In July 2016 the Medicines and Healthcare products Regulatory Agency (MHRA) made available for public consultation a document entitled “MHRA GxP Data Integrity Definitions and Guidance for Industry”. This document contains the response to that public consultation from Rammell Consulting Limited in the format requested by the MHRA.

Line number(s) of the relevant text	**Comment and rationale**	**Proposed Change (if any)**
Lines 31-33	The inclusion of a risk-based approach is welcomed. However, there would be an expectation from industry that inspectors are not critical of a more relaxed approach to data integrity issues where there is clearly minimal risk to the patient or to the environment. We have seen some inspectors being highly critical of issues associated with managing GCP trial master file documents when the risk from a patient safety or environmental perspective is almost non-existent.	Not applicable.
Lines 67-69	Systems often consist of multiple technologies (including paper) where data is transferred several times between disparate systems throughout the data lifecycle, often including different data owners (e.g. vendors, CROs). It is suggested that this complexity is mentioned in this paragraph.	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, the systems between which data is transferred throughout its lifecycle, and therefore potentially manipulated.
Lines 96-98	Consideration should be given to including in a risk assessment whether data exists in multiple records. For example, in a clinical trial, individual subject data is often recorded in multiple places. A reduced effort and/or frequency of control measures might therefore be appropriate as data amendment in one system/record can more readily be identified by comparison with other systems/records.	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 or for data that is captured and/or reported in multiple systems/records.

Line number(s) of the relevant text	Comment and rationale	Proposed Change (if any)
Lines 119-128	An important usability issue that impacts data integrity is end-user configuration of systems, for example, modifying system settings to "clean-up" data. It is important that end-user configuration options are limited, where possible, and this should be included in the examples given.	Add bullet: • Access to system configuration options that have the ability to modify data
Line 157	Data is often generated by an instrument of computerised system. In these cases, it is more appropriate that the data is attributable to the generating system. Data will often have a Uniform Resource Identifier to uniquely identify the source of the data.	A - attributable to the person or system generating the data
Line 163	Typographical error	Data governance measures should also ensure that data is complete, consistent and enduring throughout the lifecycle
Lines 163-164	Definitions are provided for ALCOA. It is recommended that definitions are also provided for 'complete', 'consistent' and 'enduring' in order to avoid ambiguity or misinterpretation.	Complete – the data must be whole, an entire set (from CDISC) Consistent - the data must be self-consistent and free from self-contradiction (from CDISC) Enduring – available throughout the data lifecycle
Lines 173-178	Line 176-178 appear to contradict lines 173-174. You state that a printout from a basic electronic instrument (e.g. pH meter) constitutes raw data but have previously stated that "paper copies of raw data generated electronically cannot be considered as 'raw data'." You also state that a "true copy" can be considered raw data. This section of definitions needs clarification.	

Line number(s) of the relevant text	**Comment and rationale**	**Proposed Change (if any)**
Lines 185-188	It might be helpful to understanding of the term “metadata” to mention that this is often known alternatively as “file properties”.	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. Metadata is sometimes known as ‘file properties’. It also permits data to be attributable to an individual (or if automatically generated, to the original data source).
Line 190	The statement that data has *no* meaning without metadata is perhaps too strong. Data without metadata is significantly less meaningful but does not always have no meaning at all in all circumstances. For example, textual data is likely to have some meaning without metadata and it could be interpreted reasonably based on the system that holds the data.	Without metadata, the data usually has no meaning or limited meaning.
Lines 214-217	In Lines 151-164 you have highlighted the importance of ALCOA and completeness, consistency and durability for data integrity, but here only make reference to completeness, consistency and accuracy. An inference may be drawn that the remaining characteristics of data are not important for data integrity.	The extent to which all data are attributable, legible, contemporaneous, original, complete, consistent and accurate throughout the data lifecycle.
Lines 247-248	The concept of direct access by competent authorities is recognised under GCP for trial master file documents; this guidance appears to extend the requirement to *all GxP-regulated data*. This may be problematic to achieve in many systems, especially automated data capture systems. Where raw data is being generated by a third party e.g. contract research organisation, is the expectation that the third party retain the raw data in the event of a request for direct access? It is often the case that the raw data transferred to the sponsor is not able to be accessed directly as access requires hardware and/or software that is not available at the sponsor.	

