AS/NZS ISO 13028:2012
ISO/TR 13028:2010

AS/NZS ISO 13028:2012

Australian/New Zealand Standard™

Information and documentation—Implementation guidelines for digitization of records

This Joint Australian/New Zealand Standard was prepared by Joint Technical Committee IT-021, Records and Document Management Systems. It was approved on behalf of the Council of Standards Australia on 3 April 2012 and on behalf of the Council of Standards New Zealand on 2 April 2012.
This Standard was published on 24 April 2012.

The following are represented on Committee IT-021:

Archives New Zealand
AMS Imaging, Australia
Australian Society of Archivists
Department of Finance and Services, NSW
Inland Revenue, New Zealand
Institute for Information Management, Australia
Monash University, Vic.
National Archives of Australia
NSW Right to Information and Privacy Practitioners' Network
Porirua City Council, New Zealand
Public Record Office Victoria
Queensland State Archives
Records and Information Management Professionals Australasia
State Records Office of Western Australia
State Records (New South Wales)
Tasmanian Archive and Heritage Office
The Institute of Internal Auditors, Australia

Additional Interests:

City of Ryde
Enterprise Knowledge
NZ Micrographic Services
Outsource Australia T/A Converga
Recordkeeping Systems
State Records of South Australia
The Caldeson Consultancy
The University of Melbourne
The University of Sydney

Keeping Standards up-to-date

Standards are living documents which reflect progress in science, technology and systems. To maintain their currency, all Standards are periodically reviewed, and new editions are published. Between editions, amendments may be issued. Standards may also be withdrawn. It is important that readers assure themselves they are using a current Standard, which should include any amendments which may have been published since the Standard was purchased.

Detailed information about joint Australian/New Zealand Standards can be found by visiting the Standards Web Shop at www.saiglobal.com.au or Standards New Zealand web site at www.standards.co.nz and looking up the relevant Standard in the on-line catalogue.

For more frequent listings or notification of revisions, amendments and withdrawals, Standards Australia and Standards New Zealand offer a number of update options. For information about these services, users should contact their respective national Standards organization.

We also welcome suggestions for improvement in our Standards, and especially encourage readers to notify us immediately of any apparent inaccuracies or ambiguities. Please address your comments to the Chief Executive of either Standards Australia or Standards New Zealand at the address shown on the back cover.

This Standard was issued in draft form for comment as DR AS/NZS ISO 13028.

Online Subscription Service PDF Terms & Conditions

- **An Authorised User may download a single copy of a document and retain that copy on their personal computer for a maximum of five working days for their internal purposes. At the expiry of five working days, the Document must be deleted from the Authorised User's computer.**
- **Each Authorised User may print one hard copy of any Document for their internal purposes. These copies may not be used to build up a hard copy reference collection. A reference collection is defined as a collection comprising more than 10% of the number of the Documents within the Authorised User's subscription portfolio. All hard copies of Documents must be destroyed on expiry of the subscription period.**
- **Copyright subsists in each of the Documents and the full title to that copyright is at all times retained by Standards New Zealand.**
- **Except as otherwise may be expressly permitted under the Copyright Act 1994 Authorised Users will undertake not to copy, modify, merge with other software or documents, or circulate electronically without securing the prior written permission of SNZ. .**
- **Under no circumstance may a Document, whether in electronic or hard copy form, be sold, or transferred to a third party.**
- **Under no circumstances may any Document be placed on a network of any sort without express permission of SNZ.**
- **Authorised Users may not modify, adapt, translate, reverse engineer, decompile, disassemble or create derivative works based on the Documents.**
- **Right of access to the Subscription Service is personal to Authorised Users and can not be transferred, sold, leased, rented or loaned via a timesharing, service bureau or other arrangement.**
- **All Authorised User identification information, including logins and passwords, are to be kept secret and secure.**
- **No Authorised User may attempt to damage, interfere or harm the SNZ website, or any network, or system underlying or connected to the Subscription Service.**

NOTES

Australian/New Zealand Standard™

Information and documentation—Implementation guidelines for digitization of records

First published as AS/NZS ISO 13028:2012.

Jointly published by SAI Global Limited under licence from Standards Australia Limited, GPO Box 476, Sydney, NSW 2001 and by Standards New Zealand, Private Bag 2439, Wellington 6140.

ISBN 978 1 74342 076 8

PREFACE

This Standard was prepared by the Standards Australia Committee IT-021, Records and Document Management Systems.

The objective of this Standard is provide implementation guidelines for processes and policies for converting hard copy or non-digital records into digital format.

This Standard is identical with, and has been reproduced from ISO/TR 13028:2010, *Information and documentation—Implementation guidelines for digitization of records.*

As this Standard is reproduced from an International Standard, the following applies:

(a) Its number appears on the cover and title page while the International Standard number appears only on the cover.

(b) In the source text 'this Technical Report' should read 'this Australian/New Zealand Standard'.

(c) Where the source document uses the term 'worm (write once, read many)' *replace* with 'WORM (write once, read many)'.

(d) Where the source document uses the term 'best practice guidelines' *replace* with 'implementation guidelines'.

(e) A full point substitutes for a comma when referring to a decimal marker.

References to International Standards should be replaced by references to Australian or Australian/New Zealand Standards, as follows:

Reference to International Standard		*Australian/New Zealand Standard*	
ISO		AS ISO	
15489	Information and documentation—Records management	15489	Information and documentation—Records management
15489-1	Part 1: General	15489.1	Part 1: General
23081	Information and documentation—Records management processes—Metadata for records—	23081	Information and documentation—Records management processes—Metadata for records
23081-1	Part 1: Principles	23081.1	Part 1: Principles
		AS/NZS ISO	
23081	Information and documentation—Managing metadata for records	23081	Information and documentation—Records management processes—Metadata for records
23081-2	Part 2: Conceptual and implementation issues	23081.2	Part 2: Conceptual and implementation issues
ISO/TR		AS ISO	
15801	Document management—Information stored electronically—Recommendations for trustworthiness and reliability	15801	Document management—Information stored electronically—Recommendations for trustworthiness and reliability

CONTENTS

INTRODUCTION

With the shift to managing records in digital systems, many organizations are digitizing paper and/or other non-digital records. To manage themselves effectively, organizations need to create full and accurate records of their activities and maintain these records over time for subsequent reference. These considerations are valid regardless of the records' storage media.

Digitization is the process of converting hard-copy, or other non-digital, records into a digital format, such as taking digital photographs of non-digital source records or imaging non-digital source records (also known as scanning).

When converting records into digital objects, they are commonly:

a) captured as static pictures (raster image) represented by pixels;

b) processed by optical character recognition technology which converts the pixels into digital representations which are searchable, editable and manipulable; or

c) captured into both formats.

Digitization can broadly be categorized into two types:

— business-process digitization: ongoing, routine digitization as part of daily business processes;

— digitization projects: project-based bulk digitization of legacy records.

AUSTRALIAN/NEW ZEALAND STANDARD

Information and documentation—Implementation guidelines for digitization of records

1 Scope

This Technical Report:

— establishes guidelines for creating and maintaining records in digital format only, where the original paper, or other non-digital source record, has been copied by digitizing;

— establishes best practice guidelines for digitization to ensure the trustworthiness and reliability of records and enable consideration of disposal of the non-digital source records;

— establishes best practice guidelines for the trustworthiness of the digitized records which may impact on the legal admissibility and evidential weight of such records;

— establishes best practice guidelines for the accessibility of digitized records for as long as they are required;

— specifies strategies to assist in creating digitized records fit for long-term retention;

— establishes best practice guidelines for the management of non-digital source records following digitization.

This Technical Report is applicable for use in the design and conduct of responsible digitization by all organizations undertaking digitization, either business process digitization or back capture digitization projects for records management purposes, as outlined in ISO 15489-1:2001 and ISO/TR 15801:2009.

This Technical Report is not applicable to:

a) capture and management of born-digital records;

b) technical specifications for the digital capture of records;

c) procedures for making decisions about records' eventual disposition;

d) technical specifications for the long-term preservation of digital records;

e) digitization of existing archival holdings for preservation purposes.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 15489-1:2001, *Information and documentation — Records management — Part 1: General*

ISO/TR 15801:2009, *Document management — Information stored electronically — Recommendations for trustworthiness and reliability*

ISO 23081-1:2006, *Information and documentation — Records management processes — Metadata for records — Part 1: Principles*

ISO 23081-2:2009, *Information and documentation — Managing metadata for records — Part 2: Conceptual and implementation issues*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 15489-1:2001, ISO/TR 15801:2009, ISO 23081-1:2006 and ISO 23081-2:2009 and the following apply.

