

Faculté de philosophie, arts et lettres

Numérisation durable aux archives de la Commission européenne.

Réduire l'empreinte du stockage des données par le choix des formats et de la compression.

Auteur : Martin Klassen
Promotrice : Bérangère Piret
Année académique 2024-2025
Master de spécialisation en archivistique, droit et gestion des données

Table des matières

Remerciements	3
Introduction	4
Numérisation et durabilité	4
Problématique et méthodologie	10
La numérisation aux Archives Historiques de la Commission européenne	12
Analyse des besoins et comparaisons des formats	15
Les besoins des utilisateurs	15
Les besoins de la préservation numérique.....	17
Comparaison entre les formats de données et les algorithmes de compression	20
Méthodologie.....	20
Résultats et recommandations pour une numérisation durable	21
Conclusion.....	30
Bibliographie	33
Sources et travaux.....	33
Travaux :	33
Ressources en ligne :	37
Annexe.....	41

Remerciements

Je tiens à remercier chaleureusement l'ensemble du Service des Archives historiques de la Commission européenne pour leur accueil et le temps qu'ils m'ont consacré afin de répondre à mes questions, tant au cours de la rédaction de ce mémoire qu'en amont.

Dans le cadre de ce travail, je souhaite tout particulièrement remercier Lieven Baert pour m'avoir accordé des entretiens et pour la qualité de ses réponses, tant à l'oral que par écrit, ainsi qu'Alexis Bigonville pour son aide précieuse dans la sélection des échantillons d'archives numérisées et pour le temps qu'il m'a consacré afin de me présenter les procédures de numérisation du SAH.

Enfin, je tiens à remercier Héloïse Cautere, Laura Nisi, et Teodora Tarnovetchi, membres de l'équipe de la préservation numérique, ainsi que l'ensemble des assistants archivistes employé par la Commission que j'ai côtoyés pour toute leur aide.

Introduction

Numérisation et durabilité

La numérisation est une pratique largement répandue au sein des organisations européennes qui conservent ou manipulent des documents patrimoniaux. Bien qu'il n'existe pas de données précises et actualisées à l'échelle européenne sur le nombre de musées, bibliothèques et centres d'archives (désignés ci-après comme *organisations culturelles*) engagés dans des activités de numérisation, l'enquête *ENUMERATE Core Survey 4* (2017, $n = 983$) indique que 82 % d'entre elles déclarent posséder une collection numérique ou mener des projets de numérisation¹. Cette proportion, stable par rapport aux éditions précédentes de l'enquête (2012, 2014 et 2015), n'a pu que s'accroître, du moins en quantité d'objets numériques, grâce au soutien actif de la Commission européenne, notamment à travers ses recommandations 2011/711/EU² et 2021/1970/EU³, ainsi qu'au déploiement de nombreux projets de financement et à la création d'un espace de données commun pour le patrimoine culturel⁴.

Selon une enquête récente menée par Michael Friedewald, Ivan Svékely et Murat Karaboga auprès de centres d'archives situés dans des pays membres du Conseil de l'Europe (2024, $n = 62$), les deux principales raisons invoquées pour justifier les efforts de numérisation sont la protection et la conservation des documents fragiles (81 %), ainsi que l'amélioration de l'accessibilité des archives (79 %), souvent en réponse aux demandes des utilisateurs⁵. Une autre enquête sur la numérisation en Europe, réalisée par les Archives

¹ JAN NAUTA G., TEUNISSE S., VAN DEN HEUVEN W., *Access to Digital Ressources of European Heritage : Report on ENUMERATE survey 4*, Europeana DSI 2, Deliverable D4.4, 31 août 2017. Europeana PRO, [en ligne], https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_DSI-2/Deliverables/d4.4-report-on-enumerate-core-survey-4.pdf.

² Commission Recommendation of 27 October 2011 on the digitisation and online accessibility of cultural material and digital preservation (OJ L 283 29.10.2011, p. 39, ELI: <http://data.europa.eu/eli/reco/2011/711/oj>).

³ Commission Recommendation (EU) 2021/1970 of 10 November 2021 on a common European data space for cultural heritage (OJ L 401 12.11.2021, p. 5, ELI: <http://data.europa.eu/eli/reco/2021/1970/oj>).

⁴ European Commission: Directorate-General for Communications Networks, Content and Technology, *The future of Europe's past – Why Member States must do more to advance digitisation of our cultural heritage – Implementation of the 2021 Commission recommendation on a common European Data Space for Cultural Heritage – Progress report 2021-2023*, Publications Office of the European Union, 2024, <https://data.europa.eu/doi/10.2759/5921820>. Pour plus d'informations sur la politique européenne, voir CAPURRO C., *Digitizing european culture : Europeana and the government of digital cultural heritage*, thèse en vue de l'obtention du grade de docteur en philosophie, Utrecht University, Utrecht, 2021.

⁵ FRIEDEWALD M., SZÉKELY I. et KARABOGA M., « Preserving the Past, Enabling the Future: Assessing the European Policy on Access to Archives in the Digital Age », in *Preservation, Digital Technology & Culture*, vol. 53 (1 juillet 2024), n° 2, p. 63-65. Le rapport complet de cette enquête est disponible sur *CULTURE AND CULTURAL HERITAGE : NEWSROOM*, « Access to archives in the digital age », *Conseil de l'Europe*, [en ligne], [Access to archives in the digital age - Culture and Cultural Heritage](https://www.culture-heritage.eu/Access-to-archives-in-the-digital-age-Culture-and-Cultural-Heritage), publié le 18 avril 2024, consulté le 25 juin 2025.

nationales de Finlande auprès d'institutions culturelles majoritairement issues du secteur public européen (2017, $n = 40$), indique que quatre institutions disposaient d'une politique autorisant l'élimination des originaux après numérisation, et que six autres étaient en train d'en élaborer une⁶.

Ainsi, dans de nombreux projets de numérisation, les copies numériques doivent être en mesure de se substituer aux originaux, que ce soit à la suite d'une élimination intentionnelle ou d'une perte accidentelle. La numérisation ne se limite donc pas à la création d'une copie numérique à partir d'un objet analogique ou de données numériques stockées sur un support externe ; elle implique également la préservation de ces copies sur le très long terme.

Pour répondre à ce besoin, des outils intellectuels et conceptuels ont été développés afin de garantir que ces substituts puissent être considérés comme des copies authentiques des originaux. Cela implique la préservation de leur identité et de leur intégrité, ainsi que la transmission de leur authenticité sur le long terme⁷. En conséquence, depuis les années 1990, de nombreuses normes, modèles, lignes directrices et recommandations ont vu le jour pour accompagner les organisations dans leurs projets de numérisation et en assurer la pérennité.

Dans un réflexe naturel d'archiviste, ces standards ont souvent été conçus et mis en œuvre selon une approche maximaliste, visant une authenticité et une intégrité optimales des copies numériques. Cela se traduit notamment par une numérisation à la meilleure qualité possible, accompagnée de systèmes de préservation numérique intégrant des vérifications régulières de l'intégrité (fixité) et une redondance des données⁸. Après tout, n'est-ce pas là la mission première de l'archiviste ? Conserver les archives dans les meilleures conditions possibles, afin que les générations futures puissent y accéder et comprendre leur passé ?

Mais cette vision entraîne des conséquences qui peuvent compromettre la durabilité même de la mission que se sont donnée les organisations culturelles. En effet, la numérisation

⁶ HIRVONEN L., « Survey of Digitization in Archives », in *The National Archives of Finland*, publié le 19 décembre 2017, [en ligne], https://kansallisarkisto.fi/documents/141232930/150411434/Liite_2_Digitization_Survey_2017.pdf, p. 12-13.

⁷ Pour plus d'informations sur ces notions d'authenticité, d'identité et d'intégrité, voir DURANTI L., « Pour une diplomatie des documents électroniques », in *Bibliothèque de l'École des chartes*, vol. 161 (2003), n° 2, p. 603-623. Voir aussi le glossaire d'InterPARES : « Terminology Database », in *InterPARES Trust*, [en ligne], <http://interparestrust.org/terminology>, consulté le 29/07/2025.

⁸ PENDERGRASS K.L., SAMPSON W., WALSH T. et ALAGNA L., « Toward Environmentally Sustainable Digital Preservation », in *The American Archivist*, vol. 82 (1 mars 2019), n° 1, p. 180-184. Pour plus d'informations sur les pratiques de fixité des centres d'archives, voir NATIONAL DIGITAL STEWARSHIP ALLIANCE, « 2021 Fixity Survey Report », octobre 2021, [en ligne], <https://osf.io/2qkea/>.

et la préservation numérique sont des activités particulièrement coûteuses, mobilisant des ressources et des compétences qui ont longtemps fait défaut dans les organisations culturelles. Elles reposent sur les technologies de l'information et de la communication (TIC), lesquelles consomment de l'énergie, requièrent un matériel informatique souvent onéreux, et sont soumises à une obsolescence rapide ainsi qu'à une dégradation matérielle inévitable.

Pour les utiliser, il faut non seulement mobiliser les compétences traditionnelles de l'archiviste, mais aussi des compétences spécifiques en matière de numérisation, de gestion des données et métadonnées, d'informatique appliquée aux archives, de cybersécurité, de gestion de projet et de planification, et de veille technologique et normative dans un domaine en constante évolution. Or, pour garantir l'authenticité des objets numériques sur le très long terme, ces compétences et les ressources associées doivent être maintenues de manière continue, potentiellement indéfiniment.

Cependant, c'est justement ces deux aspects, les ressources financières et les compétences spécialisées, qui représentent les plus grands défis pour les organisations culturelles, en particulier celles de petite taille ou disposant de ressources limitées⁹. Plusieurs enquêtes menées auprès de professionnels en témoignent. Par exemple, selon une récente enquête mondiale menée par *InterPARES Trust* sur l'utilisation de l'intelligence artificielle pour la numérisation ($n = 219$), 58 % des organisations interrogées (principalement européennes et nord-américaines) identifient le manque de financement, de temps et de compétences comme les trois principaux obstacles auxquels elles sont confrontées¹⁰. Selon une autre enquête sur la numérisation menée par le *Canadian Research Knowledge Network* (CRKN, $n = 44$) auprès de ses membres, 55,26 % des répondants indiquent que le manque de financement constitue leur principal défi (rang 1) en matière de numérisation. Au total, 84,21 % l'ont classé parmi leurs trois principaux défis (sur une échelle de 1 à 10, où 1 représente le défi le plus important). Le manque de compétences, quant à lui, a été classé parmi les trois premiers défis par 56,71 % des répondants¹¹. De manière générale, une revue de la littérature par Pandey R. et Kumar V. démontre que ces limites sont observées

⁹ RINEHART A.K., PRUD'HOMME P.-A. et HUOT A.R., « Overwhelmed to action: digital preservation challenges at the under-resourced institution », in *OCLC Systems & Services*, vol. 30 (4 février 2014), n° 1, p. 28-42.

¹⁰ BUSHEY J., DE SOUZA M., FUKUSHIMA K., IGLÉSIAS D., RIESS M., SENGSAVANG E., STANCIC H., TRBUSIC Z., *Report on the Survey "Digitization and Artificial Intelligence for Archives and Documentary Heritage Materials"*, InterPARES Trust AI, mai 2025, https://interparestrustai.org/assets/public/dissemination/RA03-InterPARESAI-Survey_Report_FINAL.pdf, p. 8.

¹¹ « Survey on Digitization, Access, and Digital Preservation », CRKN, 27 juillet 2020, https://www.crkn-rcdr.ca/sites/crkn/files/2020-10/HCPGTG%20Survey_EN%20%28Member%20responses%29.pdf, p. 18.

mondialement¹², et la tendance vers toujours plus d'austérité, souvent au détriment du monde culturel, n'annonce pas que cette situation s'améliorera dans les années à venir.

Et il ne faut pas négliger un autre aspect tout aussi crucial, longtemps resté en marge des débats sur la durabilité des projets de numérisation et de préservation numérique : le coût environnemental associé à la conservation des objets numériques, ainsi que les risques à long terme que le changement climatique pourrait faire peser sur la sauvegarde du patrimoine numérique¹³. Comme le souligne Ben Goldman, « by ignoring the environmental impact of our work we also put the long-term preservation of our holdings at risk »¹⁴. Cette inquiétude est également partagée par Keith L. Pendergrass, Walker Sampson, Tessa Walsh et Laura Alagna, qui constatent que :

*Cultural heritage organization (CHOs) rely on a stable society to fulfil their missions : from researchers who make use of collections, to funding sources, to collecting activities, CHOs' success is built on the social and physical constructions of a stable society. Climate change threatens this stability*¹⁵.

Comme mentionné précédemment, la numérisation et la préservation numérique reposent sur les TIC pour conserver et rendre accessibles les objets numériques. Or, l'usage collectif de ces technologies représentait en 2019 environ 3,7 % des émissions mondiales de CO₂, selon *The Shift Project*¹⁶, soit davantage que le secteur de l'aviation, estimé à environ 2,5 %¹⁷. Ce chiffre reste cependant partiel, car il ne tient pas compte d'autres aspects tel que l'utilisation de l'eau pour refroidir les *data centers*, ou les pollutions générées par les déchets électroniques. Étant donné que ces émissions sont directement corrélées au stockage et à la

¹² PANDEY R. et KUMAR V., « Exploring the Impediments to Digitization and Digital Preservation of Cultural Heritage Resources: A Selective Review », in *Preservation, Digital Technology & Culture*, vol. 49 (2020), n° 1, p. 26-37.

