DIGITISATION & QUALITY

GUIDE TO MANAGING AND CONTROLLING QUALITY IN A HERITAGE AND DOCUMENT DIGITISATION PROJECT

ADOCHS

Auditing Digitalization Outputs in the Cultural Heritage Sector

Author: Chloé Brault
Contributors: Florence Gillet, Nicolas Roland
Proofreading and scientific direction: Anne Chardonnens
Layout: Chloé Brault

ADOCHS project comittee: Nico Wouters (CegeSoma-AGR), Frédéric Lemmers (KBR), Ann Dooms (VUB), Tan Lu (VUB/KBR), Anne Chardonnens (CegeSoma-AGR/ULB), Nicola Roland (KBR), Florence Gillet (CegeSoma-AGR), Seth Van Hooland (ULB), Chloé Brault (CegeSoma-AGR)

Centre d'Etude Guerre et Société – Studie en Documentatiecentrum Oorlog en Hedendaagse Maatschappij State Archives of Belgium - 4th operational direction
Square de l'Aviation 29 Luchtvaartsquare 1070 Bruxelles - Brussel

The elaboration of this guide would not have been possible without the funding of the **Belgian Science Policy** under contract number BR/154/A6/ADOCHS.

Institutional partners

Digitisation and Quality

Guide to managing and controlling quality
in a heritage and document digitisation project

Version 1.1 - September 2021

By Chloé Brault

Brussels
2021

TABLE OF CONTENTS

ABOUT THIS GUIDE

Background to its creation

Since the mid-1990s, cultural institutions have undoubtedly entered the digital age. In Belgium and around the world, the number of large-scale digitisation projects implemented by governments and private companies has been growing rapidly[1] since the early 2000s, and many major institutions now have their own digitisation studios.

The issue of quality control has been one of the main obstacles in this first phase. Indeed, it appeared that many projects had underestimated the scale of this step, in both human and technical terms, in the overall digitisation process. In the majority of cases, the teams have been faced with a lack of methodological standardisation and automation tools required to carry out the work. They have therefore often had to work manually and without procedural guidelines appropriate to their specific needs.

And yet quality control is an essential component at every stage of a digitisation project if we wish to ensure the integrity and consistency of the files and data produced, and their long-term preservation. This applies to both outsourced and in-house digitisation projects.

At the same time, the growing number of stakeholders has inevitably led to differences in the final outcome of digitisation. How can we ensure that the digital file meets an expected level of quality and is a faithful and complete representation of the original document so that the user does not question its integrity? How can we ensure that similar documents have the same characteristics when scanned, regardless of the institution, the operator, the studio and even the device used?

The **ADOCHS** (*Auditing Digitalisation Outputs in the Cultural Heritage Sector*) research project was developed on the basis of these reflections, and began in 2016. The project has been created at the initiative of the Belgian State Archives – and more specifically the CEGESOMA[2] – in cooperation with the Royal Library of Belgium (KBR), the Vrije Universiteit Brussel (VUB) and the Université libre de Bruxelles (ULB), with the aim of improving the quality control of files produced in the context of heritage digitisation. In doing so, the project aims to speed up the entire digitisation process, minimise its costs and increase the value of the data produced as part of future projects.

Purpose and scope

This guide is intended primarily for the digitisation departments of Belgian federal scientific institutions.

Digitisation operators will find technical recommendations to enable collections to be digitised in an appropriate, standardised manner. Collection managers and project managers will find an overview of the digitisation process, and of the elements connecting and structuring these processes, to help them in clearly developing a digitisation project.

In other words, **the purpose of this manual is to offer organisational and methodological tools to ensure the quality of the digital data produced at each stage of the digitisation process.** Thus, everyone involved in the process can see their work as part of a production chain working towards a common goal.

The purpose of this guide is twofold :

- **Standardising quality control criteria**

- **Ensuring the preservation and long-term (re)use of digitised cultural heritage.**

This guide is therefore intended to harmonise the practices of Belgian federal scientific and cultural institutions. This harmonisation will then facilitate the exchange of information within and between these institutions. Users will also benefit from this improvement by having their search capability expanded and facilitated.

Quality control is an essential component at every stage of a digitisation project if we wish to ensure the integrity and consistency of the files and data produced, and their long-term preservation.

---

1 *Announced in 2004, the Google Print Project, in cooperation with the libraries of Harvard, Stanford and Michigan Universities in the USA, Oxford University in Great Britain and the New York Public Library, aims to digitise 15 million works. This initiative has led directly to the Europeana Project and to numerous digitisation projects being carried out by private and public organisations. In his article, Jean-Michel Saläun looks back at the origins of this Dantesque project and the powerful reactions it provoked at the time.*

Jean-Michel Salaün. Bibliothèques numériques et Google-Print. Regards sur l'actualité : mensuel de la vie publique en France, La Documentation Française, 2005. *https://archivesic.ccsd.cnrs.fr/sic_00001576/document*

2 *The CegeSoma, the Centre for Historical Research and Documentation on War and Contemporary Society, has been the fourth Operational Departement of the Belgian State Archives since 2016.*

Methodology and structure

This guide is divided into six chapters according to the following structure:

1

The first chapter provides a brief review of the context surrounding digitisation and the associated challenges. It discusses the most common objectives of digitisation and how they are changing in the light of the increasing computerisation of our modern societies. These changes result in institutional changes that are themselves based on the changing expectations and needs of users of archives and libraries.

2

The second chapter defines the notion of quality according to the ISO-9001[3] international standard[3]. This general definition is then adapted and refined in line with the specific realities of the digitisation process. The strategic and intellectual development of any project is addressed here through three fundamental documents: the data management strategy, the digitisation policy and lastly the specifications.

3/4

Chapters three and four cover the quality of deliverables, and more specifically images and associated metadata. These two notions are in turn explained and characterised in order to identify the criteria that can be used to define them as good quality.

5

The fifth chapter deals with the digitisation environment and good studio management. Heritage digitisation has the particular characteristic that it deals with precious objects – whether modern or antique – which require special measures to avoid damage during handling.

6

Lastly, the sixth and final chapter provides a series of summary sheets that can be used to guarantee the quality of a digitisation project. These sheets summarise the key concepts for each quality control step and follow the chronological sequence of the digitisation chain. Each sheet also provides practical recommendations, tools and tips along with additional information sources to enable completion of the control phase described.

The content of this guide is based on :

- **Standardised reference systems** based around quality and digitisation, and more specifically the ISO standards.
- **International methodological reference systems and good practice guides for digitisation.**
- **General monographs based around digitisation, quality and its management, digital imaging and metadata.**
- **Research carried out during the ADOCHS Project** by the team, and more specifically on the works of Tan Lu and Anne Chardonnens on improving the quality of images and metadata.
- The **skills and experience of the digitisation teams at the Belgian State Archives and the Royal Library of Belgium**, thanks to a series of interviews conducted in spring 2021 with technical operators and digitisation project managers.

3 *International Organisation for Standardisation (ISO). Quality management systems — Fundamentals and vocabulary. Genève : ISO, ISO 9000 :2015, 2015.*

01.
DIGITISATION : CHALLENGES, CHANGES AND FUTURE

Digitisation is the action of converting an analogue resource into a sequence of machine-interpretable codes, thus making it possible to reproduce any type of analogue document in digital form[1].

Behind this seemingly simple definition there are in reality a range of issues affecting not only the institutions initiating these projects but also their audiences. Digitisation projects must therefore reconcile cultural, scientific, social, economic, political and technical imperatives.

The purpose of digitisation depends very much on how the digitised content is intended to be used, and the process is generally used to preserve, disseminate and exploit heritage, archive and document collections.

Preservation	Digitisation carried out for preservation purposes involves **reproducing the original document as faithfully as possible, without seeking to make improvements or modifications.** This substitute format makes it possible to limit the handling and therefore the deterioration of the original collections.
Dissemination	Digitisation carried out for dissemination purposes aims **to make a digitised document accessible on the web or in a publication.**
Exploitation	Exploitation through digitisation takes the form, for example, of the **creation of digital corpora to facilitate teaching and research** or as part of a museum exhibition.

1 *Thierry Claerr, Isabelle Westeel, Manuel de la numérisation, Electre, Tours, 2011, page 18.*

The advent of the internet and its subsequent developments have imposed certain conditions on access to and exploitation of culture. Indeed, Web 2.0 brings together various complex social systems such as blogs, wikis and social networks and sees the web as a vast space for social interactions. It is the cooperation between individuals and platforms that ensures the continuous production of content.

In this sense, Web 2.0 sites act more as portals focused on the user whose attention must be captured rather than on traditional websites. Supported by recommendation algorithms, Web 2.0 operates on the principle of similarity between profiles and content. These similarities are defined in particular thanks to user usage data, which make it possible to create groups with relatively uniform characteristics by and to which information can be distributed.

In this context, metadata – which structure the identification and description of digital content – are clearly central to joining this circle of dissemination and finding an audience. The importance assigned to the way in which content and websites are described to optimise referencing has increased and the question of the structuring of information and its readability for machines has become a central issue in the way we approach the Web.

This is the challenge of the Web 3.0 project, more commonly known as the Semantic Web. The objective of this project is no longer to build data consultation platforms, but to make reference information accessible and readable for both machines and humans.

This objective implies, among other things, that the metadata format should be sufficiently flexible to allow consultation and dissemination of the digital content described, while promoting the cross-referencing and processing of all content present on the web.

This project makes perfect sense when we think about the digital dissemination of heritage, archive and document collections: this kind of technology would make it possible, for example, **to bring together known information about bodies of work that are often dispersed across a number of institutions nationally and internationally.**

In this context, metadata are clearly central to finding an audience.

However, this approach requires a significant degree of cohesion and a common strategy from the cultural and scientific sector. Indeed, the metadata accompanying digital content must be sufficiently complete that the content can be found[2] by the potential audience.

With this objective in mind, the European Commission recently revised the development plan[3] for its flagship project to showcase cultural heritage, **Europeana**[4]. Europeana is a digital platform launched in November 2008 providing access to the digital resources of the cultural institutions from the Member States and now

2 *The findability of digital content refers to the ease with which online content can be found by internet users, and in particular by those who are not searching specifically for the content in question . This potential depends on various factors, specifically the metadata and associated keywords, its referencing, how it is advertised, the algorithms of the search engine used and the browsing habits of internet users.*

For more information on this issue, please consult: Josée Plamondon, Êtes-vous repérables ? Documenter vos contenus pour qu'ils soient compris par des moteurs de recherche, 2019.

3 *Europeana, Strategy 2020-2025 Empowering Digital Change, mai 2020,* ***https://pro.europeana.eu/files/Europeana_Professional/Publications/EU2020StrategyDigital_May2020.pdf***

4 *Europeana Pro, Présentation, n.d.* ***https://pro.europeana.eu/about-us/mission***

has almost 3,500 participating institutions. It should be noted that Europeana does not archive works, but merely serves as a search interface. The participating institutions have therefore committed to digitising their content, making it accessible on the web and ensuring it is preserved in digital form.

The purpose of the strategic document proposed by the Commission is thus to **offer a common guideline concerning the issues associated with digitisation, and specifically to encourage participating institutions to embrace the question of metadata to increase not only their visibility but also the potential for dissemination of their collections on the Web.** The various measures therefore propose to bring about a better digital transition by working on the following points:

- ***The improvement of the quality and interoperability of metadata.***
- ***The use of new technologies such as machine learning algorithms*** *to automatically or semi-automatically enrich metadata records in a faster and more scalable way.*
- *The strengthening of national infrastructure and the* ***development of stronger inter-institutional links****.*
- ***The improvement of the technical skills of actors in the sector by setting up a knowledge centre*** *offering guidance to cultural heritage professionals on topics such as digitisation, metadata enrichment, semantic interoperability, content creation, licensing, reuse, business models and innovation.*

This perspective on the evolution of the framework for dissemination of digital content nevertheless requires adjustments for heritage, archive and library professionals[5].

Moving from restricted, individual access to a mass corpus has required, and still requires, institutions to undertake a radical restructuring of their dissemination policies and of the resources supporting those policies. This paradigm shift has also changed the status and skills required from actors in the sector: from being curators of public heritage, librarians, archivists and museologists have become providers of cultural content. This change implies the development of new technical expertise and requires particular attention to the changing expectations and needs of users of these services.

A corresponding analysis was carried out in Belgium between 2015 and 2017 as part of the **MADDLAIN** research project[6]. The aim of this project was not only to provide data on the practices and needs of users in accessing digital information, but also to move institutions forward in managing the methods and tools that will help them gain a deeper understanding of their audiences.

5 *Annabelle Boutet et Karine Roudaut, « Les enjeux de la numérisation et de l'ouverture d'archives : le point de vue des professionnels », Terminal* ***http://journals.openedition.org/terminal/1220***

6 *The MADDLAIN project (2015-2017) is a collaboration between the Royal Library of Belgium, the Belgian State Archives and the Centre for War and Society Studies (Cegesoma) and aimed to analyse the behaviour and needs of the different audiences of the institutions of the documentation division of the Federal Science Policy in terms of access to digital information.*

Two fundamental issues emerge from this study[7]: the need for **accessibility** and the need for **communication**. The final report on the project states in this regard that « the principal points of dissatisfaction concern the unclear ergonomics, the too complex general structure, the difficulties for inexperienced users, **the inadequate or incomplete description of documents and the lack of communication with regard to the documents that are entered, or not, in the digital catalogue.** At the same time, the majority of the users understand that not everything can be digitized and be available online. They do however ask [for] **« explanations with regard to the underlying technical procedures, the choices made in terms of digitization and also to the potential restrictions** » .

These conclusions highlight the importance of addressing the digitisation of heritage, archive and document collections **through the notion of quality. This of course relates to quality of deliverables, but also quality of processes and especially quality of the phase involving the intellectual and strategic development of the project.** Once these two aspects of quality have been understood together, it becomes possible to better control all digitisation processes and, ultimately, to offer reliable, relevant digital content as proposed in the following chapters.

This perspective on the evolution requires adjustments for heritage, archive and library professionals.

7 *Florence Gillet, Jill Hungenaert, Melissa Hodza,et al., Identifying Needs to Modernize Access to Digital Data in Libraries and Archives (MADDLAIN) : final report, Brussels : Belgian scientific Policy, 2018.* ***http://www.belspo.be/belspo/brain–be/projects/FinalReports/MADDLAIN_%20final%20report.pdf***

FOR FURTHER INFORMATION

PLAMONDON, J. Êtes-vous repérables ? Documenter vos contenus pour qu'ils soient compris par des moteurs de recherche, 2019.
http://ipf.ca/wp-content/uploads/sites/2/2019/05/guide-documentation-webseries_20171128_vf-maj_20191405.pdf

CHAN, L. (dir.) ; MOUNIER, P. (dir.). Connecting the Knowledge Commons — From Projects to Sustainable Infrastructure : The 22nd International Conference on Electronic Publishing – Revised Selected Papers. Marseille : OpenEdition Press, 2019.

CANDELA, G., ESCOBAR, P., CARRASCO, R.-C., MARCO-SUCH, M. A linked open data framework to enhance the discoverability and impact of culture heritage. Journal of Information Science. 45(6), 2019, pp. 756-766.

FREIRE, N., VOORBURG, R., CORNELISSEN, R., DE VALK , S., MEIJERS, E., ISAAC , A. Aggregation of Linked Data in the Cultural Heritage Domain: A Case Study in the Europeana Network. Information. 10(8), 2019.

JONES, E.; SEIKEL, M. Linked Data for Cultural Heritage; Facet Publishing: Cambridge, UK, 2016

ACCART, J.-P. Le Métier de Documentaliste. Avec la collaboration de Réthy Marie-Pierre. Éditions du Cercle de la Librairie, 2015.

THIAULT, F. Mutations des métiers de l'information-documentation : vers l'émergence d'une culture de l'information numérique. In : Les Cahiers de la SFSIC, Société française des sciences de l'information et de la communication, 2012, p. 59-62.
https://hal.univ-lille.fr/hal-00955279/document

GILLET, F., HUNGENAERT, J., HODZA, M., et al. Identifying Needs to Modernize Access to Digital Data in Libraries and Archives (MADDLAIN) : final report. Brussels : Belgian scientific Policy, 2018.
http://www.belspo.be/belspo/brain-be/projects/FinalReports/MADDLAIN_%20final%20report.pdf

02.

QUALITY : A COMMON ORGANISATIONAL THREAD

Defined by the ISO 9000 standard as 'all of the characteristics and properties of a product, a process or a service that influence its ability to meet identified or implicit requirements[1]', quality is a broad concept. In theory, all projects can adopt the concept of quality and international standards, regardless of their objectives, the technologies used, the resources employed or the nature of the organisation.

Libraries, archives and museums already rely in part on these recommendations in their activities, particularly in the context of digitisation projects. Quality control, an integral part of the process, is therefore most often at the end of the chain and concerns the control of deliverables: more specifically the evaluation of image quality, metadata and file integrity.

In short, quality here means ensuring that the initial objectives of the project are in line with the results obtained. But ensuring the quality of a digitisation project cannot be reduced to a mere a posteriori inspection of the deliverables. In the event of errors, the costs involved are too high and the scope for correcting processes too low. While this final verification stage is obviously essential to the digitisation process, it must therefore be accompanied by planning and checking throughout the digitisation chain.

In this sense, the notion of quality should be seen not as a specific stage, but as a project management tool guiding all stages, a common thread guaranteeing control of all processes.

1 *International Organisation for Standardisation (ISO). Quality management systems — Fundamentals and vocabulary. Genève : ISO, ISO 9000 :2015, 2015.*

On the basis of this definition, the notion of quality therefore implies :

- *Technical skills and in-depth knowledge of images and metadata, as well as a good understanding of digitisation equipment.*

- *An overview of the most frequent quality problems and their causes, so that a coherent methodology can be developed upstream of the project and unforeseen events during the digitisation process can be handled as effectively as possible.*

- *Management of the working environment, which includes the organisation of spaces and the control of light, as well as the measurement and analysis of atmospheric conditions, in order to make the work flow more smoothly, to guarantee the best conditions for image capture and to minimise the potential deterioration of heritage collections.*

- *The definition and use of organisational tools to coordinate the various positions in the digitisation chain and facilitate communication between the operators in the studio.*

- *More broadly, each digitisation project should be part of an overall policy for the preservation, digitisation and dissemination of its digitised content, as well as a data management strategy (Digital Data Strategy[2])*

In short, to consider the notion of quality as a driver is to « sustain an action, optimise manufacturing processes, improve the efficiency of an organisation's operations, improve the quality of its products and services, innovate and share added value [and finally] account for the controlled use of funds invested for the satisfaction of end users[3] ».

The notion of quality should be seen not as a specific stage, but as a project management tool guiding all stages, a common thread guaranteeing control of all processes.

To put it another way, it is therefore no longer a question of simply establishing the quality of a production after the fact, but of determining the quality criteria of a digital product, identifying the procedures, sequences and interactions between the various stages of the digitisation chain, developing a quality of services and, finally, anticipating the expectations and needs of users so as to offer content that is high-quality and relevant.

2 *Valentina Bachi, Antonella Fresa, Claudia Pierotti, Claudio Prandoni,The digitization age: mass culture is quality culture. challenges for cultural heritage and society. In: Ioannides, M., Magnenat-Thalmann, N., Fink, E., Žarnić, R., Yen, A.-Y., Quak, E. (eds.) EuroMed 2014, 17 p. and Mike Fleckenstein, Lorraine Fellows, Data Quality. In: Modern Data Strategy. Springer, Cham, 2018.*

3 *Thierry Claerr, Isabelle Westeel, Manuel de la numérisation, Electre, Tours, 2011, page 269.*

Properly documented, each digitisation project can thus be evaluated and enable the improvement of all protocols and processes, making the organisation's actions sustainable over time. It should be noted that this approach must obviously be designed according to the technical, human and financial resources of the institution. **To be operational, the system must be appropriate for the actual resources of the organisation initiating the project.**

2.1 Data management strategy and digitisation policy : the foundation

Operational and strategic decision-making processes are essential to the success of a project, regardless of the number of documents to be digitised or the nature of the project.