3.1
born digital records
records which are created in digital form, without a non-digital equivalent

NOTE 1 This term is used to differentiate born digital records from:

— digital materials which might have been created as a result of converting non-digital source material;

— non-digital materials which might have originated from a digital source but have been printed to paper or otherwise converted into analogue form.

NOTE 2 Adapted from Reference [21].

3.2
business information system
automated systems that create or manage data about an organization's activities

NOTE 1 Business information systems are (often multiple or related) applications whose primary purpose is to facilitate transactions between an organizational unit and its customers, e.g. an e-commerce system, client-relationship management system, purpose-built or customized database, and finance or human resources systems. Business information systems typically contain dynamic data that are commonly subject to constant updates, able to be transformed (manipulated) and hold current data. For the purposes of this Technical Report, the term business information system includes electronic records management systems. A business information system will create records, but might or might not manage them according to records management requirements. An electronic document and records management system is a specific type of business information system with the dedicated functionality of managing an organization's records and information resources.

NOTE 2 Adapted from Reference [11].

3.3
business-process digitization
routine digitization of records and incorporation into business information systems where future actions take place on the digitized record, rather than on the non-digital source record

NOTE 1 For the purposes of ongoing management of authoritative records, the version of the record on which the business action took place, or which evidences the business action, is the version that needs to be managed as the official record. In all cases, organizations need to analyse their business processes to identify and manage the record that evidences the business action. Therefore:

— where the digitized record is the record the business relied on in undertaking its business actions, or which evidences the action, the digitized version needs to be regarded as the official record for management purposes;

— where the action has been completed on a non-digital record prior to the digitization process, the non-digital record is the one on which the action has been taken and which evidences the action and the digital record is the copy;

— where the digitized record is converted back to analogue form for the purposes of further business action, the paper version of the digitized record may be required for management in specific processes in addition to the digitized record.

NOTE 2 Such digitization may take place in conjunction with the operation of an electronic document and records management system.

NOTE 3 Adapted from Reference [16].

3.4
destruction
process of eliminating or deleting records, beyond any possible reconstruction

[ISO 15489-1:2001]

3.5
digitization
means of converting hard-copy or non-digital records into digital format

NOTE Examples of digitization include scanning or imaging, taking digital photographs of the non-digital source records, or converting analogue voice recordings to digital media.

[SOURCE: Reference [16]]

3.6
digitization project
retrospective, back-capture of existing sets of non-digital records to enhance accessibility and maximize re-use

NOTE 1 In such projects, the business action has been completed on non-digital form of the record prior to digitization and for ongoing management purposes the non-digital record on which the business action took place, or which evidences the action, remains the official record of action.

NOTE 2 The non-digital source records for both forms of digitization should be subject to an assessment process to determine whether there are good reasons to retain them prior to any consideration of disposition. Once non-digital records are converted into digital records, many of the management and preservation issues for born-digital records apply.

[SOURCE: Reference [16]]

3.7
disposition
range of processes associated with implementing records retention, destruction, or transfer decisions which are documented in disposition authorities or other instruments

[SOURCE: Reference [14]]

3.8
non-digital source record
document or record that has been copied, converted or migrated or will be the input for such a process

NOTE A non-digital source record can be an original record or it can be a reproduction that was generated by an earlier copying, conversion or migration process.

[SOURCE: Reference [16]]

4 Benefits and risks of digitization

4.1 Benefits of digitization

Digitization offers the following potential benefits to organizations:

— capacity of more than one person to access the images concurrently;

— networked access enabling access from multiple locations at any time;

— greater integration with business information systems;

— capacity to transmit images within a structured workflow, thus assisting work processing;

— elimination of hybrid (both paper and digital) systems which can cause confusion to users who require access to the whole history of a matter;

— capacity to re-use existing resources limited in their re-use by their format, e.g. very large maps or material held on microfilm or magnetic tape;

— application of consistent classification and indexing for document retrieval particularly for hybrid files;

— integration with existing organizational disaster recovery and back-up regimes;

— provision of a protected and secured rendition;

— potential to reduce physical storage space occupied by hard-copy records;

— potential to increase organizational productivity.

4.2 Risks of digitization

There are a number of risks associated with implementing a digitization process:

— short-term cost savings in space may be negated when balanced with longer-term costs in maintaining the accessibility of digital images over time;

— technology and technical standards used to create digital images may significantly affect longevity and capacity to re-use the images in the future;

— legislative, regulatory or other requirements to maintain authentic and reliable representations of non-digital source records may limit the capacity to deploy commonly offered digitization features (such as image manipulation, etc.);

— it may not be appropriate to destroy the non-digital source records after the digitization process, especially where there are good reasons to retain the records in their non-digital form, e.g. records with importance for national or personal identity or other societal or cultural significance;

— it may not be permissible to destroy the non-digital source records after the digitization process for legislative reasons (by law, some specific classes of records have to be retained in their original format and in some instances non-digitized records may even need to be retained along with their digitized counterparts for a period of time).

5 Preliminary considerations

5.1 Digitization project viability assessment

The rationale for digitizing should be carefully aligned to a business case geared at improving the organization's ability to carry out its functions. The business case should clearly outline the benefits and anticipated business or cost efficiencies. The business case should take into account appropriate project budgets, resource commitments and be realistically costed. Digitization can involve extensive document preparation and indexing, which can comprise the majority of a digitization budget. Organizations should not be misled by a lack of consideration of on-going costs into thinking that digitization is a cheap option. Digitization undertaken as a space and cost saving device may not be justified, especially when the costs of future migration projects are factored in.

Annex A provides a set of questions to assist in assessing the viability of digitization for non-digital records.

5.2 Master copies and derivatives

Master copies of digitized records are those maintained as a separate and inviolable record in a safe storage environment, usually executed to the highest technical specifications available at the time and subsequently used for the production of derivatives.

Derivative versions, where required, should be made during the digitization process. Master copies should be made available for the creation of subsequent derivative images, where necessary.

A master record may not be necessary for business-process digitization, where the digitized record:

— is the version used for making business decisions;

— is evidence of a business action; or

— acts as supporting reference material.

Any decisions on creating master and derivative images are dependent on analysis of the legislative framework in which the organization operates. The record to be managed to ensure ongoing evidence of business activities is the version of the record on which the business action took place, (or the record evidencing a business action), regardless of whether this is the digitized version or the non-digital source record (or both). Before destroying masters or derivatives, organizations should conduct analyses of their business processes to ensure that the appropriate format of the record on which the business action takes place, or which evidences the action, is identified and the record in that format/s is managed according to any jurisdictional records legislation or regulatory requirements.

6 Best practice guidelines

6.1 General

A full list of the best practice guidelines are provided in the Checklist of best practice guidelines given in Annex B.

6.2 Planning processes

6.2.1 Project documentation

All digitization and digitization processes should be planned, scoped and documented. The project documentation should include:

— scope definition: with clear identification of business drivers, objectives, scale, size and constraints of the project;

— statement of the purpose and expected uses of the digitized records, illustrated if necessary with examples;

— statement of benefits: clear identification of the benefits anticipated from the digitization;

— statement of user needs and impacts: for example, how the digitized records are to be used, accessed and how this impacts on users;

— statement of technical standards adopted: including format, compression and metadata;

— equipment and resources to support the digitization;

— processes for the planning, control and execution of the digitization, including those undertaken prior to, during and after digitization;

— quality control processes;

— strategies for integrating the digitized image into work processes to support the business action taking place;

— strategies for the ongoing management of the digitized records and non-digital source records for as long as they are required to be maintained;

— strategies regarding the legal requirements for digitization of the record types in question.

6.2.2 Selection of a digitization approach

6.2.2.1 General

An appropriate digitization approach should be selected, documented and implemented. A number of approaches or a combination of approaches, to digitization can be adopted. The way these decisions are approached can be different for business-process digitization and for digitization projects and can vary across organizations.

Prior to digitization, consideration of third party copyright issues, legal requirements to retain the original paper or other non-digital records as well as other constraints inherent in the record should be resolved.

Regardless of which approach to digitization is adopted, the following should apply:

— the digitization approach selected should be documented;

— quality control processes should be implemented regardless of the digitization approach adopted;

— the digitization approach should be regularly reviewed for continuing compliance with the requirements of the legal environment, relevance and cost effectiveness.