¹³ Ironiquement, pendant l'écriture de ce mémoire, la conférence organisée par le *Digital Preservation Outreach & Education Network* organisée le 25 juin 2025, intitulée « Climate change, archives, and digital preservation », a dû être annulée en conséquence de la vague de chaleur qui touchait à ce moment-là le Midwest et la Côte est des États-Unis. La chaleur a eu en effet pour conséquence une surcharge du réseau électrique et la surchauffe des équipements électroniques de la présentatrice.

¹⁴ GOLDMAN B., « It's Not Easy Being Green(e): Digital Preservation in the Age of Climate Change », in WEIDEMAN C. et CALDERA M.A. (dir.), *Archival Values: Essays in Honor of Mark Greene*, Chicago, Society of American Archivists, 2019, p. 279.

¹⁵ PENDERGRASS K.L., SAMPSON W., WALSH T. et ALAGNA L., « Toward Environmentally Sustainable Digital Preservation », *op. cit.*, p. 166.

¹⁶ FERREBOEUF H. (dir.), « Lean ICT : Toward Digital Sobriety », in *The Shift Project*, mars 2019, https://theshiftproject.org/app/uploads/2025/02/Lean-ICT-Report_The-Shift-Project_2019.pdf, p. 4. Freitag C. et al. placent quant à eux leurs estimations entre 2,1 et 3,9 % en 2021. FREITAG C., BERNERS-LEE M., WIDDICKS K., KNOWLES B., BLAIR G. S., FRIDAY A., « The real climate and transformative impact of ICT : A critique of estimates, trends, and regulations », in *Patterns*, vol. 3 (12 août 2022), n°8, p. 3.

¹⁷ RITCHIE H., « What shares of global CO₂ emissions come from aviation? », in *Our World in Data*, [en ligne], <https://ourworldindata.org/global-aviation-emissions>, consulté le 25 juin 2025.

diffusion de données numériques, les organisations culturelles détentrices d'objets numériques portent également une part de responsabilité dans cette pollution.

On pourrait objecter que la quantité de données conservées par les organisations culturelles ne représente qu'une infime fraction des 181 zettaoctets de données qui devraient être créés dans le monde en 2025 (seul 2 % sont conservés réellement)¹⁸. Toutefois, plusieurs éléments méritent d'être soulignés.

D'abord, l'un des objectifs fondamentaux des centres d'archives est précisément de préserver les données qu'ils détiennent pour des durées indéterminées. Cela signifie que, sauf changement structurel, le volume des données archivées ne fera qu'augmenter avec le temps. On peut même spéculer qu'à l'horizon de 100 ans, les organisations culturelles figureront parmi les entités les plus consommatrices d'espace de stockage numérique dans le monde.

Cette consommation est d'ailleurs déjà significative. À titre d'exemple, les 116 organisations ayant participé au *Fixity Survey* de 2021, mené par la *National Digital Stewardship Alliance*, conservaient à elles seules environ 145 pétaoctets de données¹⁹. Par ailleurs, malgré leurs critiques à l'égard d'une conservation numérique excessive, Moss M. et Gollins T.J. estiment qu'avec les pratiques actuelles d'évaluation archivistique et les évolutions des typologies documentaires, jusqu'à 20 % des archives numériques pourraient être conservées, contre seulement 5 % dans le monde analogique²⁰.

Deuxièmement, pour la majorité des organisations culturelles, la numérisation des archives textuelles n'en est encore qu'à ses débuts. Selon l'enquête précitée de Friedewald et al., 56 % des centres d'archives rapportent avoir numérisé moins de 5 % de leurs collections, et seules deux archives nationales ainsi qu'une archive municipale indiquent avoir dépassé le seuil des 50 %. La situation est similaire pour les archives audiovisuelles. Pourtant, cette

¹⁸ Pendergrass et al. fournissent une estimation très sous-estimée de 5750 pétaoctets de données conservées par les organisations culturelles en 2018, sans compter les backups. PENDERGRASS K.L., SAMPSON W., WALSH T. et ALAGNA L., « Toward Environmentally Sustainable Digital Preservation », *op. cit.*, p. 176-177. ; « Volume of data/information created, captured, copied and consumed worldwide from 2010 to 2023 with forecasts from 2024 to 2028 », in *Statistica*, [en ligne], <https://www.statista.com/statistics/871513/worldwide-data-created/>, consulté le 26 juin 2025. ; GAUDIAUT T., « Données numériques : Le Big Bang du Big Data », in *Statistica*, [en ligne], <https://fr.statista.com/infographie/17800/big-data-evolution-volume-donnees-numeriques-genere-dans-le-monde>, consulté le 26 juin 2025.

¹⁹ Cette estimation a été calculée en reprenant la méthode proposée par Pendergrass et al., et en l'appliquant à la *2021 Fixity Survey*. PENDERGRASS K.L., SAMPSON W., WALSH T. et ALAGNA L., « Toward Environmentally Sustainable Digital Preservation », *op. cit.*, p. 176. ; NATIONAL DIGITAL STEWARDSHIP ALLIANCE, « 2021 Fixity Survey Report », *op. cit.*, octobre 2021, p. 19-20. Pour les réponses « More than 5 PB [Enter amount] », les données ont été extraites des réponses disponibles sur le site de la NDSA.

²⁰ MOSS M. et GOLLINS T., « Our Digital Legacy: an Archival Perspective », in *Journal of Contemporary Archival Studies*, vol. 4 (8 décembre 2017), n° 2, p. 7-8.

activité est aujourd'hui celle qui génère le plus de données, car, dans la plupart des organisations culturelles, la quantité de données nées numériques reste très inférieure à celle produite par la numérisation²¹. Compte tenu du soutien actif de l'Union européenne à la numérisation et des attentes croissantes des usagers en matière d'accessibilité, cette tendance ne semble pas près de s'inverser²².

Enfin, comme le souligne éloquemment Goldman, « the classic definition of sustainability – “meeting the needs of the present without compromising the needs of the future” – might comfortably represent the mission of digital preservation professionals today »²³. Bien que la contribution des organisations culturelles aux pollutions liées aux TIC reste modeste, la lutte contre le changement climatique est une responsabilité collective, et non individuelle. Les professionnels de l'archivage numérique devraient donc intégrer cette dimension environnementale à leurs pratiques et à leurs stratégies. Toute organisation qui participe au changement climatique devrait, dans un monde idéal, contribuer à atténuer les impacts environnementaux des TIC. Répondre aux besoins du présent sans compromettre ceux des générations futures implique nécessairement de prendre en compte l'empreinte écologique dans l'élaboration des stratégies de numérisation et de préservation numérique.

Pour parvenir à cet équilibre, Pendergrass et al. soutiennent qu'un changement de paradigme dans l'évaluation, la permanence et l'accessibilité des archives numériques est nécessaire. En s'inspirant des travaux de Ehrenfeld²⁴, ils estiment que si l'on peut recourir à la technologie pour réduire l'impact climatique (par exemple en améliorant l'efficacité énergétique des systèmes ou des bâtiments), bien qu'utile comme mesure provisoire, ne fera que réduire la non-durabilité sans pour autant résulter sur un système durable. Pour cela, il est nécessaire de réévaluer la manière dont les archivistes remplissent leurs missions et prennent leurs décisions, en y intégrant pleinement la dimension environnementale²⁵. Ils invitent donc les organisations culturelles à revoir leurs pratiques de numérisation :

²¹ Les archives nées numériques ont aussi une tendance à être beaucoup plus légères que les archives numérisées, surtout en ce qui concerne les simples documents de traitement de texte. Les bases de données peuvent être à contrario très lourdes, de même que pour les fichiers audiovisuels. FRIEDEWALD M., SZÉKELY I. et KARABOGA M., « Preserving the Past, Enabling the Future », *op. cit.*, p. 63-64.

²² PENDERGRASS K.L., SAMPSON W., WALSH T. et ALAGNA L., « Toward Environmentally Sustainable Digital Preservation », *op. cit.*, p. 191-192. ; FRIEDEWALD M., SZÉKELY I. et KARABOGA M., « Preserving the Past, Enabling the Future », *op. cit.*, p. 64-65.

²³ GOLDMAN B., « It's Not Easy Being Green(e) », *op. cit.*, p. 275.

²⁴ EHRENFELD J.R., *Sustainability by Design: A Subversive Strategy for Transforming Our Consumer Culture*, Yale University Press, 2008.

²⁵ PENDERGRASS K.L., SAMPSON W., WALSH T. et ALAGNA L., « Toward Environmentally Sustainable Digital Preservation », *op. cit.*, p. 166-167.

*Cultural heritage professionals should critically examine standards and practices for file formats of born-digital and digitized content at their organizations. Should every item be migrated or digitized to the highest quality possible? Are there true preservation benefits to high-quality digital surrogates?*²⁶.

Ainsi, ils posent une question cruciale : la majeure partie des coûts et des impacts environnementaux associés aux activités de préservation et d'archivage numériques (ingestion, vérification de l'intégrité, accessibilité) est directement corrélée à la quantité de données à conserver et à diffuser. Par exemple, un fichier volumineux nécessitera davantage de temps et de puissance de calcul pour être migré ou pour que son empreinte numérique soit calculée. De plus, pour permettre ce calcul, et donc garantir l'intégrité des objets numériques, chaque bit de chaque fichier doit transiter du système de stockage vers le processeur, ce qui exige une infrastructure suffisamment performante.

Bien que le *cold storage* permette de réduire les coûts financiers et la consommation d'énergie, les normes et bonnes pratiques en matière de conservation imposent souvent la duplication des données sur plusieurs sites géographiques. Cela rend les systèmes de *hot storage* ou les solutions hybrides plus adaptés, en particulier pour la synchronisation des données et les vérifications régulières d'intégrité. Ces pratiques nécessitent également une infrastructure réseau capable de soutenir des échanges à haut débit entre les systèmes de stockage, ce qui ajoute encore à l'empreinte écologique²⁷.

Comme le conclut également Goldman, l'une des premières mesures à prendre est par conséquent de réduire la quantité de données produites en réinterrogeant les théories et les pratique de la préservation numérique²⁸. Par exemple, est-il justifié de chercher à minimiser au maximum la perte d'information, au regard des coûts financiers et écologiques que cela entraîne ? Toutes les archives numériques ou numérisées doivent-elles vraiment être accessibles instantanément ? La pratique, encore largement répandue, consistant à ne pas appliquer de compression sur les fichiers master est-elle toujours pertinente ? Les tests réguliers de l'intégrité des archives sur l'ensemble des collections des organisations culturelles sont-ils réellement justifiés ?

Problématique et méthodologie

Dans ce travail, nous n'allons pas chercher à répondre à l'ensemble de ces questions, qui méritent chacune leur propre étude. Nous nous concentrerons plutôt sur un aspect

²⁶ *Ibid.*, p. 183-184.

²⁷ *Ibid.*, p. 185-186.

²⁸ GOLDMAN B., « It's Not Easy Being Green(e) », *op. cit.*, p. 285-286.

spécifique : la numérisation des archives de la Commission européenne, et plus précisément sur la manière dont le Service des Archives Historiques de la Commission européenne (SAH) peut réduire ses coûts financiers et son impact environnemental liés à ses activités de numérisation, en agissant sur les paramètres techniques de cette dernière, notamment le choix des formats de données et l'utilisation d'algorithmes de compression.

Pour ce faire, nous tenterons d'appliquer l'exercice proposé par Pendergrass et al., en adaptant leurs questionnements au contexte spécifique de la numérisation des archives historiques de la Commission européenne. Plus précisément, nous chercherons à répondre à la question suivante :

*What file format(s) are you generating? Would different formats result in smaller storage demands while satisfying preservation and user needs?*²⁹.

Dans cette perspective, et à partir de documents internes ainsi que de deux entretiens menés avec le chef de l'équipe de la préservation numérique, la première partie sera consacrée à un état des lieux de la numérisation au sein du SAH. Nous décrirons l'ampleur actuelle de cette numérisation, les spécifications techniques retenues (format, résolution, couleur, compression, etc.), ainsi que les conséquences de ces choix sur l'utilisation de l'espace de stockage des serveurs. Cette analyse permettra de répondre à la première partie de la question.

Ensuite, la seconde partie sera dédiée à la réponse à la deuxième partie de la question. Elle débutera par une définition des besoins des utilisateurs et des exigences de la préservation numérique, avant de procéder à une comparaison des caractéristiques des formats de fichiers et des algorithmes de compression. À partir d'expérimentations menées sur des échantillons d'archives numérisées, il s'agira de déterminer quels formats, associés à quelles méthodes de compression, permettraient de répondre à ces besoins tout en réduisant significativement les volumes de stockage requis — et, par conséquent, en diminuant l'empreinte écologique.

