	Computer systems should be designed to ensure that the execution of critical operations are recorded contemporaneously by the user and are not combined into a single computer system transaction with other operations. A critical processing step is a parameter that must be within an appropriate limit, range, or distribution to ensure the desired product quality. These should be reflected in the process control strategy. Examples of 'units of work': • Weighing of individual materials • Entry of process critical manufacturing / analytical parameters • Verification of the identity of each component or material that will be used in a batch • Verification of the addition of each individual raw material to a batch (e.g. when the sequence of addition is considered critical to process control – see figure 2) • Addition of multiple pre-weighed raw materials to bulk vessel when required as a single manufacturing step (e.g. when the sequence of addition is not considered critical to process control – see figure 3)

Figure 2: Logical design permitting contemporaneous recording of addition of a single material in a manufacturing 'unit of work'. This record is permanently recorded (step 2), with audit trail, before progressing to next 'unit of work'.

Allows for contemporaneous recording of the material addition by the operator and verifier.

Material Additions		
Step	Instructions	Data
1.	Scan barcode of material ABC123.	ABC123 <Barcode>
2.	Add material ABC123 to the blender.	Operator Signature Verifier Signature

Next Step

Figure 3: Logical design permitting the addition of multiple materials in a manufacturing 'unit of work' before committing the record to durable media. Steps 1, 3 and 5 are contemporaneous entries (bar code), but are not permanently recorded with audit trail until step 6.

Does not allow for contemporaneous recording of the material addition by the operator and verifier.

Material Additions		
Step	Instructions	Data
1.	Scan barcode of material ABC123.	ABC123 <Barcode>
2.	Add material ABC123 to the blender.	
3.	Scan barcode of material DEF456.	DEF456 <Barcode>
4.	Add material DEF456 to the blender.	
5.	Scan barcode of material GHI789.	GHI789 <Barcode>
6.	Add material GHI789 to the blender.	Operator Signature Verifier Signature

Next Step

Audit Trail	GMP audit trails are metadata that are a record of GMP critical information (for example the change or deletion of GMP relevant data), which permit the reconstruction of GMP activities.	Where computerised systems are used to capture, process, report or store raw data electronically, system design should always provide for the retention of full audit trails to show all changes to the data while retaining previous and original data. It should be possible to associate all changes to data with the persons making those changes, and changes should be time stamped and a reason given. Users should not have the ability to amend or switch off the audit trail. The relevance of data retained in audit trails should be considered by the company to permit robust data review / verification. The items included in audit trail should be those of relevance to permit reconstruction of the process or activity. It is not necessary for audit trail review to include every system activity (e.g. user log on/off, keystrokes etc.), and may be achieved by review of designed and validated system reports. Audit trail review should be part of the routine data review / approval process, usually performed by the operational area which has generated the data (e.g. laboratory). There should be evidence available to confirm that review of the relevant audit trails have taken place. When designing a system for review of audit trails, this may be limited to those with GMP relevance (e.g. relating to data creation, processing, modification and deletion etc). Audit trails may be reviewed as a list of relevant data, or by a validated 'exception reporting' process. QA should also review a sample of relevant audit trails, raw data and metadata as part of self inspection to ensure on-going compliance with the data governance policy / procedures.

Audit trail (continued)		If no audit trailed system exists a paper based audit trail to demonstrate changes to data will be permitted until a fully audit trailed (integrated system or independent audit software using a validated interface) system becomes available. These hybrid systems are currently permitted, where they achieve equivalence to integrated audit trail described in Annex 11 of the GMP Guide. If such equivalence cannot be demonstrated, it is expected that facilities should upgrade to an audit trailed system by the end of 2017.
Data Review		There should be a procedure which describes the process for the review and approval of data, including raw data. Data review must also include a review of relevant metadata, including audit trail. Data review must be documented. A procedure should describe the actions to be taken if data review identifies an error or omission. This procedure should enable data corrections or clarifications to be made in a GMP compliant manner, providing visibility of the original record, and audit trailed traceability of the correction, using ALCOA principles (see 'data' definition).

