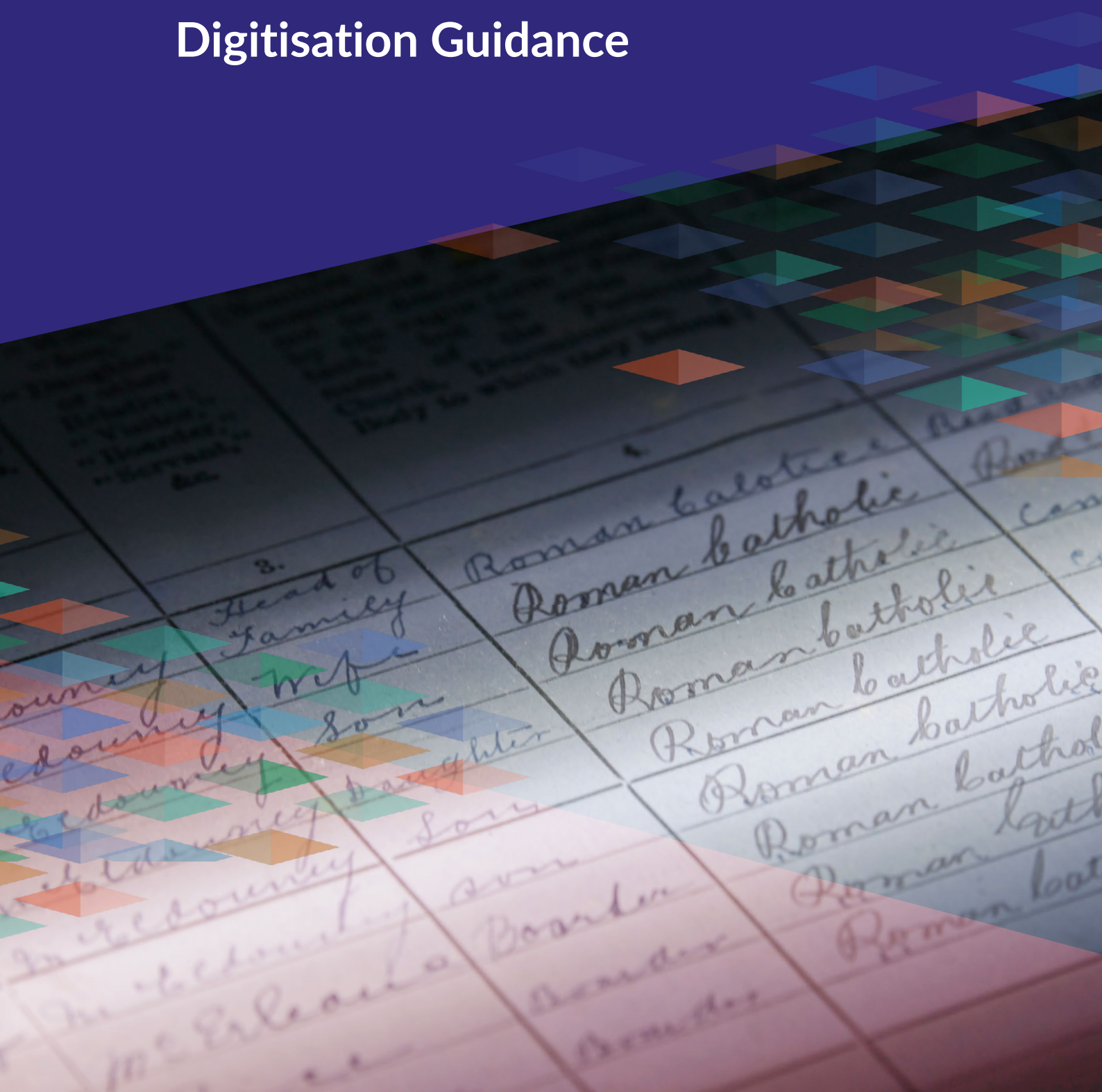




An Chartlann Náisiúnta  
National Archives

# National Centre for Research & Remembrance

## Digitisation Guidance



---

## Document Control

<b>Document title</b>	National Centre for Research and Remembrance: Digitisation Guidance
<b>Reference number</b>	
<b>Version</b>	1
<b>Date</b>	October 2024
<b>Status</b>	Final
<b>Contributors to this version</b>	Keeper and Senior Archivists, Digital Management Division
<b>Document objectives</b>	To provide guidance on high level considerations of digitisation of analogue records to departments and authorities within scope of the National Centre for Research and Remembrance.
<b>Intended recipients for final document</b>	Departments and agencies holding records that fall within scope of the National Centre for Research and Remembrance and who will transfer records in both original paper and digitised formats to the National Archives.
<b>Approving body for final document</b>	National Archives
<b>Date approved</b>	October 2024
<b>Date of issue</b>	December 2022
<b>September 2024</b>	December 2022
<b>Review date</b>	Ongoing
<b>Contact for review</b>	Keeper, Digital Management Division

---

# Contents

<b>Introduction</b>	<b>4</b>
<b>Stages of a Digitisation Project</b>	<b>6</b>
<b>01 Project planning</b>	<b>7</b>
1.1 Business case and project initiation	7
1.2 Identification and assessment of collection1	7
<b>02 Project development</b>	<b>10</b>
2.1 Development of a workflow	10
<b>03 Project implementation</b>	<b>12</b>
3.1 Cataloguing of a collection	12
3.2 Conservation assessment	14
3.3 Digitisation of documents	16
3.3.1 In-house digitisation	16
3.3.2 Contracting of digitisation services	19
3.3.3 General issues to consider	24
<b>04 Post-scanning</b>	<b>32</b>
4.1 Storage of digitised resources	32
4.1.1 Storage solutions	32
<b>GDPR and privacy</b>	<b>36</b>
<b>Copyright</b>	<b>37</b>
<b>Access to digitised resources</b>	<b>38</b>
<b>In summary</b>	<b>39</b>
Appendix 1: Naming convention considerations	41
Appendix 2: Technical specifications	42
Appendix 3: Cataloguing workflow checklist	45
Appendix 4: Digitisation workflow checklist	47
Appendix 5: Quality Assurance Checklist	48
Glossary of terms	50