Dans les deux parties, l'analyse portera spécifiquement sur les formats des fichiers master, car il s'agit des copies numériques de référence qu'une institution souhaite conserver à long terme, sans perte de qualité. C'est à partir de ces fichiers que sont générées les copies d'accès, ou, en cas d'obsolescence, de nouveaux fichiers master. De par cette fonction centrale, ils doivent répondre aux exigences les plus strictes en matière de qualité, de stabilité

²⁹ PENDERGRASS K.L., SAMPSON W., WALSH T. et ALAGNA L., « Toward Environmentally Sustainable Digital Preservation », *op. cit.*, p. 185. Ils proposent d'autres questions pour une numérisation responsable, mais l'ambition limitée de ce travail ne permet pas d'y répondre.

et de résilience face aux risques liés à la préservation numérique. Par nature, ce sont également les fichiers les plus volumineux, ce qui en fait des cibles privilégiées pour toute démarche visant à réduire l’empreinte écologique des institutions culturelles.

La numérisation aux Archives Historiques de la Commission européenne

Conformément au règlement 2015/496 du Conseil³⁰, le SAH est légalement tenu de reproduire toutes les archives analogues ouvertes au public – c’est-à-dire les documents de plus de 30 ans présentant une valeur historique ou administrative et destinés à être conservés de manière permanente³¹ – avant leur transfert définitif à l’Institut universitaire européen de Florence. Cette reproduction vise plusieurs objectifs : préserver les documents originaux en limitant leur manipulation, garantir leur sécurité grâce à la création de copies stockées en différents lieux géographiques, et en assurer la consultation aussi bien à Florence qu’à Bruxelles³². Un autre objectif, non moins important, est de permettre l’accessibilité en ligne de ces archives pour les utilisateurs³³.

C’est dans ce cadre que depuis 2012–2014, en remplacement de la reproduction sur microfiches, le SAH a engagé une campagne de numérisation de masse des archives traitées par les archivistes. Sur les 130 kilomètres linéaires actuellement conservés, environ 65 kilomètres seraient concernés par cette opération selon des estimations internes³⁴. Les paramètres techniques de cette numérisation ont été définis en 2012, à la suite d’une étude interinstitutionnelle menée conjointement par les services d’archives des institutions européennes³⁵. Cette étude a toutefois laissé ouverts certains choix techniques, tels que

³⁰ Règlement (UE) 2015/496 du Conseil du 17 mars 2015 modifiant le règlement (CEE, Euratom) n° 354/83 en ce qui concerne le dépôt des archives historiques des institutions à l’Institut universitaire européen de Florence (OJ L 79 25.03.2015, p. 1, ELI: <http://data.europa.eu/eli/reg/2015/496/oj>).

³¹ Règlement (CE, Euratom) n° 1700/2003 du Conseil du 22 septembre 2003 modifiant le règlement (CEE, Euratom) n° 354/83 concernant l’ouverture au public des archives historiques de la Communauté économique européenne et de la Communauté européenne de l’énergie atomique (OJ L 243 27.09.2003, p. 1, ELI: <http://data.europa.eu/eli/reg/2003/1700/oj>).

³² « Projet ARCHIS-scanning. Draft vision document », document interne à la Commission européenne, oib.os.1.002(2013)3193594, 04/09/2013, p. 6-7. ; Règlement (CE, Euratom), n° 1700/2003 du Conseil du 22 septembre 2003 [...], *op. cit.*, art. 9.

³³ « Draft report_19NOV12 », in *Digitisation of the archives of the EU institutions, common challenges and approaches*, document interne à la Commission européenne, Ares(2013)100912, 19/11/2012, p. 4.

³⁴ Courrier électronique de Lieven Baert, le 01/05/2025.

³⁵ Y ont participé : les Archives Historiques de l’Union européenne, le Parlement européen, le Conseil de l’Union européenne, la Commission européenne, la Cour des comptes européenne, le Comité économique et social européen, le Comité européen des régions, la Banque européenne d’investissement et la Banque centrale européenne. « Draft report_19NOV12 », *op. cit.*, 19/11/2012, p. 4-6.

l'utilisation du PDF/A-1b ou du TIFF comme format master, ou encore le recours (ou non) à des algorithmes de compression.

Le format de fichier master retenu par le SAH pour les archives numérisées est le TIFF, avec une résolution de 300 dpi, en 256 niveaux de gris (8 bits), et sans compression. Lors de la numérisation, l'objectif est de produire des copies numériques reflétant aussi fidèlement que possible la réalité matérielle des documents, en conservant notamment les taches, les déchirures ou toute autre aspérité visuelle. Cette approche vise à garantir l'intégrité des archives et à assurer une transparence maximale pour les utilisateurs³⁶. Ces paramètres sont appliqués de manière uniforme à l'ensemble des documents, quel que soit leur format ou leur type, afin de standardiser les procédures et les résultats.

Le choix de ne pas conserver les couleurs repose sur un souci d'optimisation de l'espace de stockage. En effet, considérant que la grande majorité des archives à numériser sont des documents dactylographiés en noir et blanc, il a été jugé plus pertinent de renoncer à la couleur, ce qui permet de réduire considérablement la taille des fichiers (en moyenne 26 mégaoctets par page en couleur). Quant au choix du niveau de gris plutôt que du noir et blanc (1 bit), il répond à un double objectif : préserver l'intégrité visuelle des documents et améliorer la qualité de l'OCR (reconnaissance optique de caractères)³⁷.

Enfin, le SAH a décidé de ne pas recourir à des algorithmes de compression, afin d'éviter l'ajout d'une couche de traitement supplémentaire lors de la numérisation :

*Notre principale préoccupation était d'éviter justement des algorithmes de compression dans les données [que l'on souhaite préserver]. Parce que, même si on peut avoir beaucoup de confiance dans ces algorithmes de compression, c'est quand même une manipulation supplémentaire que tu appliques sur tes fichiers, et si tu fais ça 10, 100, 1000 fois, oui ça va aller, mais on traite des millions d'images, et des erreurs vont se manifester, c'est sûr et certain*³⁸.

De plus, cela signifie qu'ils ne dépendent pas d'un *codec* (c'est-à-dire un programme permettant de COder et DECoder une suite binaire) pour compresser les fichiers, les lire ou les migrer vers d'autres formats. Ne pas utiliser de compression permet donc de garantir une plus grande indépendance technologique, tout en minimisant les risques d'erreurs lors de la numérisation ou de la manipulation des fichiers. Une autre raison avancée tient au contrôle accru sur les formats. Les formats JPEG et JPEG2000, en effet, peuvent exister sous de

³⁶ Notes prises au cours de la présentation de la procédure de numérisation du SAH par Alexis Bigonville (équipe déclassification et numérisation) le 6 mars 2025.

³⁷ Interview de Lieven Baert, le 24 avril 2025.

³⁸ *Ibidem*.

nombreuses variantes, selon les paramètres ou les logiciels utilisés. S'assurer de leur conformité stricte aux standards et de leur interopérabilité peut alors s'avérer complexe et en compliquer la gestion future ³⁹. À l'inverse, la gestion d'un fichier TIFF non compressé est beaucoup plus simple, car le format lui-même est plus simple et plus lisible. En outre, en raison de la manière dont les fichiers TIFF non compressés sont encodés, ce format offre une excellente résistance à la corruption des données ⁴⁰.

Enfin, le choix de ces paramètres leur offre une grande marge de manœuvre ainsi qu'une sécurité importante pour l'avenir. Le format TIFF, bien que propriétaire, est largement documenté, très ouvert, et couramment utilisé dans les institutions culturelles ainsi que dans de nombreux standards. Il est pris en charge par une grande variété de logiciels, ce qui limite considérablement les risques d'obsolescence⁴¹. C'est une solution stable et pérenne, qui permet d'envisager facilement une migration vers un format plus léger, ou toute autre opération, si le besoin s'en fait sentir.

Néanmoins, le principal inconvénient de ces choix réside dans leur impact sur l'intégrité des archives et sur la taille des fichiers. En décidant de ne pas numériser en couleur, l'intégrité de certaines archives peut être compromise. Par exemple, comment interpréter un graphique ou un plan reposant sur un code couleur, s'il est numérisé en niveaux de gris ? Par ailleurs, bien que cette décision ait été motivée par la volonté de limiter l'espace de stockage, chaque fichier pèse tout de même en moyenne 8 mégaoctets⁴². En avril 2025, les archives numérisées représentaient ainsi 29,53 téraoctets dans le système d'archivage électronique (SAE) du SAH, soit 84 % de l'espace total occupé, alors qu'elles ne constituaient que 23 % du nombre total d'objets numériques⁴³. Ce total ne prend pourtant pas en compte l'ensemble des archives numérisées : une quantité importante – entre 200 et 250 téraoctets⁴⁴ – attend

³⁹ « JPEG 2000 Part 1 (Core) jp2 File Format », in *Sustainability of Digital formats: Planning for Library of Congress Collections*, [en ligne], <https://www.loc.gov/preservation/digital/formats/fdd/fdd000143.shtml>, consulté le 08/07/2025. ; LEFURGY B., « Is JPEG-2000 a Preservation Risk? », in *Library of Congress Blogs*, [en ligne], <https://blogs.loc.gov/thesignal/2013/01/is-jpeg-2000-a-preservation-risk/>, consulté le 08/07/2025.

⁴⁰ Interview de Lieven Baert le 16 mai 2025.

⁴¹ « TIFF, Revision 6.0 », in *Sustainability of Digital formats: Planning for Library of Congress Collections*, [en ligne], [TIFF, Revision 6.0](#), consulté le 03/07/2025.

⁴² Interview de Lieven Baert, le 16 mai 2025.

⁴³ « 20250401_List_of_SIPs_submitted_to_a-REP_stats », document interne à la Commission, sans identifiant, avril 2025.

⁴⁴ Cette estimation a été évaluée à partir des chiffres issus du PowerPoint fourni lors de la visite du centre de Kortenberg par les étudiants en archivistique à l'UCLouvain et par les chiffres fournis par Alexis Bigonville. « European Commission Historical Archives Service, UCL Visit », PowerPoint fourni après visite du dépôt de Kortenberg par les étudiants en archivistique de l'UCL, sans identifiant, 15/10/2024, [Les outils et stratégies de préservation Kortenberg.pptx](#). ; Courrier électronique d'Alexis Bigonville, le 04/07/2025.

encore d'être ingérée dans le SAE⁴⁵. La numérisation a donc un coût économique et environnemental particulièrement disproportionné par rapport aux archives nées numériques. Lorsque l'on prend en considération que ces données sont stockées sur des serveurs en NAS (*Network Attached Storage*, du *hot storage*) au Luxembourg, avec un accès immédiat via une infrastructure réseau, la réduction de la quantité de données produites apparaît d'autant plus pertinente⁴⁶.

Analyse des besoins et comparaisons des formats

La question que posent Pendergrass et al. pour une numérisation responsable est la suivante :

1. *What file format(s) are you generating? Would different formats result in smaller storage demands while satisfying preservation and user needs?*

La première partie de cette question a déjà été traitée dans le chapitre précédent. La seconde partie invite, quant à elle, à déterminer deux éléments essentiels : d'une part, les besoins des utilisateurs, et d'autre part, les exigences spécifiques liées à la préservation numérique.

Les besoins des utilisateurs

Pour établir les besoins des utilisateurs, il est d'abord nécessaire de déterminer quels types d'archives sont numérisées. Compte tenu de la diversité des documents produits par une institution telle que la Commission européenne, établir une typologie exhaustive s'avérerait trop complexe, et ce travail n'a pas une telle ambition. Il a donc été nécessaire de procéder à une généralisation, sans doute imparfaite, mais suffisante pour couvrir la majorité des cas de figure. À cette fin, à partir de visites de terrain et de plusieurs échanges oraux et par courrier électronique avec le SAH⁴⁷, nous avons choisi d'élaborer une typologie sommaire fondée sur les caractéristiques externes des documents, afin de pouvoir en déduire les besoins potentiels des utilisateurs.

En effet, selon la définition d'InterPARES, l'intégrité d'un document est « [sa] qualité d'être complet et inaltéré dans tous ses aspects essentiels »⁴⁸. Duranti L. complète cette définition en expliquant qu'un document possède « l'attribut d'intégrité quand il est demeuré

⁴⁵ Courrier électronique de Lieven Baert, le 01/05/2025.

⁴⁶ Interview de Lieven Baert, le 16 mai 2025.

⁴⁷ Courrier électronique d'Alexis Bigonville, le 03/07/2025.

⁴⁸ Notre traduction. « Integrity », in *InterPARES Trust*, [en ligne], <http://interparestrust.org/terminology/term/integrity>, consulté le 03/07/2025.

intact et sans corruption, autrement dit quand le message qu'il est supposé transmettre afin d'accomplir sa mission demeure inaltéré »⁴⁹. Ce message dépend naturellement du type d'archive et de l'usage qu'en fera l'utilisateur. Ainsi, une copie numérique d'un document dactylographié en noir et blanc pourra être considérée comme intègre tant que le texte demeure lisible et que sa mise en page respecte l'original. Dans la majorité des cas, l'exactitude fine de sa présentation visuelle (taches, déchirures, texture du papier, etc.) n'est pas essentielle à la compréhension du contenu⁵⁰, tant que le texte est préservé dans sa forme.

À contrario, pour une carte utilisant un code couleur, la présentation visuelle, et donc la présence des couleurs, est extrêmement importante afin que l'utilisateur puisse correctement recevoir le message que le document cherche à transmettre. Ce sont aussi des documents plus susceptibles d'être réimprimés (par exemple, pour une publication ou une exposition), ce qui nécessite un haut niveau de détails.

Dès lors, selon le type de document et ses usages anticipés, il peut être pertinent d'adapter la qualité de numérisation pour trouver un équilibre entre intégrité, performance de stockage et impact environnemental. Ces principes ont été appliqués lors de l'élaboration de la typologie documentaire sommaire ci-dessous.

