Many healthcare providers digitize their paper-based patient records from routine care, destroying originals (replacing scanning). Current regulations are unclear as to which paper-based patient records may be destroyed after digitization if the respective patients participate in clinical trials. GCP-compliant destruction is possible if both sponsors and authorities recognize the digital copies as source documents. Recognition should be based on digitized paper-based patient records complying with the requirements for certified copies defined in ‘Note for Guidance – CPMP/ICH/135/95’. This paper describes principles by which digitized patient records can be recognized as GCP-compliant certified copies, allowing the paper-based originals to be destroyed. A prerequisite is written proof that the digitization process implemented is controlled and its outcome quality is permanently monitored.

Keywords: certified copy • CPMP/ICH/135/95 • digital archiving system • digitization • GCP-compliant digitization • Good Clinical Practice • paper-based patient records • source documents

Background
Medical documentation and archiving are not only immensely important, but also regulated by law in both routine care and medical research. However, in the area of conventional documentation especially, the management and maintenance of paper-based patient record archives involve high running costs. For this reason, many healthcare providers choose to digitize their captured paper records and subsequently destroy the originals.

The digitizing procedures, well established in the area of patient care, follow generally recognized methods valid across sectors. The requirements on the digitizing of healthcare records and other records being of special interest are regulated in legal decrees and administrative provisions in the various states (for instance in Germany [1]). In order to obtain legal certainty of the digitized documents, compliance with the aforementioned legal acts is a minimum requirement. Furthermore, compliance with deontological codes of conduct and sector specific laws (i.e., in the field of radiology) is required. Whereas the requirements on the quality of the digitized documents are well laid down in these decrees and administrative provisions, the details on their procedural realization and on the quality control of the digitization process of patient records are rather poor.

At many institutions, digitized paper records are saved on microfilm as an additional measure – either using a hybrid camera simultaneously with digitization or subsequent to digitization by making a copy of the digitized originals. This procedure aims to improve the reliable, long-term availability of the records, as well as their acceptance in legal disputes. In Germany, microfilm is generally equally

---


**GCP-compliant digital archiving of paper-based patient records of clinical trial subjects: a key issues paper**

Christian D Kohl*, Insa Bruns¹, Mathias Freudigmann¹, Guido Scharf⁵, Paul Schmücker⁶,⁷, Gabriele Schwarz⁸ & Sebastian C Semler⁴
¹Network of Coordinating Centres for Clinical Trials, Cologne, Germany
²Coordinating Centre for Clinical Trials Heidelberg, Heidelberg, Germany
³Institute of Medical Biometry & Informatics, Heidelberg University, Heidelberg, Germany
⁴Technology, Methods, & Infrastructure for Networked Medical Research, Berlin, Germany
⁵Health Authority, Central Inspectorate for Clinical Trial Sites in Northrhine-Westphalia, State Capital Düsseldorf, Germany
⁶German Society of Medical Informatics, Biometry & Epidemiology, Cologne, Germany
⁷Mannheim University of Applied Sciences, Mannheim, Germany
⁸Federal Institute for Drugs and Medical Devices, Bonn, Germany
*Author for correspondence:
Tel.: +49 6221 56 34510
E-mail: christian.kohl@med.uni-heidelberg.de

For reprint orders, please contact reprints@future-science.com

---

ISSN 2041-6792

451

10.4155/CLI.13.33
as acceptable as digitized documents in legal disputes, although the use of both media is not specifically regulated for such cases [3]. Each medical institution, therefore, must individually assess and decide whether or not to destroy individual documents from a patient’s record for which the written form is legally required (see also ‘Destruction of records’) [3].

As soon as an institution not only offers medical care, but also takes part in clinical trials, it must observe specific requirements, laid down in the respective legislation and international standards – that is, the EudraLex Volume 10 clinical trials guidelines especially on Good Clinical Practice (International Conference on Harmonization [ICH]-GCP [101]). According to Point 8.3.13 of the ICH-GCP guideline, patient records from routine care are an essential component of the Trial Master File to be archived by the investigators as part of the Investigator Site File. This means that all standards and guidelines formulated for the Investigator Site File and Trial Master File also apply to the patient records of trial subjects, for example, the provisions of the GCP Directive 2005/28/EC of the European Union. Chapter 4 of this Directive contains, inter alia, requirements for storage media and systems used to store essential documents [4]. These regulatory requirements are independent of the medium used and must be complied with when the data are transferred from paper to an electronic medium.

This issue has far-reaching consequences: It is currently unclear under which conditions the regulatory authorities consider digitized or microfilm-transferred patient records of study participants as source documents according to GCP requirements. If paper-based patient records are destroyed after digitization or microfilming and the digitized copy subsequently fails to qualify as source document, this implies that there are no longer any source documents. A comparison of the data collected during the study with the original data would hence no longer be possible. This would violate international standards, guidelines and eventually national legislation applicable to the execution of the underlying clinical studies and could lead to the affected study data not being usable in marketing authorization procedures. In such a case, studies might have to be repeated. This would be unethical and often impossible for economic reasons.

Taking into account this situation, the authors’ institutions initiated expert discussions and conceptual work on these issues. In the authors’ view, there are two major challenges when scanning paper-based patient records of trial participants with the aim to generate digital copies acceptable as source documents by competent authorities:

- To understand the special quality requirements, their procedural realization and the terminology (e.g., the meaning of ‘computer system validation’) associated with clinical trials;
- To find the right balance between quality assurance measures necessary to achieve an acceptance of the digitized patient records as source documents and the effort associated with these quality assurance measures.

