Medicines & Healthcare products
Regulatory Agency

MHRA
Regulating Medicines and Medical Devices

Medicines & Healthcare products Regulatory Agency (MHRA)

‘GXP’ Data Integrity Guidance and Definitions

March 2018

Medicines & Healthcare products
Regulatory Agency

MHRA
Regulating Medicines and Medical Devices

Table of contents

Medicines & Healthcare products
Regulatory Agency

MHRA
Regulating Medicines and Medical Devices

1. Background

The way regulatory data is generated has continued to evolve in line with the ongoing development of supporting technologies such as the increasing use of electronic data capture, automation of systems and use of remote technologies; and the increased complexity of supply chains and ways of working, for example, via third party service providers. Systems to support these ways of working can range from manual processes with paper records to the use of fully computerised systems. The main purpose of the regulatory requirements remains the same, i.e. having confidence in the quality and the integrity of the data generated (to ensure patient safety and quality of products) and being able to reconstruct activities.

2. Introduction

2.1 This document provides guidance for UK industry and public bodies regulated by the UK MHRA including the Good Laboratory Practice Monitoring Authority (GLPMA). Where possible the guidance has been harmonised with other published guidance. The guidance is a UK companion document to PIC/S, WHO, OECD (guidance and advisory documents on GLP) and EMA guidelines and regulations.

2.2 This guidance has been developed by the MHRA inspectorate and partners and has undergone public consultation. It is designed to help the user facilitate compliance through education, whilst clarifying the UK regulatory interpretation of existing requirements.

2.3 Users should ensure their efforts are balanced when safeguarding data from risk with their other compliance priorities.

2.4 The scope of this guidance is designated as 'GXP' in that everything contained within the guide is GXP unless stated otherwise. The lack of examples specific to a GXP does not mean it is not relevant to that GXP just that the examples given are not exhaustive. Please do however note that the guidance document does not extend to medical devices.

2.5 This guidance should be considered as a means of understanding the MHRA's position on data integrity and the minimum expectation to achieve compliance. The guidance does not describe every scenario so engagement with the MHRA is encouraged where your approach is different to that described in this guidance.

2.6 This guidance aims to promote a risk-based approach to data management that includes data risk, criticality and lifecycle. Users of this guidance need to understand their data processes (as a lifecycle) to identify data with the greatest GXP impact. From that, the identification of the most effective and efficient risk-based control and review of the data can be determined and implemented.

2.7 This guidance primarily addresses data integrity and not data quality since the controls required for integrity do not necessarily guarantee the quality of the data generated.

2.8 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.

2.9 Where terms have been defined; it is understood that other definitions may exist and these have been harmonised where possible and appropriate.

3. The principles of data integrity

3.1 The organisation needs to take responsibility for the systems used and the data they generate. The organisational culture should ensure data is complete, consistent and accurate in all its forms, i.e. paper and electronic.

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

3.3 The impact of organisational culture, the behaviour driven by performance indicators, objectives and senior management behaviour on the success of data governance measures should not be underestimated. The data governance policy (or equivalent) should be endorsed at the highest levels of the organisation.

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

3.5 Organisations are not expected to implement a forensic approach to data checking on a routine basis. Systems should maintain appropriate levels of control whilst wider data governance measures should ensure that periodic audits can detect opportunities for data integrity failures within the organisation’s systems.

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

3.7 Organisations should be aware that reverting from automated or computerised systems to paper-based manual systems or vice-versa will not in itself remove the need for appropriate data integrity controls.

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

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

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

4. Establishing data criticality and inherent integrity risk

4.1 Data has varying importance to quality, safety and efficacy decisions. Data criticality may be determined by considering how the data is used to influence the decisions made.

4.2 The risks to data are determined by the potential to be deleted, amended or excluded without authorisation and the opportunity for detection of those activities and events. The risks to data may be increased by complex, inconsistent processes with open-ended and subjective outcomes, compared to simple tasks that are undertaken consistently, are well defined and have a clear objective.