Decisions on each of the areas given in 6.2.2.2 to 6.2.2.4 will contribute to the overall approach employed.

6.2.2.2 In-house or outsourced digitization

In-house digitization requires (and gives the opportunity for) an organization to develop and acquire all the equipment and expertise necessary to digitize and integrate the digitized output into their own systems.

The alternative to this is to outsource the digitization to a third party contracted to perform this service on behalf of the organization.

6.2.2.3 Batch process digitization or on-demand digitizing

Batch processing is the collecting of source documents into sequences prior to digitizing until documents have been accumulated in numbers sufficient to provide efficiencies of scale in undertaking the digitizing process.

The alternatives to this approach are on-demand imaging or digitizing individual documents as they arrive in the digitizing facility.

6.2.2.4 Centralized or decentralized digitization

Centralized digitization involves establishing a single site for digitizing in which all records to be processed are accumulated prior to digitizing.

The alternative to this is decentralized digitization which involves placing multiple digitizing stations in different locations throughout the organization.

6.2.3 Selection of technical specifications

Technical specifications aligned to the digitization best practice guidelines should be selected, documented and implemented. A large volume of technical standards associated with digitization are available. Such standards include recommendations on:

— file formats;

— resolution;

— colour resolution or bit depth;

— compression;

— colour management.

Technical specification standards are rapidly evolving, especially in the area of technical capacity of equipment to accommodate such standards. The primary consideration in adopting technical specifications is to ensure the legibility or usability of the digitized image. The following basic criteria should be adhered to when selecting technical standards:

a) the highest quality technical specifications that can be realistically supported should be incorporated into the digitization process;

b) the formats should be open source (that is, non-proprietary) or employ open standards, have published technical specifications available in the public domain, or be widely deployed within the relevant sector;

c) the formats should not contain embedded objects, or link out to external objects beyond the specific version of the format;

d) the formats should be supported by many software applications and operating systems;

e) the formats should be able to be read by utilising a readily available viewing plug-in if the specific production software is not available to all users;

f) a body of accessible and product-independent technical expertise should be available to support the decision;

g) adequate technical support should exist to enable ongoing maintenance and migration capability when necessary;

h) the master copies should be created to the highest technical standards that can be supported;

i) the master copies should be retained inviolable in secure storage;

j) the derivative copies may be made in formats most convenient for their business purpose (e.g. thumbnails for distribution over the internet, etc.).

6.2.4 Equipment and software

Equipment and software aligned to the digitization best practice guidelines should be implemented. The quality of equipment and software used in digitizing significantly affects the capability to support appropriate technical standards and, therefore, to ensure longevity of the digital image produced. Where destruction of the non-digital source records is contemplated, organizations should be able to assert confidence in the long-term viability of those digitized images requiring on-going retention.

6.2.5 Enhancement techniques

6.2.5.1 General

Any use of enhancement techniques employed on the digitized image to create a more exact resemblance to the non-digital source record, should be well documented. The following guidelines are applicable.

6.2.5.2 Image enhancement

During the digitization process, the use of techniques that enhance the digitized image to make the image have a more exact resemblance to the non-digital source record should be documented. Such procedures may, if not undertaken in routine and documented ways, enable the challenge that the image is not an authentic copy of the non-digital source record. Such techniques include “sharpening” and/or “clipping” of highlights or shadows, “blurring” to eliminate scratches, “spotting” or “de-speckling” to touch up specific areas of a digital image.

6.2.5.3 Annotation management

Where software that is employed to manage digital images after capture enables additions of annotations to images, such as highlighting, stamps, redaction or addition of notes, these annotations should be managed as overlays that do not change the actual image. Printing of the image should be possible with or without the annotations.

6.2.5.4 Image quality

The display of the digital image should be in a manner, and to a quality, acceptable for the business being conducted. This can involve reviewing equipment specifications. For example, if the quality of the colour on a map is critical (e.g. for interpreting colour coded information illustrated on a map or chart), the quality of the equipment used to render the image needs to support the capacity to retrieve and analyse this quality. If, on the other hand, it is only essential to be able to read the contents to gain the sense of the text, the quality of display could be appropriately less critical.

6.2.5.5 Storage media

Response times required by end users can affect the storage media used for images. For example, if DVDs are used as an offline storage media, the retrieval or response time for a user to receive a copy of the requested image will be at least partly dependent on the hardware loading the relevant DVD onto the system. This timeframe may be too long for the purpose of the query.

6.3 Digitization process management

6.3.1 General

Systems to support management of the digital output of digitization as records should be in place. Before the final version is reached, the digitizing process could create a number of renditions of a digitized record – for example, a raw file, one which has been enhanced to maximize resemblance to the non-digital source record, or one to which quality control processes have been applied. Each digitization process will specify the parameters of acceptable enhancements (see 6.2.1). The final output of the digitization process should be regarded as the record for incorporation into the organization's framework for managing records.

6.3.2 Digitized records management

The business purpose of a digitization process and any legislative framework will determine the functionality and system requirements necessary to manage the digital outputs. Many organizations use electronic document and records management systems as their image management systems. Such systems should contain the records management functionality to ensure compliance with ISO 15489-1:2001.

Where the digital image is to be used as a record in current or continuing business, the system governing this business process should be integrated with other relevant business information or electronic records management systems. This ensures that the digital image inherits the business classification and metadata associated with the business process, will be lodged within its business context and will have its authenticity enhanced by integration with business information systems.

Digitization projects are often undertaken as a preservation measure to extend the life of fragile non-digital source records or to increase access by users to the informational content of records and often represent a considerable investment. Where the results of such projects are not immediately linked to a pre-existing business information system, the procurement of a suitable management system which ensures the appropriate management of processes such as identification, indexing, classification, security and access controls, rights management and preservation should be initiated.

6.3.3 Non-digital source record preparation

Non-digital source record preparation actions for digitization should be documented and implemented. The aim of producing digital images is to reproduce the non-digital source record as faithfully as possible so that the digital image can act in place of the non-digital source record, where it is required to act as evidence of business activities.

Non-digital source record preparation actions are not restricted to, but should include:

— an assessment of the non-digital source record's capability to sustain a digitization process [e.g. paper quality, creasing, stapling, condition of microform jackets, attributes of the informational content (e.g. graphics)];

— methods of digitizing non-digital source records of non-standard dimensions or handling requirements (e.g. by digitizing a photocopy of records on fragile or thin paper; creating a standard-sized document by using photocopy enlargement or reduction; enclosing fragile originals in plastic wallets or using specialty devices such as overhead scanners) and quality checks to ensure against data loss in any such processes;

- methods for dealing with non-digital source records containing handwritten annotations, marginalia, white opaque paint or other correctional fluid, tape or device or highlighted areas;
- methods of distinguishing between paper non-digital source records and photocopies;
- guidance on what types of material need not be digitized as they are of only ephemeral or short-term value;
- physical preparation for digitization (e.g. careful staple removal, alignment of single pages, batching of like documents – size, technical settings, shared indexing fields);
- processes for assigning links between associated documents to be regarded as a single item, so that the digitized image can faithfully represent the non-digital source record (e.g. a document and a self-adhesive note attached; a document enclosing an attachment, document printed on both sides of the paper or containing endorsements on the reverse side);
- processes for assigning links between the non-digital source record and the digitized copy, such links will usually be documented using identification protocols, in some applications barcode technology could be used to link paper and digitized renditions;
- procedures to enable checking and verification that all target non-digital source records have been included in the digitization process;
- principles governing the assembly of batches or groups of non-digital source records suitable for digitization at the same time (e.g. size, colour, date order, document formats, orientation, i.e. portrait or landscape, single or double sided).

6.3.4 Metadata

6.3.4.1 General

All digitized images should be assigned metadata to document digitizing processes and to support ongoing business processes. This subclause provides guidance on how to identify areas of critical importance for maintaining the authenticity and integrity of the digitized record. Organizations can adapt specific elements as required and to maximize the inheritance of data values from existing systems and equipment. Metadata management processes should maximize automatic capture of metadata, minimizing the need for manual attribution. Any consideration of metadata attribution should be undertaken with reference to ISO 23081-1:2006.

Metadata attributed to, or associated with, images is an essential component in the management and retrieval of images.

Two types of metadata should be captured:

- metadata specific to the particular image and the imaging process;
- metadata about the record, the business being transacted and the agents associated with the business.