Type de document	Caractéristiques	Besoins de l'utilisateur
Documents textuels (imprimés ou manuscrits) ⁵¹	Caractères imprimés ou notes manuscrites, noir et blanc, couleurs possibles pour bandeaux, logos ou encres.	Lisibilité du texte et des annotations. Niveau de détail faible mais suffisant pour annotations/ratures. Niveau de gris recommandé. Couleurs nécessaires uniquement pour encres différenciées (rares et accessoire).
Documents textuels avec	Caractères imprimés, images	Lisibilité du texte. Images

⁴⁹ DURANTI L., « Pour une diplomatie des documents électroniques », *op. cit.*, p. 610.

⁵⁰ La trame visuelle, comme une déchirure, peut parfois revêtir une signification historique, notamment lorsqu'elle est manifestement volontaire. Cependant, à une échelle de numérisation industrielle comme celle pratiquée au SAH, une telle évaluation au cas par cas est impraticable. Une standardisation des paramètres de numérisation s'impose donc.

⁵¹ Une limite identifiée lors de l'analyse de cette catégorie tient au fait que la Commission européenne conserve des documents rédigés dans de nombreuses langues, utilisant divers alphabets (latin, cyrillique, grec, arabe, etc.). Certains caractères de ces alphabets présentent des subtilités graphiques qui pourraient justifier une résolution plus élevée lors de la numérisation. Néanmoins, nous estimons que l'analyse demeure valable en l'état, d'autant plus que les avancées récentes dans les technologies d'OCR permettent de compenser en partie cette difficulté.

images d'illustration	avec couleurs possibles.	reconnaissables, niveau de détail faible. Couleurs nécessaires si présentes.
Documents textuels avec images essentielles à leur compréhension ⁵²	Caractères imprimés, images à grand niveau de détail, couleurs possibles.	Lisibilité du texte. Niveau de détail haut pour les images afin de préserver leur contenu probant/informatif. Couleurs nécessaires si présentes.
Documents visuels (cartes, plans, graphiques, journaux, photographies)	Caractères imprimés (souvent petits), détails complexes (lignes, symboles, images), couleurs possibles ou codes couleur.	Niveau de détail haut et couleurs nécessaires pour une lecture/reproduction fidèle.
Documents à valeur probante	Implications juridiques, détails visuels ou couleurs possibles (annotations, tampons).	Intégrité maximale, lisibilité, niveau de détail haut si annotations/tampons, couleurs nécessaires si pertinentes.

Si un document relève de plusieurs catégories (par exemple, un document dactylographié accompagné d'une carte), il devra être numérisé selon les exigences de la catégorie requérant la qualité la plus élevée.

Les besoins de la préservation numérique

Il reste maintenant à déterminer quels formats et paramètres permettent de répondre aux exigences de la préservation numérique. Selon la *Digital Preservation Coalition* (DPC)⁵³, le choix d'un format de préservation par les institutions doit se fonder sur :

- 1) **Leur obsolescence ou risque d'obsolescence** : un format peut devenir obsolète s'il cesse d'être pris en charge par les logiciels récents ou s'il ne suit pas l'évolution des technologies (évolution des besoins).
- 2) **Leur nature propriétaire ou non** : les formats propriétaires présentent certains risques, notamment la dépendance à un éditeur pour leur utilisation, leur maintenance ou leur évolution, ou encore le risque de disparition du format en cas

⁵² Un exemple serait un rapport d'enquête produit par une entité au sein de la Commission européenne. Les images contenues dans ce rapport ont une valeur probatoire et nécessaire à la compréhension des de l'enquête.

⁵³ « Formats de fichiers et normes », in *Digital Preservation Handbook*, deuxième édition traduire en français, <https://www.dpconline.org/handbook>, Digital Preservation Coalition, 2021, p. 212-217.

de faillite ou de retrait du marché. Toutefois, certains formats propriétaires, comme le PDF ou le TIFF, sont très ouverts et largement documentés. Les formats open source, quant à eux, sont plus neutres, librement modifiables et souvent compatibles avec un plus grand nombre de logiciels s'ils bénéficient d'une adoption suffisante. Ils ne sont toutefois pas à l'abri de l'obsolescence si la communauté qui les maintient se périlite ou se désengage.

- 3) **Leur documentation et standardisation** : un format dont les spécifications ne sont pas publiquement accessibles rend difficile la compréhension de sa structure interne. Cela complique, voire empêche, sa lecture, sa migration, sa restauration ou son interopérabilité logicielle. Une bonne documentation, idéalement accompagnée d'une standardisation reconnue, est donc essentielle. Le format doit de préférence figurer dans un registre reconnu comme PRONOM.
- 4) **Leur niveau d'adoption** : un format largement utilisé présente de meilleures garanties de pérennité. Il a plus de chances d'être pris en charge à long terme par divers logiciels, et de bénéficier d'un large éventail d'outils.
- 5) **Leur capacité de compression** : pour les fichiers master, lorsque l'application d'une compression est envisagée, la possibilité d'appliquer une compression sans perte est généralement considérée comme indispensable afin de garantir l'intégrité du contenu. La DPC conseille de ne pas appliquer de compression avec perte. En effet, une compression sans perte permet de reconstruire l'image originale bit à bit, sans aucune altération, contrairement à la compression avec perte, qui supprime définitivement certaines informations jugées non essentielles, ce qui rend impossible une restitution exacte de l'original.
- 6) **Leur support des métadonnées** : un format capable d'intégrer une large éventail de métadonnées (comme le JPEG2000 ou le PDF) peut être particulièrement utile, en fonction des pratiques de gestion documentaire de l'institution.
- 7) **Leurs propriétés spécifiques** : le choix du format doit également être guidé par les caractéristiques du contenu à préserver ou à rendre accessible. Par exemple, pour des documents textuels, la capacité à être OCRisé est un critère essentiel.

Heureusement, ce travail d'analyse a déjà été en grande partie réalisé par plusieurs organisations culturelles nationales, qui publient régulièrement les spécifications techniques de leurs politiques de numérisation. En complément, l'*Open Preservation Foundation* a entrepris un travail comparatif particulièrement utile : depuis 2022, elle publie chaque année

une analyse des politiques de formats acceptés dans les *Submitted Information Packages* (SIP) de différentes institutions culturelles à travers le monde. Cette publication, intitulée *International Comparison of Recommended File Formats*⁵⁴, attribue à chaque format un score en fonction de son acceptabilité en tant que format de fichier master :

- 2 = préféré,
- 1 = accepté,
- 0 = accepté, mais non recommandé,
- -1 = non accepté.

Les résultats de l'édition 2025 de cette étude permettent d'isoler quatre formats qui ressortent comme les plus largement acceptés pour la préservation des archives numérisées par les institutions interrogées :

Formats	Photographies, images (raster)	Objets numérisés (textes, images, etc.)
TIFF (version 4, 5 et 6)	64	30
PDF/A	24	24
JP2 (JPEG2000 partie 1)	49	18
PNG	39	16

Ces scores traduisent un haut niveau de confiance de la part des organisations culturelles envers ces formats. On peut ainsi en déduire qu'ils répondent de manière satisfaisante aux principaux critères de sélection pour l'archivage numérique et la préservation à long terme, notamment ceux liés à l'obsolescence, l'ouverture, la standardisation et l'adoption, soit les quatre premiers critères énoncés par la DPC.

Il convient désormais de procéder à une comparaison de ces formats, en s'attardant sur leurs caractéristiques techniques, et en évaluant leur capacité à réduire les volumes de données via l'utilisation d'algorithmes de compression.

⁵⁴ « International Comparison of Recommended File Formats », in *Open Preservation Foundation*, [en ligne], <https://openpreservation.org/resources/member-groups/international-comparison-of-recommended-file-formats/>, consulté le 05/07/2025.

Comparaison entre les formats de données et les algorithmes de compression

Méthodologie

Pour ce faire, nous avons élaboré un tableau comparatif (voir annexe « Comparaisons_formats_et_resultats ») mettant en lumière les principales caractéristiques des formats TIFF, PDF/A, JPEG2000 (JP2) et PNG, à partir de ressources en ligne. Ce tableau présente notamment la prise en charge des métadonnées, des couleurs, de la compression, les fonctionnalités spécifiques, la robustesse, les standards associés, les outils de gestion disponibles, ainsi que les inconvénients propres à chaque format.

Sur cette base, nous avons ensuite mis en correspondance ces formats avec les catégories de documents définies dans la typologie sommaire, en tenant compte à la fois des besoins des utilisateurs, des caractéristiques techniques des formats et de leur potentiel de réduction de l'espace de stockage.

Afin d'évaluer ce dernier point – exprimé en taux d'économie de stockage $((x_1 - x_2) / x_1) \times 100$ –, nous avons réalisé des tests à partir d'un échantillon de 319 pages numérisées par le SAH au format TIFF non compressé (niveau de gris, 300 dpi), à l'aide d'un scanner Kodak Alaris i4650 avec le logiciel Capture Pro 6.0. Deux documents iconographiques en couleur supplémentaires ont été intégrées : une carte (181 Mo) et un portrait photographique (25,3 Mo), tous deux en TIFF non compressé, en couleur (RGB, 24 bits).

Ces fichiers sources ont ensuite été migrés vers les formats TIFF, PDF/A-2b⁵⁵, JPEG2000 et PNG, avec diverses options de compression, à l'aide des logiciels IrfanView (4.72) et Adobe Acrobat Pro (25.1).

N'ayant pas eu accès à des exemples représentatifs de l'ensemble des catégories de notre typologie documentaire, les résultats issus des tests sur les documents iconographiques en couleur ont été utilisés comme approximation pour les documents textuels contenant des illustrations.

Pour l'analyse des images et la confirmation de la conformité des formats générés, plusieurs outils ont été mobilisés : Jpylyzer (v2.2.1), Jhove (v1.32.1), Droid (v6.8.1), VeraPDF (v1.28.1), ExifTool (v13.32), Poppler (v24.08.0-0) et ImageMagick (v7.1.1-47). Concernant la génération de fichiers JP2 (JPEG2000), il est généralement recommandé

⁵⁵ Le PDF/A-2b a été choisi car il assure un rendu visuel fiable, tout en permettant la compression avec JPEG2000. Puisqu'il s'agit d'archives numérisées, les éléments pour une fidélité des propriétés structurales et syntaxiques (PDF/A-2a) n'est pas nécessaire (voir annexe « Comparaisons_formats_et_resultats »).

d'utiliser les bibliothèques OpenJPEG ou Kakadu pour obtenir des résultats conformes au standard⁵⁶. Néanmoins, pour des raisons de simplicité et d'accessibilité, nous avons opté pour le logiciel IrfanView avec le plugin JPEG2000.dll⁵⁷, avant de valider la conformité des fichiers produits à l'aide de Jpylyser, outil développé par Johan van der Knijff (Bibliothèque royale des Pays-Bas) et l'*Open Preservation Foundation*⁵⁸.

Dès à présent, dans les lignes qui suivent, nous allons exclure le format PNG. Il est principalement optimisé pour la publication web, et n'offre aucun avantage significatif pour la numérisation par rapport aux autres formats étudiés, tout en présentant plusieurs limitations : entre autres, prise en charge limitée des métadonnées, absence du mode colorimétrique CMYK, et recours exclusif à l'algorithme de compression Deflate (sans perte)⁵⁹.

Résultats et recommandations pour une numérisation durable

À la suite de nos analyses et expérimentations, le format JP2 (JPEG2000) s'est révélé être la solution la plus polyvalente et performante. Il offre en effet les meilleurs résultats en termes de polyvalence (compression avec ou sans perte, grande robustesse aux erreurs, large éventail de métadonnées, gestion des couleurs, décodage progressif, etc.) et de taux de compression. Il permet de répondre à l'ensemble des besoins pour tous les types de documents que nous avons décrits.

Pour les documents textuels, nos essais montrent que tous les formats et algorithmes de compression testés peuvent répondre à nos besoins en termes d'intégrité et de réduction de la quantité de données. Les compressions sans perte sur ce type de documents donnent des résultats similaires, avec des taux d'économie de stockage variant entre 82,8 % (PDF/A-2b avec JPEG2000 sans perte) et 85,6 % (TIFF avec Deflate/ZIP), soit un ratio d'environ 6:1⁶⁰. Ces compressions étant sans perte, l'intégrité visuelle des archives est pleinement conservée.

Algorithmes de compression	Taux d'économie sur les 319 pages (niveau de
----------------------------	--

⁵⁶ CARON B., « Formats de données pour la préservation à long terme : la politique de la BnF », [Rapport Technique], Bibliothèque Nationale de France, Paris, 2021, identifiant : Hal-03374030v2, <https://bnf.hal.science/hal-03374030>, p. 90-91.

⁵⁷ Le plugin a été développé par la société LuraTech, qui a participé au développement du standard JPEG2000. « History of changes », in *IrfanView*, [en ligne], https://www.irfanview.com/history_old.htm, mis à jour le 11 octobre 2024, consulté le 01/07/2025.

⁵⁸ « Jpylyser », in *Openpreservation*, [en ligne], <https://jpylyzer.openpreservation.org/>, consulté le 08/07/2025.

