Medicines & Healthcare products Regulatory Agency

MHRA Regulating Medicines and Medical Devices

MHRA GxP Data Integrity Definitions and Guidance for Industry

Draft version for consultation July 2016

Background

The way in which regulatory data is generated has continued to evolve in line with the introduction and ongoing development of supporting technologies, supply chains and ways of working. Systems to support these ways of working can range from manual processes with paper records to the use of computerised systems. However the main purpose of the regulatory requirements remains the same; having confidence in the quality and the integrity of the data generated and being able to reconstruct activities remains a fundamental requirement.

Introduction:

This document provides guidance on the data integrity expectations that should be considered by organisations involved in all aspects of the chemical[1] and pharmaceutical development lifecycle.

This guidance should be read in conjunction with the applicable regulations and the general guidance specific to each GxP. Where GxP-specific references are made within this document (e.g. ICH Q9), consideration of the principles of these documents may provide guidance and further information.

Arrangements in place within an organisation with respect to people, systems and facilities should be designed, operated and where appropriate adapted to support a working environment and organisational culture that ensures data is complete consistent and accurate in all its forms, i.e. paper and electronic. The effort and resource applied to assure the validity and integrity of the data should be commensurate with the risk and impact of a data integrity failure to the patient or environment. When taken collectively these arrangements fulfil the concept of data governance.

Organisations are not expected to implement a forensic approach to data checking on a routine basis, but instead design and operate a fully documented system that provides an acceptable state of control based on the data integrity risk with supporting rationale. In addition to routine data review, the wider data governance system should ensure that periodic audits are capable of detecting opportunities for data integrity failures within the company's system, e.g. routine data review should consider the integrity of an individual data set, whereas the periodic system review might verify the effectiveness of existing control measures and consider the possibility of unauthorised activity.

It should be noted that data integrity requirements apply equally to manual (paper) and electronic data. Organisations should be aware that reverting from automated / computerised to manual / paper-based systems will not in itself remove the need for appropriate data integrity controls. Where data integrity weaknesses are identified, either as a result of audit or regulatory inspection, companies with multiple sites should ensure that appropriate corrective and preventive actions are implemented

[1] Chemical lifecycle relating to GLP studies regulated by MHRA

across the organisation. Appropriate notification to regulatory authorities should be made where applicable.

Although not included in this guidance, the impact of organisational culture and senior management behaviour on the success of data governance measures should not be underestimated.

Establishing data criticality and inherent integrity risk:

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 quality attributes.

Data may be generated by (i) manual means - a paper-based record of a manual observation, or (ii) electronic means - in terms of equipment, a spectrum of simple machines through to complex highly configurable computerised systems.

When manually recorded data requires stringent oversight, consideration should be given to risk-reducing supervisory measures. Examples include contemporaneous second person verification of data entry, or cross checks of related information sources (e.g. equipment log books).

The inherent risks to data integrity relating to equipment and computerised systems 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: Table to illustrate the spectrum of simple machine (left) to complex computerised system (right), and relevance of printouts as 'original data'

System complexity	Simple system					Complex system
	pH meter	Filter integrity test			Interactive response technology	Enterprise resource planner
		UV spec	HPLC systems	LC-MS-MS	LIMS	
	Balance	FTIR		Pharmacovigilance database		Bespoke systems
		ECG machines	Electronic trial master file		Clinical database	
		Spreadsheet			Statistical analysis tools	
	Min/Max thermometers	Data loggers	Building management system			
Software	No software	Simple software				Complex software
Printouts	Printouts may represent original data	Printouts not representative of original data				

With reference to figure 1, 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. 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 for detection (e.g. stand-alone systems with a user configurable output such as ECG machines, FTIR, UV spectrophotometers).

Different data has varying importance to quality, safety and efficacy decisions. Data criticality may be determined by considering the type of decision influenced by the data. Data risk reflects its vulnerability to unauthorised deletion or amendment, and the opportunity for detection during routine review.

Reduced effort and/or frequency of control measures may be justified for data that has a lesser impact to product and patient, if those data are obtained from a process that does not provide the opportunity for amendment without specialist software/knowledge

Data risk is typically increased by complex, inconsistent processes, with open ended and subjective outcomes compared to simple tasks that are consistent, well defined and objective.

Automation, or the use of a 'validated system' (e.g. e-CRF; analytical equipment) may not be low risk in terms of data integrity if the validated system is considered in isolation of the relevant business

process (trial subject data entry, analytical sample preparation). Where there is human intervention, particularly influencing how or what data is recorded or reported, there may be increased risk from poor organisational controls or data verification due to overreliance on the system's validated state.