The majority of this metadata can be automatically sourced from the software and hardware used to manage the digitizing process. The manual attribution or application of metadata should be minimized as far as possible.

Metadata can be embedded with the resource in header information, or can be managed in a separate system, or both, but in either case there has to be a direct relationship or association between them; i.e. while metadata may reside in a separate system, it has to link directly to the records. Metadata can also be encapsulated within the image format.

6.3.4.2 Image-level metadata

Image-level metadata should be generated automatically at the point of digital capture direct from the digitization equipment and should avoid manually-assigned data entry wherever possible.

In addition to the metadata inherited from record capture and processes for managing records, or from indexing and searching metadata, image-level metadata should include:

— unique digital image identifier;

— date and time of digitization;

— the name of the agent associated with the digitization process (e.g. name of the outsourced bureau or name of the in-house operator);

— capture device (hardware and software);

— date of last calibration (where practicable).

At the organization's discretion, additional image-level metadata may be assigned.

Recommendations on naming protocols for digital images and directories are included in Annex C.

6.3.4.3 Business-process digitization metadata considerations

Wherever possible the metadata describing the business process and the functions for managing records associated with the business process should govern, and be inherited by, the specific digital image.

This metadata should be derived from or incorporated into the organization's framework for managing digital records and be consistent with ISO 23081-1:2006. Additional metadata describing the process of digitization and specific characteristics of the digitized image should be included, as outlined above.

6.3.4.4 Digitization projects metadata considerations

Where access to content is the primary driver, more intensive attention to providing additional indexing and search entry points is usually appropriate. The images can be managed as individual items, rather than as context-linked records, particularly if intended for web access by external users. Organizations should investigate which additional indexing fields facilitate more in-depth access to content.

6.3.5 Quality control

6.3.5.1 General

Quality control procedures should be defined, documented and implemented. Quality control to ensure the digital copy of the non-digital source record is a true and accurate copy is critical to being able to assert that the records possess integrity and are authentic.

Such quality control procedures should be documented and built into the ongoing operation of the digitizing process, not only applied at the point where the digital output is produced. Quality control procedures should, at minimum, address the following issues:

— any acceptable variations from normal procedures;

— scanner operation quality control;

— verification that digital output matches the quantity of non-digital source record input;

- extent and frequency of sampling of digitized images;
- criteria for checking image quality;
- frequency and criteria for checks on metadata;
- processes for re-digitizing;
- operator training.

Quality checking should be completed before the digitized images are accepted into a business process, or as a master copy in the case of digitization projects. Quality checking should be complete before the destruction of the non-digital source records is considered.

6.3.5.2 Reviewing quality control checking

The results of quality control processes and quality checks should be documented.

A review of quality procedures for digitizing should be undertaken regularly to ensure that the procedures continue to meet the business purpose.

Appropriate training should be provided to all staff who create, manage or work with digitized records.

Documentation on the level and the frequency of training provided to those staff involved with digitization should be created and maintained.

For further information on quality control, see Annex D: Quality control recommendations, and Annex E: Recommended staff skill sets.

6.4 Management systems

6.4.1 Strategies

6.4.1.1 General

Any storage or management systems solutions should ensure:

- that the digitized records should be unalterable in all storage media;
- that security and access controls for storage media should be capable of detecting and logging unauthorized attempts at access;
- that the retrieval times implicit in off-line storage should be acceptable for the business being conducted;
- that wherever possible digitized records sharing similar retention periods should be co-located to enable execution of disposition processes.

6.4.1.2 Long-term management systems

Long-term management systems, where required, for both source and digitized records, should be documented and implemented. Storage media and procedures should be defined, documented and implemented. At this time the longer-term storage strategy most likely to ensure continuing accessibility of digital records is to maintain such records in a trustworthy and reliable a business information system or other storage environment that meets organizational business needs, such as an electronic records and document management system. However, digitized records can occupy significant storage space depending on the quality, resolution and compression ratios employed.

6.4.1.3 Short-term storage strategies

Where an appropriate trustworthy and reliable a business information system is not available or the digital output is being held temporarily before being transferred to such a system, strategies for shorter-term storage may include:

— a dedicated server or other digitized records storage solution;

— writing the digitized records to magnetic tape;

— writing the digitized records to worm (write once, read many) storage media (e.g. a CD or DVD); or

— storing the digitized records on external hard drives.

6.4.2 Back-up procedures

Back-up procedures should be defined, documented and implemented. All digitized records, and their associated metadata, should be included in the organization's back-up regime. Back-up procedures are designed to provide sufficient up-to-date copies of business records to be used in the event of loss or corruption of all or part of the data.

Back-up regimes should be documented and back-up copies maintained to a level of security that ensures the authenticity of the records used in recovery situations.

All system failures should be documented, and use of the back-up copies for restoration purposes should be accompanied by verification testing to ensure the integrity of the restored records.

Information technology professionals often use the term “archiving” to describe back-up regimes. For the purposes of managing records, conducting back-ups does not constitute an archiving or preservation strategy, it is a business continuity or disaster recovery precaution.

6.5 Records disposition

6.5.1 General

Records disposition of all records should be authorized and documented.

6.5.2 Disposition of non-digital source records

Disposition of non-digital source records should be authorized in accordance with relevant legislation and documented.

For non-digital source records to be destroyed any relevant legislation, jurisdictional authorization or other organizational requirements for their retention as evidence should be first be established. This Technical Report sets out best practice guidelines for the destruction of non-digital source records through a managed and authorized process.

In following the best practice guidelines set out in this Technical Report, an organization can be confident that it meets the requirements for trustworthiness and reliability as stated in 8.2 of ISO 15489-1:2001, and they may then, in reference to any legislative considerations mentioned above, consider retaining the record in digital form only.

All decisions to destroy non-digital source records and/or any other disposition actions should be documented and this information should be accessible and be able to be produced on request. Authorization for destruction and the instance of destruction of the non-digital source record should be documented in the metadata associated with the digitized record. Disposition actions should be documented and authorized by the relevant authority in an organization, such as the chief executive officer, legal or administrative head (or the delegated appropriate senior management representative).

Digitized records should be (re)producible in their original format on request, if so required.

Organizations should note the following criteria, exclusions and disposition considerations:

- that the digitized record is an accurate and complete rendition of the non-digital source record it replaces (including colour reproduction, where applicable);
- that appropriate quality control procedures and certifications for the digitizing processes are defined, implemented and monitored routinely;
- that the digitized record is individually identified and linked to the context of its creation and use;
- that an appropriate management system is in place to ensure the ongoing business use of the digitized record;
- that the digitized record is relied upon in the normal conduct of business;
- that the necessary metadata about the digitized record is created and maintained;
- that the disposition program in place within the organization encompasses the business information system(s) incorporating digitized records;
- that a migration and/or preservation strategy is defined, documented and implemented for digital records, including digitized records;
- that legislative or regulatory requirements to maintain the records in a specific form are not breached;
- that no known or anticipated legal action will be imperilled by the destruction of the non-digital source record;
- that the risk of challenge to the authenticity and integrity of the digitized record has been assessed and considered and is acceptable to the organization.

6.5.3 Disposition of records incorporated into business information systems

Disposition of the digitized records incorporated into business information systems should be authorized and documented. Digitized business records are subject to normal processes for disposition authorizations, established by the authority responsible for authorizing disposition in the same way as the non-digital source records. A digitized record acting in place of a non-digital source record should be retained for the authorized minimum period that was required for the non-digital source record.

Once records destruction is authorized, reasonable efforts should be taken to destroy all extant copies of the digitized record (e.g. back-up copies). It may not always be possible to identify and destroy every copy.

Authorization of destruction of the digitized copy should be documented.

Destruction should be documented in the metadata associated with the record, for example by documenting its disposal in the metadata that persists after the record has been destroyed, or a destruction register that is independent from the record being destroyed.

Long-term management systems, where required, for both source and digitized records, should be documented and implemented.

6.5.4 Management of non-digital source records

6.5.4.1 General

Non-digital source records should be managed appropriately until their authorized disposition within a broader framework for managing records.

Where non-digital source records are retained for whatever reason other than quality control, or are not authorized for destruction, systematic controls should be applied. The digital image and the non-digital source record should be persistently linked.

The non-digital source records should be organized to optimise retrieval and to enable the efficient application of management and disposition processes.