Line number(s) of the relevant text	**Comment and rationale**	**Proposed Change (if any)**
Lines 257-259	The statements regarding data destruction and data archiving are relevant but the guidance document has a section specifically discussing these topics (section 17). These two statements do not contribute to understanding of the definition of data lifecycle and should therefore be removed and incorporated, if necessary, in section 17. In the context of this definition, the important statement is that data integrity must be applied across the whole data lifecycle.	Replace existing two sentences with: Data governance, as described in the previous section, must be applied across the whole data lifecycle to provide assurance of data integrity.
Lines 281-282	The requirement to retain a record or audit trail of "all data processing activities regardless of whether the output of that processing is subsequently reported or otherwise used" appears unjustified and contrary to a risk-based approach to data integrity and record retention. If data is processed on a casual basis and is not formally used for any business purpose (e.g. processing in order to develop or perfect an analytical technique), there does not appear to be any regulatory justification to retain evidence of such processing. The requirement to retain an evidential record of transactions that are not deemed to be "official records" conflicts with most commonly-held records management principles i.e. to only retain official company records.	Audit trails and retained records should allow reconstruction of all data processing activities ~~regardless of whether~~ where the output of that processing is subsequently reported or otherwise used for regulatory or business purposes.
Line 298	The term "dynamic storage" has not been defined and is not broadly used or understood. It is used throughout this document, arguably with varying interpretations. It is recommended that a clear and unambiguous definition be provided here.	
Lines 304-305	It is assumed that "exclusion" applies to data processing and/or data reporting, though this is not explicitly stated. The statement could be misinterpreted that "excluded data" can be deleted/not saved.	Data may only be excluded from data processing and/or data reporting where it can be demonstrated through sound science that the data is anomalous or non-representative.

Line number(s) of the relevant text	**Comment and rationale**	**Proposed Change (if any)**
Lines 319-332	The example of a PDF file as static data is erroneous unless the statement is qualified as PDF/A. A user can interact with a PDF file by adding content, adding pages, filling forms, viewing embedded video objects, etc. The definition of static and dynamic data within this guidance document is not well explained, neither is the rationale for the distinction particularly clear in the context of data integrity or by the examples that are used.	
Lines 336-340	The definition of a certified copy differs from the most recent definition to be included in the ICH GCP E6 (R2) guideline. Specifically, it is not considered necessary for each copy to be accompanied by a dated signature (wet-ink or electronic) but rather that the process used for copying be validated to provide assurance that complete and accurate copies are generated. In addition, copies that have been generally accepted historically have not necessarily retained ALL of the attributes of the original, as required by this guideline. This is even demonstrated by the example given within the guideline of a digitised scan of a paper document. A digitised file loses certain attributes that are only present in the physical original. It is generally accepted that a risk-based approach should be taken to retention of attributes (or metadata) and it is only appropriate to ensure retention of attributes that are important to maintain the integrity and meaning of the content. Line 348 requires the retention of “relevant metadata”, implying that metadata that is deemed to be non-relevant need not be retained.	A paper or electronic copy of the original record that has been verified (e.g. by a dated signature) or has been generated through a validated process to produce an exact copy having all of the same information as the original and retaining attributes that are appropriate to verify its integrity. 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.
Lines 343-345	See also above comment. It is generally accepted that true copies may be generated via a validated process, meaning that the integrity of the copies does not have to be checked and documented on a per record basis.	

Line number(s) of the relevant text	Comment and rationale	Proposed Change (if any)
Line 349	In UK/European English, the word "archival" is an adjective and is therefore not the correct form to use in this sentence. "Archive" can be used as a noun or a verb or "archiving" as a verb.	It should be possible to create a true copy of electronic data, including relevant metadata, for the purposes of review, backup and archive.
Line 354	It is not understood why both "certified copy" and "true copy" are used within this document. The section heading implies that the terms are interchangeable but the text may give the impression that some "true copies" have certification and are therefore also "certified copies" whilst some are not. It is recommended that only one term is used within the text of the guideline with an explanation that the alternative term is equally acceptable (if indeed that is the case).	
Lines 354-357	As previously stated and as described in ICH GCP E6 (R2) guideline, certified copies may be generated through a validated process whereby copies are certified as complete and accurate on a process basis rather than per record. The process for certifying, including the certifying party and their authority, must accommodate this scenario.	
Line 362	The preceding text has defined the terms "true copy" and "certified copy" but a different term is now used i.e. "verified copy". Is this different from "true copy" or "certified copy"? If different, please clarify the difference. If no different, please use a single term consistently throughout the document.	
Lines 359-367	This explanation and concept may be difficult for some to fully understand. Consideration should therefore be given to inclusion of a practical example for the acceptable retention of a paper copy in lieu of dynamic electronic data.	