Setting objectives & limitations
The main objective of this Perspective Paper is to present the manner in which the existing regulatory guidelines can be implemented so that digitized paper-based patient records from routine care of trial subjects are recognized as source documents in the sense of ICH/GCP Chapter 1.52. The paper reflects the opinion of the undersigning organizations and the authority representatives as authors. Originally, electronically created documents have been expressly excluded from this paper. Nor is the archiving of other essential Trial Master File documents (e.g., contracts, device-specific parameters, certificates and the case report form pages themselves), which also must be stored in accordance with regulations, covered herein. The present Perspective Paper focuses on ‘scanned’ paper documents from patient records. However, several of the presented principles can also be applied to the digital archiving of other essential Trial Master File documents. In the following, we will generally apply the technologically neutral formulation ‘digitization’. The special requirements of microfilming will also not be covered.

If we describe a requirement in this paper and claim that it must be fulfilled, not fulfilling this requirement does not necessarily mean a violation of applicable legislative requirements. However, from the authors’ point of view, not fulfilling such a requirement bears a high risk that digitized patient records are not accepted as source documents in clinical trials.

Methods
This paper summarizes the discussions of one workshop with 52 participants in Cologne in 2011 and two workshops with 60 and 20 participants, respectively, in Berlin in 2012, which have been initiated by the authors’ institutions. The seven authors of this paper participated in these workshops as experts, collected the points mentioned there, discussed them in detail in six telephone conferences and four face-to-face meetings and developed the approach suggested in this paper based on these discussions, as well as on their own professional background and experience. The authors have been selected according to their professional background and their affiliations. To reach a common understanding among the authors with regard to the actual situation of the digitization of...
paper-based patient records and of the key points and processes necessary to reach the objective, an iterative process consisting of discussions and amicable settlements has been used. It took some further iteration to describe a digitization process, including quality assurance measures, that could be agreed upon by all authors when taking into account existing standards and legislation for clinical trials. Moreover, for some aspects the authors sought the advice of two biostatisticians and other colleagues. This paper contains the results of these discussions and of preliminary work on reliable scanning [5], the implementation of electronic signatures [3], legal and technical recommendations concerning electronic archiving [2,6] and on first GCP audits of electronic archiving at German university clinics [7].

**Regulatory requirements**

In addition to the general administrative provisions and standards governing the digitization and electronic storage of documents, special regulatory requirements apply to the records that form the source documentation of patients taking part in clinical trials. The essential regulatory standards for source documents are summarized in the “Reflection paper on expectations for electronic source data and data transcribed to electronic data collection tools in clinical trials” from the GCP Inspectors Working Group of the European Medicines Agency [8]. In Chapter 6 of this Reflection Paper, the requirements for data collection and storage in the context of clinical trials have been formulated and described extensively. In the following section, these requirements will be briefly listed and explained:

**Accurate**
The data must be accurately and carefully collected and processed. This means that the data collected must represent the observed reality and may not be manipulated.

**Legible**
The data must be legibly collected during the data collection process (this is particularly important with hand-written data). Media and storage systems and data processing must furthermore ensure that the data are legible when needed.

**Contemporaneous**
The data must be collected promptly after a patient visit. ‘Promptly after’ may only be deviated from in exceptional, duly warranted cases.

**Original**
This point is described in detail in the following section.

**Attributable**
The data must be clearly attributable to:
- The corresponding patient;
- The individual collecting the data;
- The time of collection.

**Complete**
The data must be complete with reference to the information to be recorded according to the protocol and to the diagnostic, therapeutic and other relevant measures, which were implemented.

**Consistent**
The data collected must be consistent and unambiguous.

**Enduring**
The data must be reliably stored for the prescribed period of time [9].

**Available when needed**
The data must be available on demand when needed (e.g., during inspections). This also implies that data sets can be timely and selectively retrieved.

**Originality**
The digitization of paper-based records creates electronic copies of original records. When the original and certified copies are equivalent as defined in ICH/GCP Chapter 1.51, the originals can be destroyed. ICH/GCP itself uses the term ‘certified copy’, but does not define it as such. Concerning this point, the European Medicines Agency’s Reflection Paper references the Clinical Data Interchange Standards Consortium (CDISC) glossary, which itself borrows a definition from the US FDA and expands upon it [10]. Thus, a copy is deemed certified if there is documentation showing that someone examined/ensured that it was an exact copy. Exact, in this case, means correct and complete in the sense that the copy contains the same attributes and information as the original. According to the European Medicines Agency’s Reflection Paper, the correctness and completeness of the copy can either be documented by the examiner’s dated signature or guaranteed by a validated digitization process (see ‘Validation of the digitization and archiving process’). Note, however, that the FDA has not yet adopted the expanded definition from the CDISC Glossary quoted by the European Medicines Agency, whereby certification may also take place through a validated digitization process.

**Key points for regulatory compliance**
To guarantee the regulatory standards (see the section ‘Regulatory requirements’) and prove they have been
met, the entire process – from the creation to the archiving of paper-based patient records of trial subjects – should be controlled and documented. This implies consideration of not only the actual process steps, but also the organization and personnel used, the basic process documentation and the implemented hardware and software components. In order for digitized patient records of trial subjects to be recognized as source documents, the overall concept for archiving the records must be based on a controlled process and validated electronic systems. A suitable quality management system is required to monitor the process. This includes classical elements such as process descriptions/Standard Operating Procedures (SOPs), training measures, quality and change control, as well as preventive and corrective actions for known deviations.

In the following, a process by which the paper-based records are digitized for archiving is presented. This process may consist of the following process steps:

- Approval for archiving;
- Transportation of the records to the location of digitization;
- Preparation and digitization of the records;
- Indexing and assignment of metadata;
- Import of the digitized records into a digital archiving system;
- Quality control;
- Destruction of the records;
- Access to the archived records;
- Changes to documents and metadata;
- Migration of digitally archived records;
- Deletion of digitally archived records.

In the following sections, we outline the key issues for all of the process steps that should be considered to comply with the regulatory requirements mentioned in ‘Regulatory requirements’.