Digitization projects rarely have the principle aim of enabling the destruction of the non-digital source record. After the digitization process, non-digital source records should be returned to their original context and order, reflecting the processes of their creation and management in their original format. This enables the existing finding aids to continue to function as a retrieval tool for the records.

6.5.4.2 Day boxing

Day boxing, the process of accumulating non-digital source records in chronological order, or in their digitization sequence, is not recommended. Day boxing is rarely suited to efficient management and disposition processes as it eliminates contextual linkages and mixes records intended for short-term retention with material subject to longer retention periods.

6.5.5 Digitized records obsolescence

Digitized records should be managed in a way that allows their continued existence for as long as they are required. Technology-dependent digital records are inherently vulnerable to hardware, software and media obsolescence.

Digitized records should be included in the framework for managing records adopted by the organization to support the continuing existence of records for as long as they are required.

This may entail the use of migration, conversion or other strategies for preserving digital records.

Annex A
(informative)

Issues to consider when assessing the viability of digitization

This annex gives guidance only. It provides a set of questions to assist organizations in assessing the viability of digitization for non-digital source records. It relates to the 'preliminary considerations' section of the best practice section of the Technical Report.

Table A.1 — Issues to consider

Business-process digitization	Yes	No
Will digitizing enable the organization to manage all incoming and outgoing transactions digitally, thus eliminating a hybrid paper/digital system?	☐	☐
Are the records required for transmission across internal or external networks?	☐	☐
Do the records need to be available to all staff at centralized and remote locations?	☐	☐
Are all incoming paper documents to be digitized, i.e. is the process to be consistently applied, or is a selection process required?	☐	☐
Is the digitizing taking place prior to action being initiated on the item?	☐	☐
Is there guidance for staff on what not to capture?	☐	☐
Will the digital images be incorporated into an established business information system or records management system application (e.g. EDRMS)?	☐	☐
Will all further action arising from the digital record be documented in the business information system or records management system (e.g. EDRMS) application?	☐	☐
Can adequate controls be established to ensure the reliability of the digital image?	☐	☐
Have the records to be digitized been appraised for disposition scheduling?	☐	☐
Are the records constantly or continuously requested for use by people either internal or external to the organization?	☐	☐
Is the information in the records enhanced by being available in digital form (e.g. additional indexing, sorting capability)?	☐	☐
Are the records in a form which is suited to digitization?	☐	☐
Are the items in a standard format which will enable single equipment settings to be applied?	☐	☐
Will this act as a precedent for future projects?	☐	☐
Have the records been appraised as being required for long-term retention?	☐	☐
Is there legislation that prevents these records being held only in digital form?	☐	☐
Are the records fragile or subject to damage through repeated physical handling which will be minimized after digitization?	☐	☐
Digitization projects	**Yes**	**No**
Do the records document a process or function that continues to be done?	☐	☐
Is the equivalent information in the records now recorded digitally?	☐	☐
Do the records contain a rich information source, required by current business processes?	☐	☐
Are the records a coherent and complete set?	☐	☐
Are the records required for constant or continuing access?	☐	☐

Annex B
(informative)

Checklist of best practice guidelines

This annex gives guidance only. The checklist below is the set of best practice guidelines for organizations seeking to undertake to implement the recommendations set out in this Technical Report. This process is intended as guidance to meet the requirements for trustworthiness and reliability in ISO 15489-1:2001, and 8.2 of ISO/TR 15801:2009.

Table B.1 — Checklist of best practice guidelines

See clause (subclause)	Best practice guidelines	No	Partly	Yes
Clause 5	The rationale for digitizing should be carefully aligned to a business case geared at improving the organization's ability to carry out its functions.	☐	☐	☐
	The business case should clearly outline the benefits and anticipated business or cost efficiencies.	☐	☐	☐
	The business case should take into account appropriate project budgets, resource commitments and be realistically costed.	☐	☐	☐
	Derivative renditions, where required, should be made during the digitization process.	☐	☐	☐
	Master copies should be made available for the creation of subsequent derivative images, where necessary.	☐	☐	☐
	Before destroying masters or derivatives, organizations should conduct analyses of their business processes to ensure that the appropriate format of the record on which the business action takes place, or which evidences the action, is identified and the records in that format/s is managed according to the recordkeeping requirements established by their jurisdiction.	☐	☐	☐
6.2	The project documentation should include:	☐	☐	☐
	— scope definition: with clear identification of business drivers, objectives, scale, size and constraints of the project;	☐	☐	☐
	— statement of the purpose and expected uses of the digitized records, illustrated if necessary with examples;	☐	☐	☐
	— statement of benefits: clear identification of the benefits anticipated from the digitization;	☐	☐	☐
	— statement of user needs and impacts;	☐	☐	☐
	— statement of technical standards adopted: including format, compression and metadata;	☐	☐	☐
	— equipment and resources to support the digitization;	☐	☐	☐
	— processes for the planning, control and execution of the digitization, including those undertaken prior to, during and after digitization;	☐	☐	☐
	— quality control processes;	☐	☐	☐
	— strategies for integrating the digitized image into work processes to support the business action taking place;	☐	☐	☐

Table B.1 (*continued*)

See clause (subclause)	Best practice guidelines	No	Partly	Yes
6.2	— strategies for the ongoing management of the digitized records and non-digital source records, for as long as they are required to be maintained;	☐	☐	☐
	— strategies regarding the legal requirements for digitization of the record types in question.	☐	☐	☐
	The highest technical specifications that can be realistically supported should be incorporated into the digitization process.	☐	☐	☐
	Formats should be open source or employ open standards, have published technical specifications available in the public domain, or be widely deployed within the relevant sector.	☐	☐	☐
	Formats should not contain embedded objects, or link out to external objects beyond the specific version of the format.	☐	☐	☐
	Formats should be supported by many software applications and operating systems.	☐	☐	☐
	Formats should be able to be read by utilising a readily-available viewing plug-in if the specific production software is not available to all users.	☐	☐	☐
	A body of accessible and product-independent technical expertise should be available to support the decision.	☐	☐	☐
	Adequate technical support should exist to enable ongoing maintenance and migration capability when necessary.	☐	☐	☐
	Master copies should be created to the highest technical standards can be supported.	☐	☐	☐
	Master copies should be retained inviolable in secure storage.	☐	☐	☐
	Where software that is in place to manage digital images after capture enables additions of annotations to images, these annotations should be managed as overlays that do not change the actual image.	☐	☐	☐
	Printing of the image should be possible with or without the annotations.	☐	☐	☐
	Readers available to users should support the display of the digital image in a manner, and to a quality, acceptable for the business being conducted.	☐	☐	☐
6.2.1	All digitization and digitization processes should be planned, scoped and documented.	☐	☐	☐
6.2.2	An appropriate digitization approach should be selected documented and implemented.	☐	☐	☐
	Quality control processes should be implemented regardless of the digitization approach adopted.	☐	☐	☐
	The digitization approach should be regularly reviewed for continuing compliance with the requirements of the legal environment, relevance and cost effectiveness.	☐	☐	☐
6.2.3	Technical specifications aligned to the digitization best practice guidelines should be selected, documented and implemented.	☐	☐	☐
6.2.4	Equipment and software aligned to the digitization best practice guidelines should be implemented.	☐	☐	☐
6.2.5	The use of techniques that enhance the digitized image to make the image have a more exact resemblance to the non-digital source record, should be documented.	☐	☐	☐

Table B.1 (*continued*)