4.3 Data may be generated by:

(i) Recording on paper, a paper-based record of a manual observation or of an activity or

(ii) electronically, using equipment that range from simple machines through to complex highly configurable computerised systems or

(iii) by using a hybrid system where both paper-based and electronic records constitute the original record or

(iv) by other means such as photography, imagery, chromatography plates, etc.

Paper

Data generated manually on paper may require independent verification if deemed necessary from the data integrity risk assessment or by another requirement. Consideration should be given to risk-reducing supervisory measures.

Electronic

The inherent risks to data integrity relating to equipment and computerised systems may differ depending upon the degree to which the system generating or using the data can be configured, and the potential for manipulation of data during transfer between computerised systems during the data lifecycle.

The use of available technology, suitably configured to reduce data integrity risk, should be considered.

Simple electronic systems with no configurable software and no electronic data retention (e.g. pH meters, balances and thermometers) may only require calibration, whereas complex systems require 'validation for intended purpose'.
Validation effort increases with complexity and risk (determined by software functionality, configuration, the opportunity for user intervention and data lifecycle considerations). It is important not to overlook systems of apparent lower complexity. Within these systems, it may be possible to manipulate data or repeat testing to achieve the desired outcome with limited opportunity for detection (e.g. stand-alone systems with a user-configurable output such as ECG machines, FTIR, UV spectrophotometers).

Hybrid
Where hybrid systems are used, it should be clearly documented what constitutes the whole data set and all records that are defined by the data set should be reviewed and retained. Hybrid systems should be designed to ensure they meet the desired objective.

Other
Where the data generated is captured by a photograph or imagery (or other media), the requirements for storage of that format throughout its lifecycle should follow the same considerations as for the other formats, considering any additional controls required for that format. Where the original format cannot be retained due to degradation issues, alternative mechanisms for recording (e.g. photography or digitisation) and subsequent storage may be considered and the selection rationale documented (e.g. thin layer chromatography).

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

4.5 The data integrity risk assessment (or equivalent) should consider factors required to follow a process or perform a function. It is expected to consider not only a computerised system but also the supporting people, guidance, training and quality systems. Therefore, automation or the use of a 'validated system' (e.g. e-CRF; analytical equipment) may lower but not eliminate data integrity risk. Where there is human intervention, particularly influencing how or what data is recorded, reported or retained, an increased risk may exist from poor organisational controls or data verification due to an overreliance on the system's validated state.

4.6 Where the data integrity risk assessment has highlighted areas for remediation, prioritisation of actions (including acceptance of an appropriate level of residual risk) should be documented, communicated to management, and subject to 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.

5. Designing systems and processes to assure data integrity; creating the ‘right environment’.

5.1 Systems and processes should be designed in a way that facilitates compliance with the principles of data integrity. Enablers of the desired behaviour include but are not limited to:

- At the point of use, having access to appropriately controlled/synchronised clocks for recording timed events to ensure reconstruction and traceability, knowing and specifying the time zone where this data is used across multiple sites.
- Accessibility of records at locations where activities take place so that informal data recording and later transcription to official records does not occur.
- Access to blank paper proformas for raw/source data recording should be appropriately controlled. Reconciliation, or the use of controlled books with numbered pages, may be necessary to prevent recreation of a record. There may be exceptions such as medical records (GCP) where this is not practical.
- User access rights that prevent (or audit trail, if prevention is not possible) unauthorised data amendments. Use of external devices or system interfacing methods that eliminate manual data entries and human interaction with the computerised system, such as barcode scanners, ID card readers, or printers.
- The provision of a work environment (such as adequate space, sufficient time for tasks, and properly functioning equipment) that permit performance of tasks and recording of data as required.
- Access to original records for staff performing data review activities.
- Reconciliation of controlled print-outs.
- Sufficient training in data integrity principles provided to all appropriate staff (including senior management).
- Inclusion of subject matter experts in the risk assessment process.
- Management oversight of quality metrics relevant to data governance.

5.2 The use of scribes to record activity on behalf of another operator can be considered where justified, for example:

- The act of contemporaneous recording compromises the product or activity e.g. documenting line interventions by sterile operators.
- Necropsy (GLP)
- To accommodate cultural or literacy/language limitations, for instance where an activity is performed by an operator but witnessed and recorded by a second person.