It is important to make a clear distinction between the operational level – which comprises the set of rules for handling a particular document and flows from the institution's digitisation policy – and genuinely strategic issues that fall under the management and development policies of cultural and scientific institutions, and data management and data quality strategies.

Although it is not the purpose of this guide to develop a detailed digital data management strategy or digitisation policy, it is essential to highlight their importance and differences, as some operational issues cannot be resolved without strategic trade-offs:

The data management strategy encompasses all practices for managing data, namely information assets as a valuable resource within an organisation, an institution or a company. Its objective is therefore to **define, approve and communicate the strategies, policies, standards, architectures and procedures surrounding data management and to ensure that they are monitored and applied.** The better the data created, classified and processed by the organisation, the more readable, relevant and reliable the information system will be, thus contributing to an overall improvement in the organisation's efficiency.

The digitisation policy is derived directly from the data management strategy, but has a narrower focus on the organisation's digitisation projects. **This document brings together all the principles governing the digitisation of heritage, document and archive collections. It thus makes it possible to define the institution's challenges and objectives with regard to selecting documents to be digitised and their conservation, management and dissemination in digital form.** It also covers all the procedures for the development and implementation of the projects and the role of the teams during the process.

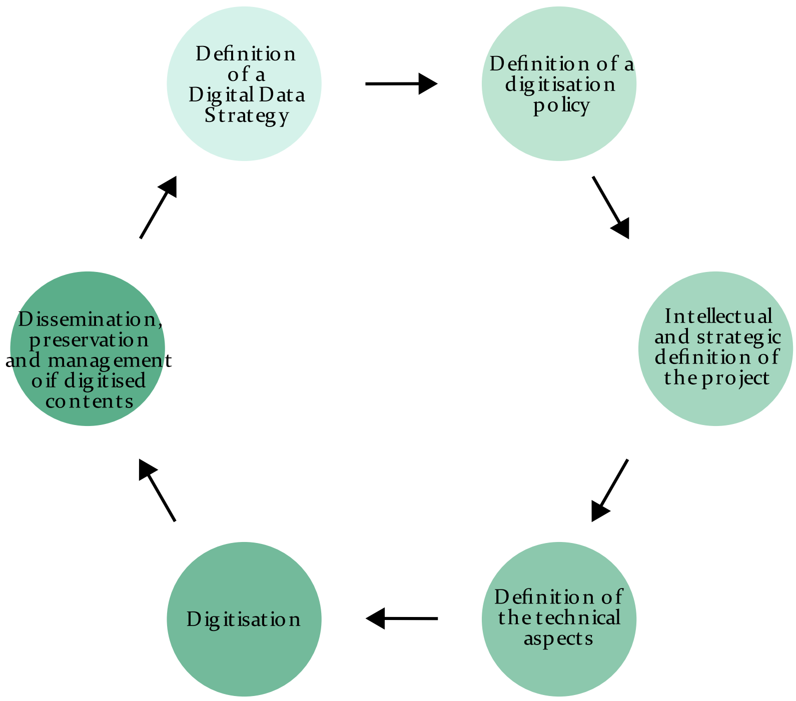


fig. 1 The complete digitisation cycle

The institutional framework of the Digital Data Strategy and the digitisation policy facilitates the strategic and intellectual definition of digitisation projects, in particular with regard to the objectives of these projects (preservation/ dissemination/exploitation) and their long-term scope.

It is important for a **legitimate, official decision-making body to be identified when the digitisation project is launched** so that any such questions can be dealt with quickly in the event of a dispute with an external service provider, for example.

The **strategic and intellectual definition of the digitisation project consists essentially in drawing up the specifications, which outline the methodological framework of the project** and in turn make it possible to define the associated technical aspects (choice of equipment, software, image capture rules, possible transport, etc.).

Once the content has been digitised, it is saved and stored in an electronic document management (EDM) system or an electronic archiving system (EAS) for the purposes of dissemination and/or conservation, with the creation, objectives and management of that system being defined directly by the data management strategy. Each deliverable resulting from a digitisation project is thus part of a much wider circle, all aspects of which need to be identified and analysed.

2.2 The specifications

The procedures for managing the quality of a digitisation project are developed as soon as the specifications are drafted.

As noted above, this document is part of a wider strategy of the commissioning organisation and derives directly from the institution's objectives in terms of conservation and dissemination of its heritage and document collections.

It must of course take into account the legal feasibility of the project based on the laws governing intellectual property. Lastly, it should reflect the material resources and organisational choices of the body initiating the project: does the institution have a digitisation studio? What is its role? Should part or all of the process be outsourced? Does the team have the necessary skills and equipment for all stages? What are the objectives of the digitisation project and what use is planned for the digitised content?

The answer to each of these questions influences the services to be provided, the quality expected, the minimum quality threshold and the management of quality control.

To better control the quality of a digitisation project, the following should be included in the specifications:

01 **General conditions for performance**

Describe the nature and general objective of the project, the types of documents involved, the expected outcome and the intended use. In short, this is a summary of the institution's digitisation policy, its resources and its ambitions for the project.

02 **Summary of the services to be performed**

Summary of the stages of implementation of the project. In the case of cooperation with a third-party company, this clause serves to define the scope of action of each party involved.

03 **Presentation of the documents to be digitised**

Detailed description of the documents to be digitised. This section should include the nature of the documents to be digitised, their material history, state of preservation, medium and size, and the quantity of documents. The specific attributes of the documents must also be stated: different formats, fragility of documents, possible bindings, etc. This section should also state whether the digitisation is to be done by double page or single page, and whether all pages of the document are to be digitised.

04 Breakdown of documents to be digitised into batches

Depending on the quantity of documents to be digitised and whether they contain documents of different formats, it is essential for the documents to be grouped into sub-sets. The size and number of these batches should take into account whether it is possible to check them within a reasonable time frame. The composition of the pre-production sample should be defined at this stage: this batch serves as a benchmark throughout the production phase and should reflect all of the various types of documents digitised.

05 Performance of the service and the time frames for completion

This involves defining the operational chain of the digitisation project, the time frame for each task and the quality controls associated with these tasks. This section should include a detailed description of the pre-production phase. If this phase is not validated, the production itself cannot be started. These tests must be carried out under actual production conditions.

06 Mapping of the digitisation chain

In the same spirit as the previous point, a graphic representation of the digitisation chain offers a valuable overview to add to the specifications: it makes it possible for everyone to better understand their roles in the process, to understand the sequence and interactions between the stages and to visualise the key stages in the quality management of the project.

07 General conditions for performance

This section specifies the ground rules to be followed throughout the process: the digitisation standards applied by the institution and some general principles on the quality of the expected results (in particular in terms of the readability or aesthetic look of the digitised content).

08 Specific description for preparation of the documents

This is a description of how documents are to be prepared (identification and marking of documents, ranking by format, packaging, etc.).

09 Definition of transport and storage procedures

The importance of this clause obviously depends on whether all or part of the digitisation project is outsourced. In the case of outsourcing, the procedures for the collection and transport of originals should be defined here. The constraints on the retention and storage of documents must be clearly stated. If the entire production chain takes place locally, this section determines the procedures for temporary storage and the conditions for handling of the original documents.

10 Calibration of the digitisation chain

This describes the type of procedure to be implemented by the digitisation studio to set up its digitisation chain before the start of production – by agreement with the commissioning organisation – and to check that the initial setting is maintained throughout the process.

The calibration of the digitisation chain covers all elements that can be automatically adjusted: resolution, colour, post-processing and file format. It is carried out, under normal production conditions, on a set of documents representative of the range of documents to be processed and then analysed on the basis of the indicators requested in the specifications, until it is validated jointly. While enabling calibration of the digitisation tool, this process will also serve as a reference throughout the operation to ensure constant quality.

11 Steps to be taken if an anomaly is identified

This may involve removing the document from the production circuit, flagging the case for analysis and resolution, defining the nature of the appropriate treatment, or feeding data into the documentation system to systematise these procedures.

12 Characteristics of the digital copies to be provided

This provides a definition of the technical characteristics (size, colour, file format, cropping, magnification ratio, resolution, colour, compression type and ratio, orientation, etc.) and the technical metadata for the images to be provided.

13 Naming scheme for all files

The naming scheme corresponds to all common rules for naming files and must be described in detail in this section.

14 Structure of the metadata inventory file

This lists all categories that must appear in the descriptive metadata file.

15 Indexing works to be performed

This defines the indexing works to be carried out, in particular the types of information to be identified. The metadata standard to be applied will be specified in this clause. The actual form of entry will depend on the envisaged access and search modes: relational databases, XML files, etc.

16 Procedures for checking and monitoring the inventory file

This defines the information that must be provided by the service provider for monitoring of production: time frame, form in which the document must be completed and delivered, etc..

17 Criteria for checking of deliverables

This section describes the final stage of quality control[4] at the end of the chain, and corresponds to the integrity check on all items produced in relation to the services. This section must specifically describe the level of quality expected of the service and the leeway available to the client to reject products that do not meet their expectations. This clause will also include the number of checks carried out during the digitisation process, as

4 *This guide distinguishes two types of quality control: control of quality as a comprehensive approach to improving and monitoring the overall quality level of the digitisation chain. and quality control as a stand-alone.*

well as the time frames for checking the intermediate and final results.

18 Delivery of digital images

In the case of outsourcing, this clause makes it possible to define the method used to deliver the digital copies, return the originals and verify any removable media on which the digital copies are stored (USB keys, external hard drives, etc.).

FOR FURTHER INFORMATION

For exhaustive information about the laws, royal decrees and circulars governing the issue of **copyright in Belgium**, see for example IPNews.be, a legal news site specialising in new technologies.
https://www.ipnews.be/legislation/legislation-belge/

The **Guide to managing a digital project** produced by the **Office of Interuniversity Cooperation (BCI) in Quebec** offers tools for evaluation, planning and monitoring that can be used to guarantee the smooth running and successful completion of a digitisation project. Although initially intended for digitisation projects on administrative documents, the digitisation approach lends itself equally well to historical or heritage archives.
https://www.enssib.fr/bibliotheque-numerique/documents/64628-guide-de-gestion-d-un-projet-de-numerisation.pdf

For specific examples of specifications, please refer to the **Technical Guide from the National Archives of France – Writing specifications for heritage digitisation.**
https://francearchives.fr/file/bf50d8fa5f554586dbf18fdc862d25970a1da0a7/static_4132.pdf

03.

QUALITY OF DELIVERABLES : IMAGE FILES

3.1 Image : definition and characteristics

A digital image means any image acquired, processed and stored in binary form that is characterised by the following elements: definition, resolution, coding depth, mode, colour space and lastly format.

RESOLUTION AND DEFINITION

Resolution is often associated with image quality because it expresses **the fineness of the digitisation analysis**, and should not be confused with image definition. Image definition corresponds **to the number of pixels making up the image and is calculated on the basis of the resolution and the size of the document.** The definition is therefore an absolute value that makes it possible to quantify an image and the sensor used to produce it. It should be noted that this value alone does not describe the dimensions of the image: a pixel has no 'physical' size per se. However, as soon as the image is displayed or printed, the display surface must be taken into account, and therefore the resolution, which is the unit of measurement expressing the fineness of analysis of the digitisation. It is expressed in ppp (*pixels per inch*) or dpi (*dots per inch*).

It is commonly accepted **that a minimum resolution of 300 dpi at 100% of the original is adequate for the majority of digitised documents**, with the exception of very large or small formats. This standard tends to evolve towards a 400 dpi resolution for better reading comfort on screen.

CODING DEPTH

The coding depth represents the quantity of data bits that define a pixel. The data bits define the number of different values that a pixel may represent.

With 8 bits, 256 different values can be reached. Each additional bit doubles the number of possible values.

MODE

For a greyscale image, a coding depth of 8 bits is adequate. For a colour image, a coding depth of 24 bits is recommended.

For a colour image, some devices scan at a higher coding depth, namely 36 bits or 48 bits (12 or 16 bits per layer).

A mode is a mathematical model designed to define the relationships of colours to each other by the number of layers present within these modes. These layers determine the perceptible level of detail of the colour and the size of the file. The table below presents the three main modes, which will be used differently depending on the type of document to be digitised :

MODE NAME	NUMBER OF LAYERS	DETAIL OF LAYERS	APPLICATION
CMJN / CMYK	4	*Cyan – Magenta – Yellow – Black* Assigns each pixel a percentage for each layer. Although the CMYK model is a standard colour model, the exact range of colours represented may vary depending on the press and printing conditions.	Preferred mode for offset printing. (This mode is not recommended for digitisation.)
RVB / RGB	3	*Red – Green – Blue* The values of each pixel are between 0 and 255 and make it possible to reproduce more than 16.7 million different colours.	Preferred mode for colour digitisation.
Niveau de gris / Gray-scale	1	The value of each pixel is between 0 and 255, which makes it possible to reproduce 256 brightness values ranging from pure white (255) to pure black (0).	Preferred mode for black-and-white digitisation.
Certain modes, such as indexed colour, bitmap or bitonal are not feasible for digitisation projects.			

tab. 1 Modes suitable for heritage digitisation

1 *Information from the BanQ digitisation guide and the Adobe website : Marie-chantal Anctil, Michel legendre, Tristan Müller, Kathleen Brosseau, Louise Renaud, Recueil de règles de numérisation, Bibliothèque et Archives nationales du Québec, Bibliothèque nationale de France et Musée canadien de l'histoire, 2014.*
http://collections.banq.qc.ca/bitstream/52327/2426216/1/4671601.pdf
https://helpx.adobe.com/fr/photoshop/using/color-modes.html

COLOUR SPACE

A colour space is **a three-dimensional mathematical model representing all colours that can be perceived, used or reproduced within a mode**. Each colour it contains is thus associated with coordinates determining a precise point and corresponding to values such as luminance, saturation and hue.

There are two types of colour spaces :

- **Device-dependent** spaces, also referred to as ICC profiles.
- **Independent spaces** that describe a set of visible colours without reference to a device, also known as LCC profiles

Regardless of their type, colour spaces are differentiated by their gamuts, namely by the number of accessible and reproducible colours within that space. The [most] common RGB colour spaces include the following, from smallest to largest gamut: sRGB, Adobe RGB (1998) and the visible spectrum, as shown in the following figure.

It should be noted that the spatial coordinates vary from one space to another and that the same set of coordinates corresponds to several colours depending on the colour space. For example, on the above illustration, the green corresponding to the RGB coordinates [0, 255, 0] corresponds to three different greens depending on whether the coordinates apply to the visible spectrum, Adobe RGB or sRGB.

In relation to digitisation, it is commonly accepted that Adobe RGB 1998 is used for a colour image and Gray Gamma 2 for a greyscale image[2].

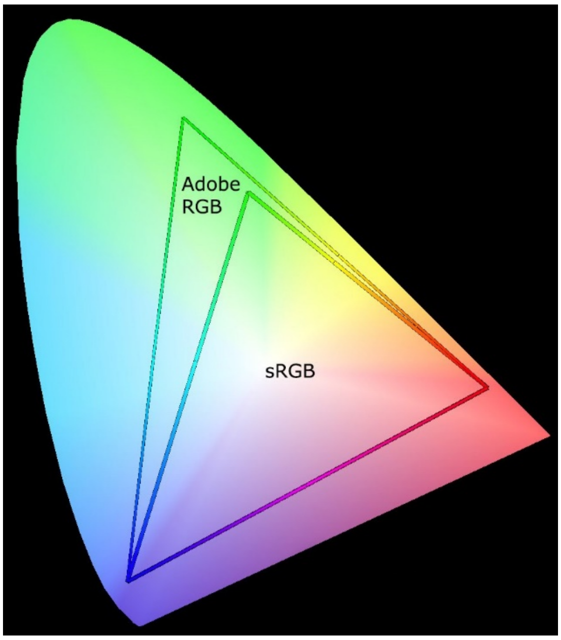


fig 2. RGB and sRGB colour spaces

FORMAT

The first image file from the scanning stage is referred to as « RAW ». This generic designation means that the image file obtained contains all the raw data from the sensor, as well as the parameters necessary to transform it into an image file that can be viewed on the screen. For it to be usable, **the file must be demosaiced, and thus converted into anoter format.**

The format in which the digitised document is saved varies according to the purpose of the digitisation project (dissemination or conservation). See the table opposoie for an overview of the most relevant formats depending on the type of operation envisaged:

2 *Marie-chantal Anctil, Michel legendre, Tristan Müller, Kathleen Brosseau, Louise Renaud, Recueil de règles de numérisation, Bibliothèque et Archives nationales du Québec, Bibliothèque nationale de France et Musée canadien de l'histoire, 2014.* http://collections.banq.qc.ca/bitstream/52327/2426216/1/4671601.pdf

DOCUMENT TYPE	DISSEMINATION FORMAT	CONSERVATION FORMAT
Text	.docx, .odt, .pdf (PDF) .epub .tei	.pdf (PDF/A)
Images	.jpg (JPEG), .jpg (JPEG2000). png	.tiff
Map / Plan	.dwg, .svg	.dxf .svg

tab 2. Formats to be used according to the possible uses

Source : Sébastien Soyez, La numérisation en marche : les étapes de la dématérialisation des processus de travail, Bruxelles : Archives générales du Royaume et Archives de l'État dans les Provinces, n.d.

http://www.arch.be/docs/brochures/la_numerisation_en_marche.pdf

For image files, **the TIFF format** is often preferred because of its excellent quality. However, this format is very large, and is not suitable for regular consultation.

It is therefore preferable to have an initial TIFF version that can then be converted into an appropriate format at the user's request or for internal use, most often in JPEG, JPEG 2000 or PNG format, where the difference is in the compression and therefore the size of the file. A JPEG causes a relative loss of quality when compressing the image to make it easily distributable, while a PNG compresses the image without alteration and without loss of quality.

For text files, **PDF format is the most commonly accepted standard**, as it is recognised by the ISO (International Organization for Standardization). It is used for both conservation and dissemination purposes because, in addition to being highly interoperable, it makes it possible to search for and/or select text in a document.

OCR (Optical Character Recognition) processing is specifically the step of converting an image into a text document. The document will then be structured according to a given standard, such as ePub or TEI (**see box below**).

It should be noted that **no format is designed to be permanent.** To guarantee the longevity of the formats used, Harvard University, in cooperation with Jstor, has developed the ***open-source software JHOVE***[3], which makes it possible to identify, analyse and characterise digital objects.

The software is therefore able to establish the properties of the file and their potential compliance with certain standards and norms selected by the user. From there, it is possible to analyse the relevance of the digital document, to correct any anomalies or to consider converting it to another format in the event of obsolescence.

3.2 Criteria for a quality image

Making images true to the original documents can, during capture, be a complex process where many variables come into play, along with the expertise and judgement of the digitisation teams. To assist institutions with the technical aspects of the process and to provide a methodological framework that is both rigorous and realistic, there are several

3 *Open Preservation Foundation, Démarrer avec JHOVE, 2015.* ***https://jhove.openpreservation.org/getting-started/***

standards and best practice documents that address the complexity of image quality.

These include the FADGI guide and the Metamorphoze guide, which is the basis for characterising the criteria for a good quality image. These are described below and include completeness, exposure, contrast, colour, focus, resolution, parallelism, flare, vignetting, noise and artefacts.