See clause (subclause)	Best practice guidelines	No	Partly	Yes
6.3	Prior to digitization, consideration of third party copyright or other constraints inherent in the record should be resolved.	☐	☐	☐
	Where an electronic document and records management system is used as the image management system, it should contain the necessary records management functionality to comply with the requirements of ISO 15489-1:2001.	☐	☐	☐
6.3	Where the digital image is to be used as a record in current or continuing business, the system governing the business process which uses the image should be integrated with systems for managing records.	☐	☐	☐
	Where digitization is undertaken for preservation purposes or primarily to enhance external access to information, care is exercised to ensure that the appropriate systematic controls are in place to manage the digital image. Consideration should be given to purchasing a management system to ensure appropriate control of processes such as identification, indexing, classification, security and access controls, rights management and preservation.	☐	☐	☐
	Non-digital source record preparation actions should include:	☐	☐	☐
	— an assessment of the non-digital source record's capability to sustain a digitization process;	☐	☐	☐
	— quality checks to ensure against data loss in digitization;	☐	☐	☐
	— methods for dealing with non-digital source records containing handwritten annotations, marginalia, white opaque paint, or highlighted areas;	☐	☐	☐
	— methods of distinguishing between non-digital source records and photocopies;	☐	☐	☐
	— guidance on what types of material need not be digitized as they are only of ephemeral or short-term value;	☐	☐	☐
	— physical preparation for digitization;	☐	☐	☐
	— processes for assigning links between associated documents to be regarded as a single item, so that the digitized image can faithfully represent the non-digital source record;	☐	☐	☐
	— processes for assigning links between the non-digital source record and the digitized copy;	☐	☐	☐
	— procedures to enable checking and verification that all target non-digital source records have been included in the digitization process;	☐	☐	☐
	— principles governing the assembly of batches or groups of non-digital source records suitable for digitization at the same time.	☐	☐	☐
	Where possible, metadata attributed to, or associated with, images should be inherited from existing systems.	☐	☐	☐
	The manual attribution or application of metadata should be minimized.	☐	☐	☐
	Choices for storage of metadata should be made according to the principles outlined in ISO 23081-1:2006 & ISO 23081-2:2009.	☐	☐	☐
	Image-level metadata should be generated automatically at the point of digital capture direct from the digitization equipment and should avoid manually-assigned data entry wherever possible.	☐	☐	☐

Table B.1 (*continued*)

See clause (subclause)	Best practice guidelines	No	Partly	Yes
6.3	In business-process digitization, wherever possible the metadata describing the business process and the processes for managing records associated with the business process should govern, and be inherited by, the specific digital image.	☐	☐	☐
	Particularly if intended for web access by external users, organizations should investigate which other relevant elements should be associated with each image.	☐	☐	☐
	Quality control procedures should be documented and built into the ongoing operation of the digitization process.	☐	☐	☐
	Quality control procedures should, at minimum, address the following issues:	☐	☐	☐
	— any acceptable variations from normal procedures;	☐	☐	☐
	— scanner operation quality control;	☐	☐	☐
	— verification that digital output matches the quantity of non-digital source record input;	☐	☐	☐
	— extent and frequency of sampling of digitized images;	☐	☐	☐
	— criteria for checking image quality;	☐	☐	☐
	— frequency and criteria for checks on metadata;	☐	☐	☐
	— processes for re-digitizing;	☐	☐	☐
	— operator training.	☐	☐	☐
	A review of quality procedures for digitizing should be undertaken regularly to ensure that the procedures continue to meet business requirements.	☐	☐	☐
	Appropriate training should be provided to all staff who create, manage or work with digitized records.	☐	☐	☐
	Documentation on the level and the frequency of training provided to all levels of staff involved with digitization should be created and maintained.	☐	☐	☐
	Retrieval times implicit in offline storage should be acceptable for the business being conducted.	☐	☐	☐
	Wherever possible digitized records sharing similar retention periods should be co-located to enable execution of disposition processes as required.	☐	☐	☐
	Back-up regimes should be documented.	☐	☐	☐
	Back-up copies should be maintained to a level of security that ensures the authenticity of the records used in recovery situations.	☐	☐	☐
	Systems to support management of the digital output of digitization as records should be in place.	☐	☐	☐
	The final output of the digitization process should be regarded as the record for incorporation into the organization's framework for managing records.	☐	☐	☐
6.3.3	Non-digital source record preparation actions for digitization should be documented and implemented.	☐	☐	☐

Table B.1 (*continued*)

See clause (subclause)	Best practice guidelines	No	Partly	Yes
6.3.4	All digitized images should be assigned metadata to document digitizing processes and to support ongoing business processes	☐	☐	☐
	Two types of metadata should be captured:	☐	☐	☐
	— metadata specific to the particular image and the imaging process;	☐	☐	☐
	— metadata about the record, the business being transacted and the agents associated with the business.	☐	☐	☐
	In addition to the metadata inherited from record capture process, image-level metadata should include:	☐	☐	☐
	— unique digital image identifier;	☐	☐	☐
	— date and time of digitization;	☐	☐	☐
	— the name of the agent associated with the digitization process (e.g. name of the outsourced bureau or name of the in-house operator);	☐	☐	☐
	— capture device (hardware and software);	☐	☐	☐
	— date of last calibration (where practicable).	☐	☐	☐
	In business-process digitization, records metadata should be incorporated into the organization's framework for managing records and be consistent with ISO 23081-1:2006 & ISO 23081-2:2009.	☐	☐	☐
6.3.5	Quality control procedures should be defined, documented and implemented.	☐	☐	☐
	Quality checking should be completed before the digitized images are accepted into a business process, or as master copies in the case of digitization projects.	☐	☐	☐
	Quality checking should be complete before the destruction of the non-digital source records is considered.	☐	☐	☐
	The results of quality control processes and quality checks should be documented.	☐	☐	☐
6.4	Storage media and back-up procedures should be defined, documented and implemented.	☐	☐	☐
	Digitized records should be unalterable in all storage media.	☐	☐	☐
	Security and access controls for controlling storage media should be capable of detecting and logging unauthorized attempts at access.	☐	☐	☐
6.4.2	Back up procedures should be defined, documented and implemented.	☐	☐	☐
	All digitized records, and their associated metadata should be included in the organization's back-up regime.	☐	☐	☐
	All system failures should be documented.	☐	☐	☐
	The use of the back-up copies for restoration purposes should be accompanied by verification testing to ensure the integrity of the restored records.	☐	☐	☐
6.5	Disposition actions should be documented and authorized by the relevant authority that is responsible for authorizing disposition of records.	☐	☐	☐
	Once authorized for destruction, all extant copies of the digitized record should be destroyed (e.g. back-ups and derivatives).	☐	☐	☐

Table B.1 (*continued*)

See clause (subclause)	Best practice guidelines	No	Partly	Yes
6.5.2	Disposition of non-digital source records should be authorized in accordance with the relevant legislation and documented.	☐	☐	☐
	All decisions to destroy non-digital source records and/or any other disposition actions should be documented.	☐	☐	☐
	Information about disposition decisions and actions should be accessible and be able to be produced on request.	☐	☐	☐
	Digitized records should be (re)producible in their original format on request.	☐	☐	☐
6.5.3	Disposition of the digitized record incorporated into the business information system should be authorized and documented.	☐	☐	☐
	A digitized record acting in place of a non-digital source record should be retained for the authorized minimum period that was required for the non-digital source record.	☐	☐	☐
	Authorization of the destruction of the digitized copy should be documented	☐	☐	☐
	Destruction should be documented in the metadata associated with the record.	☐	☐	☐
6.5.4	Long-term management systems, where required, for both source and digitized records, should be documented and implemented.	☐	☐	☐
	The digital image and the non-digital source record should be persistently linked with unique identifiers.	☐	☐	☐
	Where digitized records are incorporated into business information systems and the non-digital source records are retained for reasons other than quality control checks, the non-digital source records should be organized to optimise retrieval and to enable efficient management of retention and disposition processes.	☐	☐	☐
	In digitization projects, non-digital source records should be returned to their original context and order after the digitization process.	☐	☐	☐
	Digitized records should be included in the framework adopted by the organization to support the continuing existence of records for as long as they are required.	☐	☐	☐
	Sound management systems should be in place for non-digital source records until their authorized disposition within a framework for managing records.	☐	☐	☐
	Where non-digital source records are retained for whatever reason other than quality control, or are not authorized for destruction, systematic controls should be applied.	☐	☐	☐
6.5.5	Digitized records should be managed in a way that allows their continued existence for as long as they are required.	☐	☐	☐

Annex C
(informative)

File-naming metadata recommendations

C.1 General

This annex is for guidance only. It outlines a number of best practices in relation to determining a file naming protocol, particularly for digital images. It relates to best practice guidelines 6.3.4: All digitized images should be assigned metadata to document digitizing processes and to support ongoing business processes.

C.2 Recommendations

C.2.1 General

A file naming scheme should be established prior to capture. The development of a file naming scheme should take into account whether the identifier requires machine or human-indexing (or both, in which case, the image may have multiple identifiers). File names can either be meaningful (such as the adoption of an existing identification scheme which correlates the digital file with the source material), or non-descriptive (such as a sequential numerical string). Meaningful file names contain metadata that is self-referencing; non-descriptive file names are associated with metadata stored elsewhere that serves to identify the file. In general, smaller-scale projects may design descriptive file names that facilitate browsing and retrieval; large scale projects may use machine-generated names and rely on a database for sophisticated searching and retrieval of associated metadata.