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.

In these situations, the recording by the second person should be contemporaneous with the task being performed, and the records 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 that specifies the activities to which the process applies.

6. Definition of terms and interpretation of requirements

In the following section, definitions where applicable, are given in italic text directly below the term.

6.1. Data

Facts, figures and statistics collected together for reference or analysis. All original records and true copies of original records, including source data and metadata and all subsequent transformations and reports of these data, that are generated or recorded at the time of the GXP activity and allow full and complete reconstruction and evaluation of the GXP activity.

Data should be:

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

Data governance measures should also ensure that data is complete, consistent, enduring and available throughout the lifecycle, where;

Complete – the data must be whole; a complete set

Consistent - the data must be self-consistent

Enduring – durable; lasting throughout the data lifecycle

Available – readily available for review or inspection purposes

6.2. Raw data (synonymous with ‘source data’ which is defined in ICH GCP)

Raw data is defined as the original record (data) which can be described as the first-capture of information, whether recorded on paper or electronically. Information that is originally captured in a dynamic state should remain available in that state.

Raw data must permit full reconstruction of the activities. Where this has been captured in a dynamic state and generated electronically, paper copies cannot be considered as ‘raw data’.

In the case of basic electronic equipment that does not store electronic data, or provides only a printed data output (e.g. balances or pH meters), then the printout constitutes the raw data. Where the basic electronic equipment does store electronic data permanently and only holds a certain volume before overwriting; this data should be periodically reviewed and where necessary reconciled against paper records and extracted as electronic data where this is supported by the equipment itself.

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

6.3. Metadata

Metadata are data that describe the attributes of other data and provide context and meaning. Typically, these are data that describe the structure, data elements, inter-relationships and other characteristics of data e.g. audit trails. Metadata also permit data to be attributable to an individual (or if automatically generated, to the original data source).

Metadata form an integral part of the original record. Without the context provided by metadata the data has no meaning.

Example (i) **3.5**

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

sodium chloride batch 1234, **3.5***mg. J Smith 01/Jul/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.
3.5*mg. Analyst: J Smith 01/Jul/14*

6.4. Data Integrity

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

6.5. Data Governance

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

Data governance should address data ownership and accountability throughout the lifecycle, and consider the design, operation and monitoring of processes/systems to comply with the principles of data integrity including control over intentional and unintentional changes to data.

Data Governance systems should include staff training in the importance of data integrity principles and the creation of a working environment that enables visibility, and actively encourages reporting of errors, omissions and undesirable results.

Senior management should be accountable 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 any contract/technical agreement with a third party. The Contract Giver should also perform a data governance review as part of their vendor assurance programme.

Data governance systems should also ensure that data are readily available and directly accessible on request from national competent authorities. Electronic data should be available in human-readable form.

6.6. Data Lifecycle

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

Data governance, as described in the previous section, must be applied across the whole data lifecycle to provide assurance of data integrity. Data can be retained either in the original system, subject to suitable controls, or in an appropriate archive.

6.7. Recording and collection of data

No definition required.

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

The selected method should ensure that data of appropriate accuracy, completeness, content and meaning are collected and retained for their intended use. Where the capability of the electronic system permits dynamic storage, it is not appropriate for static (printed / manual) data to be retained in preference to dynamic (electronic) data.
As data are required to allow the full reconstruction of activities the amount and the resolution (degree of detail) of data to be collected should be justified.

When used, blank forms (including, but not limited to, worksheets, laboratory notebooks, and master production and control records) should be controlled. For example, numbered sets of blank forms may be issued and reconciled upon completion. Similarly, bound paginated notebooks, stamped or formally issued by a document control group allow detection of unofficial notebooks and any gaps in notebook pages.

6.8. Data transfer / migration

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

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

Data transfer is the process of transferring data and metadata between storage media types or computerised systems. Data migration where required may, if necessary, change the format of data to make it usable or visible on an alternative computerised system.