COMPLETENESS

The first quality criterion for a batch of images is its **completeness**. This means the completeness of the document as a whole is the priority – have all pages been scanned? – but also completeness of the elements of each image – has the page been scanned in its entirety? In the latter case, the person doing the checking must take into account the image cropping requirements. Indeed, the only way to show the reader that the whole object has been digitised is to leave a black border around the page, called the heritage frame. The archive file should therefore have enough space around the page to allow **cropping**, especially when pages are likely to require reskewing – image reskewing tools crop more or less 3% of the number of pixels per 1 degree of tilt[6].

---

4 *Thomas Rieger, Federal Agencies Digital Guidelines Initiative. Technical Guilines for Digitizing Cultural Heritage Material. Creation of Raster Image File, 2016, 99 p.* ***http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf***

5 *Hans van Dormolen, Metamorphoze Preservation Imaging Guidelines. Image Quality, 2012, version 1.0, 44 p.* ***http://www.imagingetc.com/images/Resources_Images/PDFs_DownloadFiles/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf***

6 *This proportion, given as an example, is exponential up to 45° and then decreases inversely to 0% at 90°. For an unlinked object, the Metamorfoze guidelines require a maximum tilt of 2° on the underside of the image. For a linked image, it is sometimes impossible to ensure such a degree of straightness in the image. We therefore recommend further enlarging the space around the object.*

Source : Hans van Dormolen, Metamorphoze Preservation Imaging Guidelines. Image Quality, 2012, version 1.0, 44 p. ***http://www.imagingetc.com/images/Resources_Images/PDFs_DownloadFiles/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf***

OCR

OPTICAL CHARACTER RECOGNITION

Optical Character Recognition (OCR) is an image processing technique aimed at extracting the text from an image to make it easier to read and to allow the user to perform searches easily.

The principle of OCR is divided into several steps, which include the following:

- *Deskewing of the page,*
- *Binarisation of the document, namely conversion into black-and-white,*
- *Segmentation of the image into text zones, then lines, words and characters,*
- *And finally classification of these zones for recognition of characters and then words.*

The most common OCR software applications are based on the principle of deep learning, namely improving the software's performance by learning from previously encountered characters and words. If a letter or word is not known by the software, the person in charge of the OCR operation indicates its meaning, which is then memorised and automatically recorded in the tool's dictionary. The tool thus becomes more autonomous and efficient.

There are several types of problems with OCR. First, the quality of the scanned image, which must be sufficiently deskewed and contrasted. The condition of the original document also plays a major role in the OCR process: printing defects, ink degradation or possible gaps make character recognition difficult for both the software and the operator and hinder the process. The structure can also be problematic if it is too complex – several columns – or if the page contains illustrations, very *small fonts or handwriting additions. Finally, the software is geared mainly towards recognising the Latin alphabet, which makes it difficult to transcribe, for example, the Arabic alphabet where the separation between letters and words is not as clear.*

However, constant research into the OCR suggests the following possible developments :

- *Better software automation and better management of complex documents – through the development of methods, tools and mathematical models – are refining the segmentation of the image into text zones and the management of documents containing both text and images[7].*

- *Improved handwriting recognition. As an indication, the HIMANIS[8] and CLAMM[9] research projects, carried out by the Institute for the Research of Texts and their History (IRHT-CNRS), are dedicated to image analysis and handwriting recognition for medievalists. The development of new handwriting recognition, classification and dating software has already made it possible to read and index corpora comprising several thousand pages.*

7 *Tan Lu, Homogeneity models for image processing in the cultural heritage sector, [Thèse de Doctorat non publiée], Vrie Universiteit Brussel, 2020, 245 p.*

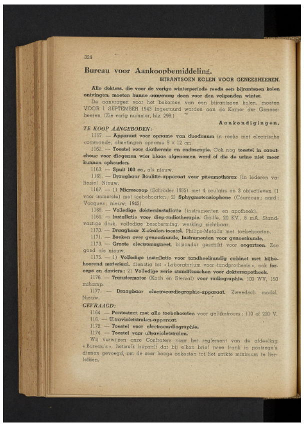
8 *IRHT - CNRS, HIstorical MANuscript Indexing for user-controlled (HIMANIS)* **https://www.irht.cnrs.fr/fr/ressources/carnet-de-recherche/himanis**

9 *IRHT - CNRS, Classification of Latin Medieval Manuscripts (CLAMM)* **https://www.irht.cnrs.fr/fr/ressources/sites-web-outils-corpus/clamm**

Where a page is missing, the only possible correction is to insert this page into the existing batch by means of a new digitisation operation.

The person performing the checking should also pay attention to **the sequence of images**, namely the order in which they appear on the screen. The sequence should reflect what the reader would find when consulting the object in the reading room. Any sequencing errors can be corrected by changing the page numbering. Objects are sometimes accompanied by loose pages, bookmarks or annotations.

The person performing the checking should then check with the curator whether these objects should be included in the sequence and where they should be placed. Is it an author's note, a page accidentally detached from its binding or a bibliographic note? Depending on the case, the images will have to be inserted or attached and described separately.



324

Bureau voor Aankoopbemiddeling.

BIJRANTSOEN KOLEN VOOR GENEESHEEREN.

Alle dokters, die voor de vorige winterperiode reeds een bijrantsoen kolen ontvingen, moeten hunne aanvraag doen voor den volgenden winter.

De aanvragen voor het bekomen van een bijrantsoen kolen, moeten VOOR 1 SEPTEMBER 1943 ingestuurd worden aan de Kamer der Geneesheeren. (Zie vorig nummer, blz. 298.)

Aankondigingen.

TE KOOP AANGEBODEN:

1157. — **Apparaat voor opname van duodenum** in reeks met electrische commande, afmetingen opname 9 × 12 cm.

1162. — **Toestel voor diathermie en endoscopie.** Ook nog **toestel in caoutchouc voor diegenen wier blaas afgenomen werd of die de urine niet meer kunnen ophouden.**

1163. — **Spuit 100 cc.,** als nieuw.

1165. — **Draagbaar Boulitte-apparaat voor pneumothorax** (in lederen valiesje). Nieuw.

1167. — 1) **Microscoop** (Schröder 1935) met 4 oculairs en 3 objectieven (1 voor immersie) met toebehoorten ; 2) **Sphygmotensiophone** (Courcoux ; aard : Vacquez ; nieuw ; 1942).

1168. — **Volledige doktersinstallatie** (instrumenten en apotheek).

1169. — **Installatie voor diep-radiotherapie.** Gaiffe, 20 KV., 8 mA. Standvastige druk, volledige bescherming, werking zichtbaar.

1170. — **Draagbaar X-stralen-toestel.** Philips-Metalix met toebehoorten.

1171. — **Boeken over geneeskunde, Instrumenten voor geneeskunde.**

1173. — **Groote electromagneet,** bijzonder geschikt voor **oogartsen.** Zoo goed als nieuw.

1175. — 1) **Volledige installatie voor tandheelkundig cabinet met bijbehoorend materiaal,** dienstig tot « Laboratorium voor tandprothesis », ook **forceps en daviers ;** 2) **Volledige serie standflesschen voor doktersapotheek.**

1176. — **Transformator** (Koch en Sterzel) **voor radiographie.** 100 WV, 150 miliamp.

1177. — **Draagbaar electrocardiographie-apparaat.** Zweedsch model. Nieuw.

GEVRAAGD:

1164. — **Pantostaat met alle toebehoorten** voor gelijkstroom ; 110 of 220 V.

116. — **Ultravioletstralen-apparaat.**

1172. — **Toestel voor electrocardiographie.**

1174. — **Toestel voor ultravioletstralen.**

Wij verwijzen onze Confraters naar het reglement van de afdeeling « Bureau's », hetwelk bepaalt dat bij elken brief twee frank in postzegels dienen gevoegd, om de zeer hooge onkosten tot het strikte minimum te herleiden.

fig.3 Example of cropping to ensure the completeness of the digitised document.

CegeSoma Collection

EXPOSURE

Exposure means the **action of light radiation on the sensor.** Specifically, an overexposed image is a digital reproduction that has captured too much light and will therefore appear too bright, while an underexposed image has only captured a limited quantity of light, and will appear too dark.

It is very rare for an exposure error to result in **underexposure,** which is manifested as poor legibility of an image. **Overexposure**, on the other hand, is more common and more serious, as it can lead to data loss: RGB values in a digital file are limited to 255. If the value for one or more channels reaches this ceiling value, it becomes impossible to correct the image, even a posteriori[10].

The exposure of an image can be adjusted **by configuring four elements: the sensitivity of the sensor, the aperture, the shutter speed and the light output.** Too much sensitivity increases the risk of digital noise . With experience, the person performing the checking should be able to detect overexposure by sight.

Fortunately, there are tools to help with this task. First, the **scanner's internal overexposure alert,** which applies a colour mask to the affected areas. By default, this alert is displayed as soon as one of the three RGB values reaches 250. If none of these values reach 255, it is still possible to recover the image, provided that it is demosaiced again – creating a new TIFF file based on the RAW file – by correcting the exposure in the capture tool.

Then, a tool such as the « eyedropper » tool provided by the Photoshop software can, for example, help to determine the extent of the problem and to refine the analysis. If it is found that one or more images are indeed overexposed and there is therefore a loss of information, the person performing the checking should reject the images and inform the team that a new capture is required.

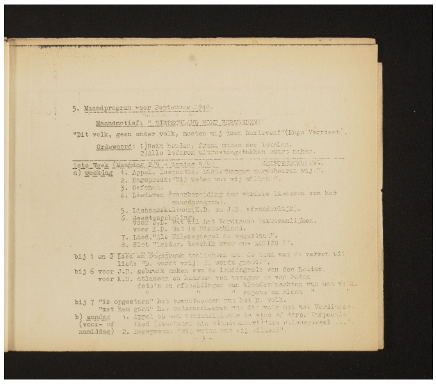


fig. 4 Example of an overexposed image
CegeSoma Collection

CONTRAST

The contrast of an image corresponds to the **difference in density between its lightest and darkest shades.** This notion does not correspond to a « light or dark » distinction, but rather to the relationship between these two types of colours.

Contrast plays an important role in the readability of the image and should therefore be carefully adjusted. **Gain modulation** is an interesting tool in this respect, and involves — using specific patches linked to the colour chart — **to check the level of contrast according to the grey levels, thus avoiding possible colour saturation.** This verification method is covered by the **ISO 19264**[11] standard, which details its

10 *This highlights the importance of retaining RAW files, at least until the preservation file is archived. There are also capture systems that do not generate RAW image files. In such a case, any error will require a new digitisation operation. It is important to consider this factor when purchasing equipment.*

11 *Organisation Internationale de Normalisation (ISO), ISO 19264-1:2021 - Photography — Archiving systems — Imaging systems quality analysis — Part 1: Reflective originals, juin 2021,* ***https://www.iso.org/obp/ui/fr/#iso:std:iso:19264:-1:ed-1:v1:en***

application and the patches and tolerance levels to be observed .

Measurement of gain modulation in the highlights therefore makes it possible **to validate the correct level of contrast in the digital image.** The aim is to ensure legibility without loss of information, soft tones and yellowed paper shades, and inks that are close to the tone of the paper.

Incorrectly set contrast levels can make ink unreadable, especially when it becomes browned over time, resembling the colour of yellowed paper. In case of doubt, the person performing the checking must, of course, verify that the physical documents are actually decipherable. If it turns out that the image is not similar to the physical object, it is necessary to check that the gain modulation function is correct.

To certify these values, it is important **for each image created to contain a standardised test target of grey values from white to black, using a Munsell or Kodak Gray Scale chart or similar.** Measurements can then be taken quickly. If the measured values are not correct, the batch of images must be rejected.

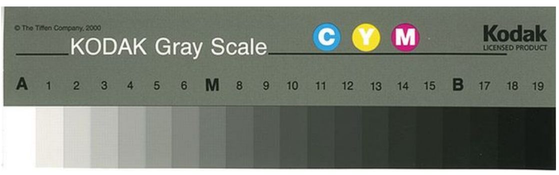


fig. 5 KODAK Gray Scale chart

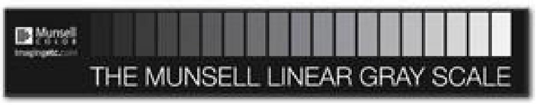


fig. 6 Munssel Linear Gray Scale chart

The **gain modulation function** is almost always represented in its graphical form and can also incorrectly be called gamma. It is commonly referred to as a **gradation curve**. By default, on digital SLR cameras, it draws a sigmoid[12], in particular to lighten the darkest parts and darken the highlights in order to preserve the details of the whole scene photographed

In the case of a reproduction, where the aim is to provide a digital version that is faithful to the original, the gradation curve must be linear. Contrary to what one might think, this setting provides better readability in soft contrasts, as the sigmoid shows very little growth in whites and deep blacks.

This measurement technique can be adapted to the entire range of neutrals up to the deepest blacks using the Munsell Linear Gray Scale chart for black-and-white digitisation – this is known as the opto-electronic conversion function analysis or OECF.

The analysis also makes it possible, by comparing the patches, to reveal a possible flare effect : it goes without saying that by deliberately choosing to prioritise contrast quality in the highlights, the limitations of the machine will be revealed in the lowlights.

If the analysis of the Munsell Linear Grayscale chart reveals non-linearity of the gradation curve in the darker tones despite a strict application of the recommendations listed above, only the qualities of the sensor can be blamed[13].

12 *The sigmoid gradation curve is a legacy of film photography and, in television, of the characteristics of the cathode ray tube. In photography, it is the qualities of the films – silver salts, filters, etc. – that determine the curve. This is interesting to know for professionals who wants a particular image result.*

13 *To check the quality of the sensor, it must be exposed to an illuminant capable – with a CRI of 95% – of producing a series of flashes for which the output must maintain an identical colour temperature close to 5000 Kelvin. The test must be repeated nine times to ensure that the results are valid, under predefined temperature and humidity conditions. The analysis of the results, documented according to a strict scheme, shows the range of densities reproducible by the sensor and the degree of compliance with the linearity of the illuminant – see ISO 7589, ISO 14524 and ISO 21550.*

COLOUR

Checking the correct reproduction of colours is probably the most delicate stage of quality control, as colour perception can vary from one individual to another[14]. In addition, the detection of a colour error is much more difficult when the series of images is distorted or when a document is the subject of only one image. Indeed, errors are easier to detect if there is an abrupt change in colour when a series of pages are reviewed.

Good colour management depends on the choice of illuminants, the choice of scanner, the choice of colour chart, the calibration of the scanning equipment and lastly the conversion of the RAW file into another format. In this case, it is possible to correct incorrect files by starting from the RAW file – on condition that the initial quality is sufficient – and by demosaicing again, while also verifying that the ICC and LCC profiles and the white balance of the images are identical to the settings of the reference image.

fig. 7 Comparison of poor and good colour reproduction.
CegeSoma Collection

14 *There are a number of colour vision disorders, the most severe of which, achromatopsia, is characterised by a total absence of colour vision. Between normal vision and achromatopsia, ophthalmologists identify a range of disorders that are sometimes very difficult to diagnose. It might therefore be interesting to test the vision of operators and checkers - the X-Rite Colour Challenge, for example, is relatively simple and quick.*

FOCUS

Unfortunately, **a blurred image is always a lost** image: it is true that many image processing software programs allow for sharpness correction, but as the algorithms used are not standardised, the process does not guarantee to restore the perfect legibility of the file. Some would argue that a legible image, even if slightly blurred, is still a legible image. But this does not consider the electronic interpretation, which the OCR software will struggle to perform due to the loss of information.

It must be possible to validate the sharpness of an image using an electronic process. To do this, each captured document should be bordered by four QA-62 targets. In practice, this is impractical, given the size of some documents in relation to the size of the sensors, and the configuration of some book cradles - V-shaped, aspirants, etc.

The same can be said the eyes of the person performing the checking. In case of doubt, it is important for that person to enlarge the image to its maximum display size, namely 100% or scale 1: 1 – which means that each pixel in the image is represented by a pixel on the screen. Often, **it is by comparing an image considered to be sharp with a dubious image that the person performing the checking will find a point of reference.**

Finally, **a capture system should be examined when it shows recurring irregularities in image sharpness.** This problem can be caused by two factors: **movement** or **unintentional adjustment.** Motion blur occurs when the shutter speed is too high.

Therefore, even the slightest vibration – the simple fact that the machine is in a busy place, for example – can have an influence and may alter the quality of the image.

The mirror movement of a poor quality DSLR unit can also be problematic. **The solution would then be to decrease the shutter speed by increasing the light output, with the use of an electronic flash being the best option.**

The second factor that can be the source of recurring blurring problems is **system misalignment**, namely when something happens between setting up and image capture. For example, some worn lenses have difficulty tolerating an upright position: because the mechanism is no longer stable, the focus changes over time and the images need to be redone. **In such cases, the person performing the checking must notify the person responsible for the equipment so that appropriate changes can be made.**

RESOLUTION

Resolution problems are usually of two kinds: **adjustment of the resolution during image capture or an error in the settings even before image capture has commenced.**

In the former case, the person performing the checking will observe **a change in the size of the object in the image**, accompanied by a **deterioration in sharpness**. The only solution is to r**e-scan the incorrectly calibrated images, but it must be possible to determine which images are corrupted.** To do this, the checker can use the colour chart associated with the image capture equipment during the calibration phase. For example, the Munsell chart includes, for example, a graduated ruler where one centimetre must correspond to a number of approximately 118 pixels per 300 PPI[15].

15 *It should be noted that this measurement cannot be done with all targets on the market: indeed, the Kodak Q13/Q14 targets, for example, do contain a ruler, but the measurements are inaccurate. This is also the case for the Color Checker charts. In such cases, the measurement can then be taken using the « Ruler » tool in Photoshop and the « Cropping » tool in any other software.*

The quality of the sharpness is an additional element that can be used to identify a change in resolution. However, it should be noted that the phenomenon is not always easily detectable, as it is rarely the consequence of a sudden change: it is more likely a component of the scanner – such as the repro stand, the tripod, etc. - which, if not properly attached, will gradually sag and progressively cause the quality of the images to deteriorate. pour 300 PPI .

In such cases, it is advisable **to compare the size of the target placed on the digitisation background of the first and last images of a series, in order to identify a possible change of size more easily**. This control point is essential, because if it is not directly identified at the end of the production phase. It is only during the cropping phase that the person responsible for editing the images will identify it and the entire digitisation chain will then be delayed.

The second cause of resolution problems, namely **an initial setting error,** is more difficult to detect visually, as it may be a bad ratio applied to a whole series of images, or even to a single image. An experienced checker may be able to detect this problem with the naked eye, if he/she is very familiar with the object and the equipment. Here again, the graduated ruler from the Munsell Chart can be used as a control point.

PARALLELISM

The person performing the checking will also have the task of verifying that the document is completely parallel to the photographic sensor during image capture. This task is easy when the scanned object has at least two right angles: the checking grids – present on all viewers – are simply applied to the image and a check is performed to ensure that the page edges adjacent to these corners are aligned.

If vanishing lines appear in the image, this means there is a divergence. If the object does not have any right angles, the checker should, in case of doubt, take at least two measurements at specific off-centre points on the object and compare these with the values shown on the image at the same points.

Another way to identify a parallelism error is to check if an area of the image is less sharp. To do so, the checker should enlarge the image to 100% and scroll it on all four sides and in the centre. This method is not always reliable, however, as some large-format sensors induce such a depth of field that a difference in orientation of a few millimetres will remain imperceptible, especially as the sensitivity of the eye varies from one individual to another.

If parallelism issues occur, the corrupted images must be rejected in all cases. A technician should be informed for analysis and the images must be rescanned under appropriate conditions.

FLARE

Flare is an **optical aberration that occurs when stray light enters the lens.** That lens, which should theoretically only capture the rays reflected by the object to guide them to the sensor, then receives other rays from a lambda source. The phenomenon is **characterised by a softening of the contrast over the whole image, sometimes by chromatic distortion, and, in the most severe cases, by a halo effect.** Once detected, the challenge is to find the origin of the problem so as to eliminate the phenomenon.