---

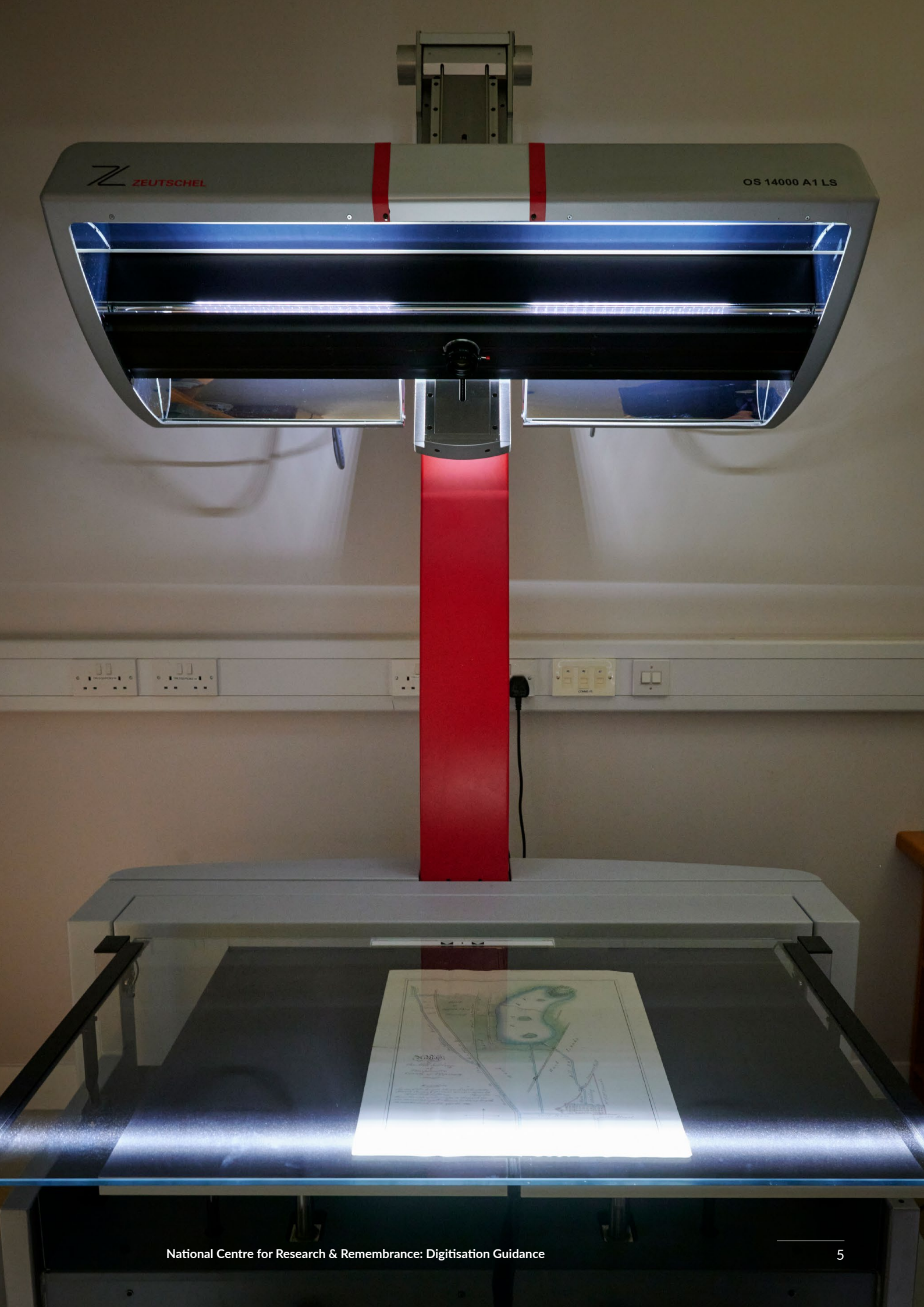
## Introduction

This document is intended to provide practical guidance to departments and agencies who intend embarking on digitisation of records as part of plans to develop an archival repository of records of institutional trauma within the National Centre for Research and Remembrance.

The guidance is issued in order to highlight some of the potential concerns and considerations required as part of any digitisation project. It is not intended to be exhaustive and will be added to as the project progresses. Guidance is issued in order to align practices within organisations that will reduce manual input and will facilitate transfer of digitised and analogue records to the National Archives in the coming years.

### **Assumptions made include:**

- Digitisation in this context applies to records already identified as warranting permanent preservation as archives;
- The original source records will be retained permanently.



---

## Stages of a Digitisation Project

A digitisation project can be broken down into a number of phases:

### 01 Project Planning

- Identification of a collection to be digitised
- Development of realistic timeframes
- Identification of end user needs
- Identification of resources required, including staffing, ICT requirements, storage requirements and specialist equipment and expertise
- Whether services will need to be procured

### 02 Project Development

- Development of cataloguing, conservation digitisation workflows
- Testing of outputs
- Configuration of scanning equipment
- Identifying requirements for inclusion in tender process

### 03 Project Implementation

- Implementation of cataloguing workflow in parallel with conservation
- Initiation of scanning of documents
- Post-scanning QA procedures, including quality control and rescanning

### 04 Post-Scanning

- Transfer of images to storage
- Production of user image
- Integration into user access system, such as website, management system or database system



## 1.1 Business Case and Project Initiation

- a This should set out the reasons for the digitisation project and why it is necessary to digitise archival records.
- b It should also highlight any potential issues arising, including costs associated with equipment, staffing, ICT requirements, including digital storage and preservation, training, procurement of specialised equipment or external expertise, and potential risks or associated pitfalls to both digitisation and failure to digitise and benefits to the organisation.
- c Resources such as the [Public Service Project Management Handbook](#) and templates may be useful as part of the planning process<sup>1</sup>.

## 1.2 Identification and Assessment of Collection

- a The starting point for any digitisation project is the identification of records to be included. The criteria for this should be set out as part of your business case and Project Initiation Document, and may include a business requirement to access records in an efficient manner to answer queries such as Subject Access Requests, the fragile condition of records and their continuing deterioration, and a wider requirement to facilitate public access.
- b Any assessment should determine whether the collection has been catalogued to archival standards, or whether any other level of intellectual control has been implemented. Without this, it will be impossible to accurately plan for timeframes, potential risks arising and the correct methodology to be used. This stage should also identify the format of records, the scale of the collection and the quantity of records to be digitised, and the condition of the records. Conservation intervention may be required to capture a clear image. Conservation work can be time-consuming and costly and should be factored in to the overall timeframe and budget.

<sup>1</sup> <https://www.gov.ie/en/publication/6748e-project-management-resources/>

- c Depending on the scale and other factors such as conservation requirements, data protection or copyright concerns, it may be decided that not all of a collection will be digitised, or digitisation will be implemented on a phased basis over a number of years. The size and format of records, along with an assessment of their condition, will determine the most suitable method of digitisation and the allocation of resources for conservation and storage of original records.