In general, file names should:

— be unique;

— be consistently structured;

— take into account the maximum number of items to be digitized and reflect that in the number of digits used (if following a numerical scheme);

— use leading zeros to facilitate sorting in numerical order (if following a numerical scheme);

— avoid using spaces within the file name, using underscore as an alternative;

— avoid an overly complex or lengthy naming scheme that is susceptible to human error during manual input;

— restrict the length of file names to under 30 characters to avoid potential problems with migration between different systems;

— use lowercase characters and file extensions;

— use numbers and/or letters but not characters such as symbols or spaces that could cause complications across operating platforms;

— record metadata embedded in file names (such as scan date, page number, etc.) in another location in addition to the file name. This provides a safety net for moving files across systems in the future, in the event that they have to be renamed. In particular, sequencing information and major structural divisions of multi-part objects should be explicitly recorded in the structural metadata and not only embedded in filenames.

C.2.2 Directory structure

Regardless of file name, files will likely be organized in some kind of file directory system that will link to metadata stored elsewhere in a database. Production master files might be stored separately from derivative files. Directories may be organized independently of the image files, such as folders arranged by date or classification structure.

They may also replicate the physical or logical organization of the non-digital source records being scanned.

The files themselves can also be organized solely by directory structure and folders rather than embedding meaning in the file name. This approach generally works well for multi-page items. Images are uniquely identified and aggregated at the level of the logical object (i.e. a document, a record, a file/folder, etc.), which requires that the folders or directories be named descriptively. The file names of the individual images themselves are unique only within each directory, but not across directories. For example, book 0001 contains image files 001.tif, 002.tif, and 003.tif. Book 0002 contains image files 001.tif, 002.tif, and 003.tif. The danger with this approach is that if individual images are separated from their parent directory, they will be indistinguishable from images in a different directory.

C.2.3 Versions

For various reasons, a single scanned object may have multiple but differing renditions associated with it (for example, the same image prepared for different output intents; versions with additional edits; layers, or alpha channels that are worth saving; versions scanned on different scanners, scanned from different original media, or scanned at different times by different scanner operators). Ideally, the description and intent of different versions should be reflected in the metadata; but, if the naming convention is consistent, distinguishing versions in the file name will allow for quick identification of a particular image. Like derivative files, this often implies the application of a qualifier to part of the file name. The reason to use qualifiers rather than entirely new names is to keep all versions associated with a logical object under the same identifier. An approach to naming versions should be well thought out; adding 001, 002 to the base file name to indicate different versions is an option; however, if 001 and 002 already denote page numbers, a different approach will be necessary.

C.2.4 Naming derivative files

The file naming scheme should also take into account the creation of derivative image files made from the production master files. In general, derivative file names are inherited from the production masters, usually with a qualifier added on to distinguish the role of the derivative from other files (i.e. “p” for published version, “t” for thumbnail). Derived files usually imply a change in image dimensions, image resolution, and/or file format from the production master. Derivative file names do not have to be descriptive as long as they can be linked back to the production master file.

For derivative files intended primarily for web display, one consideration for naming is that images may need to be cited by users in order to retrieve other higher-quality renditions. If so, the derivative file name should contain enough descriptive or numerical meaning to allow for easy retrieval of the non-digital source record or other digital renditions.

Annex D
(informative)

Quality control recommendations

D.1 General

This annex covers a number of areas relevant to quality control, including guidance on tests, standards and frequency of checking. It is intended as an aid to assist organizations in devising their own quality control processes. It relates to best practice guidelines 6.3.5: Quality control procedures should be defined, documented and implemented.

D.2 Scanner operation quality control

Scanners should be tested periodically to monitor their operational performance and check that operating performances are within agreed tolerances as set out by existing standards and benchmarks. Results of previous tests should be used as benchmarks for system performance over time.

Simple preventive measures such as ensuring the cleanliness and routine servicing of equipment should be implemented.

D.3 Validation of output

The equipment should routinely record the number of discrete documents and the number of documents comprising a record (more than one page bundled) that were scanned during a session.

D.4 Sampling

The frequency of sampling should be determined according to the system usage and expected or anticipated deterioration periods. Advice from the system vendor may assist in determining the frequency period. Initially, it is advisable that it is appropriate to scan a test target every few thousand pages. However, once benchmarks have been established and equipment and processes stabilised, this may be reduced to a random sampling of below 5 %.

D.5 Sample sets

Assemble a sample set of source documents for the purposes of evaluating scanner results against agreed quality criteria. These documents should be representative of the records to be scanned and should include examples of source documents whose quality is poor relative to the majority of the sample documents.

Test target sheets and detailed tests included in ISO 12653-1 and ISO 12653-2, may be used for black and white scanning. Where digitisation involves colour as an intrinsic part of the record, implementers should consider including a standard colour sheet with the image (such target sheets are commercially available and should comply with ISO 12641.) Colour calibration, matching and profiling should be monitored where colour scanning is used. While this may be an appropriate standard for perfect colour rendition, it is not the critical issue whereas resolution and overall quality are the significant items and are covered in ISO 29861.

D.6 Quality criteria for images

Quality criteria for images should include consideration of overall legibility:

— smallest detail legibly captured (e.g. smallest type size for text; clarity of punctuation marks, including decimal points);

— completeness of detail (e.g. acceptability of broken characters, missing segments of lines);

— dimensional accuracy compared with the non-digital source record;

— scanner-generated speckle (i.e. speckle not present on the non-digital source record);

— completeness of overall image area (i.e. missing information at the edges of the image area);

— density of solid black areas;

— colour fidelity.

D.7 Metadata

Procedures should specify the checks to be implemented to assess metadata quality assigned to images.

Issues which may be considered in the quality checking of metadata are:

— adherence to standards set by institutional policy or by the requirements of the digitization project;

— procedures for accommodating images with incomplete metadata;

— relevancy and accuracy of metadata;

— grammar – check for correct grammar, spelling and punctuation, especially for manually-keyed data;

— consistency in the creation of metadata and in interpretation of metadata;

— evaluation of the usefulness of the metadata being collected;

— synchronisation of metadata stored in more than one location, i.e. procedures should be in place to make sure metadata is updated in a synchronised manner across more than one location (e.g. information related to the image might be stored in the TIFF header, the digital asset management system, and other databases);

— completeness of metadata - all mandatory fields should be complete.

Of particular importance are the following.

a) **Verification of the accuracy of the file identifier**:

File names should consistently and uniquely identify both the digital resource and the metadata record (if it exists independently of the image). File identifiers will likely exist for the metadata record itself in addition to identifiers for the digitized resource, which may embed information such as page or piece number, date, project or institution identifier, among others. Information embedded in file identifiers for the resource should parallel metadata stored in a database record or header. Identifiers often serve as the link from the file to information stored in other databases and have to be accurate to bring together distributed metadata about a resource. Verification of identifiers across metadata in disparate locations should be made;

b) **Verification of the correct sequence and completeness of multi-page items**:

Pages should be in the correct order with no missing pages. If significant components of the resource are recorded in the metadata, such as the presence of attachments, documents with identifiable chapters or multi-page records, they should match up with the actual image files. A convention for describing these views should be followed and should match with the actual image files.

D.8 Documentation

Quality control data (such as logs, reports, decisions) should be captured in a formal system and should become an integral part of the image metadata at the file or the project level. This data may have long-term value that could have an impact on future preservation decisions.

D.9 Processes for re-digitization

If more than 1 % of the total number of images and associated metadata examined in a randomly selected sample are found to be defective for any of the reasons listed above, the entire output since the last quality check should be re-inspected. Any specific errors found in the random sampling and any additional errors found in the re-inspection should be corrected. If less than 1% of the batch is found to be defective, then only the specific defective images and metadata that are found should be redone.

D.10 Common faults

D.10.1 General

Broadly speaking, quality faults can be categorized as “implementation faults”, “process faults” or “operator faults”. Implementation faults are those that can be avoided, providing appropriate procedural controls are in place to guide the digitization. Process faults are normally out of the control of the operator and need to be addressed by a supervisor to the process. Operator faults are the day-to-day faults that are made by the operator as they work.

D.10.2 Implementation faults

There are a number of faults that can be avoided with appropriate specification of procedures to guide the implementation. These include:

— dirty non-digital source records;

— incorrect file-size and format, where files are made to wrong size or with wrong choice of file format;

— compression, where files are made with an inappropriate type or level of compression.