Data transfer/migration procedures should include a rationale, and be robustly designed and validated to ensure that data integrity is maintained during the data lifecycle. Careful consideration should be given to understanding the data format and the potential for alteration at each stage of data generation, transfer and subsequent storage. The challenges of migrating data are often underestimated, particularly regarding maintaining the full meaning of the migrated records.

Data transfer should be validated. The data should not be altered during or after it is transferred to the worksheet or other application. There should be an audit trail for this process. Appropriate Quality procedures should be followed if the data transfer during the operation has not occurred correctly. Any changes in the middle layer software should be managed through appropriate Quality Management Systems.

Electronic worksheets used in automation like paper documentation should be version controlled and any changes in the worksheet should be documented/verified appropriately.

6.9. Data Processing

A sequence of operations performed on data 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 to the raw data, including attribution to who performed the activity.

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 for regulatory or business purposes. 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 result.

6.10. Excluding Data (not applicable to GPvP):

Note: this is not applicable to GPvP; for GPvP refer to the pharmacovigilance legislation (including the GVP modules) which provide the necessary requirements and statutory guidance.

Data may only be excluded where it can be demonstrated through valid scientific justification that the data are not representative of the quantity measured, sampled or acquired.
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.

6.11. Original record and true copy

6. 11.1. Original record

The first or source capture of data or information e.g. original paper record of manual observation or electronic raw data file from a computerised system, and all subsequent data required to fully reconstruct the conduct of the GXP activity. Original records can be Static or Dynamic.

A static record format, such as a paper or electronic record, is one that is fixed and allows little or no interaction between the user and the record content. For example, once printed or converted to static electronic format chromatography records lose the capability of being reprocessed or enabling more detailed viewing of baselines.

Records in dynamic format, such as electronic records, allow an interactive relationship between the user and the record content. For example, electronic records in database formats allow the user to track, trend and query data; chromatography records maintained as electronic records allow the user or reviewer (with appropriate access permissions) to reprocess the data and expand the baseline to view the integration more clearly.
Where it is not practical or feasibly possible to retain the original copy of source data, (e.g. MRI scans, where the source machine is not under the study sponsor's control and the operator can only provide summary statistics) the risks and mitigation should be documented.

Where the data obtained requires manual observation to record (for example results of a manual titration, visual interpretation of environmental monitoring plates) the process should be risk assessed and depending on the criticality, justify if a second contemporaneous verification check is required or investigate if the result could be captured by an alternate means.

6.11.2. True copy

A copy (irrespective of the type of media used) of the original record that has been verified (i.e. by a dated signature or by generation through a validated process) to have the same information, including data that describe the context, content, and structure, as the original.

A true copy may be stored in a different electronic file format to the original record if required, but must retain the metadata and audit trail required to ensure that the full meaning of the data are kept and its history may be reconstructed.

Original records and true copies must preserve the integrity of the record. True copies of original records may be retained in place of the original record (e.g. scan of a paper record), if a documented system is in place to verify and record the integrity of the copy. Organisations should consider any risk associated with the destruction of original records.

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

Data must be retained in a dynamic form where this is critical to its integrity or later verification. If the computerised system cannot be maintained e.g., if it is no longer supported, then records should be archived according to a documented archiving strategy prior to decommissioning the computerised system. It is conceivable for some data generated by electronic means to be retained in an acceptable paper or electronic format, where it can be justified that a static record maintains the integrity of the original data. However, the data retention process must be shown to include verified copies of all raw data, metadata, relevant audit trail and result files, any variable software/system configuration settings specific to each record, and all data processing runs (including methods and audit trails) necessary for reconstruction of a given raw data set. It would also require a documented means to verify that the printed records were an accurate representation. To enable a GXP compliant record this approach is likely to be demanding in its administration.

Where manual transcriptions occur, these should be verified by a second person or validated system.

6.12. Computerised system transactions:

A computerised system transaction is a single operation or sequence of operations performed as a single logical 'unit of work'. The operation(s) that makes 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 (e.g. username, date, and time) are 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 for the record to be saved and become permanent.

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. Computer systems should be designed to ensure that the execution of critical steps is recorded contemporaneously. Where transactional systems are used, the combination of multiple unit operations into a combined single transaction should be avoided, and the time intervals before saving of data should be minimised. Systems should be designed to require saving data to permanent memory before prompting users to make changes.