## 2.1 Development of a workflow

- a Each phase of the process should develop a separate workflow document that can be accessed by staff engaged in the project. The workflow document should be drafted as part of the initial implementation of a project, but the document will evolve, particularly in the early stages of implementation.
- b A level of inter-dependency is likely at each phase of the process, so a standardised approach will help to identify any issues arising early in the project.
- b For example, the cataloguing process is the first operational phase. The information gathered as part of this will capture accurate data about the size of a collection, as well as the content and condition of the records. This will determine the development of both conservation and digitisation workflows, as well as decisions around access and security. It may also result in changes to the overall project plan, such as unrealistic budgets or storage requirements.
- d Staff across all functions involved in a digitisation project should provide input to the development of workflows to ensure all operational considerations are captured.
- e Workflow documents should be amended as necessary. However, this role should be centralised and delegated to a supervisor or manager. Staff should be informed of any changes to the workflow.
- f The workflow document should set out role profiles, reporting structures and methodology. It should be a simple, succinct document that is easily understood by all staff.





### 3.1 Cataloguing of a collection

- a Intellectual control of a collection at the beginning of a digitisation project is essential to ensure a successful outcome. This ensures that images are associated from the beginning with archival metadata produced as part of the cataloguing process. It also facilitates access to the collection and linking of images to a unique reference code in a database system, such as an electronic document management system, digital asset management system or website.
- b Where unique reference codes are not included as part of the preparation and cataloguing process, this will have a serious impact on the post-processing phase, including the ability to correctly identify the content of digitised images and the level of manual input and staffing resources required.
- c Cataloguing facilitates planning of indexing requirements to increase accessibility, which can be included as part of the cataloguing processing. The indexing requirements may differ depending on final user requirements, and whether access is required for business purposes only or publication online.
- d It is essential that naming conventions are developed at an early stage to ensure information is catalogued and captured in a standardised way. Naming conventions should include a unique identifier such as file reference and should avoid use of spaces or special characters that inhibit machine reading and accessibility, and may create complications around interoperability between systems.
- e For smaller projects descriptive names can be allocated that indicate the file series or content of the records, but for larger scale projects this will be impractical. In such cases, the use of a file reference that links to archival metadata should suffice to identify the records.

- f** The use of descriptions relating to living individuals should be avoided, or a separate anonymised field captured as part of the cataloguing process that can be used to facilitate access.
- g** The National Archives catalogues collections to ISAD(G) standard, which is the International Standard of Archival Description. All records identified for transfer to the National Archives must comply with our cataloguing template. Training can be provided to ensure the necessary fields of information are captured as part of any digitisation project.

## 3.2 Conservation Assessment

- a A conservation assessment should be included as part of the planning stage.
- b Where conservation is required, adequate resourcing should be included in the overall budget.
- c Depending on the size and nature of the collection, conservation intervention may take a significant amount of time and should be a central consideration in the development of realistic timeframes.
- d In the majority of cases, internal conservation expertise will not be available and external expertise will be required. Any need to tender for conservators should be included in the budget and timeframe.
- e A collection in fragile condition may be unsuitable for digitisation and will influence decisions on whether to facilitate offsite scanning, the equipment used and expectations around the quality of the scanning output.
- f The National Archives can provide guidance on assessing a collection's suitability for digitisation, and issues to consider when engaging with contractors, including relevant qualifications, accreditation and experience.



## 3.3 Digitisation of documents

Digitisation of records involves a number of steps. Depending on the nature of the collection and the end goal of the project, the first issue to be determined is whether capacity, equipment and skills exist in-house to undertake the project or whether it needs to be outsourced.

Project Implementation → Digitisation of Documents → In-house digitisation

### 3.3.1 In-house digitisation

*Issues to be considered as part of in-house digitisation:*

- a Whether the project includes a budget for the procurement of specialist equipment to undertake the work, including archival quality scanning equipment and digital cameras, and associated hardware and software as well as maintenance and support contracts;
- b The availability of staff with sufficient technical ability to undertake the work, including continuity of staff for the duration of the project. The rate of staff turnover will impact the project deliverables and timeframe;
- c The availability of staff to prepare records for digitisation, including removal of staples or tags and re-attachment after completion, oversight of the retrieval and return of records to storage, and ensuring records are presented in the correct order;
- d The availability of specialist knowledge and understanding of issues such as data protection, copyright, risk assessment, project management, procurement;
- e Sufficient support from ICT personnel and understanding of the requirements of a digitisation project, including hardware, software and systems requirements and updates; storage; ongoing maintenance and access, including digital preservation requirements;



- f Appropriate space to accommodate staff, records and equipment for the duration of the project, including size, light quality, requirements for darkened spaces, requirements for upgraded wiring/connectivity;
- g Development of procedures and workflows with clear identification of roles and responsibilities, including quality assurance, document preparation, document scanning, metadata capture, quality control of output, oversight and return of records to storage, tracking of records and overall management and monitoring of progress.
- h Inclusion of input from a professionally trained archivist in either a project delivery or project assurance role.
- i Clear benchmarks to be achieved, including timeframes, acceptable errors or quality assurance concerns and a reporting mechanism including a requirement for weekly reports of work carried out.



ches plantation  
 contain 11 perches  
 33 perches  
 to the arch in  
 holding length  
 is

## 3.3.2 Contracting of digitisation services

### *Issues to be considered as part of procurement of digitisation services:*

- a Whether digitisation will take place on site or offsite in the contractor's premises;
- b While the National Archives would strongly recommend onsite digitisation only, this must be considered within the overall context of the project, including the scale and nature of the records to be digitised, as well as their condition. Cost is only one of a number of factors to consider in choosing onsite or offsite digitisation. These include:
  - b.2 availability of archival quality scanning equipment that meets the requirements of the project, including the size and nature of the documents to be digitised;
  - b.3 proposed methodology, including supervisory structures and quality assurance processes;
  - b.4 factors influencing delivery within required timeframes such as office opening hours and availability of trained staff;
  - b.5 data security requirements for access by third parties to internal ICT infrastructure;
  - b.6 availability of suitable onsite accommodation.
- c Departments should assess the potential risks to a collection to be sent offsite as part of any project planning and should ensure such concerns are stipulated in contracts and agreed workflows with contractors. The removal of records offsite substantially increases the risk to the records and facilitates a lack of oversight on an operational level of scanning and document handling practices.

- d Issues to be considered as part of any tendering process should include:
  - d.2 Security measure for protecting physical documents throughout the process (e.g. document tracking, vehicles for transport, staff training, storage conditions, premises security).
  - d.3 Inspection visits in advance of awarding tender.
  - d.4 Specify length of time documents can remain at contractor premises.
  - d.5 Specify timeframe within which documents must be returned.
  - d.6 The contractor should provide details of the experience and skillsets of proposed staff, including experience of working with archival records and digitisation project management.
  - d.7 The contractor should provide details of the equipment to be used and the level of training provided to their operators, including document handling and the suitability of the digitisation space for scanning archival documents.
  - d.8 The format of the digitised output and associated metadata that will be provided with the files, including considerations such as file format, PPI that the documents will be scanned to and colour.
  - d.9 The contractor should provide details of any sub-contractors to be engaged, such as digitisation of specialised formats like audio visual or audio files, and it should be clear that any sub-contractor will be subject to the same stipulations, particularly with regard to confidentiality and security.