All steps and procedures mentioned in this chapter should be described in SOPs.

- **Approval for archiving**
  The archiving process starts when a patient record is handed over to the part of the organization responsible for archiving. In the case of in-patient care, this is the case when the patient has been discharged, all diagnostic findings requested have come in and the final discharge papers are ready. The employees working in a healthcare provider’s archive typically are not capable of evaluating whether a record sent in for archiving is complete and has been sorted according to an established and documented record structure. Often, this is only possible for the individual who created the document or to whom it belongs – typically the medical and nursing personnel in a ward, ambulance or other healthcare institution. Thus, the responsibility of this circle of individuals is to verify that a record is complete and is ready to be archived. Confirmation of the completeness and approval to archive a record should therefore be given by an authorized person and documented on a supplementary sheet, which is archived with the record. If necessary, a checklist can be printed on such a supplementary sheet in order to guarantee that a record meets all of the prerequisites for archiving of a particular healthcare institution (e.g., a particular sorting of the documents contained therein).

- **Transportation of records to the location of digitization**
  As soon as a record is approved for archiving, it should be brought to the location of digitization. During this process, suitable technical and organizational methods should be used to ensure that the records are promptly transported, excluding the possibility of loss, unauthorized access or manipulation. This can be achieved, for example, by regular retrieval times and transport in closed containers. The transport of the records should be described in the corresponding SOP. Collection points for the records to be archived should be set up, as well as times for their regular retrieval, and the wards and ambulances should be made aware of these. When choosing central collection points, careful attention should be paid to whether the records will be sufficiently protected until they are retrieved.

- **Registration of incoming records at the location of digitization**
  Incoming records delivered to the digitization facility should be registered, for example, in a central record administration system by recording the patient identification and the respective case identification. A short turnaround time is recommended. This way, if a record is needed on short notice, it can quickly be determined whether it is already in the process of being archived. A basic visual inspection of the records being archived is also recommendable. This ensures the possibility to recognize problems – for example, a cover sheet that has not been completely filled out – early and to clarify these with the sender.

- **Preparation & digitization of the records**
  Before digitization, the records should be prepared in such a way that their entire contents can be recorded and associated to the right patient.
Information covered by sticky notes or bent corners should be uncovered. The information on the sticky notes should also be recorded.

The patient records of individual patients should be separated by dividing sheets or processed independently of one another. During the digitization process, it is imperative to prevent different patients’ patient records from being mixed up inadvertently.

Different sheets from the same document (e.g., multisheeted doctor’s letters or laboratory findings) should not be separated from one another during digitization. However, sensible sorting of the documents contained in a record can take place before digitization and indexing (see the section ‘Indexing & assigning of metadata’). If the documents in a record are summarized into classes (e.g., doctor’s letters, laboratory findings, and so forth) within a digital archive, the already-existing document order of the paper record should be maintained within the document class, given that there is no indexing of the documentation times.

The front and reverse sides of a sheet should be digitized together.

Blank sides should either be saved in the system and then hidden later, or they can be removed based on a precisely controlled and documented procedure that can verifiably ensure no loss of information.

When using feed scanners, it should be ensured that no more than one sheet is pulled in at a time.

Color-coded information should also include these colors in the digitized record.

In the event that paper-based documents or other kinds of source documents (e.g., X-rays), are not digitized, but rather stored conventionally, this should be precisely regulated and described. It must be possible to locate separately archived records on short notice.

There should be a clear procedure for the dividing-up and digitization of oversized documents that ensures the documents can be unambiguously re-assigned to each other after digitization.

The digitization process should be designed in such a way that damages of the source document that might have led to information loss are recognizable as such. The digital copy may not give the impression that the affected document sections did not contain any information.

If any problems concerning the record content or structure are identified during the digitization process (e.g., different patients’ patient records being combined), the record should be returned to the document owner for clarification. Said return, along with the time, processor and reason for return should be centrally documented (see ‘Registration of incoming records at the location of digitization’).

It needs to be ensured that the entire number of documents in a record are present and have been processed completely and correctly.

Compliance with the requirements for the creation of certified copies in accordance with the section ‘Originality’ in ‘Regulatory requirements’ should be documented in an appropriate way (preferably by electronically signing every digitized record). However, a paper-based documentation for every generated record would be possible or – according to European Medicines Agency’s Reflection Paper [8] and the CDISC definition of a certified copy – a validation of the digitization process.

A procedure that makes it possible to digitize and add subsequently filed documents (e.g., from a treatment case) to an already-digitized record or treatment case should be established and described. This is relevant, for example, when a laboratory finding for a treatment case arrives at a ward after the associated record has already been approved for archiving.

Digitized documents should be secured by a digital time-stamp in order to make verifiable the integrity and authenticity of the data over an extended period of time. If algorithms used to calculate the hash values and used for encryption lose their validity, an update of the data’s time stamp is required (A hash value can be seen as a number that is the fingerprint of a digital record. This unique number is calculated based on a mathematical function and the record itself. Even a slight change of the record would lead to a different hash value. A digital time-stamp makes use of a record’s hash value and proves that the record has been unchanged since the timestamp was generated) [11].

Indexing & assigning of metadata

After paper-based records have been digitized, they should be indexed and fed into the digital archiving system along with a reference to the individual patient or treatment case, to allow the record to be found on short notice using the patient’s identification. Other possible options include searching by case identification or the name of the patient. In addition to the IDs mentioned above, it is recommended to use other metadata (e.g., treatment date, author and document class) for indexing. These must be determined before starting the digitization process. The determination as to which
metadata will be used is very important for later work with the digital records and should therefore reflect the users' requirements concerning retrieval and access to the stored documents. The indexing of electronic documents thus has the same basic purpose as placing paper documents into different folders within a paper record. Additionally, indexing allows for a deeper record structure and a greater variety of access possibilities.