Computerised system user access / system administrator roles		Full use should be made of access controls to ensure that people have access only to functionality that is appropriate for their job role, and that actions are attributable to a specific individual . Companies must be able to demonstrate the access levels granted to individual staff members and ensure that historical information regarding user access level is available. Shared logins or generic user access should not be used. Where the computerised system design supports individual user access, this function must be used. This may require the purchase of additional licences. It is acknowledged that some computerised systems support only a single user login or limited numbers of user logins. Where alternative computerised systems have the ability to provide the required number of unique logins, facilities should upgrade to an appropriate system by the end of 2017. Where no suitable alternative computerised system is available, a paper based method of providing traceability will be permitted. The lack of suitability of alternative systems should be justified based on a review of system design, and documented. System administrator access should be restricted to the minimum number of people possible taking account of the size and nature of the organisation. The generic system administrator account should not be available for use. Personnel with system administrator access should log in under unique log-ins that allow actions in the audit trail(s) to be attributed to a specific individual.

Computerised system user access / system administrator roles (continued)		System Administrator rights (permitting activities such as data deletion, database amendment or system configuration changes) should not be assigned to individuals with a direct interest in the data (data generation, data review or approval). Where this is unavoidable in the organisational structure, a similar level of control may be achieved by the use of dual user accounts with different privileges. All changes performed under system administrator access must be visible to, and approved within, the quality system. The individual should log in using the account with the appropriate access rights for the given task e.g. a laboratory manager performing data checking should not log in as system administrator where a more appropriate level of access exists for that task.
Data retention		Raw data (or a true copy thereof) generated in paper format may be retained for example by scanning, provided that there is a process in place to ensure that the copy is verified to ensure its completeness. Data retention may be classified as archive or backup Data and document retention arrangements should ensure the protection of records from deliberate or inadvertent alteration or loss. Secure controls must be in place to ensure the data Integrity of the record throughout the retention period, and validated where appropriate. Where data and document retention is contracted to a third party, particular attention should be paid to understanding the ownership and retrieval of data held under this arrangement. The physical location in which the data is held, including impact of any laws applicable to that geographic location should also be considered. The responsibilities of the contract giver and acceptor must be defined in a contract as described in Chapter 7 of the GMP Guide

• **Archive**	Long term, permanent retention of completed data and relevant metadata in its final form for the purposes of reconstruction of the process or activity.	Archive records should be locked such that they cannot be altered or deleted without detection and audit trail. The archive arrangements must be designed to permit recovery and readability of the data and metadata throughout the required retention period.
• **Backup**	A copy of current (editable) data, metadata and system configuration settings (variable settings which relate to an analytical run) maintained for the purpose of disaster recovery.	Backup and recovery processes must be validated .
File structure		File structure has a significant impact on the inherent data integrity risks. The ability to manipulate or delete flat files requires a higher level of logical and procedural control over data generation, review and storage.
• **Flat files:**	A 'flat file' is an individual record which may not carry with it all relevant metadata (e.g. pdf, dat, doc).	Flat files may carry basic metadata relating to file creation and date of last amendment, but may not audit trail the type and sequence of amendments. When creating flat file reports from electronic data, the metadata and audit trails relating to the generation of the raw data may be lost, unless these are retained as a 'true copy'. Consideration also needs to be given to the 'dynamic' nature of the data, where appropriate (see 'true copy' definition) There is an inherently greater data integrity risk with flat files (e.g. when compared to data contained within a relational database), in that these are easier to manipulate and delete as a single file.

• **Relational database:**	A relational database stores different components of associated data and metadata in different places. Each individual record is created and retrieved by compiling the data and metadata for review	This file structure is inherently more secure, as the data is held in a large file format which preserves the relationship between data and metadata. This is more resilient to attempts to selectively delete, amend or recreate data and the metadata trail of actions, compared to a flat file system. Retrieval of information from a relational database requires a database search tool, or the original application which created the record.
Validation - for intended purpose (See also Annex 15 and GAMP 5)		Computerised systems should comply with the requirements of EU GMP Annex 11 and be validated for their intended purpose. This requires an understanding of the computerised system's function within a process. For this reason, the acceptance of vendor-supplied validation data in isolation of system configuration and intended use is not acceptable. In isolation from the intended process or end user IT infrastructure, vendor testing is likely to be limited to functional verification only, and may not fulfil the requirements for performance qualification. For example - validation of computerised system audit trail • A custom report generated from a relational database may be used as a GMP system audit trail. • SOPs should be drafted during OQ to describe the process for audit trail verification, including definition of the data to be reviewed. • 'Validation for intended use' would include testing during PQ to confirm that the required data is correctly extracted by the custom report, and presented in a manner which is aligned with the data review process described in the SOP.

Revision History

Revision	Publication Month	Reason for changes
Revision 1	January 2015	None. First issue.
Revision 1.1	March 2015	Added clarifications in response to stakeholder questions.