- d.10 Determination of the most efficient process for the transfer of images, such as secure file transfer system or encrypted hard drives. This will be impacted by the project size and scale.
- d.11 Weighting of scoring requirements in favour of proposed methodology, including pre and post-processing requirements and file transfer; relevant skills and experience; quality management; security of documents and data.
- d.12 Makeup of the assessment panel, including personnel with relevant expertise of digitisation projects for archival documents.
- e Responsibility for provision of equipment, and proof that any equipment provided meets the required standards for the digitisation of archival records;
- f Clear identification of roles and responsibilities, including whether the contractor will be responsible for labour intensive activities such as removal of records from storage, document preparation and post-processing tasks such as batch file renaming, reformatting, cropping of images, deletion of blank pages;
- g Clear indication and assurance of the security and confidentiality of the records;
- h Clear benchmarks to be achieved by the contractor, including timeframes, acceptable errors or quality assurance concerns and a reporting mechanism, including a requirement for weekly reports of work carried out;

- i Sufficient support from internal ICT personnel and understanding of the requirements of a digitisation project, including hardware, software and systems requirements and updates; storage; ongoing maintenance and access, including digital preservation requirements. Regardless of whether a project is outsourced, there will be internal ICT and other staff requirements to manage outputs in the short and longer term;
- j Planning of procedures and processes for the transfer of any images to Departmental servers, including a dedicated point of contact within ICT Unit;
- k Adequate space to accommodate staff, records and equipment for the duration of the project;
- l Identification of Departmental staff responsible for liaison with the contractor, including tracking of documents to digitisation for both onsite and offsite projects; management of benchmarks and quality assurance oversight; and winding up of a project to ensure all images are transferred in the agreed formats;



### 3.3.3 General Issues to Consider — Project Planning Stage

- a Expertise required to manage digitisation of archival documents, including availability of qualified archivists and staff with relevant experience in cataloguing, digitisation and project management.
- b Potential delays in the procurement of goods and services and the potential impact on expected timeframes.
- c Staff training, including specialised scanning and photography training and document handling. The National Archives can provide training in document handling to staff engaged in digitisation projects.
- e The highest quality and resolution possible should be captured. The resolution should be determined at the planning stage and should be applied consistently throughout the process. This will increase the longevity of the digitised file and facilitate greater reuse.
- f The resolution chosen will depend on the project budget, storage availability, the nature and format of the original records and the expected use to which the digitised records will be put. The greater the resolution, the larger the digitised file or image will be, which may significantly impact the overall storage requirements and costs.
- g The National Archives recommends digitisation of records at a minimum resolution of 300ppi in line with accepted international standards. A higher resolution may be required, and should be considered as part of the procurement of suitable equipment and the final output required.



### 3.3.3 General Issues to Consider — Project Planning Stage

- h Not all digitisation equipment is capable of capturing archival quality resolution, or is suitable for the digitisation of archival documents.
- i Where original records are to be destroyed, other factors need to be considered such as whether authorisation for disposal is required under section 7 of the National Archives Act, 1986. The National Archives is unlikely to authorise disposal of archival records in analogue formats in the absence of a Civil Service wide records management and digital preservation strategy.
- j For records that have no archival value and are of a supporting or processing nature, departments and agencies should decide the required level of quality for digitised information sufficient to support their business needs.
- k The best equipment available at the time should be used to digitise records, in line with budgetary limitations. The standard of equipment used will determine the quality of output and is one of the main factors in identifying the likelihood of accessibility issues over time. Other factors include storage conditions and continued monitoring of accessibility and image quality.
- l The choice of equipment will depend on the nature of the records to be digitised and should be determined at an early stage as part of project planning. The use of specialised contractors with archival quality scanning equipment may be a more feasible option and will influence decisions around onsite or offsite digitisation. Further guidance on scanning equipment can be provided by the National Archives.

### 3.3.3 General Issues to Consider — Project Development Stage

- a If automated transcription processes are envisaged, testing of resolution should be undertaken as part of the digitisation project planning. While 300ppi is considered the minimum preservation standard, automated transcription is currently more accurate at 400ppi for handwritten documents. If transcribing printed documents, a lower resolution may be sufficient.
- b Testing will allow decisions on resolution to be made at an early stage, which is crucial for determining requirements for storage and how scanning equipment will be configured.
- c Higher resolution images may result in a greater accuracy level for transcription, but this should be balanced against the overall scale, quality control requirements, available resources, including personnel, and intended use, particularly if publication or widespread use by the public is envisaged. A 1% accuracy difference when working on a large scale transcription or digitisation project may substantially increase the level of manual input and staffing resources required. This is a major consideration when working with limited resources to tight or immovable deadlines.
- d If generating an OCR transcription of the document as part of the digitisation process, consideration should be given to what metadata will be required to bundle the transcription with the digitised document.
- e Metadata fields captured should conform to recognised international standards. The National Archives applies ISO 15836, also known as Dublin Core metadata standard, to our digitisation process. This is separate to the cataloguing data generated as part of the archival processing. Some of the information captured as part of the cataloguing process will feed into the digitisation metadata process

### 3.3.3 General Issues to Consider — Project Development Stage

- f The metadata standard used will depend on the format of the original records and should be adapted to fit requirements for different types and formats of records, business process and system capabilities. This may result in the use of elements of various metadata standards.
- g The National Archives can provide guidance to organisations on correct listing standards and metadata to be applied to both the cataloguing and digitisation processes. We would encourage organisations to contact the National Archives at the planning phase of any project. This is particularly important where third party contractors are to be engaged. Such specifications should be included in any requests for tender as part of the procurement process.
- h Testing of equipment settings should take place to ensure they are adequate for the capture of clear images and different document formats. Settings may need to be adjusted depending on the nature of the records in question. For example, typed documents generally require less adjustment than handwritten documents.
- i Equipment should be adjusted to ensure automatic generation of as much technical metadata as possible at the image capture level to reduce the level of manual input required. Automation and standardisation of processes will reduce staffing requirements and will increase accuracy.
- j Where possible, the output from digitisation should be configured to reduce the level of manual input and potential for human error. Some scanning equipment can be configured to map output to specified locations such as dedicated network drives and/or to pre-configured naming conventions and folder structures.