As manual indexing is very time-consuming and expensive, automated indexing usually makes sense. Information such as patient identification, document type or author can be encoded in the documents in advance or read using barcode recognition during indexing. Using optical character recognition technology, textual information could also be digitized for use in indexing. Furthermore, it is common practice to use dividing sheets containing encoded information within records, which can be recognized using barcodes or optical character recognition during digitization and used to create document classes within a digital record.

■ Import of the digitized records into a digital archiving system

Import of the digitized records into a digital archiving system should take place by means of an automated process, as manual processes are too time-consuming and susceptible to error. During import, the integrity of the data being imported should be verified. If documents are imported by batch processing, it should be ensured that all of the documents have been imported, for example, by comparing with a delivery list.

■ Quality Control

The quality of digitization and indexing should be regularly verified and monitored. This random checking for the purpose of process quality control should be distinguished from visual inspection as a single step in the digitization process, during which the processor verifies that individual copies were created both correctly and completely.

Final verification of the process quality by means of random checking is required and should preferably be undertaken by the healthcare institution itself (even if digitization of the records has been contracted out to a service provider). During this quality check, a defined random sample of digital copies is compared with the original paper records before they are destroyed in order to verify the quality of the archiving process. This check should not be performed by the same individual who digitized or indexed the records.

The size and frequency of the random checks should orient themselves on the given conditions (see the section 'Risk-based approach'), particularly on the quality of the process used. Different methods may be used to
determine the sample size, for example, the standard ISO 2859-1. In-process inspections can lead to improved quality of the final checking process; however, they do not need to be statistically planned. The main challenge lies in using a procedure that allows for sufficiently reliable outcome quality at a maintainable level of effort.

In the following section, we will present such a procedure, based on ISO 2859-1. The digitization process here is assumed to be a 'black box' – to assess the quality of the results, only the input (paper-based patient records) and the output (electronic copies in the digital archive) are taken into consideration. The quality control and assurance measures taken during the process are not directly taken into account, but rather only make their way into consideration by means of their effect on the quality of the results (potentially fewer errors).

Digitization of patient records is generally a continual process. In order to verify the outcomes' quality, all of the digital copies generated during the digitization process over a set period of time are grouped together into one lot. The failure rate of this lot is estimated using the failure rate of a representative random sample taken from that lot. The following assumptions are made:

■ The quality is defined by the number of incorrectly digitized pages per lot. The lower the failure rate, the higher the quality.

■ The unit being observed is the page and not the sheet, document or record. A sheet of paper printed on both sides, therefore, is made up of two pages. A document consists of one or more pages. In turn, a record consists of one or more documents. As individual pages and not entire records are being observed, the scope of the records processed is irrelevant to further observations.

■ A page is considered erroneous as soon as one of the following errors is identified, although this list does not purport to be exhaustive:

- Page was not digitized/cannot be located in the digital archive;
- Page was digitized illegibly (e.g., with low contrast);
- Page was not digitized completely (e.g., because of a folded corner);
- Information on the page was lost in the digitization process (e.g., black and white scan of a page with color-coded information);
- Pages or documents were assigned the wrong metadata (e.g., assigned to the wrong patient/treatment case or to the wrong document class);
A binary value (correctly/incorrectly processed) is determined for each page.

- If the same error occurs in all pages of a document or all documents within a record (e.g., assignment of all documents within a record to the wrong patient), it is just counted once. The intention for this is that multiple errors may be the result of one single cause (e.g., the selection of the wrong patient). Counting all resulting errors would lead to a biased recognition of the process quality. The remaining ‘associated pages’ are not counted as follow-up errors. However, if the same error occurs again after at least one page was correctly digitized, then the error must be counted anew.

- As soon as one page shows a different type of error than the previous page, the current page is likewise considered erroneous. A page also counts as erroneous if it not only contains the same error as the previous page but at least one other error in addition (e.g., all pages of a record are assigned to the wrong patient and one page shows low-contrast digitization).

- Furthermore, it is assumed that the quality of the digitization process remains more or less constant during the observed time period/the ‘production’ of a lot. The time periods chosen, therefore, should not be too long. This is also true for practical reasons: the larger the time period observed, the more paper records will be digitized and must be stored/redigitized if the observed error rate is too high.

- In order to ensure digitization quality, the machines used should be continuously maintained and (new) personnel consistently trained.

ISO 2859-1 can be used complementarily to these preliminary considerations to determine the minimum number of pages in a lot (which is the sample size) that must be checked in order to ensure that the true error rate of the digitization process is less than the maximum tolerable error rate, taking statistical fluctuations into consideration. The standard was specifically developed for evaluation of a continuous series of lots according to the following procedure:

- First, the size ($N$) of the lot being observed is determined. A lot could consist of all of the records that were digitized over the course of one week, for example. This would mean that a new lot is generated and reviewed every week. For the following example, it is assumed that, in a given clinic, 800 paper records with an average of 20 pages are digitized in the corresponding week. The size of the observed lot is therefore $N = 16,000$ pages.

- Additionally, the general inspection level of the sample survey must be determined. Three inspection levels are differentiated: I, II and III. The higher the inspection level, the greater the ability of the procedure to differentiate whether a lot is good enough or not. However, the effort expended to perform the random checks rises along with this improved ability to differentiate. As a considerable amount of selectivity is required due to the importance of the process, inspection level III is considered necessary.

- The size of the lot and the inspection level combine to form the sample size ($n$) – the values for $n$ can be taken from tables included in the standard (Tables 1 & 2). In the example mentioned, the lot size of $N = 16,000$ and an inspection level III leads to a sample size of $n = 500$ pages to be examined within the framework of a normal inspection. In the chosen sample, therefore, a minimum of 500 pages would have to be inspected. At an average of 20 pages per record, approximately 25 records would have to be checked.