⁵⁹ « PNG, Portable Network Graphics », in *Sustainability of Digital formats: Planning for Library of Congress Collections*, [en ligne], [PNG, Portable Network Graphics](https://www.archimag.com/demat-cloud/2021/03/29/numerisation-format-conserver-documents), consulté le 08/07/2025. ; THOMAS M., « Numérisation : sous quel format conserver les documents ? », in *Archimag*, [en ligne], <https://www.archimag.com/demat-cloud/2021/03/29/numerisation-format-conserver-documents>, publié le 29 mars 2021, consulté le 08/07/2025. Voir notre analyse dans l'annexe « Comparaisons_formats_et_resultats ».

⁶⁰ Annexe « Comparaisons_formats_et_resultats ».

	gris)
JPEG2000 (lossless)	84 %
Deflate (PNG, échelle = 6)	85,6 %
LZW (TIFF)	84,5 %

Néanmoins, la compression JPEG2000 reste selon nous à privilégier, car elle offre de meilleurs taux de compression pour les documents iconographiques contenant une grande quantité d'informations (par exemple, une photographie aux détails complexes), comparativement aux algorithmes LZW (Lempel-Ziv-Welch) et Deflate (LZ77 + codage de Huffman) utilisés dans les formats TIFF, PDF et PNG.⁶¹ Dans certains cas, lorsque l'image est particulièrement riche en détails, l'algorithme LZW peut même produire des fichiers plus volumineux que les originaux⁶². Ainsi, la compression sans perte JPEG2000 permet généralement d'obtenir des taux de compression supérieurs ou légèrement supérieurs à ceux de LZW et Deflate (y compris en ajustant les paramètres de compression) pour les images d'illustrations en couleur⁶³.

Algorithmes de compression	Taux d'économie sur l'image du plan (couleurs)	Taux d'économie sur l'image du portrait (couleurs)
JPEG2000 (lossless)	55,77 %	37,69 %
Deflate (PNG, échelle = 6)	36,64 %	35,88 %
LZW (TIFF)	37,62 %	13,07 %

De plus, la compression JPEG2000 est bien plus paramétrable que ses concurrentes et offre une large gamme de fonctionnalités⁶⁴. Si une standardisation stricte est nécessaire, elle est ainsi une solution plus polyvalente. Par ailleurs, la compression JPEG2000 (avec et sans

⁶¹ PILLAY R., « TIFF Image Encoding: Optimizing for Size, Speed and Quality », in *IIP*, [en ligne], <https://iipimage.sourceforge.io/2024/12/tiff-image-encoding-optimizing-for-size-speed-and-quality>, publié le 3 décembre 2024, consulté le 26/07/2025. ; COSSU S., PILLAY R., ROBSON G., SMITH M.D., « Evaluating HTJ2K as a Drop-In Replacement for JPEG2000 with IIF », in *Code4lib Journal*, n°57 (29 août 2023), <https://journal.code4lib.org/articles/17596>.

⁶² « Comparing TIFF Compression Algorithms », in *IDR Solutions*, [en ligne], <https://blog.idrsolutions.com/comparing-tiff-compression-algorithms/>, publié le 7 mai 2025, consulté le 07/07/2025. ; DAVIES K., MAY P., « Practical Analysis of TIFF file size reduction achievable through compression », article présenté à la conférence *iPRES 2016*, Swiss National Library, Bern, 2016, p. 4.

⁶³ Voir annexe « Comparaisons formats et résultats ».

⁶⁴ Compression avec et sans perte, possibilité de fixer un taux de qualité (comme JPEG) ou une taille cible pour le produit final, riches métadonnées personnalisables, décodage progressif, etc. Voir annexe « Comparaisons formats et résultats ». OTEMUYIWA P., « The Pros and Cons of JPEG 2000 and Other Common Image Formats », in *Cloudinary*, [en ligne], https://cloudinary.com/blog/the_great_jpeg_2000_debate_analyzing_the_pros_and_cons_to_widespread_adoption, publié le 6 septembre 2020, consulté le 25 juillet 2025.

perte) se montre plus résiliente à la corruption des données (*bit rot*) que les formats TIFF (avec ou sans compression) et JPEG⁶⁵.

Malgré les performances intéressantes de la compression sans perte, plusieurs études menées par la *California Digital Library*, l'*Internet Archive*, la *Harvard University Library* et la Bibliothèque nationale de France (BnF), concluent qu'une compression JPEG2000 avec perte est préférable à une compression sans perte. Effectivement, cette dernière s'avère moins efficace pour les images à haute valeur informationnelle, et ne permet d'atteindre que des ratios d'environ 2:1 ou moins⁶⁶. Cela s'explique par le fonctionnement du codec JPEG2000, qui repose sur une transformation en ondelette (*wavelet transform*) et l'algorithme EBCOT (*Embedded Block Coding With Optimized Truncation of the embedded bit stream*) : les zones de l'image moins riches en information sont compressées de manière plus agressive. Par conséquent, pour les images complexes et bruitées, le taux de compression sans perte peut être relativement limité⁶⁷.

Duploux L. et ses collègues de la BnF ont par exemple conclu qu'un ratio de compression avec perte de 4:1 permettait d'obtenir une compression efficace des documents iconographiques à haute valeur informationnelle (environ 75 % d'économie de stockage), sans introduction d'artefacts visuels perceptibles, ceux-ci étant indiscernables du bruit déjà présent lors de la numérisation, dû aux capteurs⁶⁸. Pour les documents transparents en niveaux de gris (par exemple les microfilms), un ratio de 16:1 permet également d'obtenir un rendu visuellement sans perte⁶⁹.

La *Federal Agencies Digital Guidelines Initiative* (États-Unis) indique quant à elle que les documents textuels peuvent être compressés jusqu'à un ratio de 20:1 (soit 95 % de

⁶⁵ À condition d'intégrer les marqueurs SOP (*Start of Packet*), EPH (*End of packet header*) et PLT (*packet length, tile-part header*). « JPEG2000 » in *Référentiel de format de fichier image*, deuxième version, Bibliothèque Nationale de France, service numérisation, 2015, https://www.bnf.fr/sites/default/files/2018-11/ref_num_fichier_image_v2.pdf, p. 5. ; BUONORA P. et LIBERATI F., « A Format for Digital Preservation of Images: A Study on JPEG 2000 File Robustness », in *D-Lib Magazine*, vol. 14 (juillet 2008), n° 7/8, <https://www.dlib.org/dlib/july08/buonora/07buonora.html>.

⁶⁶ CHAPMAN S., DUPLOUX L., KUNZE J. BLAIR S., INTERNET ARCHIVE, ABRAMS S., LUPOVICI C., JENSEN A., JOHNSTON D., « Page Image Compression for Mass Digitization », in *Proc. IS&T Archiving 2007*, 2007, p. 37-42.

⁶⁷ CARON B., « Formats de données pour la préservation à long terme : la politique de la BnF », *op.cit.*, 2021, p. 90. Pour plus d'informations, voir BAKAL J.W., PAWADSHETTY J. S., « JPEG2000 Region of Interest Coding Methods », in *International Journal of Engineering Research and Applications*, vol. 3 (Janvier-février 2013), n°1, p. 1184-1188.

⁶⁸ DUPLOUX L., « JPEG2000 as a preservation format for digitization: lessons learned from a library », in *Proc. IS&T Archiving 2017*, 2017, p. 157-158. Selon nos expériences, le ratio 4:1 résulte en effet sur des images indiscernables de l'original à l'œil nu. Voir annexe « Comparaisons_formats_et_resultats »

⁶⁹ « JPEG2000 » in *Référentiel de format de fichier image*, *op. cit.*, 2015, p. 4.

réduction) tout en restant visuellement sans perte⁷⁰. Nos propres expériences confirment la validité de ce ratio, qui ne génère ni différences visuelles notables ni artefacts perceptibles⁷¹.

Dans la lignée des travaux de Goldman, Pendergrass et al., nous considérons que cette perte de données est acceptable, dans la mesure où elle est sans conséquence pour l'intégrité fonctionnelle des archives : les documents sont visuellement sans perte, et donc utilisables comme s'ils l'étaient effectivement. Cette approche, bien qu'en contradiction avec certaines recommandations strictes⁷², repose sur l'idée que la qualité des images des archives numérisées doit avant tout répondre aux besoins des utilisateurs. Dès lors, la conservation intégrale ou la possibilité de recréer les flux binaires originaux perd de sa pertinence. Rien n'indique en effet que les utilisateurs aient besoin d'un tel niveau de fidélité. Le fait que le SAH ne reçoive pratiquement aucune demande visant à obtenir les fichiers sources des numérisations, les usagers se satisfaisant d'une version d'accès en PDF (compression JPEG, noir et blanc), tend à le confirmer⁷³.

Quant au risque de pertes cumulées liées à des migrations successives de formats dans le futur, il nous semble limité. Un tel scénario ne se produirait que si les logiciels permettant la lecture et l'encodage des fichiers JP2 devenaient obsolètes, ce qui, comme nous le verrons plus loin, est improbable.

Ces résultats tendent à démontrer que le format PDF/A-2b avec compression JPEG avec perte (niveau de qualité 80, également disponible pour le format TIFF) n'est pas le plus pertinent, bien qu'il respecte les critères d'intégrité définis dans notre typologie documentaire sommaire (lisibilité du texte et des images, respect des couleurs, niveau de détail faible). Effectivement, cette solution produit des résultats tout à fait acceptables pour la lecture des documents de l'échantillon. La compression JPEG permet une économie de stockage de 94,5 % pour les documents en niveaux de gris, de 93,7 % pour les plans, et de 95,1 % pour les

⁷⁰ STELMACH M., WILLIAMS D., « Technical Guidelines for Digitizing Cultural Heritage Materials », FADGI, 2023, <https://www.digitizationguidelines.gov/guidelines/digitize-technical.html>, p. 41.

⁷¹ Nous avons obtenu ce résultat en fixant la qualité de l'image au niveau 70, ce qui correspond approximativement à un ratio de compression de 20:1. Annexe « Comparaisons formats et résultats ».

⁷² L'édition de 2008 du *Digital Preservation Handbook* indique par exemple : « Do not use encryption or compression for archival files if possible ». *Digital Preservation Handbook*, première édition, <https://www.dpconline.org/docman/digital-preservation/handbook/superseded/304-digital-preservation-handbook-media-and-formats/file>, Digital Preservation Coalition, 2008, p. 157. Pour Wasim Ahmad Bhat, « lossy algorithms promise high compression ratios but do occur loss of information and hence are not applicable ». BHAT W.A., « Long-term preservation of big data: prospects of current storage technologies in digital libraries », in *Library Hi Tech*, vol. 36 (2018), n°3, p. 548.

⁷³ Il n'y a pas plus de trois demandes des scans originaux par an, et certaines années n'enregistrent aucune demande. Courriel électronique de Sylvia Pérez, le 04/07/2025.

portraits⁷⁴. Les artefacts visuels générés par la compression JPEG, notamment autour des caractères imprimés ou des annotations manuscrites, ne nuisent pas à la consultation des documents : le contenu reste lisible et compréhensible. Il en va de même pour les images d'illustration en couleur, pour lesquelles la compression est perceptible, mais n'entrave ni la lecture ni l'interprétation⁷⁵.

Ce format peut donc être envisagé comme format master potentiel pour des documents textuels simples, voire pour des documents textuels moins précieux contenant des illustrations. C'est d'ailleurs le format recommandé par le magazine *Archimag* pour la numérisation, de préférence au TIFF⁷⁶. Cependant, la perte de données plus importante⁷⁷ et la création d'artefacts visuels que cette compression induit ne semblent pas justifiées si la compression JPEG2000 permet d'obtenir une qualité d'image supérieure avec une économie d'espace comparable, que ce soit pour les documents textuels ou pour les documents iconographiques⁷⁸. Par conséquent, le PDF/A-2b apparaît davantage comme un format d'accès dans un contexte de numérisation, plutôt que comme un format de conservation⁷⁹.

Cependant, le format PDF/A-2b (avec compression JPEG) présente un avantage notable : il bénéficie d'un écosystème logiciel très vaste. En tant que format ouvert, standardisé, et spécifiquement conçu pour l'archivage à long terme, il est pris en charge par quasiment toutes les visionneuses modernes. Cela s'explique notamment par son caractère *self-contained*, c'est-à-dire qu'il embarque en lui-même les éléments nécessaires à sa lecture (polices, informations sur les couleurs, etc.)⁸⁰. Tant le format PDF que la compression JPEG font partie des technologies les plus répandues au monde. La même observation peut être faite pour les formats TIFF et PNG, qui bénéficient également d'une large compatibilité avec les logiciels de visualisation et de traitement d'image. En revanche, le format JP2 ne bénéficie

⁷⁴ Annexe « Comparaisons formats et résultats ».

⁷⁵ Information obtenue après discussions et visite de l'équipe de numérisation du SAH.

⁷⁶ THOMAS M., « Numérisation : sous quel format conserver les documents ? », *op. cit.*, 29 mars 2021.

⁷⁷ La compression JPEG2000 avec perte permet de compresser jusqu'à 200 % en plus que JPEG pour une même qualité d'image. « JPEG vs JPEG2000 », in *Adobe*, [en ligne], <https://www.adobe.com/mx/creativecloud/file-types/image/comparison/jpeg-vs-jpeg-2000.html>, consulté le 31/07/2025.