### 3.3.3 General Issues to Consider — Project Implementation Stage

- a Once digitisation commences, the quality assurance process should begin immediately. Regular checks of different batches and format types will capture potential issues arising, and procedures may need to be adjusted accordingly.
- b For large-scale digitisation projects, a testing phase to configure settings and adjust initial workflows should be included.
- c Documents should be digitised in the correct order to reduce the level of manual input required at the end of the processing. Digitisation will automatically generate a sequential number for each image. If the documents have not been placed in the correct order, a large amount of manual input will be required to replicate the original order.
- d Workflows should include a document preparation and post-scanning phase to ensure intellectual control and protection of the records, including return to storage. These operational tasks should be factored in to the work day and sufficient time allocated. This may impact the staffing levels required and/or the amount of digitisation possible within a particular timeframe.
- e The quality control process should ensure preservation of the original order, which is key to preserving the authenticity and integrity of the scanned record. The digitisation process should maintain the archival hierarchy developed as part of the cataloguing process. Any deviation from the original order may lead to loss of original context and misrepresentation of the records.

### 3.3.3 General Issues to Consider — Project Implementation Stage

- f The suspension of digitisation in the middle of the process or unnatural break in work should be avoided, where possible. This will reduce the risk to the records and will increase continuity in the level of quality assurance achieved.
- g Any updates to software or ICT processes should be scheduled outside of scanning times.
- h Blank pages should not be digitised, but blank forms etc. should be included. Where a decision is taken to only partially digitise certain classes of records, or to omit or remove individual documents from files, this should be clearly stated in the project plan and methodology statement. Where documents have been omitted from a file for any reason, this should be clearly indicated on a cover sheet captured as part of the digitisation process.

### 3.3.3 General Issues to Consider — Post-Scanning Stage

- a All records should be restricted and made read only to avoid any alteration, including additions or deletions. This is particularly important where archival records are digitised to facilitate ongoing business requirements, such as Subject Access Requests.
- b Where additions or amending statements are to be appended to archival records they should be classed as a separate sub-series and stored separately, but linked to the original record through metadata capture. The evidential integrity of the digitised record will be undermined if the original record is altered in any way.





## 4.1 Storage of Digitised Resources

Storage requirements should be included as part of the scoping of any digitisation project. The type and level of storage will depend on the nature of the records, the scale and quantity of images captured, the resolution at which they were captured, whether it is necessary to create a lower resolution access copy and access requirements for business purposes.

Ongoing storage costs will arise regardless of the type of storage solution used, and should be included in annual budget planning. Digitised resources should also be included in disaster preparedness and business continuity plans.

### Storage solutions include:

A dedicated server or space on an existing server on site in an organisation (often referred to as an On-Prem solution). This solution allows for greatest control of the records and may be most suitable for records of a highly sensitive nature. It requires significant support from in-house ICT units, including responsibility for maintaining accessibility and regular checking of digitised resources as part of a digital preservation strategy.

### Two main types of Off-Prem solutions exist:

- a Storage on a government server offsite. This allows for storage within a secure government network, but reduces the level of access and direct control of the records.
- b A Cloud solution involves the outsourcing of storage to a third party private operator, who manages all aspects of storage, including responsibility for maintenance and access to records. This option substantially reduces the level of access and direct control by an organisation and its staff, but may be appropriate for projects of a large scale. Where cloud solutions are chosen, an exit strategy should be put in place as part of any contract arrangement to ensure continuing access to your data. Issues of ownership and access when required should also be clearly understood as part of any contract arrangement.



- c The location of the Cloud may also influence the decision. Cloud storage for sensitive and confidential records, or records subject to GDPR containing personal data of living individuals, should be located within the jurisdiction. The government's National Digital Strategy and guidance issued by the Office of the Government Chief Information Officer and the National Cyber Security Centre should be consulted. The National Archives has no role in prescribing the location or quality of digital storage.
- d Where records contain identifiable personal information, particularly special category personal data, of living individuals it would be advisable to include storage requirements as a risk factor when undertaking a data protection impact assessment to determine the potential risks of each solution. While outsourcing of storage brings inherent risks, inadequate internal expertise and resourcing of ICT infrastructures also creates risk.

### 4.2 Storage of Original Paper Records

Storage requirements for original records should also be considered as part of any project. Good quality storage in a stable, dry environment will substantially reduce the rate of deterioration and risk to analogue collections. This will be reinforced by the use of archival quality materials, such as acid free boxes and folders. Organisations should also have disaster plans and business continuity plans in place for archival and vital records. Digitised records may form part of a business continuity plan.

Storage facilities should be dry, secure and access should be limited. Where storage of records has been outsourced, the department or agency should ensure regular on-site visits take place and areas where records are stored are inspected. This requirement should be stipulated as part of any contract.

Departmental records, including archival records, should not be stored in commercial facilities outside the jurisdiction.



---

## GDPR and privacy

Digitisation constitutes a form of processing for the purposes of data protection. A data protection impact assessment should be included at the project planning stage for any collections involving personal identifiable information of living individuals.

The end user requirements, including facilitating access by individuals to their own information, may determine the nature and level of digitisation undertaken.

Where the immediate need is not related to access by living individuals, the risk associated with digitisation may be reduced by prioritising older collections, or parts of collections, and undertaking digitisation on a phased basis over a number of years.

Where a decision is taken to proceed with digitisation to facilitate access to records containing information about living individuals, strict protocols should be developed to limit access to individuals who have a clear requirement to view the records and security measures are put in place. This should include the creation of an access audit trail and restrictions on rights to amend or delete records. Access should be based solely on business need and not on other factors such as seniority or grade.

---

## Copyright

Copyright is an intellectual property right which protects the owner's creative skills and labour. A work subject to copyright can include literary, artistic, dramatic, musical, photographic, film or sound recording. It gives the copyright owner or holder the right to control the use of the work. These rights include the right to copy, publish, reproduce, broadcast or perform the work in public. **As copyright is an intellectual right, ownership is separate to ownership of the physical item.**

When assessing collections to be digitised, potential copyright concerns should be noted. This is particularly important for records created by third parties. The presence of a record within a file that falls within the definition of a departmental record may include records where copyright is owned by a third party. This is especially relevant where publication of the digitised record is a possibility. An example of this found in the National Archives is a letter on a Department of Foreign Affairs file written by James Joyce. The letter remains the property of the State, but the ability of the State to publish that letter is limited as the copyright remains the property of the Joyce Estate.

Copyright may also apply to work undertaken by contractors as part of the digitisation process. It is essential that any contract for service clearly states that all intellectual property rights and all copies of digitised resources remain the property of the contracting authority.

A failure to clearly indicate this may lead to disputes around ownership and may limit the use to which the images are put, including publication as part of any outreach and education activities or sharing with third parties, including transfer of digitised records to the National Archives.