When flare seems to be apparent, the checker should compare patches 0 and 5 of the Munsell chart present on the image. If the difference in RGB value between these two measurements is not significant, the technician should be informed so that he/she can determine the origin of the problem and correct it before continuing with the digitisation.

Flare is most often **caused by incorrect positioning of the illuminant** – less than 30° in relation to the image capture plane. In this case, it can be easily compensated for by the use of a covering light shade. If the phenomenon persists, it is worth checking other parameters, such as the **cleanliness of the lens** – the trace of a finger on a lens is enough to cause flare –, **the type of floor of the digitisation room** – a glossy coating, for example – **or even the accidental presence of a reflective surface close to the object.**

Images with a flare effect should be rejected and the document scanned again.

VIGNETTING

Vignetting is **characterised by a darkening of the corners of an image. It is therefore a phenomenon that alters the uniformity of the light on the image capture plane.**

It is relatively difficult to detect, as the object rarely covers the entire digital background. At the same time, because this background is black to improve the readability of the image and the corners of a book page are the most exposed to handling and therefore to wear, these areas may already show a perceptible darkening on the physical object, making it more difficult to identify a vignetting problem. The checker must therefore pay particular attention to the condition of the object in case of uncertainty.

Vignetting can have various causes. The most common is an **incompatibility between the lens and the sensor: because of its shape, the lens sends a larger or smaller circular image to the rectangular sensor.**

This image is called the image circle. If the image circle is not able to cover the entire sensor uniformly, the brightness in the corners of the image is reduced. However, in the context of heritage digitisation, the lens is assumed to have been tested and to be free of this defect. If vignetting is detected, **it is more likely that a lamp has been moved by accident**. The checker will then notify the person responsible for setting up the machines to organise an analysis. By checking the qualities of the illuminant and demosaicing some images again, that individual will be able to identify the source of the problem.

NOISE

Digital noise is defined as **the presence of information in the digital image that does not appear in the original document.** This information is derived from the digitisation process. As the signal-to-noise ratio is strongest in low light and most visible in areas of uniform colour, **it is by observing the black background of an image that the person performing the checking will best be able to detect the phenomenon.**

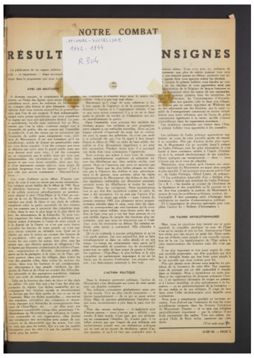
In concrete terms, this translates into **sets of pixels whose colours alter the legibility and sharpness of the image.** Most often, the noise is due to the sensor being too sensitive – ASA or ISO – and is referred to as **read noise**. The first step taken by the person performing the checking will therefore be **to check in the technical metadata of the image to see whether the ISO value is set to the minimum.**

As an indication, the majority of sensors coming off an assembly line have an optimal sensitivity of between ISO 50 and 200. The latest machines can deliver even smoother results up to ISO 400[16].

Since there is no system to correct them, images with digital noise should be rejected and re-scanned.

ARTEFACTS

Finally, it can happen that an artefact accidentally disrupts the correct reading of the image or simply spoils its aesthetic. This could be dust on a screen, a checking target present on the cropped image, or an object designed to hold open a page in a book.


NOTRE COMBAT

RESULT … NSIGNES

fig. 8 Example of an artefact
CegeSoma Collection

In such cases, the action to be taken depends on the digitisation policy put in place by the institution. Thus, a visible tool for straightening the lines of a text will make sense in the case of a printed page intended for OCR, but should be excluded in the production of a facsimile of a manuscript.

In order to provide a guideline, however, it is possible to establish that everything that can be easily removed from the image and is not useful for its reading should be deleted. For example, a checking target that encroaches on an object may disrupt the scan – as it contains text boxes.

16 *However, it is not advisable to set the value close to this, as the repeated electrical charges on the photodiodes and the non-stop computer processing during scanning can cause the system to heat up and alter the output signal. This is referred to as thermal noise. It should be noted that a sensor left in live view mode for too long will start to heat up. It is therefore possible that a series of images may be affected by thermal noise before the phenomenon fades away as the sensor temperature falls.*

In the same vein, depending on the severity of the situation and the operating objectives, dust or paper residues on the cradles, screens or background may result in a new digitisation operation.

The person performing the checking may draw the operator's attention to the problem in order to prevent it from recurring. If the deterioration in the image is due to an equipment defect, he/she will also inform the technical operator, who will take the necessary steps to get the machine back into production.

FOR FURTHER INFORMATION

BIBLIOTHÈQUE NATIONALE DE FRANCE (BnF), *Techniques et formats de conversion en mode texte*, n.d, [en ligne] ,
https://www.bnf.fr/fr/techniques-et-formats-de-conversion-en-mode-texte

CHIRON, G., DOUCET, A., COUSTATY, M., VISANI, M., et MOREUX, J.-P., *Impact of OCR Errors on the Use of Digital Libraries: Towards a Better Access to Information*, 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL), june 2017, Toronto, Canada, 4 p.
https://hal.archives-ouvertes.fr/hal-03025508/document

LU, T., *Homogeneity models for image processing in the cultural heritage sector*, [Thèse de Doctorat non publiée], Vrie Universiteit Brussel, 2020, 245 p.

MEMON, J., SAMI, M., KHAN R.-A., et UDDIN M., Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR). In : *IEEE Access*, vol. 8, pp. 142642-142668, 2020.

RIEGER, T., *Federal Agencies Digital Guidelines Initiative. Technical Guilines for Digitizing Cultural Heritage Material. Creation of Raster Image File*, 2016, 99 p.
http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf

VAN DORMOLEN, H., *Metamorphoze Preservation Imaging Guidelines. Image Quality*, 2012.
http://www.imagingetc.com/images/Resources_Images/PDFs_DownloadFiles/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf

04.

QUALITY OF DELIVERABLES : METADATA

4.1 Metadata : definition and characteristics

In the broad sense, metadata **are data about data.** Their purpose is **to define or describe other data, regardless of their medium** – paper or digital.

The term « metadata » was already used before the digital age by libraries and museums in the context of paper catalogues and inventories, but has become more widespread with the increasing computerisation of our societies. Today, the term corresponds to « **all the information describing any digital or other resource, [and whose] primary function is to describe the content of the resource, while making it possible to identify, qualify and enrich it**[1]».

Just as there is a wide range of heritage and document content with different characteristics, there are various types of metadata.

The primary function is to describe the content of the resource, while making it possible to identify, qualify and enrich it.

These data are written **internally** or **externally** to digital files. Metadata are said to be internal when they are **encapsulated within the same file as the digitised source.** Conversely, so-called external metadata are **contained in a file or database separate from the scanned document.** Seven categories can be distinguished :

---

1 *Observatoire de la culture et des communications du Québec, État des lieux sur les métadonnées relatives aux contenus culturels, Québec : Institut de la statistique du Québec, 2017, 118 p.* https://bdso.gouv.qc.ca/docs-ken/multimedia/PB01600FR_MetadonneesCulturel2017H00F00.pdf

CATEGORY	DESCRIPTION	EXAMPLES
Identification	Formally identify the scanned content, most often as an alphanumeric string.	• ISBN • ISNI • URI
Descriptive	Describe the scanned document as accurately and objectively as possible.	• Name of the author • Title of the work • Materials • Date created
Administrative	Provide information about the date the file was created and any subsequent changes.	• Date that the file was created • Name of the institution creating the file • Identification of the source document
Legal	Specify all legal terms and conditions to which the digitised content is subject: copyright, rights holder, licence to use, etc.	• Name of the owner or manager of the copyright • Indication of sources that must appear with the title of the work
Enriching	Provide additional information about the work or document.	• Biography of the artist • Photographs • Notes
Technical (or structural)	Provide information about how the file was created, its format and how it can be used.	• Identifier of the software and relevant version • Digitisation mode • Compression format • File size
Usage	Provide information about the consumption of a particular cultural content (most often automatically created by the platform distributing the scanned content).	• Number of views • Recommendations made by users

tab. 3 Different types of metadata

Source : Observatoire de la culture et des communications du Québec, État des lieux sur les métadonnées relatives aux contenus culturels, Québec : Institut de la statistique du Québec, 2017, p. 19 https://bdso.gouv.qc.ca/docs-ken/multimedia/PB01600FR_MetadonneesCulturel2017H00F00.pdf

4.2 Role and function of metadata

Metadata are used to **organise and sort information**. Using metadata, it is possible to search catalogues, directories and databases to find the information required. **The production of metadata is therefore neither a strategy nor an end in itself**, but rather a means of meeting the needs and expectations of the various stakeholders in the value chain of cultural products.

The more structured the information is, the easier it is to process and make it relevant. Given the ever-increasing flow of content, publications and information in general, metadata also ensure that information is not simply lost.

Before producing data, it is therefore essential to **know the context in which the information will be published and the purpose of the publication**. This makes it possible to deploy links that connect our information to all other relevant information. The more links there are between works, people, organisations, places and events, the greater the potential for all this information to be found and (re)used.

To meet these needs in a manner that is relevant, metadata must satisfy certain prerequisites: they must be **interoperable, meet an acceptable level of quality, and, finally, result from a vision shared by the various stakeholders and users involved**.

4.3 Prerequisite for the use of metadata: interoperability

The need to structure information first appeared in the 1970s, with the widespread use of computers in our societies[2]. One of the key elements in the evolution of metadata has been the need to have **information that is readable and usable by the various actors in a given sector, whether they are humans or machines**.

This is the challenge of **interoperability**, defined as « the ability of a product or system, with interfaces that are fully known, to work with other existing or future products or systems, without restrictions on access or implementation[3]».

Before going further, it is important to **differentiate between interoperability and compatibility**. Compatibility allows two systems to communicate with each other, while interoperability allows two systems to work together.

Compatibility is between a format A and a format B, and involves two elements understanding each other: format A of the data of a software makes it possible to receive the information contained in format B of the data of another software and vice versa. In short, compatibility is the result of translation from one format to another, on a case-by-case basis, especially when the formats are closed.

2 *Claerr, Thierry, et Isabelle Westeel, ed. Numériser et mettre en ligne. Villeurbanne : Presses de l'enssib, 2010. **http://books.openedition.org/pressesenssib/414***

3 *Définition de l'interopérabilité, n.d., **http://definition-interoperabilite.info/***

Interoperability, on the other hand, is general, since it **is based on one or more open formats** (such as HTML), **irrespective of the software used to create and use the data**. These standards have a dual role: they indicate how the dialogue between the different elements will operate and they allow a communication gateway that will be able to adapt to the changing needs of the elements.

More specifically, interoperability concerns three aspects of information exchange:

TECHNICAL INTEROPERABILITY

Technical or IT interoperability concerns the way in which systems are linked, the definition of interfaces, data formats and protocols. It describes the ability for different technologies to communicate and exchange data based on unambiguous interface standards clearly defined in advance (for example: HTTP, Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)).

SEMANTIC INTEROPERABILITY

Semantic interoperability ensures that the exact meaning of the information exchanged is understood by any other application, even if it was not originally designed for that specific purpose. To achieve this, systems need to share a common information exchange model, such as the Dublin Core or MODS metadata models.

SYNTACTIC INTEROPERABILITY

Finally, syntactic interoperability concerns the way in which data are coded and formatted, defining in particular the nature, type and format of messages exchanged (for example: XML, RDF).

INTEROPERABILITY OF IMAGES

THE IIIF PROJECT

*While many schemes and protocols have been developed to meet the needs of metadata exchange, **images have long remained outside these considerations**. Indeed, putting quality images online is still a complex, technical and expensive process. While cultural institutions have, for example, created tools for visualising digital content, these are only rarely interoperable, trapping the images on their original site and making them viewable only through this medium, which makes them difficult to share and reuse.*

*It was in response to this problem that **the IIIF Project** (International Image Interoperability Framework) **was created, with the objective of defining an interoperability framework for digital libraries.** This set of standards then allows scientific and cultural institutions to promote their content « in a standardised manner on the Web and thus make them available for searching, manipulating and annotating by any compatible application or software[4]».*

This interoperability framework is based on two pillars:

- *A data model, the **Shared Canvas**[5], based on linked data to describe digitised content in a cooperative way.*
- ***APIs** (Application Programmable Interfaces), namely sets of functions permitting access to the services of an application, through a programming language.*

These

APIs enable, for example, the definition of a standardised URL syntax making it possible to manipulate an image remotely or even to enable delivery of standardised presentation and structure information for a digital object.

*All this information is gathered in a single file, the « **IIIF manifesto** », which not only describes the digital image but also specifically describes the internal structure of the digitised content. Within this manifesto, this information is coupled with a canvas, namely an empty space representing in abstract a particular view of digitised content – for example, a page in a book. This canvas will serve as a receptacle for all sorts of digital resources designed as annotations on the same canvas.*

*This manifesto must then be read by an IIIF-compatible image viewer[6], which will then be **able to remotely load the information contained in the manifesto to reassemble the digital content in a web interface other than the original one, and to then be able to manipulate and annotate it at will.***

IIIF technologies therefore open up new horizons for exploiting digitised cultural heritage in an environment that is by nature decentralised. For cultural and scientific institutions, the advantages of such technologies are numerous: economically the project is not overwhelming since it is supported by a large international community. Dependence on a specific tool or software is reduced and it becomes easier to change tools and choose the best tool in its class. More broadly, IIIF standards increase the potential of digital library content.

4 *Régis Robineau, Comprendre IIIF et l'interopérabilité des bibliothèques numériques | Insula, 8 novembre 2016.* **https://insula.univ-lille3.fr/2016/11/comprendre-iiif-interoperabilite-bibliotheques-numeriques/**

5 *Robert Sanderson, Benjamin Albritton, Shared Canvas Data Model, n.d.* **https://iiif.io/model/shared-canvas/1.0/**

6 *The IIIF community has posted on GitHub all image viewers compatible with IIIF technologies:* **https://github.com/IIIF/awesome-iiif#image-viewers**

4.4 Criteria for quality metadata

As with images, digitisation professionals and scientists have identified criteria for defining quality metadata. The categorisation described below is directly inspired by the works of the Europeana Task Force[7], the FORCE11 Research Group and its FAIR Principles[8], the works of the Quebec Observatory of Culture and Communication[9], the Quality Assurance Framework (QAF) for statistical data[10], developed by Statistics Canada (STC), and lastly the works of the scientists Thomas R. Bruce and Diane I. Hillmann on the subject[11].

To be of good quality, metadata must be:

01 **Complete**

Metadata should describe the digital object as precisely as possible. These categories should be designed to apply as much as possible to all digital files in the collection.

7 *Marie-Claire Dangerfield, Lisette Kalshoven, Report and Recommendations from the Europeana Task Force on Metadata Quality, mai 2015, 54 p.* **https://pro.europeana.eu/files/Europeana_Professional/Publications/Metadata%20Quality%20Report.pdf**

8 *Force 11, Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0* **https://www.force11.org/fairprinciples**

9 *Observatoire de la culture et des communications du Québec, État des lieux sur les métadonnées relatives aux contenus culturels, Québec : Institut de la statistique du Québec, 2017, 118 p.* **https://bdso.gouv.qc.ca/docs-ken/multimedia/PB01600FR_MetadonneesCulturel2017H00F00.pdf**

10 *European Statiscal System, Quality Assurance Framework, n.d., version 2.0, 58 p.* **https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646**

11 *Thomas Bruce, Diane Hillmann, The Continuum of Metadata Quality: Defining, Expressing, Exploiting, 2004.* **https://www.researchgate.net/publication/247818823_The_Continuum_of_Metadata_Quality_Defining_Expressing_**

02 **Accurate**

Metadata must be correct and based on objective and factual knowledge of the document. Accuracy problems can take the form of typographical errors, incorrect use of names or places, or the extensive use of abbreviations.

03 **Reliable due to their origin**

The reliability of the metadata depends partly on the person who encoded them, that person's degree of competence, the methodology used to create and process the metadata, and finally that person's experience. In some cases, someone else changes the metadata after the object has been created.

However, there are tools and methodologies to ensure the reliability of metadata: in the libraries sector, for example, the Americans have developed the Anglo-American Cataloging Rules (AACR2).

04 **In line with expectations**

In order to be meaningful, the metadata must reflect the expectations and needs of the users of the digital object described. To do this, they must provide a precise definition of the operational objective of the digitisation project and identify the needs of the project's target audience.

05 **Consistent and coherent**

To ensure the consistency of digitised documents, the vocabulary used must be governed by precise rules to avoid inaccuracies.

06 **Accessible**

Metadata must be readable by both machines and humans. Using a standard that is too complex or difficult to read will hamper the encoding and processing of the document described. A fortiori, this difficulty could lead to errors of consistency and accuracy.

Accessibility is also characterised by how easy it is to find the object online, and in fact corresponds to the problem of the identifier. This must be unique and durable. As an indication, the identification standard URI[12] *(Uniform Resource Identifier) meets this dual requirement.*

4.5 A common vision for better interoperability

There is no universal standard that is sufficiently exhaustive to meet all of the previously mentioned quality criteria while adapting perfectly to the needs of cultural and scientific institutions and their audiences. In practice, metadata schemes often have multiple functions, as each type of cultural institution – library, archive and museum – has its own needs for identifying and cataloguing its collections.

Libraries have a long tradition of cataloguing their collections. Each work must be precisely identified using a bibliographic notice, but these metadata are often not very detailed or

12 *A URI is a short string of characters identifying a resource on a physical or abstract network (for example, a Web resource), whose syntax follows an internet standard established for the World Wide Web. The acronym URI is generally used to designate such string of characters. URIs are the basic technology of the World Wide Web because all hyperlinks are expressed in this form. They fall into two categories: URNs and URLs. For an overview of the W3C-recommended Web identifiers, see:* ***https://www.w3.org/Addressing/***

contextualised. Eulalia Roel points out in this regard that « the strength of the library model is that a minimal amount of metadata are assigned to a very high percentage of objects in the collection[13]».

One of the major challenges for libraries has been to facilitate the exchange of bibliographic data between libraries and archive services. This problem has been addressed in particular through the MARC format from the US Library of Congress, whose architecture and operation facilitate these exchanges. Several variants have been developed using this model: for example, the National Library of France has developed the InterMarc format, while Canada has created CAN/Marc. In reality, these individual efforts have paradoxically made data exchange more complex, resulting in a need to develop conversion programmes between metadata standards.

Archives differ from libraries in that metadata must first and foremost contextualise the various objects in the collection by accurately identifying the relationships of a collection to people, places, etc. The major difficulty in such an approach stems, according to Eulalia Roel, from the « major inconsistency in the details and variables of description, and from the widely disparate levels of granularity in the hierarchy of the structure of a collection to which metadata are assigned[14]». These disparities contribute to misunderstandings between types of institutions when they try to use metadata in several disciplines.

Museums are more reserved in their use of descriptive standards for their collections, partly because the culture of museums does not favour making all information about a collection freely available to the public. At the same time, the eclectic nature of museum collections often makes the exercise extremely complex: the rigidity of metadata would create an obstacle in terms of the nuanced information required for each object in a collection[15].

While interoperability has long been identified as the key element in data exchange between institutions and between sectors[16], the means required to achieve this level of interoperability remains uncertain.

As pointed out by Milena Dobreva[17] and Philip Hider[18], this difficulty stems from the fact that **interoperability is approached from two angles** that are not easy to combine: on the one hand, some institutions focus on **internal interoperability**, while others concentrate on **inter-institutional development**. On the other hand, in terms of how to achieve interoperability, some institutions emphasise compliance with metadata standards, while others focus on how new technologies can work with divergent metadata formats and content[19].