⁷⁸ Selon nos expériences, et GILLESSE R., ROG J., VERHEUSSEN A., « Life Beyond Uncompressed TIFF : Alternative File Formats for the Storage of Master Image Files », in *Proc. IS&T Archiving 2008*, 2008, p. 41-46.

⁷⁹ KIDD T., « Choosing What Digital Image Format Works Best for You », in *Backstage Library Works*, [en ligne], <https://bslw.com/choosing-a-digital-image-format/>, publié le 16 mars 2021, consulté le 27/07/2025.

⁸⁰ « PDF/A-2b, PDF for Long-term Preservation, Use of ISO 32000-1 (PDF 1.7), Level B Conformance », in *Sustainability of Digital formats: Planning for Library of Congress Collections*, [en ligne], <https://www.loc.gov/preservation/digital/formats/fdd/fdd000322.shtml>, consulté le 08/07/2025.

pas d'un niveau d'adoption aussi étendu, ce qui suscite, chez certains, des inquiétudes quant à une potentielle obsolescence logicielle à long terme⁸¹.

L'algorithme JPEG2000 a en effet longtemps été au cœur des débats parmi les spécialistes de la numérisation et de la préservation numérique. Développé entre 1997 et 2000 par le *Joint Photographic Experts Group* (ISO/CEI 15444-1), il avait pour ambition de remplacer le JPEG classique en introduisant la compression par transformée en ondelettes. Toutefois, sa complexité a freiné son adoption, notamment chez les fabricants d'appareils photo de l'époque, dont les capacités mémoire étaient limitées. À cette première barrière technique sont venus s'ajouter d'autres freins : la question des brevets, la concurrence de formats alternatifs, le manque de support dans les navigateurs web, ainsi qu'un effet d'attente réciproque entre développeurs logiciels et constructeurs. En outre, le format JPEG était alors perçu comme suffisamment performant pour la majorité des usages et jouissait déjà d'une très large implantation⁸².

Cela a conduit à une adoption relativement faible du format, et par conséquent à un nombre limité d'outils permettant sa lecture, sa création et sa validation, majoritairement maintenus par un petit nombre de communautés spécialisées. Cela présentait naturellement des risques au niveau de la préservation numérique, ce qui a freiné son adoption par les institutions culturelles⁸³.

Néanmoins, la diffusion du format s'est largement renforcée ces dernières années. Il est désormais utilisé dans des projets de numérisation de masse par la *Library of Congress*, à la BnF, à la *British Library*, à la *Koninklijke Bibliotheek, Netherland*, pour le projet *League of Nations Total Digital Access* (LONTAD), et bien d'autres encore⁸⁴. Même lorsqu'il n'est pas

⁸¹ KYLE R., « JPEG2000 and Digitisation : Export round table », in *TownswWeb Archiving*, [en ligne], <https://blog.townswwebarchiving.com/2015/10/jpeg2000-and-digitisation-expert-round-table>, publié le 19 octobre 2015, mis à jour le 20 janvier 2020, consulté le 27/07/2025.

⁸² KELECHAVA B., « Why JPEG 2000 Never Took Off », in *American National Standard Institute*, [en ligne], <https://blog.ansi.org/ansi/why-jpeg-2000-never-used-standard-iso-iec/>, publié le 10 juillet 2018, consulté le 27/07/2025. ; BOXER A., « The Rise and Fall and Rise of JPEG2000 », conférence présentée au *Free and open source software developers' European meeting* (FOSDEM), Bruxelles, 2 février 2020, <https://youtu.be/UGXeRx0Tic4?si=-BB7QJONzG9f6Hhj>.

⁸³ MAY P., PENNOCK M., WHEATLEY P., « Preservation Assessment: JP2 Format », version 1.3., *Digital Preservation Team*, British Library, 03/09/2015. ; CARON B., « Formats de données pour la préservation à long terme : la politique de la BnF », *op. cit.*, 2021, p. 157. ; BARINA D., KLIMA O., « JPEG 2000 : guide for digital libraries », in *Digital Library Perspectives*, vol. 36 (2020), n°3, p. 249-263.

⁸⁴ « Setting Standards (JPEG 2000) », in *Digital Preservation, Library of Congress*, [en ligne], https://www.digitalpreservation.gov/series/challenge/formats2_challenge.html, consulté le 08/07/2025. ; DUPOUY L., « JPEG2000 as a preservation format for digitization: lessons learned from a library », *op.cit.*, 2017, p. 157-159. ; BROWN T., « Digital Imaging : is JP2 the file format of choice for the museum and heritage sector? », in *Museum+ Heritage Advisor*, [en ligne], <https://museumsandheritage.com/advisor/insights/digital-imaging-is-jp2-the-file-format-of-choice-for-the-museum-and-heritage/>, publié le 18 novembre 2014, consulté le

utilisé comme format master, il peut servir de format d'accès pour certaines organisations culturelles, grâce au décodage progressif qui permet de fournir des copies à la résolution demandée par les utilisateurs⁸⁵. D'autres secteurs ont également largement adopté le format et la compression JPEG2000, notamment le cinéma numérique, la diffusion vidéo, l'imagerie médicale, la télédétection et l'imagerie spatiale⁸⁶.

En ce qui concerne les logiciels prenant en charge la lecture, la création et la validation de fichiers JP2, leur nombre reste plus limité que pour les formats JPG, PDF ou TIFF, mais il s'est développé parallèlement à l'élargissement de l'adoption. Plusieurs bibliothèques logicielles libres ou open source, telles que OpenJPEG⁸⁷, Grok⁸⁸, JasPer⁸⁹, ainsi que des solutions propriétaires comme Kakadu⁹⁰ sont maintenues et disponibles⁹¹. Enfin, l'outil de validation Jpylyzer⁹², et dans une moindre mesure JHOVE⁹³, permettent de vérifier la conformité des fichiers JP2 et d'en extraire les caractéristiques techniques afin de garantir leur validité.

Nous estimons par conséquent que les outils disponibles, ainsi que le nombre croissant d'institutions et de secteurs ayant un intérêt à les maintenir, permettent d'avoir une confiance raisonnable dans sa pérennité. Les scores obtenus par le format dans l'*International Comparison of Recommended File Formats* de l'*Open Preservation Foundation* semblent d'ailleurs confirmer cette évaluation⁹⁴.

Par ailleurs, comme l'a souligné David Rosenthal, l'obsolescence des formats n'est pas aussi fréquente qu'on le redoutait au début des années 2000⁹⁵, en particulier lorsqu'il

08/07/2025. ; FAASE J., GRAVENHORST C., « Quantity Meets Quality : Towards a digital library », [IFLA International Newspaper Conference 2011- Papers.pdf](#), conférence présentée à l'*International Newspaper Conference*, 25-27 avril 2011, Kuala Lumpur, p. 2-10. ; BUSSARD S., « Les archives numérisées de la SdN vont-elles permettre une réécriture de l'histoire ? », in *Le Temps*, [en ligne], <https://www.letemps.ch/monde/archives-numerisees-sdn-vontelles-permettre-une-reecriture-lhistoire>, publié le 30 janvier 2022, consulté le 08/07/2025.

⁸⁵ THE FADGI STILL IMAGE WORKING GROUP, « Raster Still Images for Digitization. A Comparison of File Formats. Part 3. Narrative and Summary Table », FADGI, 29 août 2014, p. 5-6.

⁸⁶ « Application for JPEG 2000 », in *Jpeg.org*, [en ligne], <https://jpeg.org/jpeg2000/applications.html>, consulté le 27/07/2025. ; « What is JPEG 2000 ? », in *AIRBUS*, [en ligne], <https://www.intelligence-airbusds.com/en/8724-jpeg-2000-format>, consulté le 27/07/2025. ; « JPEG 2000 Part 1 (Core) jp2 File Format », op. cit.

⁸⁷ Disponible à l'adresse : <https://www.openjpeg.org/>.

⁸⁸ Disponible à l'adresse : <https://github.com/GrokImageCompression/grok>.

⁸⁹ Disponible à l'adresse : <https://github.com/jasper-software/jasper>.

⁹⁰ Disponible à l'adresse : <https://kakadusoftware.com/>.

⁹¹ Pour une liste complète des logiciels pouvant lire et écrire le format JP2, voir la page Wikipedia de l'algorithme : https://en.wikipedia.org/wiki/JPEG_2000.

⁹² Disponible à l'adresse : <https://jpylyzer.openpreservation.org/>.

⁹³ À l'aide d'un module complémentaire : <https://jhove.openpreservation.org/modules/jpeg2000/>.

⁹⁴ « International Comparison of Recommended File Formats », op. cit.

⁹⁵ ROSENTHAL D., « The Prostate Cancer of Preservation re-examined », in *DSHR's Blog*, [en ligne], <https://blog.dshr.org/2015/09/the-prostate-cancer-of-preservation-re.html>, publié le 16 septembre 2015, consulté

s'agit d'un format standardisé (ISO/CEI 15444-1) et open source, comme c'est le cas du JP2. Si une obsolescence devait se produire, elle serait probablement lente et prévisible.

Ce qui importe avant tout, c'est la disponibilité de logiciels capables de lire, d'écrire et de valider le format. Or, depuis plus de dix ans, de nombreuses institutions de grande envergure utilisent le JP2 pour la numérisation de masse, ce qui renforce la probabilité que ces outils continuent d'être maintenus à long terme, comme c'est le cas d'OpenJPEG, actif depuis 2005, ou de Jpylyzer, maintenu depuis 2012.

Quant à la complexité technique de l'implémentation du format, bien qu'elle soit réelle, elle est désormais atténuée par l'existence de guides et de ressources d'accompagnement, destinés à aider les institutions souhaitant l'adopter⁹⁶.

Cela permet de répondre à plusieurs des inquiétudes exprimées par le SAH, notamment le fait qu'une compression constitue une couche de traitement supplémentaire, augmentant la complexité du fichier et créant une dépendance à un codec externe, ainsi que la difficulté potentielle de produire des fichiers JP2 conformes aux standards.

Comme le souligne Rog J., la complexité d'un format ou de son algorithme de compression, tout comme la dépendance à un codec, ne posent réellement problème que dans l'éventualité où un codec devrait être entièrement réécrit. Or, selon elle, et nous partageons cet avis, un tel scénario est improbable⁹⁷. Cela supposerait en effet une situation extrême où aucune application ne serait plus capable d'ouvrir les fichiers JP2 et où aucun décodeur ne serait disponible.

En réalité, la standardisation du format, sa diffusion dans de nombreux secteurs, ainsi que l'existence d'outils open source garantissent que ni sa complexité, ni sa dépendance à un codec ne constituent une menace pour la préservation numérique. Même dans l'hypothèse où les structures actuellement en charge du développement de ces outils open source venaient à disparaître, les institutions culturelles disposeraient du temps et des ressources nécessaires

le 28/07/2025. ; COCHRANE E., « Obsolescence », in *Digital Preservation : A Critical Vocabulary*, [en ligne], <https://digital-preservation-a-critical-vocabulary.pubpub.org/pub/ajzjdqk4>, publié le 28 septembre 2024, consulté le 28/07/2025. ; *Digital Preservation Handbook*, deuxième édition traduite en français, <https://www.dpconline.org/handbook>, Digital Preservation Coalition, 2021, p. 181.

⁹⁶ Notamment BARINA D., KLIMA O., « JPEG 2000 : guide for digital libraries », *op. cit.*, 2020, p. 249-263. D'autres guides se trouvent sur les sites internet de la *Library of Congress*, de la BnF, de la FADGI, de HathiTrust, etc.

⁹⁷ ROG J., « Compression and digital preservation: do they go together? », in *Proc. IS&T Archiving 2007*, 2007, p. 81.

pour s'adapter ou reprendre leur développement⁹⁸. En effet, le code source de ces programmes étant accessible publiquement, n'importe quelle organisation ou communauté technique pourrait en assurer la continuité. En dernière mesure, les formats pourront aussi toujours être migrés.

L'existence de l'outil Jpylyzer permet quant à lui de s'assurer de la conformité des fichiers produits aux paramètres et aux spécifications retenues.

Concernant les erreurs potentielles pouvant survenir au cours des étapes de numérisation, de compression, de décompression ou de conservation, nos travaux n'ont pas permis d'évaluer leur fréquence. Néanmoins, selon Paolo Buonora et Franco Liberati, le format JP2 figure parmi les plus résistants aux erreurs⁹⁹. Judith Rog ajoute qu'il n'existe aucune preuve indiquant que les algorithmes de compression introduiraient plus d'erreurs que ceux utilisés pour les fichiers non compressés¹⁰⁰.

De son côté, Robert Buckley, fait valoir que les mécanismes intégrés de détection et de correction d'erreurs, présents dans les systèmes informatiques et les pratiques de préservation numérique, ainsi que des facteurs comme le bruit inhérent aux processus de numérisation, permettent de considérer les risques liés à la compression comme négligeables¹⁰¹. En cas d'erreurs affectant toutes les copies d'un fichier, un contrôle de qualité rigoureux permet de les détecter et de limiter leur impact. Et, si nécessaire, les originaux physiques conservés à Florence peuvent toujours être renumérisés.

Dès lors, les bénéfices offerts par la compression en termes de réduction des coûts de stockage et d'impact environnemental paraissent largement supérieurs aux risques résiduels liés à la complexité technique ou aux erreurs potentielles.