The organisation should define during the development of the system (e.g. via the user requirements specification) what critical steps are appropriate based on the functionality of the system and the level of risk associated. Critical steps should be documented with process controls that consider system design (prevention), together with monitoring and review processes. Oversight of activities should alert to failures that are not addressed by the process design.

6.13. Audit Trail

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

Where computerised systems are used to capture, process, report, store or archive raw data electronically, system design should always provide for the retention of audit trails to show all changes to, or deletion of data while retaining previous and original data. It should be possible to associate all data and changes to data with the persons making those changes, and changes should be dated and time stamped (time and time zone where applicable). The reason for any change, should also be recorded. The items included in the audit trail should be those of relevance to permit reconstruction of the process or activity.

Audit trails (identified by risk assessment as required) should be switched on. Users should not be able to amend or switch off the audit trail. Where a system administrator amends, or switches off the audit trail a record of that action should be retained.

The relevance of data retained in audit trails should be considered by the organisation 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.).

Where relevant audit trail functionality does not exist (e.g. within legacy systems) an alternative control may be achieved for example defining the process in an SOP, and use of log books. Alternative controls should be proven to be effective.

Where add-on software or a compliant system does not currently exist, continued use of the legacy system may be justified by documented evidence that a compliant solution is being sought and that mitigation measures temporarily support the continued use. [1]

Routine data review should include a documented audit trail review where this is determined by a risk assessment. When designing a system for review of audit trails, this may be limited to those with GXP relevance. Audit trails may be reviewed as a list of relevant data, or by an ‘exception reporting' process. An exception report is a validated search tool that identifies and documents predetermined ‘abnormal’ data or actions, that require further attention or investigation by the data reviewer.

Reviewers should have sufficient knowledge and system access to review relevant audit trails, raw data and metadata (see also ‘data governance’).

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

6.14. Electronic signatures

A signature in digital form (bio-metric or non-biometric) that represents the signatory. This should be equivalent in legal terms to the handwritten signature of the signatory.

The use of electronic signatures should be appropriately controlled with consideration given to:

- How the signature is attributable to an individual.

[1] It is expected that GMP facilities with industrial automation and control equipment/ systems such as programmable logic controllers should be able to demonstrate working towards system upgrades with individual login and audit trails (reference: Art 23 of Directive 2001/83/EC).

- How the act of ‘signing’ is recorded within the system so that it cannot be altered or manipulated without invalidating the signature or status of the entry.
- How the record of the signature will be associated with the entry made and how this can be verified.
- The security of the electronic signature i.e. so that it can only be applied by the ‘owner’ of that signature.

It is expected that appropriate validation of the signature process associated with a system is undertaken to demonstrate suitability and that control over signed records is maintained. Where a paper or pdf copy of an electronically signed document is produced, the metadata associated with an electronic signature should be maintained with the associated document.

The use of electronic signatures should be compliant with the requirements of international standards. The use of advanced electronic signatures should be considered where this method of authentication is required by the risk assessment. Electronic signature or E-signature systems must provide for “signature manifestations” i.e. a display within the viewable record that defines who signed it, their title, and the date (and time, if significant) and the meaning of the signature (e.g. verified or approved).

An inserted image of a signature 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 adequate. Where a document is electronically signed then the metadata associated with the signature should be retained.

For printed copies of electronically signed documents refer to True Copy section.

Expectations for electronic signatures associated with informed consent (GCP) are covered in alternative guidance (MHRA/HRA DRAFT Guidance on the use of electronic consent).

6.15. Data review and approval

The approach to reviewing specific record content, such as critical data and metadata, cross-outs (paper records) and audit trails (electronic records) should meet all applicable regulatory requirements and be risk-based.

There should be a procedure that describes the process for review and approval of data. Data review should also include a risk-based review of relevant metadata, including relevant audit trails records. Data review should be documented and the record should include a positive statement regarding whether issues were found or not, the date that review was performed and the signature of the reviewer.

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 provide visibility of the original record, and traceability of the correction, using ALCOA principles (see ‘data’ definition).