---

## Access to digitised resources

The access requirements will be intrinsically linked to the overall aims of the project and should be assessed as part of the project planning stage. This assessment should also consider the level of indexing required as this can be undertaken as part of the cataloguing process. Indexing of the source data increases the findability and usability of the digitised resources, for both business needs and wider researcher and access requirements, especially for resources published online.

Access for business purposes may require the development of a user interface linked to the digitised images held on a server. This may involve procurement of a specialist Digital Asset Management System using proprietary software, or development of a custom built solution in-house. A number of factors will influence the final decision, including the availability of in-house ICT support; procurement and maintenance costs associated with proprietary systems, including the number of licenses required; the scale and searchability of the digitised assets; sensitivity of the assets and the level of access controls required.

Where wider public access is envisaged, it may be necessary to include the development of a website into any project plan. If the records are openly accessible, a digital asset management solution may be integrated into the website. This can be used to facilitate public access online and staff access on a more secure internal structure.

Access should be provided to a user copy only in a format that allows for optimal usability. Access should be restricted to the source data and preservation copy of digitised resources. PDF is the format most widely used to facilitate access to text documents and images, but access to more specialised formats such as audio recordings may also be required. The level of functionality provided will depend on the budget available, the nature of access required, the quality of the metadata captured as part of the cataloguing and digitisation processes and the quality of the digitisation output.

---

## In summary

- 1** Project planning is key to understanding the overall aims of the digitisation project and potential issues and risks associated with the collection in question.
- 2** The end user should be considered at the beginning of the process, in order to ensure issues around accessibility, privacy, security and usability of the digitised resources is integral to all stages of the process.
- 3** Quality assurance procedures should be defined, documented and implemented in order to ensure the integrity and accuracy of the digitised resource, including retention of the original order.
- 4** Procedures should be developed with clear identification of roles and responsibilities, including document preparation, document scanning, metadata capture, oversight and return of records to storage, tracking of records and overall management and monitoring of progress.
- 5** Digitisation projects will require an input of resources, including ICT resources, staff time and management, regardless of whether the processing is outsourced.
- 6** Digitised resources should be included in any disaster preparedness and business continuity planning, including ongoing storage requirements and procedures for the regular checking of resources as part of digital preservation policy.

---

## In summary

- 7** Metadata capture and cataloguing of records should conform to recognised international standards to increase interoperability and access to records.
- 8** Indexing requirements to increase searchability should be integrated into the cataloguing workflow.
- 9** Digitisation equipment should be configured to automatically capture technical metadata to reduce the level of manual input required.
- 10** Digitisation of resources should be captured at the highest resolution possible, in line with overall project aims, storage and budget.
- 11** Images should be captured at a minimum of 300 ppi, but testing should be undertaken in the early stages to ensure the output aligns with expected uses and technical requirements.
- 12** Naming conventions should be developed to ensure consistency in the level and type of data captured, in order to facilitate access and linking of data.
- 13** A data protection impact assessment should be carried out prior to commencement of digitisation of collections containing personal identifiable information of living individuals.



---

# Appendix

Appendix → Appendix 1

## Appendix 1

### Naming convention considerations

- Naming conventions should be logical and relate to the collection to be catalogued and digitised.
- Scanners will automatically generate a unique identifier for each image, but this image needs to be associated with metadata captured as part of the cataloguing process.
- Batch scanning may result in the same numbers produced across different batches. If these images do not have a unique identifier it may appear that duplication has occurred when in fact all the images are related to different source material. For example, a collection arranged by County that does not include a unique identifier for each county may result in 26 variations of the same number, whereas a naming convention that includes a unique identifier for each county will associate the scanning output to each county.

For example, STAT/1/3/ is the county code for the 1926 census for Carlow. If all output from digitisation omitted the number 3 (Carlow is county 3 in alphabetical order), and was STAT/1, images from all 26 counties would have no associated structure. This would mean aside from the image number automatically generated as part of the digitisation process it would be a massive task to link the cataloguing metadata.

	Reference code
1926 Census	STAT/1
County (Carlow):	STAT/1/3
DED (Agha):	STAT/1/3/1
Townland (Cloneen):	STAT/1/3/1/Cloneen

- Naming conventions can be numerical, alpha-numerical or solely text. Where a collection is well known through abbreviations these can be used in Alpha-numerical codes to facilitate easy identification.

---

# Appendix

Appendix → Appendix 2

## Appendix 2

### Technical specifications

- Technical specifications help ensure the legibility and usability of the digitised information for as long as is required. These specifications are primarily for information that has long term/high value to the organisation or is required to be transferred to the National Archives.
- Optional Character Resolution (OCR) applications work best on documents scanned at 400 ppi or higher.
- PDF format is widely used for web access as it is a compressed format and quicker to download online. PDF files are normally generated from a higher resolution scanning output such as TIFF or lossless JPEG.

### Technical specifications applicable to the digitisation of text and photographic prints

Bit depth	Resolution	File Format	Compression
8 bit Greyscale or bi-tonal	300 ppi	PDF/A, TIFF, JPEG 2000	Lossless compression
8 bit Greyscale or bi-tonal	300 ppi	PDF/A, TIFF, JPEG 2000	Lossless compression

### Technical specifications applicable to the digitisation of audio recordings

The Digital Preservation Coalition has developed guidance on preserving audio files as part of its Technology Watch series. This brings together guidance available from a range of sources. The guidance listed below has been extracted from the DPC note. The FADGI (Federal Agencies Guidelines Initiative) should also be consulted for guidance on preserving audio, audio visual and still image formats.

There is no single perfect format for the preservation and future use of audio. Decisions made on file formats should be dependent on the features and functionality to be preserved and the future use cases to be supported. Note that the tables below do not provide an exhaustive list of formats suitable for preservation and access.

<sup>2</sup> <https://www.dpconline.org/docs/technology-watch-reports/2476-preserving-audio/file>

<sup>3</sup> <https://www.digitizationguidelines.gov/>

---

## Appendix

Appendix → Appendix 2

The most suitable format for preserving the important features and functionality of a file may be the original format that it was created in. It is recommended that careful research and analysis is carried out before migrating files to a new format.