Ces éléments nous permettent de formuler des recommandations partielles concernant les paramètres de numérisation des archives du SAH, en nous fondant sur notre typologie documentaire sommaire et les besoins des utilisateurs.

Type de document	Paramètres de numérisation	Taux d'économie de stockage
------------------	----------------------------	-----------------------------

⁹⁸ ROSENTHAL D., « The Prostate Cancer of Preservation re-examined », *op. cit.*

⁹⁹ BUONORA P. et LIBERATI F., « A Format for Digital Preservation of Images: A Study on JPEG 2000 File Robustness », *op. cit.*, juillet 2008.

¹⁰⁰ ROG J., « Compression and digital preservation: do they go together? », *op. cit.*, 2007, p. 82.

¹⁰¹ BUCKLEY R., « Error Correction and Recovery of JPEG 2000 Compressed Images in Digital Preservation », in *Proc. IS&T Archiving 2012*, 2012, p. 270-274.

Documents textuels (imprimés ou manuscrits) ¹⁰²	JP2 (JPEG2000, lossy), 300 dpi, niveau de gris, ratio 20:1 (visuellement sans perte).	95 %
Documents textuels avec images d'illustration	JP2 (JPEG2000, lossy), 300 dpi, couleurs, ratio 16:1 ¹⁰³ . Si documents exceptionnels, ratio 4:1.	75-95 %
Documents textuels avec images essentielles à leur compréhension ¹⁰⁴	JP2 (JPEG2000, lossy), 300 dpi, couleurs, ratio 4:1.	75 %
Documents visuels (cartes, plans, graphiques, journaux, photographies)	JP2 (JPEG2000, lossy), 300 dpi, couleurs, ratio 4:1.	75 %
Documents à valeur probante	JP2 (JPEG2000, lossless), 300 dpi, couleurs.	50 % en moyenne (dépend du contenu)

Conclusion

En tant que consommatrices des technologies de l'information et de la communication, les organisations culturelles impliquées dans la numérisation ou la préservation d'objets numériques participent à leur impact environnemental, estimé à 3,7 % des émissions mondiales de CO₂. Si leur responsabilité peut encore sembler modeste aujourd'hui, le développement de la numérisation des archives, l'avènement du tout numérique et le caractère cumulatif des collections numériques ne feront qu'amplifier leur rôle.

C'est pourquoi Ben Goldman, Keith L. Pendergrass, Walker Sampson, Tessa Walsh et Laura Alagna soutiennent qu'un changement de paradigme est nécessaire dans les pratiques et les théories des professionnels de la conservation du patrimoine numérique. Plutôt que de viser systématiquement la maximisation de l'intégrité et de l'accessibilité des archives, les organisations devraient repenser leurs stratégies d'évaluation, de pérennisation et de diffusion

¹⁰² Une limite identifiée lors de l'analyse de cette catégorie est que la Commission européenne conserve des documents dans de nombreuses langues différentes, et notamment dans des alphabets différents (latin, cyrillique, grec, arabe, etc.). Certains caractères de ces alphabets sont particulièrement subtils, et pourraient demander une plus haute résolution lors de la numérisation. Je pense cependant que l'analyse reste valable en l'état, et que l'évolution dans les techniques d'océrisation permet de compenser.

¹⁰³ Ratios issus du référentiel de la BnF. « JPEG2000 » in *Référentiel de format de fichier image*, op. cit., p. 4.

¹⁰⁴ Un exemple serait un rapport d'enquête produit par une entité au sein de la Commission européenne. Les images contenues dans ce rapport ont une valeur probatoire et nécessaire à la compréhension des de l'enquête.

des contenus, afin de réduire la quantité de données stockées et, par conséquent, diminuer l’empreinte environnementale des technologies de l’information.

La Commission européenne, en tant que l’un des principaux moteurs de la numérisation du patrimoine culturel européen et acteur engagé dans la lutte contre le changement climatique, avec pour objectif d’atteindre la neutralité carbone d’ici 2050¹⁰⁵, devrait voir son service d’archives montrer l’exemple en matière de numérisation responsable.

Pour ce faire, ce travail s’est attaché à explorer les pratiques de numérisation du Service des Archives Historiques de la Commission européenne (SAH) dans une perspective de durabilité, en répondant à la question de Pendergrass et al. : *What file format(s) are you generating ? Would different formats result in smaller storage demands while satisfying preservation and user needs ?* À travers une analyse fonctionnaliste de l’intégrité des archives et des besoins des utilisateurs, et une comparaison des formats TIFF, PDF/A-2b, JP2 (JPEG2000) et PNG, nous avons conclu que le format JP2, avec une compression avec perte, offre le meilleur équilibre entre intégrité des archives, besoins des utilisateurs et réduction de l’empreinte écologique.

La perte de données induite par la compression JPEG2000 avec perte, bien qu’en contradiction avec certaines recommandations traditionnelles relatives aux formats de fichiers master, lesquelles préconisent une numérisation en qualité maximale, sans compression, apparaît comme une option préférable. Selon la littérature et nos propres tests effectués sur 319 pages numérisées ainsi que deux documents iconographiques complexes en couleur, cette méthode de compression permet des économies de stockage significatives : jusqu’à 95 % pour les documents textuels avec une compression à 20:1, et environ 75 % pour les images en couleur, tout en préservant la lisibilité et l’intégrité fonctionnelle des documents. L’intégrité selon InterPARES est conservée. Ces gains de stockage contribuent à réduire l’impact environnemental ainsi que les coûts économiques liés aux activités de numérisation, notamment dans le contexte du SAH, où les fichiers master, stockés en *hot storage* (NAS).

Ces résultats demeurent toutefois exploratoires, et restent contextuels. Par exemple, il n’est pas certain que l’application de la compression soit pertinente dans toutes les situations. Dans les cas où les données sont conservées dans un système de *cold storage*, les coûts financiers et environnementaux associés aux ressources de calcul supplémentaires nécessaires

¹⁰⁵ « 2050 long-term strategy », in *Climate Action, European Commission*, [en ligne], https://climate.ec.europa.eu/eu-action/climate-strategies-targets/2050-long-term-strategy_en, consulté le 26/07/2025.

à la compression et à la décompression pourraient ne pas être justifiés par les économies de stockage obtenues¹⁰⁶. Par ailleurs, d'autres formats ou méthodes de compression émergentes, tels que WebP, AVIF ou JPEG XL, auraient également pu être envisagés, bien qu'ils sont récents et demeurent encore peu étudiés dans le domaine de la numérisation patrimoniale. Enfin, pour les institutions disposant de ressources techniques et financières limitées, le PDF/A-2b peut constituer une solution plus accessible pour les documents textuels, tandis que le TIFF non compressé reste un choix prudent et éprouvé pour les documents iconographiques.

¹⁰⁶ BHAT W.A., « Long-term preservation of big data: prospects of current storage technologies in digital libraries », *op. cit.*, 2018, p. 548-560.

Bibliographie

Sources et travaux

Travaux :

- BAKAL J.W., PAWADSHETTY J. S., « JPEG2000 Region of Interest Coding Methods », in *International Journal of Engineering Research and Applications*, vol. 3 (Janvier-février 2013), n°1, p. 1184-1188.
- BARINA D., KLIMA O., « JPEG 2000 : guide for digital libraries », in *Digital Library Perspectives*, vol. 36 (2020), n°3, p. 249-263.
- BHAT W.A., « Long-term preservation of big data: prospects of current storage technologies in digital libraries », in *Library Hi Tech*, vol. 36 (2018), n°3, p. 539-555.
- BOXER A., « The Rise and Fall and Rise of JPEG2000 », conférence présentée au *Free and open source software developers' European meeting (FOSDEM)*, Bruxelles, 2 février 2020, <https://youtu.be/UGXeRx0Tic4?si=-BB7QJONzG9f6Hhj>.
- « Breaking Down PNG vs TIFF », in *Cloudinary*, [en ligne], <https://cloudinary.com/guides/image-formats/png-vs-tiff>, mis à jour le 9 mai 2025, consulté le 08/07/2025.
- BROWN T., « Digital Imaging : is JP2 the file format of choice for the museum and heritage sector? », in *Museum+ Heritage Advisor*, [en ligne], <https://museumsandheritage.com/advisor/insights/digital-imaging-is-jp2-the-file-format-of-choice-for-the-museum-and-heritag/>, publié le 18 novembre 2014, consulté le 08/07/2025.
- BUCKLEY R., « Error Correction and Recovery of JPEG 2000 Compressed Images in Digital Preservation », in *Proc. IS&T Archiving 2012*, 2012, p. 270-274.
- BUCKLEY R., « Using Lossy JPEG 2000 Compression For Archival Master Files », version 1.1, Library of Congress - Office of Strategic Initiatives, 12 mars 2013.
- BUONORA P. et LIBERATI F., « A Format for Digital Preservation of Images: A Study on JPEG 2000 File Robustness », in *D-Lib Magazine*, vol. 14 (juillet 2008), n° 7/8, <https://www.dlib.org/dlib/july08/buonora/07buonora.html>.
- BUSHEY J., DE SOUZA M., FUKUSHIMA K., IGLÉSIAS D., RIESS M., SENGSAVANG E., STANCIC H., TRBUSIC Z., *Report on the Survey "Digitization and Artificial Intelligence for Archives and Documentary Heritage Materials"*, InterPARES Trust AI, mai 2025, https://interparestrustai.org/assets/public/dissemination/RA03-InterPARES-AI-Survey_Report_FINAL.pdf.

- BUSSARD S., « Les archives numérisées de la SdN vont-elles permettre une réécriture de l'histoire ? », in *Le Temps*, [en ligne], <https://www.letemps.ch/monde/archives-numerisees-sdn-vontelles-permettre-une-reecriture-lhistoire>, publié le 30 janvier 2022, consulté le 08/07/2025.
- CAPURRO C., *Digitizing european culture : Europeana and the government of digital cultural heritage*, thèse en vue de l'obtention du grade de docteur en philosophie, Utrecht University, Utrecht, 2021.
- CARON B., « Formats de données pour la préservation à long terme : la politique de la BnF », [Rapport Technique], Bibliothèque Nationale de France, Paris, 2021, identifiant : Hal-03374030v2, <https://bnf.hal.science/hal-03374030>.
- CHAPMAN S., DUPLOUY L., KUNZE J. BLAIR S., INTERNET ARCHIVE, ABRAMS S., LUPOVICI C., JENSEN A., JOHNSTON D., « Page Image Compression for Mass Digitization », in *Proc. IS&T Archiving 2007*, 2007, p. 37-42.
- COCHRANE E., « Obsolescence », in *Digital Preservation : A Critical Vocabulary*, [en ligne], <https://digital-preservation-a-critical-vocabulary.pubpub.org/pub/ajzjdqk4>, publié le 28 septembre 2024, consulté le 28/07/2025.
- « Comparing TIFF Compression Algorithms », in *IDR Solutions*, [en ligne], <https://blog.idrsolutions.com/comparing-tiff-compression-algorithms/>, publié le 7 mai 2025, consulté le 07/07/2025.
- COSSU S., PILLAY R., ROBSON G., SMITH M.D., « Evaluating HTJ2K as a Drop-In Replacement for JPEG2000 with IIF », in *Code4lib Journal*, n°57 (29 août 2023), <https://journal.code4lib.org/articles/17596>.
- DAVIES K., MAY P., « Practical Analysis of TIFF file size reduction achievable through compression », article présenté à la conference *iPRES 2016*, Swiss National Library, Bern, 2016.
- DAVIS K., « TIFF compression? Use Discretion! », in *Open Preservation Foundation*, [en ligne], <https://openpreservation.org/blogs/compression-at-your-discretion/>, publié le 9 février 2016, consulté le 07/07/2025.
- *Digital Preservation Handbook*, deuxième édition traduire en français, <https://www.dpconline.org/handbook>, Digital Preservation Coalition, 2021.
- DUPLOUY L., « JPEG2000 as a preservation format for digitization: lessons learned from a library », in *Proc. IS&T Archiving 2017*, 2017, p. 157-158.