The objective today is therefore to develop metadata models that allow digitised content to be shared across the three sectors. This pooling of content requires more flexibility and inclusiveness in interoperability. Without claiming to be an exhaustive overview of the initiatives developed in recent years, the following are a few models that are moving in this direction:

---

13 *Eulalia Roel, The MOSC project: Using the OAI-PMH to bridge metadatacultural differences across museums, archives, and libraries. Information Technology and Libraries, 24 (1), 2005, 22-24.*

14 *Idem*

15 *Idem*

16 *Kalina Sotirova, Juliana Peneva, Stanislav Ivanov, Rositza Doneva, Milena Dobreva, Digitisation of Cultural Heritage : Standards, Insitutions, Initiatives, In : Access to Digital Cultual Heritage : Innovative Applications of Automated Metadata Generation, Plovdiv : Plovdiv Universtity Publishing House, 2012. http://www.math.bas.bg/infres/book-ADCH/ADCH-ch1*

17 *Idem*

18 *Philip Hider, Australian digital collections: Metadata standards and interoperability, Australian Academic & Research Libraries, 35, 2004.*

19 *Idem*

The CIDOC Conceptual Reference Model[20] **(CIDOC CRM)** is a conceptual model for describing museum objects developed since 2006 by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM). It aims to **make available all the information necessary to document and manage cultural heritage by providing a common and scalable semantic framework**, with which all cultural heritage information can be matched. It is still evolving, currently includes 86 classes and 137 properties and is centred around the notion of events, namely the type of relationship that unites an object and its characteristic - such as its creation date, for example.

Developed in 2017 by the International Federation of Library Associations and Institutions (IFLA), the **IFLA LRM**[21] provides a new model for libraries. It is focused around the needs of users in the sense that it **promotes the structuring of data for searching bibliographic information on the Web** to enable for instance users to identify works or places as a priority before locating a resource.

To achieve this objective, the model is based on three concepts: the **entity**, namely the object described, the **attribute**, which corresponds to the different characteristics of that object, and the **relationships**. These relationships concern an entity's relationships with its own attributes and relationships between several entities of one or more systems.

In the long term, **the aim is to replace bibliographic notices by a network of relationships between entities and thus to promote their visibility on the Web and their use by machines.**

Records in Contexts[22] (RiC) is a **project to overhaul archival description standards** initiated in 2016 by the International Council on Archives. **The objective is to improve and increase the interoperability of the archival descriptions published on the Web by creating a model that brings together existing archival description standards – ISAD(G)**[23], **ISAAR (CPF)**[24], **ISDF**[25] **and ISDIAH**[26].

Like the two models mentioned above, RiC also proposes the establishment of a relational system between entities and their characteristics, as well as between entities.

The objective today is therefore to develop metadata models that allow digitised content to be shared across the sectors

20 *Home | CIDOC CRM,*
http://www.cidoc-crm.org/

21 *IFLA, IFLA Library Reference Model (LRM),*
https://www.ifla.org/publications/node/11412

22 *International Council on Archives, Records in Contexts, Conceptual Model, n.d.*
https://www.ica.org/en/egad-ric-conceptual-model

23 *A standard that ensures the writing of compatible, relevant and explicit archival descriptions, while allowing the use of common authority data to make possible the integration of descriptions from different repositories.*

Source : Observatoire de la culture et des communications du Québec, État des lieux sur les métadonnées relatives aux contenus culturels, Québec : Institut de la statistique du Québec, 2017, p.98.
https://bdso.gouv.qc.ca/docs-ken/multimedia/PB01600FR_MetadonneesCulturel2017H00F00.pdf

24 *International Standard Archival Authority Record for Corporate Bodies, Persons and Families.*

Source : International Council on Archives, ISAAR (CPF), 2011.
https://www.ica.org/sites/default/files/CBPS_Guidelines_ISAAR_Second-edition_EN.pdf

25 *International standard for the description of functions and communities associated with the production and management of archives.*

Source : International Council on Archives, ISDF, 2011.
https://www.ica.org/sites/default/files/CBPS_2007_Guidelines_ISDF_First-edition_EN.pdf

26 *International Standard for Describing Institutions with Archival Holdings.*

Source : International Council on Archives, ISDIAH, 2011.
https://www.ica.org/sites/default/files/CBPS_2008_Guidelines_ISDIAH_First-edition_EN.pdf

The value of these new metadata models lies in **their approach to digital objects, considered not as isolated elements** to which a set of descriptive characteristics should simply be assigned, but **as a heterogeneous set of objects that only really makes sense through the relationships linking these objects to each other**. By bringing together existing standards, based on the language RDF (Resource Description Framework) and finally by relying on the use of Linked Open Data (LOD), these new models are fully in line with the transition to the Semantic Web.

Indeed, this approach not only makes it possible to build a more coherent and meaningful network of digital objects, but also to bring more visibility to these objects. For cultural and scientific institutions, this provides the prospect of creating and pooling rich and complex bodies of work, paving the way for increased exploitation of their collections.

For cultural and scientific institutions, this provides the prospect of creating and pooling rich and complex bodies of work, paving the way for increased exploitation of their collections.

However, while these models work towards greater linkages and interoperability, the way in which the different sectors will be able to combine and interact with each other has yet to be defined. Initial harmonisation efforts have been made in this direction with, for example, the **FRBRoo project**[27], which seeks to align or even merge the CIDOC CRM model, designed mainly for museums, and FRBR, developed for libraries.

27 *Introduction à FRBRoo | Documentation, n.d.* ***https://doc.biblissima.fr/introduction-a-frbroo***

FOR FURTHER INFORMATION

CARBONE, F., OOSTERBEEK, L., COSTA, F., Extending and adapting the concept of quality management for museums and cultural heritage attractions: A comparative study of southern European cultural heritage managers' perceptions, In: *Tourism Management Perspective*, vol. 35, juillet 2020.

FLECKENSTEIN M., FELLOWS, L., Data Quality. In: *Modern Data Strategy*, Cham: Springer, 2018.

BACHI, V., FRESA, A., PIEROTTI, C., PRANDONI, C., The digitization age: mass culture is quality culture. challenges for cultural heritage and society. In: IOANNIDES, M., MAGNENAT-THALMANN, N., FINK, E., ŽARNIĆ, R., YEN, A.-Y., QUAK, E., *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*, 5th International EuroMed 2014, Limassol, Cyprus, November 3-8, 2014.
https://www.digitalmeetsculture.net/wp-content/uploads/2014/12/The-Digitization-Age-Mass-Culture-is-Quality-Culture.pdf

GUEGUEN, G., MARQUES DA FONSECA, V.-M., D., PITTI, V. et SIBILLE, C., *Vers un modèle conceptuel international pour la description archivistique. Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description*, 2013.
https://www.ica.org/sites/default/files/EGAD_French.pdf

CHARDONNENS, A., La gestion des données d'autorité archivistiques dans le cadre du Web de données, [Thèse de doctorat non publiée], Université Libre de Bruxelles, 2021, 420 p. **https://dipot.ulb.ac.be/dspace/bitstream/2013/315804/4/These_ChardonnensAnne.pdf**

CARON, P., DEFIOLLE, R., et Lay, M.-H., *L'enjeu des métadonnées dans les corpus textuels : Un défi pour les sciences humaines*, Presses universitaires de Rennes, 2019.

BRUCE, T., HILLMANN, D., *The Continuum of Metadata Quality: Defining, Expressing, Exploiting*, 2004.
https://www.researchgate.net/publication/247818823_The_Continuum_of_Metadata_Quality_Defining_Expressing_Exploiting

05.
MANAGEMENT OF THE DIGITISATION STUDIO

While this guide has so far focused on digital documents, it is also worth looking at digitised originals. These precious and often fragile objects – in the case of ancient manuscripts, for example – require particular attention to be paid to their handling. Consideration should also be given to the environment around them in order to limit their deterioration.

This is what **preventive conservation** is all about. This concept is defined by the ICCROM[1] as « all measures and actions aimed at avoiding and minimising future deterioration or loss of an object of art or a heritage collection[2]». **These measures and actions are indirect: they do not interfere with the materials and structures of the items and do not change their appearance.**

1 *The International Centre for the Study of the Preservation and Restoration of Cultural Property (ICCROM) is an intergovernmental organisation working in the service of its member states to promote the conservation of all forms of cultural heritage, in all regions of the world.*

2 *ICCROM, Preventive conservation, n.d.* ***https://www.iccrom.org/fr/section/conservation-preventive***

5.1 Preserving heritage and documentary collections

There are ten main factors involved in heritage deterioration[3] :

- Temperature,
- Humidity
- Light
- Atmospheric pollution
- Dust and mildew
- Insects and animals
- The human factor – including theft and vandalism
- Disasters – including fires, floods and armed conflicts

Organic and cellulose materials such as paper, parchment or leather are among the most

3 *Canadian Conservation Institute, Agents of deterioration, 2017* ***https://www.canada.ca/en/conservation-institute/services/agents-deterioration.html***

sensitive to relative humidity[4], because they are hygroscopic, and thus capable of absorbing and desorbing moisture. Too much variation in relative humidity can lead, in particular, to dimensional changes, which in turn lead to cracking, tearing, lifting, deformation[5], etc.

The table of environmental recommendations[6] below summarises the environmental measurements and permissible variations for the optimal conservation of the most common materials found in archive and library collections.

The difficulty of preventive conservation lies in the heterogeneity of the objects and documents gathered in a single space: the optimal conditions for the conservation of paper are not the same as for a photograph or a leather object. **However, these figures are only indicators, because more than focusing on the figures, it is essentially a question of creating the most stable environment possible for heritage collections while maintaining the comfort of the teams.** The excellent *Vade-mecum de la conservation préventive* from the **Centre for Research and Restoration of the Museums of France (C2RMF)** includes recommendations on how best to control the environmental conditions for collections and how to respond in the case of anomalies .

The most important are the following:

- Measuring the relative humidity, temperature and luminosity of the environment by setting up thermo-hygrographs and light meters
- Making a climate map of the building by noting the measurements on a plan
- Identifying areas at risk
- Identifying the sources of the problems
- Minimising variations in temperature and humidity
- If necessary, reviewing the distribution of objects according to climate zones
- Carrying out regular checks and maintenance of measuring instruments and specific equipment
- Determining the most stable areas for storage of the most fragile collections
- Changing the climatic conditions in certain areas if necessary in order to determine when and how to move the collections
- Designing an emergency plan and an evacuation plan in case of a disaster

4 *Humidity is the water vapour content of air, and its content in the ambient atmosphere is measured in relative humidity, which expresses « the amount of water vapour contained in a given volume of air in relation to the maximum it could contain at a given temperature and pressure ».*

Source : Etienne Fehau, Nathalie Le Dantec, Vade-mecum de la conservation préventive, C2RMF, version du 18 novembre 2013, p.10.
https://c2rmf.fr/sites/c2rmf.fr/files/vademecum_cc.pdf

5 *Jean-Paul Oddos (dir.), La conservation : principes et réalités, Paris : Éd. du Cercle de la Librairie, Coll. Bibliothèques, 1995, 405 p.*

6 *This table summarises the information found in the following documents:*

Conseil Canadien des Archives, Manuel de conservation des documents d'archives, chapitre 3 - Environnement, Conseil canadien des archives, 2003, 18 p.
http://www.cdncouncilarchives.ca/RBch3_fr.pdf

Pierre Diaz Pedregal, Climat des magasins d'archives : objectifs, moyens et méthodes - Petit manuel de climatologie appliquée à la conception des bâtiments d'archives, Direction des Archives de France, 2009, 159 p.
https://francearchives.fr/file/47cc986ffbb4b6b02960a1bcbd2b4a6a890cc098/Manuel%20de%20climatologie.pdf

Etienne Fehau, Nathalie Le Dantec, Vade-mecum de la conservation préventive, C2RMF, version du 18 novembre 2013, p.10.
https://c2rmf.fr/sites/c2rmf.fr/files/vademecum_cc.pdf

7 *Etienne Fehau, Nathalie Le Dantec, Vade-mecum de la conservation préventive, C2RMF, version du 18 novembre 2013, p.10.*
https://c2rmf.fr/sites/c2rmf.fr/files/vademecum_cc.pdf

TYPE	TEMPERATURE	RELATIVE HUMIDITY
Paper	• 17 °C to 18,3 °C *Variation : 2°C over 24h*	• 45-55% *Variation : 3% over 24h*
Parchment and leather	• 17 °C to 18,3 °C *Variation : 2°C over 24h*	• 45-55% *Variation : 3% over 24h*
Textiles	• 17 °C to 18,3 °C *Variation : 2°C over 24h*	• 45-55% *Variation : 3% over 24h*
Photographs		
Black-and-white prints	• Below 18°C	• 30 - 50 %
Glass negatives	• 18°C	• 30 - 40 %
Black-and-white silver gelatin negatives on a polyester base	• 21°C +/- 2° over 24h	• 20 - 50 %
Cellulose acetate and cellulose nitrate negatives	• 18°C	• 18°C
Black-and-white silver gelatin negatives on a triacetate film base	• 2°C	• 2°C
Colour films	• 2°C	• 2°C
Colour prints	• Inférieure à 2°C	• Inférieure à 2°C
Microfilms *Black-and-white silver gelatin negatives on a polyester base* *Heat-treated silver films* *Vesicular films on polyester base* *Bleach-processed silver images on a polyester base*	• Maximum 21°C	• 20-50%

tab. 4 Recommended temperature and relative humidity for the preventive conservation of paper, textiles and leather

5.2 Managing light

Light is the small part of the electromagnetic spectrum that is detected by our visual system. It is called visible radiation to indicate that there are other types of radiation that are invisible to our visual system, but which accompany visible radiation in varying proportions: ultraviolet and infrared radiation. Indeed, all light sources, from the sun to fluorescent tubes and halogen lamps, emit a very large quantity of radiation, in both the visible and invisible ranges. Jean-Jacques Ezrati has listed the spectral composition of the most common light sources :

It is essentially a question of creating the most stable environment possible for heritage collections while maintaining the comfort of the teams.

LIGHT SOURCES	ULTRAVIOLET RADIATION	VISIBLE RADIATION	INFRARED RADIATION	GLOBAL
Light of an average day	6%	44%	50%	100%
Tungsten-halogen lamp	1%	9.5%	90%	100%
Fluorescent tube 3000 K	1%	89%	10%	100%
Fluorescent tube 5000 K	2%	88%	10%	100%
Electroluminescente diode	0%	100%	0%	100%

tab.5 The different types of illuminants and their radiation
Source : Jean-Jacques Ezrati, Éclairage d'exposition, musée et autres espaces, Paris : Eyrolles, 2015, p.97

This radiation **exerts photochemical and/or thermal actions on organic materials, leading to irreversible deterioration of the works.** Ultraviolet light has a photochemical action and is responsible for colour changes, browning of the paper and general mechanical weakening of the works. For its part, heat-producing infrared has a direct softening, drying or igniting effect on organic materials. More broadly, it accelerates their chemical degradation. Light-sensitive works deteriorate each time they are exposed to light. This damage is cumulative and irreversible. **The aim is therefore to eliminate ultraviolet radiation, to reduce infrared radiation, and to monitor visible radiation.**

To do this, the installation of blinds and UV filters on windows, monitoring using light meters and the installation of light-emitting diodes are potentially interesting solutions.

In short, protecting heritage objects from light **involves limiting the amount of light and reducing the intensity or duration of exposure.**

At the same time, light management is a particularly sensitive issue for digitisation studios. As noted above, the challenge of digitisation is to recreate the original document as faithfully as possible. However, too much light would impair the perception of colours and therefore the reliability of the digital reproduction.

Controlling light implies that **no other source interferes with the lighting installed, so it is essential to work in an environment that cannot be penetrated by daylight**.

Otherwise, a machine that has been calibrated during the winter would overexpose all the images during brighter periods, thus causing an irreversible loss of information. Other light sources – additional lighting for reading administrative documents or to avoid accidents, computer screens – can be added in the digitisation room provided they do not influence the light reflected by the object.

It is therefore essential that these sources are not directed towards the object and that their power level is very low - maximum 32 lux on the object.

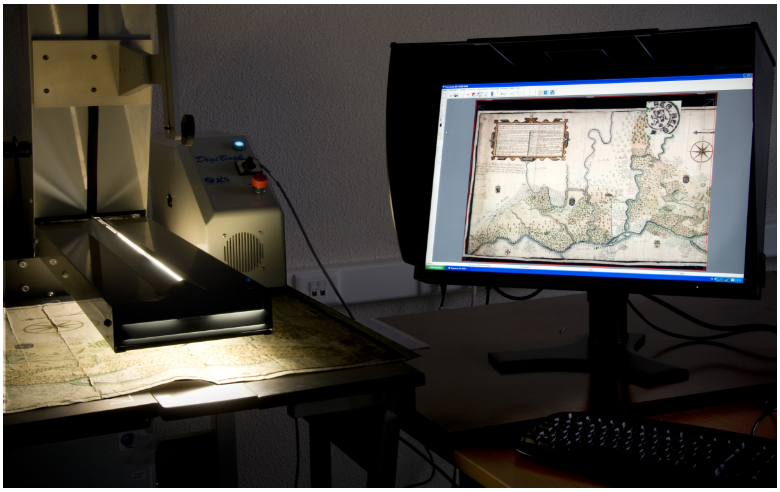


fig. 8 Digitisation studio
National Archives of Belgium

FOR FURTHER INFORMATION

ICCROM, an intergovernmental organisation at the service of its member states, promotes the conservation of cultural heritage in all its forms and throughout the world. In this regard, several tools are available to cultural institutions to support them in this endeavour:

- A risk management guide :
  https://www.iccrom.org/sites/default/files/2017-12/risk_management_guide_english_web.pdf

- The ABC method, which consists of a methodology for identifying, evaluating and correcting the risks and factors contributing to the deterioration of heritage collections:
  https://www.iccrom.org/sites/default/files/2017-12/risk_manual_2016-eng.pdf

The **Canadian Conservation Institute (CCI)** offers a quick reference tool providing heritage professionals with practical advice on how to protect collections from agents of deterioration. These recommendations are divided into three sections: building and facilities, equipment and materials, and procedures.
https://www.canada.ca/fr/institut-conservation/services/conservation-preventive/plan-preservation-collections-patrimoniales.html

The Centre de Recherche et de Restauration des Musées de France (C2RMF) has been conducting and publishing research on heritage conservation and restoration as well as on archive management and the use of new technologies for this purpose. On its website, the centre proposes a methodology to assist museums and cultural institutions in drawing up their conservation and safeguarding plan.
https://c2rmf.fr/conserver/fiches-techniques

In Belgium, the Royal Institute for Artistic Heritage (IRPA) has a unit dedicated to preventive conservation research. The institute offers advice and analysis to cultural and scientific institutions wishing to improve their approach to preventive conservation.
https://www.kikirpa.be/fr/conservation-restauration/cellule-conservation-pr%C3%A9ventive

EZRATI, J.-J. *Éclairage d'exposition, musée et autres espaces*, Paris : Eyrolles, 2015.

CONSEIL CANADIEN DES ARCHIVES, *Manuel de conservation des documents d'archives. Chapitre 3 – Environnement*. 2003.
http://www.cdncouncilarchives.ca/RBch3_fr.pdf

FEHAU E., LE DANTEC N., *Vade-mecum de la conservation préventive*, C2RMF, version du 18 novembre 2013, 46 p.
https://c2rmf.fr/sites/c2rmf.fr/files/vademecum_cc.pdf

06.