---

### Table 1. Code letters for sample sizes according to ISO 2859-1:2004-01.

<table>
<thead>
<tr>
<th>Lot size</th>
<th>General inspection levels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
<tr>
<td>2–8</td>
<td>A</td>
</tr>
<tr>
<td>9–15</td>
<td>A</td>
</tr>
<tr>
<td>16–25</td>
<td>B</td>
</tr>
<tr>
<td>26–50</td>
<td>C</td>
</tr>
<tr>
<td>51–90</td>
<td>C</td>
</tr>
<tr>
<td>91–150</td>
<td>D</td>
</tr>
<tr>
<td>151–280</td>
<td>E</td>
</tr>
<tr>
<td>281–500</td>
<td>F</td>
</tr>
<tr>
<td>501–1200</td>
<td>G</td>
</tr>
<tr>
<td>1201–3200</td>
<td>H</td>
</tr>
<tr>
<td>3201–10,000</td>
<td>J</td>
</tr>
<tr>
<td>10,001–35,000</td>
<td>K</td>
</tr>
<tr>
<td>35,001–150,000</td>
<td>L</td>
</tr>
<tr>
<td>150,001–500,000</td>
<td>M</td>
</tr>
<tr>
<td>500,001 and over</td>
<td>N</td>
</tr>
</tbody>
</table>

The definitive version for the implementation of this standard is the edition bearing the most recent date of issue, obtainable from Beuth Verlag GmbH, Burggrafenstraße 6, 10787 Berlin, Germany. Reproduced by permission from [15].
In addition to the size of the lot and the inspection level, the Acceptance Quality Limit (AQL) must be determined. The AQL is the percentage of erroneous or undigitized pages per lot, which is still acceptable. Assuming the AQL is 0.25%, then according to the standard, the sample of 500 pages would only be allowed to contain a maximum of three erroneous or undigitized pages for the corresponding lot to be acceptable (Tables 1 & 2). As soon as the sample contains four or more erroneous pages, the quality level of the lot is lower than the determined AQL. In this case, the lot would be unacceptable – the affected paper records would have to be redigitized. According to ISO 2859-1, extending the sample size is not an option in that case. However, ISO 2859-1 also provides double and multiple sampling plans. An example for such a double sampling plan is given below.

If two out of five consecutively drawn lots are rejected, the ISO 2859-1 standard requires switching to a tightened inspection. In that case, only two erroneous pages per sample of 500 pages are permitted in the observed example – a sample will be rejected after three erroneous pages. If five consecutively drawn lots are rejected during the tightened inspection, then inspection of the lots based on representative sampling is suspended and all digitized pages in the lot(s) rejected must be subjected to a final check until the production process is improved. However, manually checking every single page for completeness and correctness might not be feasible in practice because of the lot size. An alternative would be to completely suspend the destruction of paper originals until the production process has improved. Consequently, the originals that were not destroyed should be stored conventionally or reprocessed in an improved manner. After the process has improved, representative sampling can recommence with a tightened inspection. The tightened inspection must be carried out until five consecutive lots have been accepted. Only then can a normal inspection be reintroduced.

If normal inspection proves that the process in question is consistently compliant with ISO 2859-1 (this is systematically determined in the form of a ‘switching score’), then a reduced inspection can be

Table 2. Sampling plan for a normal inspection according to ISO 2859-1:2004-01.

<table>
<thead>
<tr>
<th>Code letters for sample size†</th>
<th>Sample size</th>
<th>Acceptance Quality Limit of nonconforming items (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.10 (AN‡)</td>
<td>0.10 (RN§)</td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>315</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>800</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>1,250</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>2,000</td>
<td></td>
</tr>
</tbody>
</table>

†Indicates the maximum number of erroneous or undigitized pages a sample may contain for the digitization quality of the associated lot to be accepted.
‡Indicates the cut-off point at which the digitization quality is no longer acceptable.
§Indicates the maximum number of erroneous or undigitized pages a sample may contain for the digitization quality of the associated lot to be accepted.

From Table 1.

The definitive version for the implementation of this standard is the edition bearing the most recent date of issue, obtainable from Beuth Verlag GmbH, Burggrafenstraße 6, 10787 Berlin, Germany.

Reproduced by permission from [15].
employed. In such a case, only a sample taken from 200 pages would have to be inspected; whereby the lot would have to be rejected if there are more than two erroneous pages in the sample.

- The AQL is not to be understood as the desired level of quality, but rather, as a minimum level of quality. The production process used should always deliver a level of quality superior to the AQL. Otherwise, there is a danger of having to ‘tightly’ the inspection procedure, which in turn can quickly lead to having to suspend the representative sampling procedure.

- By switching between reduced, normal and tightened inspections, the general inspection level determined remains unchanged.

Alternatively to the procedure presented here, the standard also allows for the possibility to take double and multiple samples. In this procedure, first a smaller sample is inspected to check whether the error rate is under the AQL (this would be n = 315 pages for the normal inspection if we were to continue the previous example). If the predetermined number of allowed errors is exceeded (one error in the example), but a second limit, starting at which the lot must be rejected (three errors in the example) is not reached, then an additional sample may be taken to complement the first (the scope of the second sample also being equal to 315 pages). The errors found in the two samples are added together and compared with the maximum allowable number of errors for the second sample (four errors in the example).

Currently, it cannot be definitively said which upper limits for the error rate are practical and acceptable from a regulatory standpoint. The creation of error classes, to which different AQLs could be assigned and inspected, is also conceivable. The error classes could be created based on a severity classification in accordance with Failure Mode and Effects Analysis (see the section ‘Risk-based approach’ for more on this subject). In this way, errors such as missing or illegible pages, for example, could be given lower tolerable error rates than pages that were correctly digitized and assigned to the correct patient, but to the wrong document class. The authors encourage a review of the feasibility and significance of different methods with regards to:

- Upper limits for tolerable error rates;
- Determination of the AQL;
- Creation of severity classifications and error classes in accordance with Failure Mode and Effects Analysis approach;
- Required sample size;
- The required statistical power of the random sampling test;
- Within the framework of evaluation projects, and to examine their results with progressive technological advancements.