Companies should balance data risk with other quality and compliance priorities. Prioritisation of actions, including acceptance of an appropriate level of residual risk should be documented, communicated to senior management, and kept under review. In situations where long-term remediation actions are identified, risk reducing short-term measures should be implemented to provide acceptable data governance in the interim.

Designing systems to assure data quality and integrity

Systems and processes should be designed in a way that encourages compliance with the principles of data integrity. Consideration should be given to ease of access, usability and location whilst ensuring appropriate control of the activity guided by the criticality of the data. Examples include:

- Access to appropriately controlled / synchronised clocks for recording timed events.
- Accessibility of records at locations where activities take place so that ad hoc data recording and later transcription to official records is not necessary
- 'Free access' to blank paper proformae for raw/source data recording should be controlled where this is appropriate. Reconciliation may be necessary to prevent recreation of a record.
- User access rights that prevent (or audit trail) unauthorised data amendments
- Automated data capture or printers attached to equipment such as balances
- Control of physical parameters (time, space, equipment) that permit performance of tasks and recording of data as required.
- 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 contemporaneous recording compromises the product or activity 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 should be contemporaneous with the task being performed, and should identify both the person performing the task and the person completing the record. The person performing the 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.

Definitions and guidance

In the following section, definitions are given in italic text.

1. Data

Facts and statistics collected together for reference or analysis

Data should be:

A - attributable to the person generating the data
L – legible and permanent
C – contemporaneous
O – original record (or true copy)
A - accurate

Data governance measures should also ensure that data is compete, consistent and enduring throughout the lifecycle

2. Raw data (GCP: synonymous with 'source data')

Original records, 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.

The definition of 'original records' currently varies across regulatory documents. By its nature, paper copies of raw data generated electronically cannot be considered as 'raw data'.

Raw data must permit the full reconstruction of the activities resulting in the generation of the data. 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.

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

3. 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 (or if automatically generated, to the original data source).

Metadata forms an integral part of the original record. Without metadata, the data has no meaning.

See also 'flat files'

Example (i) **3.5**

metadata, giving context and meaning, (italic text) are:

sodium chloride batch 1234, **3.5***mg. J Smith 01/07/14*

Example (ii) **3.5**

metadata, giving context and meaning, (italic text) are:

Trial subject A123, sample ref X789 taken 30/06/14 at 1456hrs.
INR, **3.5***mg. Analyst: J Smith 01/07/14*

4. 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.

5. 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 enables visibility of 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 risk management techniques such as the principles of ICH Q9. Contract Givers should ensure that data ownership, governance and accessibility are included in a contract/technical agreement. The Contract Giver should also perform a data governance review as part of their vendor assurance programme.

Routine data review should evaluate the integrity of an individual data set, compliance with established organisational and technical measures and any data risk indicators (e.g. data amendment). Periodic review of data governance measures (for example audit) should assess

effectiveness of established organisational and technical measures, and also consider the possibility of unauthorised activity.

Data governance systems should also ensure that data are readily available and directly accessible on request from national competent authorities.

6. Data Lifecycle

All phases in the life of the data (including raw data) from initial generation and recording through processing (including analysis, transformation or migration), use, data retention, archive / retrieval and destruction.

The procedures for destruction of data should consider data criticality and where applicable legislative retention requirements. Archival arrangements should be in place for long term retention of relevant data in compliance with legislation.

7. Data transfer / migration

Data transfer is the process of transferring data and metadata between storage media types or computer systems. Data migration changes the format of data to make it usable or visible on an alternative computerised system.

Data transfer/migration should be designed and validated to ensure that data integrity principles are maintained.

8. Data Processing

A sequence of operations performed on data in order to extract, present or obtain information in a defined format. Examples might include: statistical analysis of individual patient data to present trends or conversion of a raw electronic signal to a chromatogram and subsequently a calculated numerical result

There should be adequate traceability of any user defined parameters used within data processing activities. Audit trails and retained records should allow reconstruction of all data processing activities regardless of whether the output of that processing is subsequently reported or otherwise used. If data processing has been repeated with progressive modification of processing parameters this should be visible to ensure that the processing parameters are not being manipulated to achieve a more desirable end point.

9. Recording data:

Companies should have an appropriate level of process understanding and technical knowledge of systems used for data recording, including their capabilities, limitations and vulnerabilities.

The selected method should ensure that data of appropriate accuracy, completeness, content and meaning is collected and retained for its intended use. Where the capability of the electronic system permits dynamic storage it is not appropriate for low-resolution or static (printed / manual) data to be collected in preference to high resolution or dynamic (electronic) data.

10. Excluding Data:

Data may only be excluded where it can be demonstrated through sound science that the data is anomalous or non-representative. In all cases, this justification should be documented and considered during data review and reporting. All data (even if excluded) should be retained with the original data set, and be available for review in a format that allows the validity of the decision to exclude the data to be confirmed.