METHODOLOGY SHEETS TO ENSURE THE QUALITY OF A DIGITISATION PROJECT

The sheets presented below follow the chronological order of the digitisation chain and are ***summaries of the elements to be taken into consideration during the various stages of the digitisation process.***

They are each made up of three categories: ***key concepts****, which briefly summarise the important theoretical elements about the stage developed,* ***recommendations*** *in the form of lists, and finally* ***tools*** *for implementing these recommendations.*

BEFORE	Defining your digital data strategy Defining your digitisation policy
INTELLECTUAL AND STRATEGIC DEFINITION OF THE PROJECT	Defining your system reference Establishing and implementing a quality approach Defining the quality of the deliverables How and when to involve quality control Controlling quality in case of outsourcing
DEFINITION OF THE TECHNICAL ASPECTS	Choosing digitisation equipment Choosing illuminants Calibrating the digitisation chain
DIGITISATION	General rules for image capture Creating and filling the metadata inventory file Saving and naming of files
POST-PRODUCTION	Retouching image files OCR scanning of text documents Encapsulating metadata Final quality control
DISSEMINATION PRESERVATION	Choosing your digital data management system

01/ DEFINING YOUR REFERENCE SYSTEMS

KEY CONCEPT

A reference system is **a coherent or well-resourced set of data or information shared by a community of stakeholders.**

It is characterised by its stability, reliability, unity of meaning and interoperability. This type of document is crucial to the large-scale development of information exchange and communication and stems from the need for a common language for these exchanges. The use of reference systems is part of a dynamic approach used, for example, to identify, analyse, classify, compare or update processes and products.

According to this definition, the reference system can take several forms. In the context of a digitisation project, three types of reference systems are relevant :

- **Standards**
- **In-house document reference systems**
- **Methodological reference systems.**

Standards

Standards are standardised, documented definitions that centralise good practices and processes applied to various fields[1]. At international level, the ISO (International Organisation for Standardisation) is the reference in this field. There are also standardisation bodies dedicated to specific areas: This is the case, for example, of the World Wide Web Consortium (W3C) or the Organization for the Advancement of Structured Information Standards (OASIS), which focus on information technology standards.

1 *If the specifications of a standard are fully adopted by an organisation, it is possible to apply for certification, which underlines the seriousness, technical competence and reliability of the organisation. Although this is interesting in many respects, it remains optional. In Belgium, BELAC has been the only organisation in charge of these certifications since 2006.*

In-house document reference systems

In-house document reference systems most often take the form of audits. These are more or less elaborate documents specific to an organisation, bringing together the processes, resources and important information that needs to be known on a given subject.

This type of document tends to be increasingly standardised, in particular thanks to the ISO 19001 standard, which provides a rigorous and comprehensive framework for the drafting of these documents. On its website, the National Library of France (BnF) also provides a methodologicalschemeforthe preparation of audits.

Methodological reference systems

Lastly, methodological reference systems are, as their name suggests, corpora of methods and good practice on a given subject.

These documents change with each version and can be of any nature: general or specific to a domain or process, internal or not. Most often, this kind of document needs to be adapted to the specific realities of the organisation or institution using it.

RECOMMENDATIONS

- **ANCTIL, M.-C, LEGENDRE, M., MÜLLER, T ., MAILLET, D., BROSSEAU, K., RENAUD, L.,** *Recueil de règles de numérisation*, Bibliothèque et Archives nationales du Québec, Bibliothèque nationale de France et Musée canadien de l'histoire, 2014. ***https://numerique.banq.qc.ca/patrimoine/details/52327/2426216***

- **BIZINGRE, J., PAUMIER, J., RIVIÈRE, P.**, *Les référentiels du système d'information*, Paris : Dunod, 2013.

- **BURESI, C., et CÉDELLE-JOUBERT, L.**, *Conduire un projet de numérisation*, Paris : Tec et Doc, 2002, 326 p.

- **CLAERR, T., WESTEEL, I., et MELOT, M.**, *Manuel de la numérisation*, Paris : Éd. du Cercle de la Librairie, 2011. 317 p.

- **CLAERR, T., et WESTEEL, I.**, *Numériser et mettre en ligne*, Villeurbanne : Presses de l'enssib, 2010. ***http://books.openedition.org/pressesenssib/414.***

- **CLAERR, T., WESTEEL, I., ORY-LAVOLÉE, B., GEORGES, N.,** *Manuel de constitution de bibliothèques numériques*, Paris : Éd. du Cercle de la Librairie, 2013. 407 p.

- **DIRECTION DES ARCHIVES DE FRANCE. MINISTÈRE DE LA CULTURE ET DE LA COMMUNICATION - COMITÉ DE PILOTAGE NUMÉRISATION – MISSION DE LA RECHERCHE ET DE LA TECHNOLOGIE (DDAI)**, *Écrire un cahier des charges de numérisation du patrimoine. Guide technique. Documents reliés- manuscrits- plans- dessins- photographies- microformes*, Paris : Archives de France, 2008, 62p.
http://www.archivesdefrance.culture.gouv.fr/static/4132

- **DIRECTION INTERMINISTÉRIELLE DU NUMÉRIQUE ET DU SYSTÈME D'INFORMATION ET DE COMMUNICATION DE L'ETAT**, *Référentiel général d'interopérabilité*, 2015, version 2, 84 p.
https://www.numerique.gouv.fr/uploads/Referentiel_General_Interoperabilite_V2.pdf

- **FEHAU E., LE DANTEC N.**, *Vade-mecum de la conservation préventive*, C2RMF, version du 18 novembre 2013, 46 p.
https://c2rmf.fr/sites/c2rmf.fr/files/vademecum_cc.pdf

- **FORCE 11,** *Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0*, N.d.
https://www.force11.org/group/fairgroup/fairprinciples

- **PLAN D'EXPLOITATION ET DE PRÉSERVATION DES PATRIMOINES (PEP'S).** *Normes et lignes directrices techniques et organisationnelles pour les initiatives de numérisation des patrimoines culturels soutenues par la Communauté française, ministère de la communauté française*, Bruxelles : Secrétariat général – Préservation et exploitation des patrimoines, 2009, 61 p.
http://www.peps.cfwb.be/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/sites/numpat/upload/numpat_super_editor/numpat_editor/documents/CFWB/Normes.pdf&hash=a2d0e740e46e34b8c3475c604c3e96edbb8a8712

- **RIEGER, T.,** *Federal Agencies Digital Guidelines Initiative. Technical Guilines for Digitizing Cultural Heritage Material. Creation of Raster Image File*, 2016, 99 p.
http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf

- **VAN DORMOLEN, H.,** *Metamorphoze Preservation Imaging Guidelines. Image Quality*, 2012.
http://www.imagingetc.com/images/Resources_Images/PDFs_DownloadFiles/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf

ISO 9000 :2015	**Quality management systems** *Fundamentals and vocabulary*
ISO 9001 :2015	**Quality management systems** *Requirements*
ISO 9004 :2018	**Quality of an organisation** *Guidelines for optimum performance*
ISO 19011 :2018	**Guidelines for auditing management systems**
ISO 31000	*Risk Management*
ISO 15489-1 :2016	**Information and documentation — Records management** *Part 1: Concepts and principles*
ISO/IEC 27001:2013	**Information technology — Security techniques — Information security management systems** *Requirements*
NF Z42-026	**Definition and specification of services for the faithful digitisation of paper documents and monitoring of these services**
ISO 2859-1	**Sampling procedures for inspection by attributes** *Part 1: Sampling schemes indexed by acceptance quality limit (AQL) for lot-by-lot inspection*
FD Z42-017: 2009	**Electronic imaging – Digitisation of documents – Subcontracting digitisation operations** *Guide for drawing up technical specifications*
ISO 12641-1 :2016	**Graphic technology — Prepress digital data exchange** *Colour targets for input scanner calibration*
ISO 19264-1 :2021	**Photography — Archiving systems — Imaging systems quality analysis** *Part 1: Reflective originals*
ISO 7589 :2002	**Photography — Illuminants for sensitometry** *Specifications for daylight, incandescent tungsten and printer*
ISO 14524 :2009	**Photography — Electronic still-picture cameras** *Methods for measuring opto-electronic conversion functions* (OECFs)
ISO 21550 : 2004	**Photography — Electronic scanners for photographic images** *Dynamic range measurements*
ISO 25964-2:2013	**Information et documentation** *Thesauri and interoperability with other vocabularies*

02 / ESTABLISHING AND IMPLEMENTING A QUALITY APPROACH

KEY CONCEPT

A quality approach or quality management is a managerial concept enabling better understanding and analysis of work processes and their results.

This concept is based on the PCDA (Plan – Do – Change – Act) quality management method, aiming to create a virtuous circle that works constantly to improve the quality of projects and institutions.

The heir to this method, the ISO 9000 family of standards, offers a general methodological framework for improving work processes and their results.

Objectives et advantages

- Increasing the individual performance of the digitisation studio by **improving its organisation**
- Taking better account of **user** demand and offering users **content in line with their expectations**
- Including the institution in **the European digital transition and sharing an over-arching organisational culture**
- Improving **the quality of deliverables**

RECOMMENDATIONS

- Develop **a data management strategy.**
- Establish **a digitisation policy.**
- Draft **specifications** for each digitisation project.
- Create a **workflow** for the digitisation process.
- Define precisely the role and tasks of each team member.
- Use a common project monitoring tool for the whole team to facilitate communication.
- During the course of the project, **keep a log of errors, anomalies and malfunctions and how these difficulties were addressed.**
- At the end of the digitisation project, **evaluate the project with the whole team**, on its own and in comparison with past digitisation projects: have some of the same errors been encountered? Has the response to these malfunctions been appropriate? Have the expected results been achieved?
- **Adapt the strategic documents and workflow as necessary to improve subsequent projects.**

TOOLS

- **Standard ISO 9000 :2015 - Quality Management Systems**

  Essential principles and vocabulary

  Requirements

- **Standard ISO 9000 :2015 - Quality Management Systems**

  Quality of an organization — Guidance to achieve sustained success

- **Project monitoring tools : Projeqtor et Redmine**

  *The open-source tools **Projector** and **Redmine** offer many features for tracking tasks, visualising the resources available for a project or for managing the planning within a team.*

FOR FURTHER INFORMATION

BAROUCH, G., *Révolution des pratiques managériales : le changement par la qualité*, La Plaine Saint-Denis, Afnor Éditions, 2013

GIESEN, E., *Démarche qualité et norme ISO 9001 : Une culture managériale appliquée à la recherche*, Marseille : IRD Éditions, 2018, 180 p.

CARBONE F., OOSTERBEEK L., COSTA C., FERREIRA AM., Extending and adapting the concept of quality management for museums and cultural heritage attractions: A comparative study of southern European cultural heritage managers' perceptions, In: *Tourism Management Perspective*, vol.35, juillet 2020.

03 / DEFINING THE QUALITY OF THE DELIVERABLES

KEY CONCEPTS

Quality: All of the characteristics and properties of a product, a process or a service that influence its ability to meet identified or implicit requirements.

Sub-quality or non-quality: Non-compliance of the product with the minimum requirements expected of a digitised document, failing to ensure its legibility or aesthetic integrity and thus effectively hindering its use.

Over-quality: Excessive optimisation of the quality of a product that results in the use of too many resources and costs to improve the product without real added value.

Minimum or acceptable quality threshold: Margin of error permitted in the production of a digitisation project. These errors should be minor and should not interfere with the readability, aesthetic integrity and use of the document.

Objectives

- Quality is defined in terms of the intended use of the digitised documents (preservation / dissemination / exploitation).
- Three milestones are to be identified during the development of the project: non-quality, over-quality and the acceptable quality threshold.
- For each criterion defined as a key element of a production quality analysis, it is necessary to define thresholds above or below which images will be declared compliant or non-compliant.
- These thresholds can be based on one or more values for each criterion.

RECOMMENDATIONS

EXAMPLES OF MAJOR ERRORS

- No inventory file finalised by the service provider
- Non-compliance of digital file formats and media
- Non-compliance of the naming applied to media, folders, files and images
- Non-compliance of file structures and directories
- Inconsistency in the content of the main fields of the inventory file
- Absence of indexing requested
- Non-compliance of the main indexing with the rules provided
- Incomplete digitisation (documents not reproduced)

EXAMPLES OF MINOR ERRORS

- Inconsistency in the content of secondary fields in the inventory file
- Non-compliance of indexing entry
- Non-compliance of the order of images
- Poor cropping with no impact on reading, and no loss of information
- Minor deviation in colour and/or density compared to the reference files
- Incorrect orientation of images
- Impaired fidelity to the original document (problems with light distribution, flatness, perspective, sharpness, scratches)
- Use of dirty glass to flatten a document, for example

04 / WHEN AND HOW TO INVOLVE QUALITY CONTROL

KEY CONCEPT

There are **three types of quality control** to be carried out during the digitisation phase:

- **Visual control**, which allows you to check the legibility and aesthetics of the image.
Examples: margins, background colour, blur, etc.

- **Technical control**, which involves the verification of the technical elements of the digitised content.
Examples: resolution/definition, file formats, metadata (EXIF, IPTC XMP model or others), presence of colour profile, test pattern analysis, etc.).

- **Integrity or consistency control**, which focuses on verifying compliance with models that affect the integration of images into the digital document management tool
Examples: tree structure, file name, directory name, completeness, etc.

Theoretically, each stage of the digitisation process can be subject to quality control: depending on the result, the process can be continued or partially or completely redone.

Pragmatically, **the earlier non-compliance is detected, the lower the cost of redoing the process.** A non-compliant product detected by the end user costs at least twice as much as a product properly manufactured in the first instance.

The cost of redoing the work is sometimes much higher in reality. A non-compliant product discovered during production represents a significant additional cost, but less than a complete reworking. Quality is therefore financially justified.

RECOMMENDATIONS

STAGE IN THE	FACTORS TO BE CHECKED	RECOMMENDED TOOLS
Preparation of the environment	• Temperature • Level of relative humidity • Brightness	• Thermohygrometers • Light meters
Choice of illuminants	• Arrangement and tilting of the lights around the image capture equipment	
Selection and preparation of the documents	• Condition of the original documents • External elements to be removed for digitisation (paper clips, loose pages, bookmarks, etc.)	
Creation and filling of the descriptive metadata inventory file	• Consistency between the data entered and the standard used • Completeness and accuracy of the metadata entered • Consistency of the vocabulary used	• OpenRefine
Calibration of the chain	• Condition of the equipment: stability, cleanliness of the lens • Impact of brightness on the image capture equipment • Colour space used • Consistency of the ICC profiles applied to the peripherals	• AutoSFR • OpenDice • Universal Test Target (UTT)
Pre-production & Production	• Completeness • Exposure • Contrast • Colour • Focus • Resolution • Parallelism • Flare • Vignetting • Noise • Artefacts	
Saving and naming of files	• Consistency between the names of files and the institution's naming plan	• File Renamer Basic • BulkRename

STAGE IN THE DIGITISATION CHAIN	FACTORS TO BE CHECKED	RECOMMENDED TOOLS
Retouching of image documents	• Consistency between the ICC profiles of the image capture equipment and the program • Presence and accuracy of technical metadata Technical image characteristics	• Photoshop • Lightroom • Gimp
OCR processing of text documents	• Omissions • OCR software classification issues • Segmentation problems	• ABBY Fine Reader • Tesseract • Layout Evaluation
Encapsulation of technical and administrative metadata	• Completeness of the metadata present: descriptive/administrative/technical • Accuracy of the paths between the various metadata files	• METS Standard
Final quality control	• Consistency between the criteria described in the specifications and the results obtained: • Compliance of formats • Compliance of file naming • Completeness of images • Technical image characteristics • Presence and accuracy of descriptive, technical and administrative metadata • Completeness and accuracy of the METS file	

05 / CONTROLLING QUALITY IN THE CASE OF OUTSOURCING

KEY CONCEPT

Outsourcing of production means **the possibility for an organisation to entrust all or some of the digitisation process to an external service provider**. Outsourcing makes it possible to respond to a lack of technical resources or to manage a large flow of documents to be digitised. It can also be used as a strategy to control the costs associated with an operation or as an option to gain greater expertise. However, outsourcing requires additional steps to ensure that control of the digitisation chain is not lost and to protect the quality of the deliverables.

The scope and nature of the tasks to be entrusted to the service provider are determined when the specifications are drafted. The control procedures must also be defined precisely.

RECOMMENDATIONS

- **Determine the scope of action of all employees** : type of tasks, process to be applied, deadline.
- **Define the general conditions for performance** (standards to be applied, methodological reference systems, etc.).
- **Define the procedures for transporting and handling original documents.**

- **Define the method for monitoring** during production, specifically the type of reports and their frequency.
- **Define the steps to be taken if anomalies are identified** during production.
- **Provide the service provider with the naming scheme to be applied**, the metadata standard to be applied and the indexing to be performed.
- **Share responsibility for quality control of deliverables.**
- **Define** simple, reliable and relevant **control monitoring indicators**.
- **Set a permitted percentage of errors.**

Example: The National Library of France (BnF)

Since its first large-scale digitisation projects, the **BnF** has chosen to outsource part of its digitisation projects, and therefore to revise its quality control policy to ensure a smooth and transparent dialogue with external providers.

The solution adopted was **to develop a quality approach based on the principle of fluidity** with the following fundamental principles:

The involvement of the service provider from the point when the specifications are drawn up, and in particular agreement on the working methods, the quality of the digital document, the processes in the digitisation chain, the indicators to be monitored, etc.

The establishment of a constant dialogue, in particular in the form of weekly reports drawn up by the service provider indicating the number of documents entering and leaving each studio, as well as the number of pages entering and leaving each studio. These reports make it possible to identify the problematic stages, to analyse the causes of these bottlenecks and to correct the digitisation chain appropriately during production.

Sharing of responsibility for quality control.

For more information please consult:

BELLIER, L., Numérisation, pour une nouvelle approche de la qualité, In: *Bulletin des bibliothèques de France*, Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB), 2014.

https://bbf.enssib.fr/contributions/numerisation-pour-une-nouvelle-approche-de-la-qualite#xd_co_f=NTllNDJiZTMtY2YyYi00ZDI5LWI3OGUtNzkwOTQzYzAyNjE2~

06 / CHOOSING DIGITISATION EQUIPMENT

RECOMMENDATIONS

The range of scanners for heritage and document digitisation has never been greater. In the face of the overkill from manufacturers in terms of technical innovations, it is sometimes difficult to identify the basic selection criteria for digitisation equipment.

Choosing digitisation equipment means first and foremost looking at the following :

- The **type** of documents to be scanned.
- Their **formats.**
- The **quantity** of documents and **how precious they are.**
- Their **condition**, and how they can be handled in the process.
- The available **budget.**
- **The level of qualification** of the teams.
- **The degree of complexity in using the equipment** and whether the teams are trained to the systems concerned.
- **The space** required to store the machines.
- **The environmental conditions** of the digitisation studio.

Once these elements have been determined, the choice of image capture equipment **is based purely on technical characteristics:**

- Resolution, or the accuracy of the image capture.
- The depth of analysis, or the number of colours and greyscales the device is capable of producing.
- **The interface**, the presence of USB ports.
- **The digitisation speed.**

Suitable and recommended image capture systems for heritage and document digitisation are:

- **Cameras.**
- **Drum scanners.**
- **Process cameras.**
- **Scanners specialising in the digitisation of bindings.**

Please note

- Over time, or as a result of a handling error, **the elements of a system can deteriorate or lose their settings** without this necessarily being noticeable to the naked eye.
- This is why it is important **to carry out regular quality control of the images produced. The frequency of this type of control depends on the manufacturer's recommendations and how intensively the machine is used**. This type of control also makes it possible to validate the images produced between two tests and to participate in the optimisation of potential printing work.

07 / CHOOSING ILLUMINANTS

KEY CONCEPTS

As pointed out above, **natural light is damaging to heritage objects**. The choice of illuminants and their output and positioning is therefore essential in optimising the quality of the digitisation process.