The example given above serves the purpose of illustrating the method and effect of target sizes on the sample size; however, it also reflects the order of magnitude the authors believe the error rate should move in.

Until empirically determined data and guidelines from public authorities on accepted target values are available, the commissioning entities themselves will have to make justifiable assumptions to determine sample size. When defining and measuring error rates for the digitization process, only those errors that are most probably caused by the process should be taken into consideration. If, for example, two documents from two different patients are filed into the same record during the creation of the paper-based record, then this must be corrected upon discovery; however, this does not constitute an error caused by the digitization process itself. The error would have just as easily occurred and presumably not been discovered if the paper record had been conventionally archived. Nonetheless, such errors should be recorded, and measures should be taken to prevent them if they are discovered during the digitization and associated quality assurance processes.

The approach a healthcare institution ultimately decides on, as well as the inspection level and the execution of sample surveys, should be documented in detail. It is equally as necessary to document errors identified during random inspections, as it is to document the measures taken to correct them. It is recommendable to use systematic corrective and preventive actions. The authors also recommend that error statistics be tracked to identify systematic errors and find suitable solutions. Errors discovered, for example, by the user while accessing the digital archive, should also be fed into these statistics. In addition to the sample size itself and specification of the basic population the sample was drawn from, the following parameters could also be examined for each sample:

- Number of missing/undigitized pages and documents;
- Number of incompletely digitized pages and documents (e.g., missing margins of a page in the digital image);
- Number of incorrectly digitized pages (e.g., black and white digitization despite relevant color information in the original);
Destruction of records
After the documents have been digitized and stored in a digital archive, the original paper documents should be retained for a restricted period of three to six months, but at minimum until the quality control measures described in the section ‘Quality control’ have been completed. If no issues are identified during this time, the paper-based patient records can be destroyed in compliance with data protection requirements. Taking into account the implications for providing legally recognized evidence and after consultation with its liability insurer, the healthcare institution should decide whether documents legally requiring a signature should be sorted out and retained instead of being destroyed (see ‘Background’) [3].

Occasionally, informed consent forms of trial subjects are not filed and stored in the investigator’s site file for the clinical trial, but rather as part of the patient record of the patient. In these cases, depending on national guidelines and legislation, it might be necessary that the digitization process takes place in such a manner as to completely prevent these documents from being destroyed with others in the process of replacing scanning. The informed consent forms of the patients are to be stored as legally valid instruments in their original paper form.

Deletion of paper-based patient records of trial subjects and other original source documents after digitization requires the sponsor’s written agreement. This agreement covenant allows the sponsor to fulfill its organizational responsibility for proper conduct of the clinical trial at the investigator’s site.

Access to the archived records
Normal users accessing digital archiving systems from a ward or outpatient department should exclusively be granted read-only access. The digital archiving system should also be capable of restricting access to the digitized records to authorized individuals; that is, primarily to medical and nursing personnel in the context of treatment. Additionally, it should also be possible to provide study-related access rights, so that the access rights of an inspector or monitor can be restricted to those patients who are participating, or have participated, in a clinical trial. In general, national data protection laws and recommendations have to be taken into consideration.

Changes to documents & metadata
If it becomes necessary to modify metadata of a document in the digital archive (e.g., because of incorrect indexing), then the archive system should have the ability to document which changes were made by whom and why, and to do so in a traceable and manipulation-proof manner via audit trail. These changes may only be made by a small group of designated individuals expressly given the right to do so. Furthermore, it should be considered that a document may be printed out from the digital archive, modified by hand and then redigitized. In such a case, it is important to ensure that the new digital copy is recognized as a new version of a document already existing in the archive, and that it is linked to the old document by means of some kind of ‘electronic staple’.

A user searching for the corresponding document should only be able to find the most recent version of the document in the archive, with a reference to the fact that there is an older version available.

Migration of digitally archived records
Rapid technological advancements generally require the migration of digitally archived records to other storage media and technologies after 5 or 10 years. Further clarification of this will follow in the section ‘Technical Aspects’.

Deletion of digitally archived records
Under no circumstances may records in the electronic archive be deleted within the respective legally required archiving periods. On the other hand, digitally archived documents may be deleted after their legally required storage periods have ended, in accordance with data protection legislation. If the patient records are source documents for clinical trials, then contractual arrangements with the sponsor also have to be observed.

Deletion should take place based on the ‘four eyes principle’. From a legal, data protection standpoint, complete anonymization is equivalent to deletion of the documents [12].

Relevance of this process & the listed requirements
A digitized patient record that has not been generated in accordance with the aforementioned requirements might be acceptable for routine care purposes but not as a source document for a clinical trial, as GCP provides for specific standards and requirements (e.g., the ‘certified copy’). Furthermore, at least in some countries, for example, Germany, although digitizing procedures for patient records from routine care follow generally recognized methods, no legally binding detailed standards and requirements have yet been introduced. This leaves grounds for uncertainty with regard to the legal
Validation of the digitization & archiving process

In order for digitally archived paper-based patient records of trial subjects to be recognized as source documents, the associated digitization and archiving processes, as well as the systems implemented, should be validated. Validation establishes documented proof that a specific process consistently creates products that meet the predetermined specifications and quality characteristics with a high level of reliability [102].

Thus, validation should provide documented evidence that the digitization process:

- Does not add, modify or lose data during the entire process;
- Allows the timely digital reproduction of a document and that the digital reproduction is visually consistent with the original document;
- Makes it apparent who the author of the document is.