11. Original record / True Copy(also referred to as 'certified copy' or 'verified copy'):

11.1. 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

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). An analogy being 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 1: An electronic monitoring system records temperatures every 5 minutes, providing the ability to interrogate data to investigate excursions (magnitude and duration). This ability is compromised when working from a summary graph.

Example 2: Chromatography systems provide dynamic electronic records in database format with the ability to track, trend, and query data. This allows the reviewer (with proper access permissions) to interact with the data (e.g. view hidden fields, and expand the baseline) to view the integration more clearly. Once printed or converted to static file format (e.g. .pdfs), chromatography records lose the interaction capability.

11.2. True Copy:

A copy of original information that been verified as an exact (accurate and complete) copy having all of the same attributes and information as the original. The copy may be verified by dated signature or by a validated electronic signature. A true copy may be retained in a different electronic file format to the original record, if required, but must retain the equivalent static/dynamic nature of the original record.

Original records and true copies must preserve the integrity (accuracy, completeness, content and meaning) of the record. 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. Companies should consider any risk associated with the destruction of original records.

It should be possible to create a true copy of electronic data, including relevant metadata, for the purposes of review, backup and archival. Accurate and complete copies for certification should include the meaning of the data (e.g. date formats, context, layout, electronic signature and authorisations), as well as the full audit trail. Consideration should be given to the dynamic functionality of a 'true copy' throughout the retention period (see 'archive').

Where certified copies are made, the process for certification should be described, including the process for ensuring that the copy is complete and accurate and for identifying the certifying party and their authority for making that copy. The process of making a true copy of electronic data should be validated.

Data must be retained in a dynamic form where this is critical to its integrity or later verification. It is conceivable for some 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, any variable software/system configuration settings specific to each record, and all data processing runs (including methods and audit trails) necessary for reconstruction of a given raw data set. It would also require a documented means to verify that the printed records were an accurate representation. This approach is likely to be onerous in its administration to enable a GxP compliant record.

12. 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 saves the transaction to durable storage. In computerised systems, an electronic signature may be required in order for the record to be saved and become permanent.

Computer systems should be designed to ensure that the execution of critical steps are recorded contemporaneously by the user and are not combined into a single computer system transaction with other operations. A critical step is a parameter that must be within an appropriate limit, range, or distribution to ensure the safety of the subject or quality of the product or data.

Computerised systems should enforce saving immediately after critical data entry. Data entry prior to saving to permanent memory with audit trail (server, database) is considered to be temporary memory. These data are at risk of amendment or deletion without audit trail visibility. The length of time that data is held in temporary memory should be minimised.

Critical steps should be documented with process controls that consider system design (prevention), together with monitoring and review processes (surveillance). Surveillance activities should alert to failures that are not addressed by the process design.

Example:
Computerised system may be configured to prevent data manipulation (prevention). This does not restrict a person from repeating the process by manipulating data to achieve a desired result. Periodic reviews (surveillance) for undisclosed data may reduce risk from repeated events.

13. Audit Trail

Audit trails are metadata that are a record of critical information (for example the change or deletion of relevant data) that permit the reconstruction of activities.

Where computerised systems are used to capture, process, report, store and archive raw data electronically, system design should always provide for the retention of 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. The items included in the audit trail should be those of relevance to permit reconstruction of the process or activity.

Audit trails should be switched on. Users (with the exception of system administrator) 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. 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 appropriately designed and validated system reports.

Where relevant audit trail functionality does not exist (e.g. within legacy systems and spreadsheets) an equivalent level of control may be achieved for example by the use of log books, protecting each version and change control.

Routine data review should include a documented audit trail review. When designing a system for review of audit trails, this may be limited to those with GxP 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 'exception reporting' process. An exception report is a validated search tool that identifies and documents predetermined 'abnormal' data or actions, which requires further attention or investigation by the data reviewer.

QA should have sufficient knowledge and system access to review relevant audit trails, raw data and metadata as part of audits to ensure on-going compliance with the company's data governance policy and regulatory requirements. See also 'data governance'.

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 acceptable, where they achieve equivalence to integrated audit trail, such as described in Chapter 4 of the GMP Guide. If such

equivalence cannot be demonstrated, it is expected that GMP facilities should upgrade to an audit trailed system by the end of 2017 (reference: Art 23 of Directive 2001/83/EC).

14. Electronic signatures

The use of electronic signatures should be compliant with the requirements of international standards such as Directive 1999/93/EC (requirements relevant to 'advanced electronic signature').

Where a paper or pdf copy of an electronically signed document is produced the metadata associated with an electronic signature should be maintained together with the associated document.

An inserted image of a signature alone. or a footnote indicating that the document has been electronically signed (where this has been entered by a means other than the validated electronic signature process) is not sufficient.