D.10.3 Process faults

There are a wide variety of process faults that can be caused by many problems within the workflow. These problems can include:

— incomplete or inaccurate specifications or process documentation;

— faulty capture hardware (incorrectly calibrated and characterised devices);

— faulty software (inaccurate image processing or faulty image links within database);

— incorrectly established colour management systems;

- low quality original data (either non-digital surrogates or legacy digital image files);
- inaccurate source metadata.

D.10.4 Operator faults

These faults are caused by some form of operator error within the workflow and can include:

- basic capture faults;
- cropping that has cut into the image, is too loose, or is uneven;
- orientation of the image is the wrong way around, or upside down;
- exposure of the image is too light or too dark;
- focus, where the image is out of focus;
- daily calibration, where the capture device has not been calibrated;
- basic image processing faults;
- File Optimisation Faults, where incorrect adjustments are made to the colour, contrast and brightness of the image during processing;
- incorrect file-naming, where image files are incorrectly named or use non-unique names;
- basic metadata attribution faults;
- incorrect data entry, where data is incorrectly entered into the management system;
- incorrect use of controlled vocabulary, the use of words not established within scope notes.

Annex E
(informative)

Recommended staff skill sets

E.1 General

This annex outlines the types of skill areas and tasks that staff engaged in digitization should be expected to possess. It is included as an aid to organizations in meeting best practice guidelines 6.3.5. Quality control procedures should be defined, documented and implemented.

E.2 Recommended skills set

Staff training to support digitization should include the areas, and be provided to all levels of skill expected within the organization, as shown in Table E.1.

Table E.1 — Areas and levels of recommended skills

Skill area	Tasks
Management	— Assessing the business case for digitization. — Negotiating purchase, ongoing service, maintenance of equipment and supplies. — Digitization contractor management skills or supervision of staff undertaking digitization processes dependent on whether an in-house or outsourced model is followed.
Business analysis	— Defining the workflow for the digitization process. — Defining the integration of digitized records into existing business information systems/workflows. — Selecting image format. — Determining image enhancement requirements. — Identifying information architecture for business-process support.
Systems analysis	— Selection of scanner hardware. — Defining storage requirements. — Integration of hardware, imaging equipment and software. — Integration of digitization requirements into existing organizational IT infrastructure. — Compliance with national and organizational IT standards. — Testing of configurations. — Ongoing support of digitization equipment (where necessary). — Definition of policies and procedures to ensure authenticity and integrity of digital images.

Table E.1 (*continued*)

Skill area	Tasks
Managing records	— Ensuring legislative compliance. — Integration with organizational records and business information systems. — Integration with existing classification and disposition regimes. — Defining file naming conventions. — Defining and implementing disposition process. — Defining metadata. — Monitoring of quality of metadata. — Managing the non-digital source records after digitization.
Equipment Operations	— Operating scanners. — Applying any defined selection criteria. — Carrying out quality checking on digital records. — Adding metadata to digital records.

Bibliography

[1] ISO/TS 12032, *Document imaging — Statistical sampling for document images*

[2] ISO 12033, *Document management — Electronic imaging — Guidance for the selection of document management compression methods*

[3] ISO 12653-1:2000, *Electronic imaging — Test targets for the black and white scanning of office documents — Part 1: Characteristics*

[4] ISO 12653-2:2000, *Electronic imaging — Test targets for black and white scanning of office documents —Part 2: Methods of use*

[5] ISO 12641:1997, *Graphic technology — Prepress digital data exchange — Colour targets for input scanner calibration [using lossy compression]*

[6] ISO/TR 15489-2:2001, *Information and documentation — Records management — Part 2: Guidelines*

[7] ISO 19005-1:2005, *Document management — Electronic document file format for long-term preservation — Part 1: Use of PDF 1.4 (PDF/A-1)*

[8] ISO 29861:2009, *Document management applications — Quality control for scanning office documents in colour*

[9] IEC 82045-2:2004, *Document management — Part 2: Metadata elements and information reference model*

[10] International Council on Archives, *Principles and Functional requirements for records in Electronic Office Environments, Module 1 — Overview and Statement of Principles*, 2008

[11] International Council on Archives, *Principles and Functional requirements for records in Electronic Office Environments, Module 2 — Guidelines and Functional Requirements for ERMS*, 2008

[12] International Council on Archives, *Principles and Functional requirements for records in Electronic Office Environments, Module 3 — Guidelines and Functional Requirements for Records in Business information systems*, 2008

[13] The InterPARES 2. Project dictionary 2009, Available at http://www.interpares.org/ip2/display_file.cfm?doc=ip2_dictionary.pdf&CFID=563511&CFTOKEN=89427946

[14] ANSI/AIIM/ARMA TR48-2004. Technical Report. *Framework for integration of electronic document management systems and electronic records management systems*, Lenexas: ARMA International

[15] ANSI/NISO Z39.87-2006, *Data Dictionary — Technical Metadata for Digital Still Images*

[16] Archives New Zealand's Recordkeeping Standard S6: Digitization Standard, January 2006, Available at http://continuum.archives.govt.nz/files/file/standards/s6/index.html

[17] Queensland State Archives. “Guidelines for the Digitization of Paper Records”, v. 2, April 2006. Available at http://www.archives.qld.gov.au/publications/digitization/digiguideline.pdf

[18] Queensland State Archives, “Digitization Disposal Authority” April 2006. Available at http://www.archives.qld.gov.au/publications/digitization/QSA6814%20DisposalPolicy.pdf

[19] National Archives of Australia. Glossary of Records Management terms. Available at http://www.naa.gov.au/records-management/glossary/index.aspx

[20] New Zealand State Services Commission "Trusted Computing and Digital Rights Management Principles and Policies", September 2006. Available at http://www.e.govt.nz/policy/tc-and-drm/principles-policies-06/tc-drm-0906.pdf

[21] The Preservation Management of Digital Material Handbook, Digital Preservation Coalition, November 2008. Available at www.dpconline.org/graphics/handbook/

[22] Technical Advisory Service for Images (TASI), *In-depth Report: Quality Assurance*. Available at http://www.tasi.ac.uk/advice/creating/qassurance.html

[23] Technical Advisory Service for Images (TASI). Available at http://www.tasi.ac.uk/advice/creating/qassurance.html

[24] United States National Archives and Records Administration, *Technical Guidelines for digitizing Archival Materials for Electronic Access: Creation of Production Master Files — Raster Images*, June 2004. Available at http://www.archives.gov/preservation/technical/guidelines.pdf

Standards Australia

Standards Australia is an independent company, limited by guarantee, which prepares and publishes most of the voluntary technical and commercial standards used in Australia. These standards are developed through an open process of consultation and consensus, in which all interested parties are invited to participate. Through a Memorandum of Understanding with the Commonwealth government, Standards Australia is recognized as Australia's peak national standards body.

Standards New Zealand

The first national Standards organization was created in New Zealand in 1932. The Standards Council of New Zealand is the national authority responsible for the production of Standards. Standards New Zealand is the trading arm of the Standards Council established under the Standards Act 1988.

Australian/New Zealand Standards

Under a Memorandum of Understanding between Standards Australia and Standards New Zealand, Australian/New Zealand Standards are prepared by committees of experts from industry, governments, consumers and other sectors. The requirements or recommendations contained in published Standards are a consensus of the views of representative interests and also take account of comments received from other sources. They reflect the latest scientific and industry experience. Australian/New Zealand Standards are kept under continuous review after publication and are updated regularly to take account of changing technology.

International Involvement

Standards Australia and Standards New Zealand are responsible for ensuring that the Australian and New Zealand viewpoints are considered in the formulation of international Standards and that the latest international experience is incorporated in national and Joint Standards. This role is vital in assisting local industry to compete in international markets. Both organizations are the national members of ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission).

Visit our web sites

www.standards.org.au www.standards.co.nz

GPO Box 476 Sydney NSW 2001
Phone (02) 9237 6000
Fax (02) 9237 6010
Email mail@standards.org.au
Internet www.standards.org.au
SAI Global Customer Service
Phone 13 12 42
Fax 1300 65 49 49
Email sales@saiglobal.com

Level 10 Radio New Zealand House
155 The Terrace Wellington 6011
(Private Bag 2439 Wellington 6140)
Phone (04) 498 5990
Fax (04) 498 5994
Customer Services (04) 498 5991
Information Service (04) 498 5992
Email snz@standards.co.nz
Internet www.standards.co.nz

ISBN 978 1 74342 076 8

Printed in Australia