Within the context of prospective validation, the entire process, including all of the systems involved and their interfaces, must be considered. Not a guideline but a good aid to orientation is the Good Automated Manufacturing Practice. In the following, an outline of different aspects of validation will be presented to facilitate an introduction to the subject.

■ Validation plan

Validation should be based on a predetermined plan. A validation plan is a plan put down in writing that describes how validation should be carried out in order to determine whether a process or product is appropriate for a defined purpose. The validation plan includes product characteristics, a list of the hardware and software used (including their integration), process descriptions, test parameters and decision criteria with respect to the acceptance of test results [13]. The validation plan can also reference other documents, such as test plans or system and interface descriptions. The execution of validation in accordance with the validation plan is documented in writing. This documentation includes completed test logs and reports, as well as system and process approval (in the case of successful validation).

In the following, we will describe important points that should be part of the validation process in the case of digitization and archiving of paper-based patient records and should be considered in the validation plan.

■ System description & IT security

The hardware and software components used, as well as the existing interfaces and their interaction, should be described in detail in a technically oriented system description. The backup and security measures taken (e.g., measures to protect against unauthorized access) and emergency plans in the event of the failure of system components should also be explained. The chosen role and rights concept should be described (who has what kind of access to which data?).

■ Qualification of hardware, software & personnel

Prerequisite to a controlled process and valid systems is the appropriateness of the hardware and software used, as well as the qualification of the personnel involved. Appropriateness in the context of computerized systems refers to the production of proof, especially by means of checks and tests, that the systems were designed according to the requirements, correctly installed, function correctly after installation, and ultimately perform under load in real operating conditions [103]. Furthermore, all maintenance and inspection work must be documented. Similarly, the qualifications of the employed personnel must be documented by keeping their short CVs, initial training plans and training certificates on file.

■ Archive framework manual

The archive framework manual is the overarching document organizing, among other things, the digitization and archiving process of paper-based patient records into the entirety of the archiving-related processes. It also describes the archive organization, structure and workflows and defines global standards and/or legislation, such as storage periods and access rights and responsibilities.
SOPs
In order for the archiving process and, especially, the steps to digitize and index paper-based patient records, to be controllable and comprehensible to third parties, they must be described with the help of SOPs. Responsibilities and competencies must be clearly defined. For example, it is necessary to define who may approve a record for archiving or who is responsible for monitoring outcome quality. For the control of outcome quality, sample size, acceptance criteria and error limits must also be determined, as well as the measures to be taken in the case of error (see the section 'Quality control').

To enable monitoring of the digitization and archiving process, validation of the systems utilized as well as to maintain their validation status, supplementary SOPs are particularly necessary on the following topics:
- The creation, introduction and repeal of SOPs;
- Validation and risk assessment;
- Approval, release, implementation and inspection of changes (Change Management).

All SOPs must be set out in writing.

Risk-based approach
Practicability and financial viability play an important role in the implementation of the guidelines presented. A risk-based approach should help to take account of this. The risk associated with every process step within the entire process must be evaluated. The risk can be seen as a product of three factors [104]:
- Probability of occurrence of an error or damage;
- Probability that the error will not be detected and corrected during the process itself;
- Severity of the error or damage.

All potential errors that occur during the digitization process should be detectable, although certain sources of error can be excluded as early as during the process planning stage (e.g., by selecting the optimal scanning resolution before scanning). It is also important to decide what is considered an error. Possible errors have already been described in the sections 'Preparation & digitization of the records' and 'Quality control'.

The determined risks dictate the scope of the testing and qualification measures to be performed: before operational start-up, whenever changes are made to the process or system, and while operations are running. For example, it stands to reason that the process of scanning from the conventional to the digital medium bears many sources of error (change of media, use of complex technology) and therefore requires special measures to be taken to ensure that the copies created exactly match the originals.

Technical aspects
Due to rapid technological progress and the limited durability of systems, software and storage media, migration concepts are an integral component of digital archiving systems. A migration concept ensures that:

- Only those technologies that allow later migration are chosen (e.g., by means of standardized user data formats, readable/exportable metadata and standardized signature procedures);
- Migration projects are planned and implemented in a timely manner by those responsible, before the technology 'expires', and based on the current state of technology at that time.

In addition to the migration of archive contents to newer hardware and software systems, the required transformation of data formats and updating of electronic time stamps and signatures must be possible and carried out on time, if necessary. In general, the same guidelines as for the creation of the original copy in the digitization process apply to validation and visual conformity. Caution is particularly important when converting data formats, as visual conformity is not necessarily achievable after conversion [14,105].

Microfilming does not produce any additional legal quality. However, microfilm, when correctly stored, offers an analog medium for backup security that is stable for decades and readable with simple equipment.

Generally, a digital archiving solution should fit into the general IT strategy of the respective institution and be selected to suit its archiving concept.

Involving external service providers
Many healthcare providers have their paper-based patient records digitized and archived by service providers. The quality standards presented in the paper are also to be upheld by contractually obligated service providers and it is therefore necessary to sign contracts with the service providers that precisely define the services to be performed and the expected process and outcome quality. It is also necessary to determine how and who is responsible for monitoring process and outcome quality. It is highly recommended that the contractors themselves perform the random inspections of the outcome quality of the digitization and archiving process.

If digitization does not take place on site at the healthcare institution, but rather at the service provider’s place...
of business, then the arrangements governing the secure transport to, and storage of the records with the service provider must be agreed upon. This also includes checking that all of the delivered patient records actually arrived at their destination. In general, provisions on data protection and IT security must be agreed in writing by all of the parties involved. It is also important to define the procedure to be followed in the case that a record sent in for digitization is urgently required.