The quality of lighting is assessed by its ability to make the entire colour spectrum perceptible, which is defined as the discriminating capacity. This capacity is measured by the Colour Rendering Index (CRI) using a dedicated colour temperature meter for digital images. The maximum CRI value is 100 and corresponds to ideal daylight. Its minimum value is 0 and corresponds to a light that would make it impossible to distinguish colours. **The minimum CRI value for good colour reproduction is 95.**

To achieve this threshold and ensure uniformity of lighting, **the illuminants must be of the same quality and output, positioned symmetrically on either side of the capture plane, and directed towards the centre of the capture plane at an angle of 30° to 45°**. Their height must be at least equal to that of the lens, which should be equipped with a light shade, in particular to avoid **flare**.

RECOMMENDATIONS

- **An electronic flash** is the best solution to ensure that the appearance of an object is rendered correctly. With a CRI of around 98%, the flash light has a light output that improves contrast values.

- **Metal halide lamps** – HMI, MSR, GEMI, etc. – are an interesting alternative. Initially designed specifically for cinema and television, they produce a stable light with a wide and complete spectrum. Calibrated between 5000 and 5500 Kelvins, their CRI is close to 95%.

- **Types of lighting not recommended for digitisation include fluorescent lamps – for which the CRI is too low – and LEDs – which do not provide optimal colour rendition.**

- **The two illuminants used should be positioned on either side of the digitisation equipmen**t. In the case of a large document, it is best to opt for four light sources.

- The light should be **softened and constant**. Maximum recommended brightness: **32 lux.**

- **The screen should be protected from daylight** or direct glare from illuminants to prevent light from reflecting off the digitisation equipment.

08 / CALIBRATING THE DIGITISATION CHAIN

KEY CONCEPTS

Calibration involves setting the parameters for all the image capture equipment and peripherals according to the technical requirements stated in the specifications. This makes it possible to set the following parameters:

- Resolution
- White balance
- Exposure
- Focus
- Contrast
- Colours and how they are rendered

Consistent, accurate colour management requires **the use of reliable ICC profiles for all colour devices.** If these profiles are not controlled, a scanned image may show colour differences from one device to another, due to a simple difference between the scanner and the graphics display software. With a reliable profile, the program importing the image can correct differences between devices and thus display the true colours of a scanned image. It is therefore essential – regardless of the colour profile chosen – **to align these profiles across the different devices and to calibrate them in the same way.**

Please note

Every scanner has its weaknesses in colour rendering. It is therefore essential **to calibrate the image capture equipment regularly by scanning a document and comparing the colours of the digital file and the original to ensure the reliability of the rendering**. A colour chart – or colour test pattern – placed next to the scanned object enables a better assessment of the colour rendering.

TOOLS

Standards and reference systems

- **Standard ISO 12641-1 :2016 - Graphic technology**

  Prepress digital data exchange — Colour targets for input scanner calibration

- **Standard ISO 7589 :2002 - Photography**

  Illuminants for sensitometry — Specifications for daylight, incandescent tungsten and printers

- **Standard ISO 14524 :2009 - Photography**

  Electronic still-picture cameras — Methods for measuring opto-electronic conversion functions (OECFs)

- **Standard ISO 21550 : 2004 - Photography**

  Electronic scanners for photographic images — Dynamic range measurements

- **Metamorfoze Guide**

  http://www.imagingetc.com/images/Resources_Images/PDFs_DownloadFiles/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf

- **FADGI Guide**

  http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-

Colour spaces and ICC profiles

- ## Adobe RGB 1998

  Preferred colour space for colour digitisation consisting of three levels (Red–Green–Blue).

- ## Adobe sRGB

  Alternative colour space for colour digitisation consisting of three levels (Red–Green–Blue).

- ## Gray Gamma 2.2

  Preferred ICC profile for greyscale. Makes it possible to set the balance for whites and greys.

- ## Kodac Gray Scale

  Alternative ICC profile for greyscale. Makes it possible to set the balance for whites and greys.

- ## Digital Color Checker

  Preferred ICC profile for colour digitisation.

Evaluation tools

- ## Open DICE

  Free software for measuring and analysing the technical criteria of scanners. This automatic control software uses several ISO standards to analyse the quality of images produced by scanners and the technical components of the imaging equipment.
  www.digitizationguidelines.gov/guidelines/OpenDICE/OpenDICE_manual_Command_v1.docx

- ## Auto SFR

  Free program developed to help imaging professionals determine the actual resolution of images and set the appropriate resolution for the documents to be scanned according to their type.
  http://www.digitizationguidelines.gov/guidelines/OpenDICE/AutoSFR_manual.pdf

- ## UTT

  Method for checking and controlling all the parameters of an image capture system, developed by the National Library of the Netherlands. This standardised test target – available in formats ranging from DIN A4 to DIN A0 – makes it possible to validate parameters such as: resolution, contrast, white balance, gain modulation, uniformity of light on the object, noise, colour rendition, geometric distortion or the parallelism of a capture solution.
  http://universaltesttarget.com/about.php

RECOMMENDATIONS

➡ **The adjustment of the illuminants must be carried out before the calibration** so as not to distort that process. Maximum permitted brightness: 32 lux..

➡ Recommended resolution:
- **300 DPI** for documents between DIN5 and DIN A2
- **400 DPI** for other formats

➡ Use of **Adobe RGB 1998 colour spaces** should be preferred for colour digitisation. For greyscale images, opt for **Gray Gamma 2.2**.

➡ To ensure the quality of reproduction of the tonality and hue of the original document, **the reference colour target should be scanned under the same conditions as the reproduced documents**.

➡ This **test target should be scanned each day** to ensure that the machine settings remain consistent.

09 / GENERAL RULES FOR IMAGE CAPTURE

RECOMMENDATIONS

- Documents must be scanned in full: **the image should reflect the entire document and include all pages** in the case of digitisation of a work.
- To certify that the document is complete, **leave a margin of 0.25 to 1 cm all around the document.** The larger the document, the larger this margin should be.
- If the outline of the document to be reproduced is irregular, leave a border wide enough to enclose the entire document on all sides.
- It is recommended that documents be scanned in **colour**.
- Documents should be **laid flat as much as possible**, within the constraints of the document, its condition and how precious it is.
- This flattening must be **uniform and homogeneous** to avoid any distortion of the document.
- The orientation of the document must be in line **with the reading direction**. If the format of the document does not allow this, it will need to be oriented in this direction during post-production retouching.
- For documents containing printed text, **the text line must be horizontal with a tolerance of 1.30 degre.**

- Focus should be given to **the straightness of the text** rather than the page to avoid compromising the document character recognition stage (OCR).

- Position **a neutral digitisation background** – grey, black or white – to increase the contrast and therefore the readability of the document.

- For bound documents with gaps, **position a matt cardboard cover in a neutral colour** to increase legibility and avoid compromising the OCR.

- To be able to correct errors identified during the procedure, **keep the RAW file** so that you can demosaic it again and correct the anomaly.

10 / CREATING AND FILLING THE METADATA INVENTORY FILE

KEY CONCEPT

The **inventory file** is a file containing the descriptive metadata of the digitised document in order to identify the author, volume, dimensions, conservation characteristics, and more generally, all essential elements required by digitisation operators in view of its dissemination and cataloguing.

The definition of these categories again depends **on the intended use of the digitised documents,** but also on whether a standardised metadata model is used. The categories of that model are already defined.

In the case of outsourcing, the inventory file must be created in advance and transmitted to the service provider. Once completed and submitted, this file should be integrity-checked to ensure the accuracy of the information created.

RECOMMENDATIONS

The inventory file must include at least the following information:

- The **descriptive categories** of the digitised objects (unique identifier, author, title, creation date, subject, dimensions, description, etc.)
- A **definition** of these categories

- The **degree of obligation of the metadata categories** (mandatory / optional / recommended)
- The **type of data** – quantity, date, controlled vocabulary, free text, etc.
- The **input rules for free text** and the reference system document for the metadata standard applied by the institution
- The **name of the encoder and the changes made** in order to track changes and identify potential reliability issues.

TOOL

- ## Open Refine

Open Refine is an open-source program that cleans and corrects a set of data, making it more consistent. It therefore enables the following:

- *The importing of data in different formats*
- *The application of cell transformation*
- *The processing of cells containing several values*
- *The creation of links between data sets*

https://openrefine.org/documentation.html

11 / SAVING AND NAMING OF FILES

KEY CONCEPT

Any digitisation project raises the question of the naming of the files that will be produced as a result of the operations. This problem resonates with the context of the Semantic Web, where one of the fundamental principles is the use of perennial identifiers.

An identifier is said to be perennial when it uniquely designates a resource and all the metadata associated with it, thus making it easier to find on the Web. This identifier – such as an URI, ARK or DOI – also guarantees the reuse of digital resources and their visibility by offering the possibility of linking them to other resources and thus increasing the number of reading paths or file paths.

Given the large number of documents to be managed in a digitisation project, it is not feasible to name the files individually. **The use of semi-automated programs allows mass management of file identifiers, regardless of their type, size or location.**

RECOMMENDATIONS

- Whatever the management program chosen, it is essential to define a file naming convention to facilitate digitisation work and reduce the risk of losing information.

There is no perfect way to structure and order the elements in a file name. The important thing is to be **consistent.** Some of the most common categories used to identify documents in cultural and scientific institutions are the following:

- Alphanumeric identifier.
- Document type.
- Document creation date (date format YYYYMMDD).
- Document purpose.
- Recipient in the case of a request for reproductions.
- Version.

Style rule to be observed:
- Separation of words in a capital or underscore field.

Prohibited characters:
- Accented characters.
- Punctuation marks.
- Full stops.
- Spaces.

It is preferable **to comply with the naming plan in force within your institution** or, if one does not yet exist, to define one that is specific to digitisation operations.

Guarantee the uniqueness of the identifiers.

In the event of deletion or depublication, provide information about the circumstances and, if necessary, **point to a substitute resource.**

Publish your **identifier management policy.**

TOOLS

File Renamer Basic

File Renamer Basic is a file renaming program for Windows. It can apply a naming format to all files in a folder and is able to batch edit the technical metadata of image files (EXIF).

https://download.cnet.com/File-Renamer-Basic/3000-2248_4-10306538.html

- ## Bulk Rename Utility

*Bulk Rename Utility is a file renaming program for **Windows**. It offers the same functionality as File Renamer Basic, but has been created specifically to support folders containing several thousand files.*

https://www.bulkrenameutility.co.uk/

- ## Name Changer

*Name Changer is an excellent alternative for **Mac.***

https://www.macupdate.com/app/mac/21516/namechanger

12 / RETOUCHING IMAGE FILES

KEY CONCEPT

Despite all the precautions taken during image capture, it is not uncommon to have to do some **retouching in post-production to refine the image produced or correct minor anomalies.** Not all the operations listed below are necessary. It is the responsibility of the person performing the checking – on the basis of the specifications or a summary document – to determine the elements to be corrected.

This reference document should include **all the parameters to be checked** at this stage and the results to be achieved for each type of document. This process-related information will then be integrated into the metadata accompanying the images, making it possible to reconstruct their history.

Retouching of image files may include:

- **Rotation of the image** to place it in the reading direction. This operation is particularly important in the case of a text document that is then to be scanned by OCR. The text line must be horizontal, even if the binding or document is not quite horizontal.
- **Trimming**, which involves removing the pixels outside the area of the scanned object.
- **Saturation correction**.
- **Noise reduction**.
- **Sharpening or softening of the image**.
- **Contrast correction**.
- **Gamma correction or modification of the luminance values o**f the pixels.

RECOMMENDATIONS

- Before retouching, make sure that the **ICC profile of the image processing software is the same as that of the scanner.**
- **Work on a copy of the scanned file**: it is not possible to cancel all operations easily.

TOOLS

- ### Photoshop

  Photoshop is probably the best known image retouching and processing software. Developed by Adobe, this software has a vast range of functionalities and is probably the most complete tool for working on a raster image.

- ### Lightroom

  Lightroom is an interesting alternative as it has been developed specifically for photo retouching. It also offers a collaborative editing mode, allowing several people to work on the same image if necessary.

- ### Gimp

  GIMP is a cross-platform image editor available for GNU/Linux, OS X, Windows and other operating systems. Although less sophisticated than the two programs mentioned above, it nevertheless has the advantage of being free and adaptable to all operating systems.

  https://www.gimp.org/

13 / OCR SCANNING OF TEXT DOCUMENTS

KEY CONCEPT

Optical Character Recognition (OCR) is an image processing technique aimed at extracting the text from an image to make it easier to read and to allow the user to perform searches easily.

The principle of OCR is divided into several steps, which include the following:

- **Deskewing** of the page.
- **Binarisation of the document**, namely conversion into black-and-white.
- **Segmentation of the image into text zones**, then lines, words and characters.
- And finally **classification of these zones** for recognition of characters and then words.

RECOMMENDATIONS

- To check the quality of the scan, **use a reference or ground truth** OCR document.
- To create a ground truth, perform a « manual » scan on an image document:
  - Binarise the image
  - Remove noise automatically and correct any errors or inaccuracies in this automatic correction manually
  - Remove the borders
  - Create text zones manually or semi-automatically

• Classify the zones according to their type (image, text, handwritten notation, drawing, etc.)
• Manually define the reading order between text blocks in the document
• Define the lines of text and manually enter a representative sentence, namely the most varied set of normal and special characters.

This reference document can then be used to assess, on the basis of factual criteria, the software's ability to recognise text zones, lines and characters, and in particular:
• Omissions
• Classification problems
• Segmentation problems, problems breaking down the whole image into sub-parts – lines, words, characters.

OCR checking can be automated using the LayoutEval software, which uses the reference document loaded to perform a rapid comparison of quality differences.

TOOLS

- ## ABBY FineReader

ABBY FineReader – Proprietary software
https://www.abbyy.com/media/6652/guide_english.pdf

- ## Tesseract

Tesseract – Open source software
https://ichi.pro/fr/guide-du-debutant-sur-tesseract-ocr-18608940018749

- ## Layout Evaluation

Layout Evaluation
https://www.primaresearch.org/tools/PerformanceEvaluation

14 / ENCAPSULATING METADATA

KEY CONCEPT

To be usable and distributable, any digital file must include a set of metadata describing the digital content as precisely as possible. These metadata can be external or internal but must be collected in a single document during the post-production phase, in order to optimise their conservation and use.

- Descriptive metadata include all the information describing the digital content – author, date of creation of the document, subject, etc. – and are recorded in a separate document.

- Technical and administrative metadata reflect all the information related to the process used to digitise the object: date of capture, model of scanner or camera used, colour space of the file, ICC profile, resolution, size in pixels, coding depth, etc. These metadata are usually internal and are automatically generated by the image capture equipment.

RECOMMENDATIONS

The **METS** (Metadata Encoding and Transmission Standard) standard developed by the Digital Library Federation makes it possible to express in detail any information related to the internal structure of documents. Based on the XML schema, this document makes it possible to group together and link all the elements related to the description of the object and the information concerning the methods used to create and use the digitised content. The METS file should therefore be seen as a metadata container, pointing to various related files. It has the advantage of being compatible with other metadata standards such as MARC, EAD and MODS.

A METS file has seven categories:

1. **The header**, describing the METS document itself.
2. **The descriptive metadata** or the path to the file containing this information.
3. **The administrative metadata**, which provide information about how the files were created and saved, and about the intellectual property rights of the original object.
4. **The files section**, which provides information about electronic versions of the digital object.
5. **The structure map**, which, as its name suggests, draws a hierarchical structure for the digital object, and links each element of this structure to the related content files and metadata.
6. **The structural links**, describing the existence of hyperlinks between elements within the structure map.
7. And finally **the behaviours**, which define how the different files linked to the METS document are run.

TOOLS

- General documentation and user guide for the METS standard
  http://www.loc.gov/standards/mets/METSOverview.v3_en.html
- Current outline of the METS standard
  http://www.loc.gov/standards/mets/mets.xsd
- METS file creation and management programs
  http://www.loc.gov/standards/mets/mets-tools.html
- Full guide to creating and processing METS files developed by the National Library of France
  https://www.bnf.fr/sites/default/files/2021-01/ref_num_metadonnees_mets.pdf

15 / FINAL QUALITY CONTROL

RECOMMENDATIONS

At this stage, **quality control concerns the digital images, but also all products included within the service:**

- Compliance of media and file formats.
- **Naming** of media, folders, files and images.
- **Completeness of the image files** in relation to the number of originals to be reproduced and/or the accompanying indexing.
- **Integrity of the indexing** requested on the basis of the input rules formulated.
- **Compliance of the technical characteristics of the image files** with the expected objectives, which must be assessed objectively.

➡ An exhaustive check of the results obtained is only possible when the number of documents to be scanned is limited. In most cases, **sampling is the preferred checking method.** The NF ISO 2859-1 standard provides a methodology for this type of checking.

➡ Checking by sampling is always carried out **in relation to the minimum quality threshold**, namely a maximum percentage of defective objects in a lot for that lot to be considered satisfactory.

➡ The verification of indicators is carried out **on the basis of errors considered major and minor.**

16 / CHOOSING YOUR DIGITAL DATA MANAGEMENT SYSTEM

KEY CONCEPT

Long-term preservation of electronic content is the series of actions, tools and methods implemented to collect, identify, select, classify, destroy and preserve electronic content, on a secure medium, with the aim of using it and making it accessible over time. Electronic archiving is not only about storing and sharing information, but must also make it possible to ensure the integrity and confidentiality of digitised content.

The choice of the system that will manage and/or preserve the digitised documents is therefore fundamental. First, the most appropriate type of system should be chosen according to the documents being archived:

- If the system is intended to manage live, editable documents, an electronic document management system (EDMS) should be chosen.
- On the contrary, if the documents are fixed and validated, it is better to use a legal electronic archiving system (EAS).

A digital safe is an additional element of an EAS. It is more common to speak of a digital safe component (CCFN), a term used in the NF Z 42-020 standard. Care should be taken to avoid confusing the two concepts: the digital safe component ensures the integrity and traceability of the digital objects deposited in it, while the EAS offers specific archival functionalities: format sustainability, document typology, life cycle management, archiving profiles, etc.

EDMS

ELECTRONIC DOCUMENT MANAGEMENT SYSTEM

OBJECTIVES:

- Centralise all documents, whatever their origin and format, in a single place.
- Harmonise processes to facilitate document management within an institution.
- Provide day-to-day document management in an institution.

FUNCTIONALITIES:

- Makes it possible to modify and/or destroy documents.
- Makes it possible to modify and/or destroy documents.
- May include an organised storage structure, under the control of users.
- Document digitisation (if paper), capture, indexing, filing, storage, access and dissemination.

EAS

ELECTRONIC ARCHIVING SYSTEM

OBJECTIVES:

- Ensure the access, confidentiality, integrity, durability and traceability of each document.
- Guarantee archiving with evidential value, namely archiving that ensures the long-term preservation of a document, in compliance with legal constraints and archiving standards.

FUNCTIONALITIES:

- Prevents the modification and/or destruction of documents.
- Guarantees mandatory monitoring of retention periods.
- Must include a rigorous filing structure.

RECOMMENDATIONS

When deploying an Electronic Archiving System, it is advisable **to set up an archiving policy that defines the associated legal, operational, technical and functional constraints and the necessary access rights.** This further strengthens the security function of the EAS.

EAS and EDMS are **complementary solutions** for the management of heritage archives: a EDMS facilitates the management of current documents and collaborative work in the context of internal digitisation projects, for example, while an EAS provides a permanent preservation solution for digitised content.

- Give preference to an EAS that complies with the NF-Z42-013 standard. Because of the completeness of the evidence file it provides (traceability of actions carried out, proof of integrity, existence of the document on a given date), an EAS guarantees the authenticity and compliance of the document.

TOOLS

- ## Standard AFNOR NF Z 42-013

Published by AFNOR in 1999 and revised in 2001, 2009 and most recently 2020, the NF Z 42-013 standard defines a series of functional, organisational and infrastructure requirements and recommendations for the design and operation of an electronic archiving system (EAS).