Where data review is not conducted by the organisation that generated the data, the responsibilities for data review must be documented and agreed by both parties. Summary reports of data are often supplied between organisations (contract givers and acceptors). It must be acknowledged that summary reports are limited and critical supporting data and metadata may not be included.

Many software packages allow configuration of customised reports. Key actions may be incorporated into such reports provided they are validated and locked to prevent changes. Automated reporting tools and reports may reduce the checks required to assure the integrity of the data.

Where summary reports are supplied by a different organisation, the organisation receiving and using the data should evaluate the data provider's data integrity controls and processes prior to using the information.

- Routine data review should consider the integrity of an individual data set e.g. is this the only data generated as part of this activity? Has the data been generated and maintained correctly? Are there indicators of unauthorised changes?

- Periodic audit of the data generated (encompassing both a review of electronically generated data and the broader organisational review) might verify the effectiveness of existing control measures and consider the possibility of unauthorised activity at all interfaces, e.g. have there been IT requests to amend any data post review? Have there been any system maintenance activities and has the impact of that activity been assessed?

6.16. 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. Where the system does not capture this data, then a record must be maintained outside of the system. Access controls should be applied to both the operating system and application levels. Individual login at operating system level may not be required if appropriate controls are in place to ensure data integrity (e.g. no modification, deletion or creation of data outside the application is possible).

For systems generating, amending or storing GXP data 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. Systems (such as MRP systems) that are not used in their entirety for GXP purposes but do have elements within them, such as approved suppliers, stock status, location and transaction histories that are GXP applicable require appropriate assessment and control.

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 that comply with current regulatory expectations[2].

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 routine 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. The intent of this is to prevent giving access to users with potentially a conflict of interest so that they can make unauthorised changes that would not be traceable to that person.

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).

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.

6.17. Data retention

Data retention may be for archiving (protected data for long-term storage) or backup (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 should be validated where appropriate (see also data transfer/migration).

Data (or a true copy) generated in paper format may be retained by using a validated scanning process provided there is a documented process in place to ensure that the outcome is a true copy.

Procedures for destruction of data should consider data criticality and where applicable legislative retention requirements.

[2] It is expected that GMP facilities with industrial automation and control equipment/ systems such as programmable logic controllers should be able to demonstrate working towards system upgrades with individual login and audit trails (reference: Art 23 of Directive 2001/83/EC).

6.17.1. Archive

A designated secure area or facility (e.g. cabinet, room, building or computerised system) for the long term, retention of data and metadata for the purposes of verification of the process or activity.

Archived records may be the original record or a 'true copy' and should be protected so they cannot be altered or deleted without detection and protected against any accidental damage such as fire or pest.

Archive arrangements must be designed to permit recovery and readability of the data and metadata throughout the required retention period. In the case of archiving of electronic data, 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). Where hybrid records are stored, references between physical and electronic records must be maintained such that full verification of events is possible throughout the retention period.

When legacy systems can no longer be supported, consideration should be given to maintaining the software for data accessibility purposes (for as long possible depending upon the specific retention requirements). This may be achieved by maintaining software in a virtual environment.
Migration to an alternative file format that retains as much as possible of the 'true copy' attributes of the data may be necessary with increasing age of the legacy data. Where migration with full original data functionality is not technically possible, options should be assessed based on risk and the importance of the data over time. The migration file format should be selected considering the balance of risk between long-term accessibility versus the possibility of reduced dynamic data functionality (e.g. data interrogation, trending, re-processing etc). It is recognised that the need to maintain accessibility may require migration to a file format that loses some attributes and/or dynamic data functionality (see also 'Data Migration').

6.17.2. Backup

A copy of current (editable) data, metadata and system configuration settings maintained for recovery including disaster recovery.

Backup and recovery processes should be validated and periodically tested. Each back up should be verified to ensure that it has functioned correctly e.g. by confirming that the data size transferred matches that of the original record.

The backup strategies for the data owners should be documented.

Backups for recovery purposes do not replace the need for the long term, retention of data and metadata in its final form for the purposes of verification of the process or activity.