15. Data Review

There should be a procedure that describes the process for the review and approval of data. Data review should also include a review of relevant metadata, including audit trails.

Review should be based upon original data or a true copy. Summary reports of data are often supplied between companies (contract givers and acceptors). However, it must be acknowledged that summary reports are limited, in that critical supporting data and metadata are often not included.

Prior to acceptance of summary reports, a risk-based evaluation of the contract acceptor's quality system including compliance with data integrity principles should be established.

Where data review is not conducted by the company that generated the data, the responsibilities for data review must be documented and agreed by both parties.

Data review should 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 GxP compliant manner, providing visibility of the original record, and audit trailed traceability of the correction, using ALCOA principles (see 'data' definition).

16. 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. Controls should be applied at both the operating system and application levels.

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 no suitable alternative computerised system is available, equivalent control may be provided by third party software, or a paper based method of providing traceability (with version control). The suitability of alternative systems should be justified and documented. Increased data review is likely to be required for hybrid systems because they are vulnerable to non-attributable data changes. It is expected that companies should be implementing systems which comply with current regulatory expectations. It is expected that GMP facilities should upgrade to system with individual login and audit trails by the end of 2017 (reference: Art 23 of Directive 2001/83/EC).

System administrator access should be restricted to the minimum number of people possible taking account of the size and nature of the company. The generic system administrator account should not be available for use. Personnel with system administrator access should log in with unique credentials that allow actions in the audit trail(s) to be attributed to a specific individual.

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 company 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 should 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.

Individuals may require changes in their access rights depending on the status of clinical trial data. For example, once data management processes are complete the data is ‘locked’ by removing editing access rights. This should be able to be demonstrated within the system.

17. Data retention

Data retention may be classified as either archive (protected data for long term storage) or backup (dynamic data for the purposes of disaster recovery).

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. See also data transfer / migration.

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 certified.

17.1. Archive

A designated secure area or facility (e.g. cabinet, room, building or computerised system) for the long term, permanent retention of complete data and relevant metadata in its final form for the purposes of reconstruction of the process or activity.

Archive records may be the original data or a 'true copy', and should be protected such that they cannot be altered or deleted without detection.

The archive arrangements must be designed to permit recovery and readability of the data and metadata throughout the required retention period. In the case of electronic data archival, this process should be validated, and in the case of legacy systems the ability to review data periodically verified (i.e. to confirm the continued support of legacy computerised systems).

When legacy systems can no longer be supported, consideration should be given to maintaining the software for data accessibility purposes as long as reasonably practicable. This may be achieved by maintaining software in a virtual environment (e.g. Cloud or SaaS). Migration to an alternative file format which retains the 'true copy' attributes of the data may be necessary with increasing age of the legacy data. The migration file format should be selected taking into account the balance of risk between long term accessibility versus possibility of reduced dynamic data functionality (e.g. data interrogation, trending, re-processing etc).

17.2. Backup

A copy of current (editable) data, metadata and system configuration settings (variable settings which relate to a record or analytical run) maintained for the purpose of disaster recovery.

Backup and recovery processes should be validated and periodically tested.

18. File structure

18.1. Flat files:

A 'flat file' is an individual record which may not carry any additional metadata with it, other than that included in the file itself

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 when compared to data contained within a relational database in that they are easier to manipulate and delete as a single file.

18.2. 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 using a database reporting tool.

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 reporting tool, or the original application which created the record.

Access rights for database entry or amendment should be controlled, and consistent with the requirements for computerised system user access / system administrator roles.

19. Validation - for intended purpose (See also GMP Annex 15 and GAMP 5)

Computerised systems should comply with regulatory requirements and associated guidance 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 - A custom report generated from a relational database, used as a system audit trail. Functional verification demonstrates that the required information is consistently and completely presented. Validation for intended purpose ensures that the steps for generating the custom report accurately reflect those described in the data checking SOP, and that the report output is consistent with the procedural steps for performing the subsequent review.

20. Cloud providers and virtual service / platforms (also referred to as software as a service SaaS / platform as a service (PaaS) / infrastructure as a service (IaaS))

Where 'cloud' or 'virtual' services are used, particular attention should be paid to understanding the service provided, ownership, retrieval, retention and security of data.
The physical location where the data is held, including impact of any laws applicable to that geographic location should be considered. The responsibilities of the contract giver and acceptor should be defined in a technical agreement or contract. This should ensure timely access to data (including metadata and audit trails) to the data owner and national competent authorities upon request. Contracts with providers should define responsibilities for archiving and continued readability of the data throughout the retention period (see archive). Appropriate arrangements must exist for the restoration of the software/system as per its original interactive validated state, including validation and change control information to permit this restoration.

Business continuity arrangements should be included in the contract, and tested.