The text addresses the organisation based around the functions of electronic archiving, deposit, preservation, accessibility, communication and disposal management, EAS interoperability, retrieval and reversibility.

- ## Standard AFNOR NF Z 42-020

Published in 2012 by AFNOR, the Z 42-020 standard, referred to as « Functional specifications for a digital safe component intended for the preservation of digital information under conditions that guarantee its integrity over time ». This standard is limited to addressing the storage layer of the digital safe and does not cover issues such as authentication, confidentiality, metadata or interfacing with the business process layers.

- ## Standard ISO 14641-1

Based entirely on the NF Z42-013 standard, the international ISO 14641-1 standard also defines the specifications for implementing an electronic archiving system. Its evolution is, however, independent of the AFNOR standard.

- ## Standard ISO 14721

Initially published in 2003 and revised in 2012, the ISO 14721 standard is a transposed version of the OAIS (Open Archive Information System), a conceptual model intended for the management, archiving and long-term preservation of digital documents.

- ## Standard ISO 15489-1

Revised in 2016, this standard details the concepts and principles for the management of documents of evidential value for natural or legal persons, both private and public, and defines the characteristics that give a document its authoritative value, its evidential value: authenticity, integrity, reliability and usability.

BIBLIOGRAPHY

ANCTIL, M.-C, LEGENDRE, M., MÜLLER, T., MAILLET, D., BROSSEAU, K., RENAUD, L., *Recueil de règles de numérisation*, Bibliothèque et Archives nationales du Québec, Bibliothèque nationale de France et Musée canadien de l'histoire, 2014, [Online], **http://collections.banq.qc.ca/bitstream/52327/2426216/1/4671601.pdf**

ACCART, J.-P., *Le Métier de Documentaliste*. Avec la collaboration de Réthy Marie-Pierre. Éditions du Cercle de la Librairie, 2015.

BACHI, V., FRESA, A., PIEROTTI, C., PRANDONI, C., *The digitization age: mass culture is quality culture. challenges for cultural heritage and society*. In: Ioannides, M., Magnenat-Thalmann, N., Fink, E., Žarnić, R., Yen, A.-Y., Quak, E. (eds.) EuroMed 2014.

BAROUCH, G., *Révolution des pratiques managériales : le changement par la qualité*, La Plaine Saint-Denis : Afnor Éditions, 2013.

BIBLIOTHÈQUE ET ARCHIVES NATIONALES DU QUÉBEC., *La numérisation des documents administratifs : Méthodes et recommandations : version revue et corrigée*, Montréal : Bibliothèque et Archives nationales du Québec, mai 2012, 35 p., [Online], **https://docplayer.fr/79033783-La-numerisation-des-documents-administratifs.html**

BIBLIOTHÈQUE NATIONALE DE FRANCE., *Techniques et formats de conversion en mode texte* | BnF - Site institutionnel, n.d, [Online], **https://www.bnf.fr/fr/techniques-et-formats-de-conversion-en-mode-texte**

BIZINGRE, J., PAUMIER, J., RIVIÈRE, P., *Les référentiels du système d'information*, Paris : Dunod, 2013.

BOUTET, A. and ROUDAUT, K., « Les enjeux de la numérisation et de l'ouverture d'archives : le point de vue des professionnels » *In*: *Terminal*, 110 | 2012, p. 27-37 pp.

BRUCE, T. and HILLMANN, D., *The Continuum of Metadata Quality: Defining, Expressing, Exploiting*, 2004, [Online], **https://www.researchgate.net/publication/247818823_The_Continuum_of_Metadata_Quality_Defining_Expressing_Exploiting**

BURESI, C., and Cédelle-Joubert, L., *Conduire un projet de numérisation*, Paris : Tec et Doc, 2002, 326 p.

CARBONE, F., OOSTERBEEK, L., COSTA, F., *Extending and adapting the concept of quality management for museums and cultural heritage attractions: A comparative study of southern European cultural heritage managers' perceptions*, Tour Manage Perspect, 2020, 10 p.

CARON, P., DEFIOLLE, R., and Lay, M.-H., *L'enjeu des métadonnées dans les corpus textuels : Un défi pour les sciences humaines*, Presses universitaires de Rennes, 2019, 206 p.

CHARDONNENS, A., *La gestion des données d'autorité archivistiques dans le cadre du Web de données*, [Thèse de doctorat non publiée], Université Libre de Bruxelles, 2021, 420 p, [Online], **https://difusion.ulb.ac.be/vufind/Record/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/315804/Holdings**

CHAUMIER, J., *Document et numérisation. Enjeux techniques, économiques, culturels et sociaux*, Paris : AdBS éditions, 2006, 119 p.

CHIRON, G., DOUCET, A., COUSTATY, M., VISANI, M., and MOREUX, J.-P., « Impact of OCR Errors on the Use of Digital Libraries: Towards a Better Access to Information », *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 2017, pp. 1-4.

CLAERR, T., WESTEEL, I., and MELOT, M., *Manuel de la numérisation*, Paris : Éd. du Cercle de la, 2011. 317 p.

CLAERR, T., WESTEEL, I., *Numériser et mettre en ligne*, Villeurbanne : Presses de l'Enssib, 2010, [Online], **http://books.openedition.org/pressesenssib/414**.

CLAERR, T., WESTEEL, I., ORY-LAVOLÉE, B., GEORGES, N., *Manuel de constitution de bibliothèques numériques*, Paris : Éd. du Cercle de la librairie, 2013. 407 p.

CHEVRY, E., *Stratégies numériques*, Paris : Hermès Science publications-Lavoisier, 2011, 272 p.

COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS, *Comment permettre à l'homme de garder la main? Les enjeux éthiques des algorithmes et de l'intelligence artificielle. Synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la Loi pour une république numérique*, December 2017, [Online], **https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf**

CONSEIL CANADIEN DES ARCHIVES. *Manuel de conservation des documents d'archive, Chapitre 3 – Environnement*, 2003, [Online], **http://www.cdncouncilarchives.ca/RBch3_fr.pdf**

DANGERFIELD, M.-C. and KALSHOVEN, L., *Report and Recommendations from the Task Force on Metadata Quality*, May 2015, 54 p., [Online], **http://pro.europeana.eu/files/Europeana_Professional/Publications/Metadata%20Quality%20Report.pdf**

DIRECTION DES ARCHIVES DE FRANCE. MINISTÈRE DE LA CULTURE ET DE LA COMMUNICATION - COMITÉ DE PILOTAGE NUMÉRISATION – MISSION DE LA RECHERCHE ET DE LA TECHNOLOGIE (DDAI). *Écrire un cahier des charges de numérisation du patrimoine. Guide technique. Documents reliés- manuscrits- plans- dessins- photographies- microformes*, Paris : Archives de France, 2008, 62 p., [Online], **https://francearchives.fr/file/bf50d8fa5f554586dbf18fdc862d25970a1da0a7/static_4132.pdf**

DIRECTION INTERMINISTÉRIELLE DU NUMÉRIQUE ET DU SYSTÈME D'INFORMATION ET DE COMMUNICATION DE L'ETAT. *Référentiel général d'interopérabilité*, 2015, version 2, [Online], **https://www.numerique.gouv.fr/uploads/Referentiel_General_Interoperabilite_V2.pdf**

DUFRÊNE, B., IHADJADENE, M., DENIS, B., BRUCKMANN, D., BARBIER, B., et RACINE, B., *Numérisation du patrimoine : quelles médiations ? quels accès ? quelles cultures ?*, Paris : Hermann, 2013, 311 p.

D'HOORE, M., *La numérisation du patrimoine écrit : du virtuel à la réalité*, Bruxelles : Archives et bibliothèques de Belgique, 2011, 280 p.

EAD, *EAD en bibliothèque, guide des bonnes pratiques*, *Origine - origination*. 2020, [Online], **https://www.ead-bibliotheque.fr/guide/donnees-du-did/origination/**

EZRATI, J.-J., *Éclairage d'exposition, musée et autres espaces*, Paris : Eyrolles, 2015.

FEHAU E., LE DANTEC N., *Vade-mecum de la conservation préventive*, C2RMF, version of 18 November 2013, [Online], **https://c2rmf.fr/sites/c2rmf.fr/files/vademecum_cc.pdf**

FEYLER, F., « De la compatibilité à l'interopérabilité en matière de repérage d'information pertinente. La problématique et l'exemple d'OTAREN », In : *Documentaliste-Sciences de l'Information*, vol. 44, no. 1, 2007, pp. 84-92.

FLECKENSTEIN, M., FELLOWS, L., « Data Quality », In: *Modern Data Strategy*, Cham: Springer, 2018.

FONDS INDÉPENDANT DE PRODUCTION. *Êtes-vous repérables? Guide pratique pour documenter vos contenus*, November 2017, [Online], **http://ipf.ca/FIP/ressources/etes-vous-reperables/#1512498362849-3b320f4f-a27f**

FONDS DES MÉDIAS DU CANADA. *Découvrabilité : vers un cadre de référence commun*, 2016, [Online], **https://trends.cmf-fmc.ca/fr/rapports-de-recherche/decouvrabilite-vers-un-cadre-de-reference-commun**

FORCE 11, *The Fair Data Principles*. N.d. [Online], **https://www.force11.org/group/fairgroup/fairprinciples**

GIESEN, E., *Démarche qualité et norme ISO 9001 : Une culture managériale appliquée à la recherche*, Marseille : IRD Éditions, 2018, [Online], **http://books.openedition.org/irdeditions/20609**

GILLET, F., HUNGENAERT, J., HODZA, M., et al., *Identifying Needs to Modernize Access to Digital Data in Libraries and Archives (MADDLAIN) : final report*, Brussels : Belgian scientific Policy, 2018, [Online], **http://www.belspo.be/belspo/brain-be/projects/FinalReports/MADDLAIN_%20final%20report.pdf**

GUEGUEN, G., MARQUES DA FONSECA, V.-M., D., PITTI, V. et SIBILLE, C., *Vers un modèle conceptuel international pour la description archivistique. Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description*, 2013, [Online], **https://www.ica.org/sites/default/files/EGAD_French.pdf**

INTERNATIONAL CENTRE FOR THE STUDY OF THE PRESERVATION AND RESTORATION OF CULTURAL PROPERTY (ICCROM), *A Guide to Risk Management of Cultural Heritage*, 2016, [Online], **https://www.iccrom.org/sites/default/files/2017-12/risk_management_guide_english_web.pdf**

INTERNATIONAL CENTRE FOR THE STUDY OF THE PRESERVATION AND RESTORATION OF CULTURAL PROPERTY (ICCROM), *The ABC Method. A risk management approach to the preservation of cultural heritage*, 2016, [Online], **https://www.iccrom.org/sites/default/files/2017-12/risk_manual_2016-eng.pdf**

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO), *Information and Documentation: Part 1: Guidelines*, Geneva: ISO, ISO 15489-1: 2001, 2001.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO), *Information and Documentation: Part 2: Practical Guide*, Geneva: ISO, ISO 15489-2: 2001, 2001.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO), *Photography - Sensitometric Illuminants - Specifications for Daylight, Artificial Light and Shutte*r, Geneva: ISO, ISO 7589:2002, 2002.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO), *Photography - Electronic scanners for photographic images - Dynamic interval measurements*, Geneva: ISO, ISO 21550: 2004, 2004.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO), *Document Management : Electronic Document File Format for Long-term Preservation: Part 1: Use of PDF 1.4 (PDF/A-1)*, Geneva: ISO, ISO 19005-1, 2005.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO), *Photography - Electronic Cameras - Methods of Measurement of Optoelectronic Conversion Functions*, Geneva: ISO, ISO 14524: 2009, 2009.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO), *Space data and information transfer systems - Open archival information system (OAIS) - Reference model*, Geneva: ISO, ISO 14721: 2012, 2012.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO), *Quality management systems - Fundamentals and vocabulary*, Geneva: ISO, ISO 9001:2015, 2015.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO), *Information and documentation - Records management - Part 1: Concepts and principles*, Geneva: ISO, ISO 15489:2016, 2016.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO), *Quality management systems - Requirements*, Geneva: ISO, ISO 9004:2018, 2018.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO), *Quality management - Quality of an organization - Guidance to achieve sustained success*, Geneva: ISO, ISO 9004:2018, 2018.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO), *Electronic document management - Design and operation of an information system for the preservation of electronic documents - Specifications*, Geneva: ISO, ISO 114641:2018, 2018.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (ISO), *Information and documentation - The Dublin Core set of metadata elements - Part 2*, Geneva: ISO, ISO 15836-2:2019, 2019.

INSTITUT DE CONSERVATION DU CANADA (ICC), *Plan de préservation des collections patrimoniales : Stratégies pour éviter ou limiter les dommages*, N.d., [Online], **https://www.canada.ca/fr/institut-conservation/services/conservation-preventive/plan-preservation-collections-patrimoniales.html**

INSTITUT ROYAL DU PATRIMOINE ARTISTIQUE (IRPA), *Cellule de conservation préventive* | KIK-IRPA, N.d., [Online], **https://www.kikirpa.be/fr/conservation-restauration/cellule-conservation-pr%C3%A9ventive**

KULESZ, O., *Culture, machines et plateformes : L'intelligence artificielle et son impact sur la diversité des expressions culturelles*, Comité intergouvernemental pour la protection et la promotion de la diversité des expressions culturelles, Douzième session, Paris, siège de l'UNESCO, 11-14 décembre 2018, DCE/18/12.IGC/INF.4, 13 November 2018, [Online], **https://en.unesco.org/creativity/sites/creativity/files/12igc_inf4_fr.pdf**

LU, T., *Homogeneity models for image processing in the cultural heritage sector*, [Unpublished PhD], Vrie Universiteit Brussel, 2020, 245 p.

MANDERLKERN, D., *Diffusion des données publiques et révolution numérique*, Rapport rédigé à la suite des travaux de l'Atelier présidé par D. Mandelkern au Commissariat au Plan, Paris : La Documentation Française, 1999, 124 p.

MEMON, J., SAMI, M., KHAN R.-A., and UDDIN M., « Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR) », *In* : *IEEE Access*, vol. 8, pp. 142642-142668, 2020.

MICKELVEY, F., et HUNT, R., *Responsabilité algorithmique et découverte de contenus numériques, document de réflexion, Rencontre internationale sur la diversité des contenus à l'ère numérique*, Patrimoine canadien, February 2019, [Online], **https://fr.readkong.com/page/responsabilit-algorithmique-et-d-couverte-de-contenus-2696781**

MINISTÈRE DE LA CULTURE ET DE LA COMMUNICATION, *Feuille de route stratégique : Métadonnées culturelles et transition Web 3.0*, janvier 2014, 38 p. [Online], **https://www.enssib.fr/bibliotheque-numerique/documents/64776-feuille-de-route-strategique-metadonnees-culturelles-et-transition-web-3-0.pdf**

NAPOLI, P., *Diversité de contenus à l'ère numérique : découvrabilité de contenu diversifié aux échelons local, régional et national, document de réflexion, Rencontre internationale sur la diversité des contenus à l'ère numérique*, Patrimoine canadien. Février 2019, [Online], **https://culturenumeriqc.qcnum.com/wp-content/uploads/2019/03/Napoli-De%CC%81couvrabilite%CC%81-de-contenu-diversifie%CC%81-aux-e%CC%81chelons-local-re%CC%81gional-et-national.pdf**

PLAN D'EXPLOITATION ET DE PRÉSERVATION DES PATRIMOINES (PEP'S), *Normes et lignes directrices techniques et organisationnelles pour les initiatives de numérisation des patrimoines culturels soutenues par la Communauté française*, ministère de la communauté française Bruxelles : Secrétariat général – Préservation et exploitation des patrimoines, 2009, 61 p., [Online], **http://www.peps.cfwb.be/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/sites/numpat/upload/numpat_super_editor/numpat_editor/documents/CFWB/Normes.pdf&hash=a2d0e740e46e34b8c3475c604c3e96edbb8a8712**

OBSERVATOIRE DE LA CULTURE ET DES COMMUNICATIONS DU QUÉBEC. *État des lieux sur les métadonnées relatives aux contenus culturels*, Québec : Institut de la statistique du Québec, Observatoire de la culture et des communications du Québec, 2017, 118 p., [Online], **https://statistique.quebec.ca/fr/fichier/etat-des-lieux-sur-metadonnees-relatives-aux-contenus-culturels.pdf**

PLAMONDON, J., *Bien documenter pour favoriser la découverte en ligne. Travailler avec des données*, 2019, [Online], **https://espaceschoregraphiques2.com/wp-content/uploads/2019/12/Guide-me%CC%81tadonne%CC%81es-FJPP.pdf**

PLAMONDON, J., *Données ouvertes et liées : le Web comme base de données*, March 2018, [Online], **https://joseeplamondon.com/donnees-ouvertes-liees-web-base-de-donnees/**

PLAMONDON, J., *Etes-vous repérables ? Documenter vos contenus pour qu'ils soient compris par des moteurs de recherche*, 2019, [Online], **http://ipf.ca/wp-content/uploads/sites/2/2019/05/guide-documentation-webseries_20171128_vf-maj_20191405.pdf**

PLAMONDON, J., *Produire des données : entre outils de marketing et bases de connaissances*, 21 August 2019, [Online], **https://joseeplamondon.com/produire-donnees-culture-marketing-ou-bases-connaissances**

RIEGER, T., *Federal Agencies Digital Guidelines Initiative. Technical Guilines for Digitizing Cultural Heritage Material. Creation of Raster Image File*, 2016, 99 p. [Online], **http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf**

SALAÜN. J.-M., *Bibliothèques numériques et Google-Print*, Regards sur l'actualité : mensuel de la vie publique en France, La Documentation Française, 2005, [Online], **https://archivesic.ccsd.cnrs.fr/sic_00001576/document**

SIBILLE, C., Avantages et inconvénients d'un encodage en EAD, In: *La Gazette des archives*, n°220, 2010-4. Les instruments de recherche : évolutions, publics et stratégies. pp. 179-194.

SOYEZ, S., *La numérisation en marche : les étapes de la dématérialisation des processus de travail*, Bruxelles : Archives générales du Royaume et Archives de l'État dans les Provinces, n.d., [Online], **http://www.arch.be/docs/brochures/la_numerisation_en_marche.pdf**

TCHÉHOUALI, D., « Note d'orientation pour une étude sur les enjeux, défis et opportunités de la découvrabilité en ligne des contenus culturels francophones », *In:* Dispositif d'observation des dynamiques culturelles et linguistiques, Comité d'orientation des Dynamiques culturelles et linguistiques, April 2018, [Online], **https://www.francophonie.org/sites/default/files/2020-01/perspectives_de_recherche_web-2.pdf**

TESSIER, M., FERRY, L., RACINE, B., JEANNENEY, J.-N., *La révolution du livre numérique : état des lieux, débats, enjeux*, Paris : Odile Jacob, 2011. 208 p.

THIAULT, F., « Mutations des métiers de l'information-documentation : vers l'émergence d'une culture de l'information numérique », *In*: *Les Cahiers de la SFSIC*, Société française des sciences de l'information et de la communication, 2012, p. 59-62.

VAN DORMOLEN, H., *Metamorphoze Preservation Imaging Guidelines. Image Quality*, 2012, version 1.0, [Online], **http://www.imagingetc.com/images/Resources_Images/PDFs_DownloadFiles/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf**

VILLANUEVA, E., et SHIRI, A., « Methodological Diversity in the Evaluation of Cultural Heritage Digital Libraries and Archives: An Analysis of Frameworks and Methods / Diversité méthodologique dans l'evaluation des bibliothèques et les archives numériques du patrimoine culturel : Une analyse des cadres et des méthodes », *In*: *Canadian Journal of Information and Library Science*, vol. 43 no. 3, 2021, pp. 316-342.