- DURANTI L., « Pour une diplomatie des documents électroniques », in *Bibliothèque de l'École des chartes*, vol. 161 (2003), n° 2, p. 603-623.
- EHRENFELD J.R., *Sustainability by Design: A Subversive Strategy for Transforming Our Consumer Culture*, Yale University Press, 2008.
- FAASE J., GRAVENHORST C., « Quantity Meets Quality : Towards a digital library », [IFLA International Newspaper Conference 2011- Papers.pdf](#), conférence présentée à l'*International Newspaper Conference*, 25-27 avril 2011, Kuala Lumpur, p. 2-10.
- FERREBOEUF H. (dir.), « Lean ICT : Toward Digital Sobriety », *The Shift Project*, mars 2019, https://theshiftproject.org/app/uploads/2025/02/Lean-ICT-Report_The-Shift-Project_2019.pdf, p. 4.
- FREITAG C., BERNERS-LEE M., WIDDICKS K., KNOWLES B., BLAIR G. S., FRIDAY A., « The real climate and transformative impact of ICT : A critique of estimates, trends, and regulations », in *Patterns*, vol. 3 (12 août 2022), n°8, p. 1-18.
- FRIEDEWALD M., SZÉKELY I. et KARABOGA M., « Preserving the Past, Enabling the Future: Assessing the European Policy on Access to Archives in the Digital Age », in *Preservation, Digital Technology & Culture*, vol. 53 (1 juillet 2024), n° 2, p. 63-65.
- GAUDIAUT T., « Données numériques : Le Big Bang du Big Data », in *Statistica*, [en ligne], <https://fr.statista.com/infographie/17800/big-data-evolution-volume-donnees-numeriques-genere-dans-le-monde>, consulté le 26 juin 2025.
- GILLESSE R., ROG J., VERHEUSSEN A., « Life Beyond Uncompressed TIFF : Alternative File Formats for the Storage of Master Image Files », in *Proc. IS&T Archiving 2008*, 2008, p. 41-46.
- GOLDMAN B., « It's Not Easy Being Green(e): Digital Preservation in the Age of Climate Change », in WEIDEMAN C. et CALDERA M.A. (dir.), *Archival Values: Essays in Honor of Mark Greene*, Chicago, Society of American Archivists, 2019, p. 279.
- HAN Y., XUHENG W., « Digitization of Text Documents Using PDF/A », in *Information Technology and Libraries*, vol. 37 (2018), n°1, p. 52-64.
- HIRVONEN L., « Survey of Digitization in Archives », *The National Archives of Finland*, publié le 19 décembre 2017, [en ligne], https://kansallisarkisto.fi/documents/141232930/150411434/Liite_2_Digitization_Survey_2017.pdf.
- JAN NAUTA G., TEUNISSE S., VAN DEN HEUVEN W., *Access to Digital Ressources of European Heritage : Report on ENUMERATE survey 4*, Europeana DSI 2, Deliverable

- D4.4, 31 août 2017. Europeana PRO, [en ligne], https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_DSI-2/Deliverables/d4.4-report-on-enumerate-core-survey-4.pdf.
- KELECHAVA B., « Why JPEG 2000 Never Took Off », in *American National Standard Institute*, [en ligne], <https://blog.ansi.org/ansi/why-jpeg-2000-never-used-standard-iso-iec/>, publié le 10 juillet 2018, consulté le 27/07/2025.
 - KIDD T., « Choosing What Digital Image Format Works Best for You », in *Backstage Library Works*, [en ligne], <https://bslw.com/choosing-a-digital-image-format/>, publié le 16 mars 2021, consulté le 27/07/2025.
 - KYLE R., « JPEG2000 and Digitisation : Export round table », in *TownsWeb Archiving*, [en ligne], <https://blog.townswebarchiving.com/2015/10/jpeg2000-and-digitisation-expert-round-table>, publié le 19 octobre 2015, mis à jour le 20 janvier 2020, consulté le 27/07/2025.
 - LEFURGY B., « Is JPEG-2000 a Preservation Risk? », in *Library of Congress Blogs*, [en ligne], <https://blogs.loc.gov/thesignal/2013/01/is-jpeg-2000-a-preservation-risk/>, consulté le 08/07/2025.
 - MOSS M. et GOLLINS T., « Our Digital Legacy: an Archival Perspective », in *Journal of Contemporary Archival Studies*, vol. 4 (8 décembre 2017), n° 2, p. 7-8.
 - NATIONAL DIGITAL STEWARSHIP ALLIANCE, « 2021 Fixity Survey Report », octobre 2021, [en ligne], <https://osf.io/2qkea/>.
 - OTEMUYIWA P., « The Pros and Cons of JPEG 2000 and Other Common Image Formats », in *Cloudinary*, [en ligne], https://cloudinary.com/blog/the_great_jpeg_2000_debate_analyzing_the_pros_and_cons_to_widespread_adoption, publié le 6 septembre 2020, consulté le 25 juillet 2025.
 - PANDEY R. et KUMAR V., « Exploring the Impediments to Digitization and Digital Preservation of Cultural Heritage Resources: A Selective Review », in *Preservation, Digital Technology & Culture*, vol. 49 (2020), n° 1, p. 26-37.
 - PENDERGRASS K.L., SAMPSON W., WALSH T. et ALAGNA L., « Toward Environmentally Sustainable Digital Preservation », in *The American Archivist*, vol. 82 (1 mars 2019), n° 1, p. 180-184.
 - PILLAY R., « TIFF Image Encoding: Optimizing for Size, Speed and Quality », in *IIP*, [en ligne], <https://iipimage.sourceforge.io/2024/12/tiff-image-encoding-optimizing-for-size-speed-and-quality>, publié le 3 décembre 2024, consulté le 26/07/2025.

- RINEHART A.K., PRUD'HOMME P.-A. et HUOT A.R., « Overwhelmed to action: digital preservation challenges at the under-resourced institution », in *OCLC Systems & Services*, vol. 30 (4 février 2014), n° 1, p. 28-42.
- ROG J., « Compression and digital preservation: do they go together? », in *Proc. IS&T Archiving 2007*, 2007, p. 81.
- ROSENTHAL D., « The Prostate Cancer of Preservation re-examined », in *DSHR's Blog*, [en ligne], <https://blog.dshr.org/2015/09/the-prostate-cancer-of-preservation-re.html>, publié le 16 septembre 2015, consulté le 28/07/2025.
- « Survey on Digitization, Access, and Digital Preservation », CRKN, 27 juillet 2020, https://www.crkn-rcdr.ca/sites/crkn/files/2020-10/HCPGTG%20Survey_EN%20%28Member%20responses%29.pdf, p. 18.
- THOMAS M., « Numérisation : sous quel format conserver les documents ? », in *Archimag*, [en ligne], <https://www.archimag.com/demat-cloud/2021/03/29/numerisation-format-conserver-documents>, publié le 29 mars 2021, consulté le 08/07/2025.
- « What Are the Different Versions of PDF/A? », in *Apryse*, [en ligne], <https://apryse.com/blog/pdfa-format/what-are-the-different-types-of-pdfa>, publié en janvier 2019, consulté le 08/07/2025.

Ressources en ligne :

- « Application for JPEG 2000 », in *Jpeg.org*, [en ligne], <https://jpeg.org/jpeg2000/applications.html>, consulté le 27/07/2025.
- « International Comparison of Recommended File Formats », in *Open Preservation Foundation*, [en ligne], <https://openpreservation.org/resources/member-groups/international-comparison-of-recommended-file-formats/>, consulté le 05/07/2025.
- Sites de l'Union européenne :
 - « 2050 long-term strategy », in *Climate Action, European Commission*, [en ligne], https://climate.ec.europa.eu/eu-action/climate-strategies-targets/2050-long-term-strategy_en, consulté le 26/07/2025.
 - European Commission: Directorate-General for Communications Networks, Content and Technology, *The future of Europe's past – Why Member States must do more to advance digitisation of our cultural heritage – Implementation of the 2021 Commission recommendation on a common European Data Space*

for Cultural Heritage – Progress report 2021-2023, Publications Office of the European Union, 2024, <https://data.europa.eu/doi/10.2759/5921820>.

- Commission Recommendation of 27 October 2011 on the digitisation and online accessibility of cultural material and digital preservation (OJ L 283 29.10.2011, p. 39, ELI: <http://data.europa.eu/eli/reco/2011/711/oj>).
 - Commission Recommendation (EU) 2021/1970 of 10 November 2021 on a common European data space for cultural heritage (OJ L 401 12.11.2021, p. 5, ELI: <http://data.europa.eu/eli/reco/2021/1970/oj>).
 - *CULTURE AND CULTURAL HERITAGE : NEWSROOM*, « Access to archives in the digital age », Conseil de l'Europe, [en ligne], [Access to archives in the digital age - Culture and Cultural Heritage](#), publié le 18 avril 2024, consulté le 25 juin 2025.
 - Règlement (CE, Euratom) n° 1700/2003 du Conseil du 22 septembre 2003 modifiant le règlement (CEE, Euratom) n° 354/83 concernant l'ouverture au public des archives historiques de la Communauté économique européenne et de la Communauté européenne de l'énergie atomique (OJ L 243 27.09.2003, p. 1, ELI: <http://data.europa.eu/eli/reg/2003/1700/oj>).
 - Règlement (UE) 2015/496 du Conseil du 17 mars 2015 modifiant le règlement (CEE, Euratom) n° 354/83 en ce qui concerne le dépôt des archives historiques des institutions à l'Institut universitaire européen de Florence (OJ L 79 25.03.2015, p. 1, ELI: <http://data.europa.eu/eli/reg/2015/496/oj>).
- « JPEG vs JPEG2000 », in *Adobe*, [en ligne], <https://www.adobe.com/mx/creativecloud/file-types/image/comparison/jpeg-vs-jpeg-2000.html>, consulté le 31/07/2025.
 - « JPEG2000 » in *Référentiel de format de fichier image*, deuxième version, Bibliothèque Nationale de France, service numérisation, 2015, https://www.bnf.fr/sites/default/files/2018-11/ref_num_fichier_image_v2.pdf.
 - MAY P., PENNOCK M., WHEATLEY P., « Preservation Assessment: JP2 Format », version 1.3., *Digital Preservation Team*, British Library, 03/09/2015.
 - « Setting Standards (JPEG 2000) », in *Digital Preservation, Library of Congress*, [en ligne], https://www.digitalpreservation.gov/series/challenge/formats2_challenge.html, consulté le 08/07/2025.

- STELMACH M., WILLIAMS D., « Technical Guidelines for Digitizing Cultural Heritage Materials », FADGI, 2023, <https://www.digitizationguidelines.gov/guidelines/digitize-technical.html>, p. 41.
- « Sustainability of Digital Formats : Planning for Library of Congress Collections », in *Library of Congress*, [en ligne], <https://www.loc.gov/preservation/digital/formats/index.html>, consulté entre le 03/07/2025 et le 28/07/2025.
 - « JPEG 2000 Part 1 (Core) jp2 File Format », in *Sustainability of Digital formats: Planning for Library of Congress Collections*, [en ligne], <https://www.loc.gov/preservation/digital/formats/fdd/fdd000143.shtml>, consulté le 08/07/2025.
 - « PDF/A-2b, PDF for Long-term Preservation, Use of ISO 32000-1 (PDF 1.7), Level B Conformance », in *Sustainability of Digital formats: Planning for Library of Congress Collections*, [en ligne], <https://www.loc.gov/preservation/digital/formats/fdd/fdd000322.shtml>, consulté le 08/07/2025.
 - « PNG, Portable Network Graphics », in *Sustainability of Digital formats: Planning for Library of Congress Collections*, [en ligne], [PNG, Portable Network Graphics](https://www.loc.gov/preservation/digital/formats/fdd/fdd000074.shtml), consulté le 08/07/2025.
 - « TIFF Bitmap with LZW Compression », in *Sustainability of Digital formats: Planning for Library of Congress Collections*, [en ligne], <https://www.loc.gov/preservation/digital/formats/fdd/fdd000074.shtml>, consulté le 06/07/2025.
 - « TIFF, Revision 6.0 », in *Sustainability of Digital formats: Planning for Library of Congress Collections*, [en ligne], [TIFF, Revision 6.0](https://www.loc.gov/preservation/digital/formats/fdd/fdd000074.shtml), consulté le 03/07/2025.
- « Terminology Database », in *InterPARES Trust*, [en ligne], <http://interparestrust.org/terminology>, consulté le 29/07/2025.
 - « Integrity », in *InterPARES Trust*, [en ligne], <http://interparestrust.org/terminology/term/integrity>, consulté le 03/07/2025.
- THE FADGI STILL IMAGE WORKING GROUP, « Raster Still Images for Digitization. A Comparison of File Formats. Part 3. Narrative and Summary Table », FADGI, 29 août 2014.

- « Volume of data/information created, captured, copied and consumed worldwide from 2010 to 2023 with forecasts from 2024 to 2028 », in *Statistica*, [en ligne], <https://www.statista.com/statistics/871513/worldwide-data-created/>, consulté le 26 juin 2025.
- « What is JPEG 2000 ? », in *AIRBUS*, [en ligne], <https://www.intelligence-airbusds.com/en/8724-jpeg-2000-format>, consulté le 27/07/2025.

Sources :

Courriers électroniques :

- Courrier électronique de Lieven Baert, le 01/05/2025.
- Courrier électronique d'Alexis Bigonville, le 03 et 04/07/2025.
- Courrier électronique de Sylvia Pérez, le 04/07/2025.

Documents internes à la Commission européenne :

- « 20250401_List_of_SIPs_submitted_to_a-REP_stats », document interne à la Commission, sans identifiant, avril 2025.
- « Draft report_19NOV12 », in *Digitisation of the archives of the EU institutions, common challenges and approaches*, document interne à la Commission européenne, Ares(2013)100912, 19/11/2012.
- « European Commission Historical Archives Service, UCL Visit », PowerPoint fourni après visite du dépôt de Kortenberg par les étudiants en archivistique de l'UCL, sans identifiant, 15/10/2024, [Les outils et stratégies de préservation Kortenberg.pptx](#).
- « Projet ARCHIS-scanning. Draft vision document », document interne à la Commission européenne, oib.os.1.002(2013)3193594, 04/09/2013.

Interviews :

- Interview de Lieven Baert, le 24 avril 2025.
- Interview de Lieven Baert, le 16 mai 2025.

Notes de terrain :

- Notes prises au cours de la présentation de la procédure de numérisation du SAH par Alexis Bigonville le 6 mars 2025.

Annexe

- Annexe « Comparaisons_formats_et_resultats ».



Comparaisons_formats_et_resultats.xlsx

■ *Voir fichiers attachés à ce PDF.*