6.18. File structure

Data Integrity risk assessment requires a clear understanding of file structure. The way data is structured within the GXP environment will depend on what the data will be used for and the end user may have this dictated to them by the software/computerised system(s) available.
There are many types of file structure, the most common being flat files and relational databases.

Different file structures due to their attributes may require different controls and data review methods and may retain meta data in different ways.

6.19. Validation – for intended purpose (GMP; See also Annex 11, 15)

Computerised systems should comply with regulatory requirements and associated guidance. These should be validated for their intended purpose which 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 users 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.

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.

6.20. IT Suppliers and Service Providers (including 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, 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 the 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 validated state, including validation and change control information to permit this restoration.

Business continuity arrangements should be included in the contract, and tested. The need for an audit of the service provider should be based upon risk.

7. Glossary

Acronym or word or phrase	Definition
eCRF	Electronic Case Report Form
ECG	Electrocardiogram
GXP	Good ‘X’ Practice where ‘X’ is used as a collective term for GDP – Good Distribution Practice, GCP – Good Clinical practice, GLP – Good Laboratory Practice GMP – Good Manufacturing Practice GPvP – Good Pharmacovigilance Practice
Data Quality	The assurance that data produced is exactly what was intended to be produced and fit for its intended purpose. This incorporates ALCOA
ALCOA	Acronym referring to Attributable, Legible, Contemporaneous, Original and Accurate.
ALCOA +	Acronym referring to Attributable, Legible, Contemporaneous, Original and Accurate ‘plus’ Complete, Consistent, Enduring, and Available.
DIRA	Data Integrity Risk Assessment
Terminology	The body of terms used with a particular technical application in a subject of study, profession, etc.
Data cleaning	The process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.
Format	The something is arranged or set out
Directly accessible	At once; without delay
Procedures	Written instructions or other documentation describing process i.e. standard operating procedures (SOP)
Advanced electronic signatures	an electronic signature based upon cryptographic methods of originator authentication, computed by using a set of rules and a set of parameters such that the identity of the signer and the integrity of the data can be verified.
Validated scanning process	A process whereby documents / items are scanned as a process with added controls such as location identifiers and OCR so that each page duplicated does not have to be further checked by a human.

8. References

Computerised systems. In: The rules governing medicinal products in the European Union. Volume 4: Good manufacturing practice (GMP) guidelines: Annex 11. Brussels: European Commission. (http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/vol-4/pdfs-en/anx11en.pdf).

OECD series on principles of good laboratory practice (GLP) and compliance monitoring. Paris: Organisation for Economic Co-operation and Development. (http://www.oecd.org/chemicalsafety/testing/oecdseriesonprinciplesofgoodlaboratorypracticeglpandcompliancemonitoring.htm).

Good Clinical Practice (GCP) ICH E6(R2) November 2016 (http://www.ich.org/products/guidelines/efficacy/article/efficacy-guidelines.html).

Guidance on good data and record management practices; World Health Organisation, WHO Technical Report Series, No.996, Annex 5; 2016. (http://apps.who.int/medicinedocs/en/m/abstract/Js22402en/).

Good Practices For Data Management And Integrity In Regulated GMP/GDP Environments – PIC/S; PI041-1(draft 2); August 2016. (https://picscheme.org/en/news?itemid=33).

MHRA GMP data integrity definitions and guidance for industry. London: Medicines and Healthcare Products Regulatory Agency; March 2015. (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/412735/Data_integrity_definitions_and_guidance_v2.pdf).

MHRA/HRA *DRAFT* Guidance on the use of electronic consent (http://www.hra-decisiontools.org.uk/consent/)

EU Pharmacovigilance legislation: http://ec.europa.eu/health/human-use/pharmacovigilance

The Human Medicines Regulations 2012 (Statutory Instrument 2012 No. 1916): http://www.legislation.gov.uk/uksi/2012/1916/contents/made

EU Good Pharmacovigilance Practice Modules: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/document_listing/document_listing_000345.jsp&mid=WC0b01ac058058f32c

Revision History

Revision	Publication Month	Reason for changes
Revision 1	March 2018	None. First issue.