Line number(s) of the relevant text	Comment and rationale	Proposed Change (if any)
Lines 372-398	In many GxP systems, data may be captured over a period of time as individual "computer system transactions" but only become an official "regulated record" following a formal, documented approval process. This applies for the development of a GCP trial document within a TMF electronic authoring tool, for example. It could be inferred from section 12 that all recorded data prior to formal approval of the final data must be retained permanently and with full audit trails, whereas there are many scenarios (like the TMF example given here) where only the final approved record needs to be captured. Consideration should therefore be given to accommodating these scenarios within section 12.	
Lines 374 & 378	The term "durable storage" is not defined within this document and is not a commonly used or accepted term. It is recommended that a definition be provided here for what is meant within the context of computer system transactions.	
Lines 403-404	The proposed definition of "audit trail" is a little weak and omits the concept of "change" to a record of "record transactions".	Audit trails are metadata that provide a record of the sequence of activities or transactions that have changed the content, characteristics or properties of data.
Lines 413-414	Modification of the audit trail by a system administrator is accepted under defined circumstances. Nonetheless, there should still be a record maintained of the changes made through this route and/or that a system administrator change was undertaken.	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. Where a system administrator amends or switches of the audit trail, a record should be retained of changes undertaken.

Line number(s) of the relevant text	**Comment and rationale**	**Proposed Change (if any)**
Line 446	Directive 1999/93/EC is repealed effective from 1 July 2016. The Directive is replaced by Regulation (EU) 910/2014.	The use of electronic signatures should be compliant with the requirements of international standards such as Regulation (EU) 910/2014 (requirements relevant to 'advanced electronic signature').
Lines 448-449	In the context of this guideline, records are often signed in addition to documents. It is recommended that the word "document" is replaced by the more generic term "record" to avoid any uncertainty. It should additionally be clarified that **all** relevant characteristics of the electronic signature must be preserved in the copy (paper or pdf copy) to meet the previously stated requirements for a true copy. Typically, a text-based signature page is added to a copy of an electronically signed document that omits most of the metadata associated with the electronic signature, especially so when an advanced signature is used (e.g. hash algorithm used, certificate issuer details, revocation status of signature). Finally, it should be noted that there is potentially a conflict between the text stated here and the requirement in section 11.2 to maintain the dynamic nature of the original record when creating a true copy. A document signed with an advanced digital signature should be considered a dynamic record and cannot be accurately copied via printing or by rendering to pdf, since the dynamic nature of an advanced signature (e.g. revocation check) is lost. Whilst alternative text has been suggested, guidance on data integrity issues when dealing with various forms of electronic signatures – including advanced digital signatures – should be reconsidered.	Where a paper or pdf copy of an electronically signed record is produced the metadata associated with an electronic signature should be maintained together with the associated record. The copy should have all relevant attributes and information associated with the original electronic signature to be considered a true copy.

Line number(s) of the relevant text	Comment and rationale	Proposed Change (if any)
Line 487	We feel that taking a risk-based approach, there are likely to be circumstances within the scope of this guidance (i.e. covering all aspects of GxP) where use of generic user accounts would be acceptable. For example, an unannounced regulatory inspection may require use of an “inspector” login account to facilitate system access in a short period of time. Furthermore, the draft text used in this paragraph implies that this might be the case: *“Where the computerised system design supports individual user access…*”, suggesting that there will be cases when systems do NOT support individual user access. The following paragraph also provides for use of shared logins or generic user access in certain circumstances. The initial statement on line 487 is therefore too strong.	Shared logins or generic user access should not be used unless the specific circumstances make individual logins impossible or impractical. Where the computerised system design supports individual user access, this function must be used. This may require the purchase of additional licences.
Line 538	There is no general requirement to retain GxP data permanently, therefore the word “permanently” should be removed.	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.
Line 545	In UK/European English, the word “archival” is an adjective and is therefore not the correct form to use in this sentence. “Archive” can be used as a noun or a verb or “archiving” as a verb.	In the case of archiving of electronic data ~~archival~~, this process should be validated,

Line number(s) of the relevant text	Comment and rationale	Proposed Change (if any)
Lines 549-555	It is generally agreed that maintenance of the dynamic nature of many records is impossible to achieve in the long-term. Whilst this paragraph appears to acknowledge that this is not a mandatory requirement (i.e. take a risk-based approach), the wording and tone of the paragraph does not reflect reality. Migration to alternative file formats to ensure long term accessibility and readability invariably results in loss of some attributes. We feel that the guidance should provide more pragmatic and practical advice here. In addition, the text implies that a 'virtual environment' is simply storing data in a Cloud or SaaS environment; this is not the case.	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). 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.
Lines 567-599	The definitions provided for "flat file" and "relational database" are not consistent with the definitions typically used and understood. A flat file is simply one having no internal hierarchy; it can still be extremely complex and contain a full audit trail of changes to records. Similarly, a relational database is not always inherently resilient to change, simply by virtue of it being a relational database rather than a flat file.	