File format	Extensions	Brief summary
Broadcast Wave Format (BWF)	.wav	<p>Broadcast Wave Format (BWF) is an extension of the WAVE format, standardized as EBU-TECH 3285 (EBU, 2011). BWF allows for additional metadata to be embedded into the file, making the file self-describing. This includes metadata about the original broadcast extension, quality, the MPEG audio extension, descriptive XML, and other data. BWF files are limited to 4GB of audio data.</p> <p>Numerous organizations consider BWF to be a preferred preservation format (LC, 2020-2021; NEDCC, 2020; NARA, 2020; IASA, 2009; ARSC, 2015).</p>
Free Lossless Audio Codec (FLAC)	.flac	<p>Free Lossless Audio Codec (FLAC) is an open lossless audio encoding standard. FLAC represents both the container and encoding. FLAC produces an audio file that is lossless in integrity but a smaller file size than uncompressed audio, without size limitation, and supports</p> <p>metadata tagging, embedded images (e.g. for an album's cover art), and CRC-32 checksums per frame (Rice, 2013).</p> <p>FLAC is not as widely used as WAV due to the fact that it is a newer format, is less established in a preservation context, and has less explicit tool support. However, it is considered an acceptable preservation format (LC, 2020-2021; NARA, 2020; ARSC, 2015).</p>
IRENE	.tiff	<p>IRENE is a digital imaging technology that produces high-quality digital image files representing the grooved structures of analogue media (NEDCC, 2020). This is appropriate for unique wax cylinders, shellac discs, tin foils, lacquer discs, and other rare formats.</p>

---

# Appendix

Appendix → Appendix 2

File format	Extensions	Brief summary
RF64	.rf64	RF64 is an extension of the WAVE format that allows file sizes to exceed 4GB ( <a href="#">EBU, 2007</a> ). While structurally similar to WAV, RF64 is not as ubiquitous and may not be as widely supported.
Waveform Audio File Format (WAVE)	.wav	<p>WAVE is a wrapper format that can contain audio streams and other data, up to a maximum of 4 GB. The general recommendation is to store uncompressed audio sampled at 96 kHz with 24-bit audio sampling in a WAVE container.</p> <p>WAVE lacks the additional file header metadata supported by Broadcast Wave, but is still considered an acceptable preservation format (<a href="#">LC, 2020-2021</a>; <a href="#">NARA, 2020</a>).</p>
AAC	.aac	AAC is a lossy audio compression algorithm. It was designed to be the successor of the popular MP3, having higher sound quality while using the same bit rate ( <a href="#">Wikipedia, 2020</a> ).
MP3 (MPEG-1 Audio Layer III)	.mp3	MP3 (Standards <a href="#">ISO/IEC 11172-3</a> ; <a href="#">ISO/IEC 13818-3</a> ) is a popular format for storing audio data that has had huge commercial success. MP3 describes both the container and the encoding. MP3 can store data up to 16-bit depth only.
Ogg Vorbis	.ogg	Ogg Vorbis is an open-source audio codec ( <a href="#">xiph, 2016</a> ). The format is considered to have better compression ratios than MP3, creating a better-sounding access derivative while taking up less space. However, it is not as commonly used or as well-known as the older MP3 format.

---

# Appendix

Appendix → Appendix 3

## Cataloguing workflow checklist

- The collection has been identified and conditioned checked to determine suitability for digitisation, including completion of a Data Protection Impact Assessment, where necessary.
- The size and scale of the collection has been identified and documented, including any non-standard sizes and formats such as volumes or audio files.
- Realistic timelines have been developed to ensure adequate allocation of staff and space.
- Roles have been defined, including operational areas such as responsibility for retrieval of documents from storage, cataloguing and return to storage.
- Security and restricted access requirements have been implemented, where necessary.
- Cataloguing templates have been defined following testing to ensure capture of required fields of information, including any indexing requirements, and setting of realistic targets and timelines. Further guidance can be provided by the National Archives.
- Adequate training has been undertaken for project staff, including document handling and cataloguing requirements.
- Requirements such as sub-numbering of documents have been integrated into the cataloguing workflow and timelines.
- Quality assurance processes have been developed and integrated into the cataloguing workflow and timelines.
- Naming conventions have been determined based on organisational structures and functions and the context of the collection to be catalogued.
- Style guides have been developed to ensure consistency in how archives are described, including how to deal with names, placenames, abbreviations and capitalisation.

---

# Appendix

Appendix → Appendix 3

## Cataloguing workflow checklist

- Folder structure has been developed for capturing the outputs of the cataloguing process.
- Folder structure to be integrated into the scanning output structure to ensure consistency and association of scanned images with relevant metadata with minimal manual input.
- Regular checking has been undertaken to ensure the workflow remains current, and any changes required integrated at the earliest possible opportunity, including procedures for the conservation of damaged documents.
- Procedures are in place for the movement of catalogued records to the digitisation stage.

---

# Appendix

Appendix → Appendix 4

## Appendix 4

### Digitisation workflow checklist

- The size and scale of the collection has been identified and documented, including any non-standard sizes and formats such as volumes or audio files.
- Realistic timelines have been developed to ensure sufficient time for the procurement of specialised equipment or request for services.
- Adequate ICT and digital infrastructure has been developed to accommodate an in-house digitisation project, or the transfer of digital output from third party service providers.
- Project staff and space have been identified, including adequate ICT support.
- Adequate training has been undertaken for project staff, including document handling and scanner operator roles, where relevant.
- Roles have been defined, including responsibility for operation of scanning equipment, retrieval of documents from storage, quality assurance and return to storage.
- Scanning equipment has been configured following testing and format outputs and resolution defined.
- Processes, including metadata capture, have been automated to ensure standardisation and increased accuracy.
- Quality assurance and post-scanning processes have been developed and integrated into the digitisation workflow and timelines, including rescanning, image cropping, renaming.
- A folder structure has been developed for capturing the outputs of the digitisation process that associates the image with the correct metadata and reduces the level of manual post-processing required.
- Regular checking has been undertaken to ensure the workflow remains current, and any changes required integrated at the earliest possible opportunity, including procedures for the scanning and/or rescanning of damaged documents.

---

# Appendix

Appendix → Appendix 5

## Appendix 5

### Quality Assurance Checklist

Quality assurance procedures should be documented and built into the ongoing operation of the digitising process. Quality Assurance should begin once digitisation commences, and procedures amended where necessary. The extent and frequency of QA will depend on the resources available, scale and final use and is usually calculated as a percentage of the overall scanning output.

Quality assurance 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 the source input
- extent and frequency of sampling of digitised images
- criteria for checking image quality, including requirements around expected levels of legibility and intended use
- frequency and criteria for checks on metadata
- checks that digitised information is (re)producible in its original format
- short-term retention of the source information in case re-digitising is required due to faults in the original scanning
- processes for re-digitising and
- Operator training on scanning equipment and document handling



---

# Appendix

Appendix → Appendix 5

If outsourcing digitisation, relevant documentation about benchmarks should be agreed with service providers.

Periodic review of quality assurance procedures is important to ensure that benchmarks and quality assurance measures continue to meet business needs, timelines and expected outcomes.

Workflow documents should be amended as necessary to reflect findings throughout the quality assurance process. Failure to integrate ongoing learnings into workflows may result in unnecessary rescanning, which increases the risk to both the documents and project timelines.