The qualification of the chosen service provider must be verified at their place of business by means of regular ‘contractor audits’. It is recommended that these audits be performed personally by the healthcare provider and not be contracted out, as this is the only way for said healthcare provider to get a representative impression of the service provider. If necessary, additional external auditors can be involved. Furthermore, it is highly recommended that a procedure be agreed in the event that the service provider goes out of business or declares bankruptcy.

**Conclusion**

A healthcare institution involved in a clinical trial is responsible for proper archiving, and thus for compliance with existing regulatory standards. To assist these institutions in fulfilling their responsibilities, the authors’ objective with this paper is to give an overview of the existing requirements and, with it, instructions as to how these can be implemented in practice. If the principles described in this paper are observed, the authors believe that the digitized patient records of trial subjects can be recognized as source documents and the original, paper-based patient records can be destroyed. A prerequisite to this is that the respective institution having its paper records digitized be able to provide written proof that the digitization process implemented is clearly controlled and that its outcome quality has been checked and is known. If tasks are delegated to a service provider, this includes that compliance with the associated regulations and provisions is to be verified by the commissioning healthcare institution in the form of regular audits at the place of business of the service provider.

According to ICH-GCP, a ‘certified copy’ can serve as source document. Therefore, it is necessary to ensure that the digitized records meet the criteria for a ‘certified copy’. According to the European Medicines Agency reflection paper, either the check of every single page that has been digitized (not practical in most cases) or the validation of the entire digitization and archiving process would be the prerequisite for recognition of the digitized record as certified copy [8]. The requirements that would have to be observed for the validation of the process have been summarized, elaborated and discussed in this paper. Especially for the process of digitization, the type and scope of the validation/testing measures is highly dependent on the process design and the technology used. If each document is checked after digitization when doing individual scans, less laborious quality control measures are necessary than when scanning stacks of documents. There are different approaches to determining sample size; the algorithm presented in this paper represents one possible approach. As already mentioned, there is a current lack of empirical data, as well as guidance from public authorities, on sample size. Until regulatory authorities set guidelines on accepted target values, justifiable assumptions have to be made by the institutions themselves. Based on their current state of knowledge, the authors are of the opinion that a maximum of 0.25% is acceptable for the AQL.

Digital archiving offers the great advantage that records are available simultaneously in different places. When properly organized, the data can be found more quickly and easily than in paper-based archives. Furthermore, if the technology is implemented effectively, it is almost impossible to ‘lose’ a patient record. These aspects should be considered when evaluating digitization and archiving processes to identify potential errors that may be inherent to the digitization process itself.

Many of today’s healthcare institutional archives are indeed qualitatively well structured and organized, for example, with reliable digital archiving systems. In order to make those digital archiving systems GCP-compliant as well, it is sensible to first check which of the structures and documents required by ICH/GCP are already in place. Although existing elements (for instance documents such as process descriptions) may occasionally have different names than those used in the context of GCP, their content and functionality are compatible with their GCP equivalents. The second step, then, would be simply to put in place the elements still missing.

In the event of legal disputes, juridical uncertainty may not be avoidable – despite validated processes and systems employed for digitizing documents. However, this uncertainty can be counteracted by implementing technically secure, quality-assured digitizing and indexing processes, and designing and maintaining digital archiving systems in a manner that is both technically and organizationally GCP-compliant and reliable. Validation of processes and systems can be of great assistance in this endeavor.

**Future perspective**

Several of the principles presented in this paper can also be applied to the digital archiving of other essential Trial Master File documents. Originally electronically created documents (e.g., digital patient records) and their archiving were not among the topics discussed in this Perspective Paper, although the authors are aware that these documents will become increasingly prominent within the next 5–10 years. The associated consequences should be dealt with in a separate paper.
Managing and maintaining archives for paper-based patient records is costly. Therefore, many healthcare providers digitize their paper records, destroying originals thereafter.

Current regulatory requirements governing digitization of paper-based patient records of trial subjects are unclear and should be formulated in unequivocal terms.

Recognition/compliance should be based on CPMP/ICH/135/95.

GCP-compliant reliable digital archiving systems will be important for managing and maintaining clinical trial archiving documents in the coming years.

Aim of this paper

This paper should help staff of digitization facilities and digital archives in hospitals to understand relevant GCP requirements.

This paper should also help to find the right balance between quality assurance measures necessary to achieve an acceptance of the digitized patient records as source documents and the effort that is associated with these quality assurance measures.

This paper will provide a basis for discussions with the aim to establish clear and binding rules for the digitization of patient records of trial participants.

If the digitization principles described herein are observed, it should be possible that digitized documents can be recognized as GCP-compliant certified copy and the original, paper-based patient records destroyed.

Executive summary

Current & future situation

- Managing and maintaining archives for paper-based patient records is costly. Therefore, many healthcare providers digitize their paper records, destroying originals thereafter.
- Current regulatory requirements governing digitization of paper-based patient records of trial subjects are unclear and should be formulated in unequivocal terms.
- Recognition/compliance should be based on CPMP/ICH/135/95.
- GCP-compliant reliable digital archiving systems will be important for managing and maintaining clinical trial archiving documents in the coming years.

Aim of this paper

- This paper should help staff of digitization facilities and digital archives in hospitals to understand relevant GCP requirements.
- This paper should also help to find the right balance between quality assurance measures necessary to achieve an acceptance of the digitized patient records as source documents and the effort that is associated with these quality assurance measures.
- This paper will provide a basis for discussions with the aim to establish clear and binding rules for the digitization of patient records of trial participants.
- If the digitization principles described herein are observed, it should be possible that digitized documents can be recognized as GCP-compliant certified copy and the original, paper-based patient records destroyed.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as:
- of interest
- of considerable interest


This US FDA guidance summaries central requirements on computerized systems used in clinical investigations. These requirements also have to be considered when assessing systems used to digitize and electronically archive paper-based patient records that serve as source documents.
GCP-compliant digital archiving of paper-based patient records

Clinical Trial Perspective


Websites