---

## Glossary of terms <sup>4</sup>

### Glossary of terms

This glossary is not exhaustive. Further information on archival terminology is available on the International Council on Archives [Multilingual Archival Terminology database](#) <sup>7</sup>.

#### **Access copy**

A duplication, in whole or in part, of an original document. A digitised copy of an original record created to facilitate access.

#### **Archivist**

An individual responsible for appraising, acquiring, arranging, describing, preserving, and providing access to records of enduring value, according to the principles of provenance, original order, and collective control to protect the materials' authenticity and context.

#### **Bit depth**

Number of bits (zeros or ones) used to describe the colour of each pixel. Bit depth can range from 1 bit up to 48 bits. Greater bit depth allows a greater range of colours or shades of grey to be represented by a pixel.

#### **24 bit colour**

Enables the storage of 8 bits of information for each red, green and blue component of each pixel. 24 bit colour, also known as True Colour, enables millions of colour variations.

#### **Bi-tonal**

Black and white, two tone scan.

#### **Cataloguing**

A process for the creation of finding aids that describe records giving the repository physical and intellectual control over the records, which assists users to gain access to and understand the records.

#### **Compression**

Compression is a set of algorithms designed to reduce the size of an image for storage or transmission.

#### **Descriptive metadata**

Descriptive metadata enables the discovery, identification and sorting of data. Attributes include but are not limited to: title, date, creator/author, subject etc.

<sup>4</sup> This glossary is not exhaustive. Further information on archival terminology is available on the International Council on Archives Multilingual Archival Terminology database.

---

# Glossary of terms

## Glossary of terms

### **Digital Preservation**

An essential and necessary component of digital archiving ensuring longevity of an electronic object. Digital preservation covers the processes and operations involved in ensuring the technical and intellectual survival of authentic electronic records over time (such as the ongoing monitoring, migration and storage of records and managing the metadata which describes the origin and successive treatment of the record).

### **DPI**

Dots per inch describes the amount of ink dots on a printed image.

### **File Format**

The way in which the information in a file is encoded. There are many proprietary formats as well as standard file formats such as RTF, TIFF, and EPS. In some systems, such as Apple Macintosh, the information about file format and originating application is part of the file, but in other systems it is up to the user to know what the format is, although there are more-or-less strict file-naming conventions.

### **Greyscale**

Black and white in addition to a range of intermediate greys, where 8 bits are required to describe each pixel.

### **Indexing**

The process of compiling an index, which adds detail to a collection catalogue and makes the digitised documents more discoverable.

### **Intellectual control**

The control established over archival material by documenting in finding aids its provenance, arrangement, composition, scope, informational content and internal and external relationships.

### **JPEG**

Joint Photographic Experts Group graphic image file compressed with lossy compression.

### **JPEG 2000**

JPEG 2000 is an image coding system that uses state-of-the-art compression techniques; its architecture is useful for many diverse applications, including Internet image distribution, security systems, digital photography, and medical imaging

---

# Glossary of terms

Glossary of terms

## **Lossless compression**

Digital compression techniques in which no information is lost; an object is identical before and after being compressed and restored.

## **Lossy compression**

Digital compression techniques in which information is lost; an object is altered after being compressed and restored.

## **Metadata**

Set of data that describes and provides information about other data.

## **Master copy**

See preservation copy.

## **Physical control**

The control established over the physical aspects (such as format, quantity or location) of documents in the physical custody of a records centre or archives. See also: intellectual control.

## **PDF**

Portable Document Format developed by Adobe and now standardised as ISO 32000-2:2020.

## **PDF/A**

A self-contained Portable Document Format widely used in archiving and standardised as ISO 19005.

## **PPI**

Pixels per inch describes the resolution in pixels of a digital image. The fixed number of pixels that a screen can display and the density of pixels within a digital image. PPI affects the print size of a design and the quality of the output. See also DPI.

## **Preservation copy**

A digital copy of an analogue record captured through the digitisation process and operation involved in ensuring the technical and intellectual survival of authentic records through time.

The obligation to protect records and other materials potentially relevant to litigation and subject to discovery.

---

## Glossary of terms

### Glossary of terms

#### **Resolution**

A measure of the ability to capture detail in the original work, and only one aspect of quality in an image. It is often quantified in pixels per inch (ppi). The optimum resolution depends on the nature of the documents being scanned and the smallest meaningful element in them being able to be rendered legible in the scanned image.

#### **Technical metadata**

Represents technical aspects of data, either inherent or manually input. Attributes include but are not limited to: file format, bit-depth, resolution, creation date, alteration date.

#### **TIFF**

Tagged Image File Format is a common file format for graphics and photographs. TIFF, with lossless compression, has been extensively used and remains a default preservation imaging format. The considerable size of lossless TIFF images can present significant storage issues.

---

## Bibliography

[National Archives of Australia Guidelines for Handling, Preparing and Digitising Archival Paper and Printed Material ↗](#)

[National Archives of New Zealand Best Practice Guidance on Digital Storage and Preservation ↗](#)

[National Archives of New Zealand Destruction of Source Information After Digitisation ↗](#)

[Federal Agencies Digital Guidelines Initiative ↗](#)

[Federal Agencies Technical Guidelines for Digitising Cultural Heritage Materials](#)

[Library of Congress Data Dictionary for Administrative Metadata for Audio, Image, Text and Video Content ↗](#)

[Sustainable Heritage Network introduction to audio digitisation ↗](#)

[University of Santa Cruz Scanning Sizes Explained ↗](#)

[European Commission Basic Principles and Tips for 3D Digitisation of Cultural Heritage ↗](#)

[National Archives, Copyright Guidance ↗](#)

[National Archives, Data Protection Guidance ↗](#)

ISO 13028 Information and documentation: Implementation guidelines for digitisation of records

ISO 23081-1: 2017 Information and Documentation: records management processes – metadata for records part 1: Principles

ISO 15836-2: 2019 Information and Documentation: Dublin Core Metadata Element Set

ISO 13028: 2010 Information and Documentation: Implementation Guidelines for Digitisation of Records

ISO 15801: 2017 Document Management – electronically stored information – Recommendations for Trustworthiness and Reliability

BS 10008 Evidential Weight and Legal Admissibility of Electronic Information



An Chartlann Náisiúnta  
National Archives

**The National Archives**  
Bishop Street, Dublin 8  
D08 DF85, Ireland  
[www.nationalarchives.ie](http://www.nationalarchives.ie)

**Phone:** +353 (0)1 407 2300  
**Lo Call:** +353 1890 252424  
**Fax:** +353 (0)1 407 2333  
**Email:** [query@nationalarchives.ie](mailto:query@nationalarchives.ie)

Image: Scanning of a 19th century volume

