



HAL
open science

Définir une politique formats : les neuf critères essentiels

Bertrand Caron, Marion Humbert, Anne-Flora Jolly, Isabelle Josse, Émeline Levasseur, Erwann Ramondenc, Martine Sin Blima Sin Blima-Barru

► To cite this version:

Bertrand Caron, Marion Humbert, Anne-Flora Jolly, Isabelle Josse, Émeline Levasseur, et al.. Définir une politique formats : les neuf critères essentiels. [Rapport de recherche] Association Aristote. 2020. hal-03182307

HAL Id: hal-03182307

<https://bnf.hal.science/hal-03182307>

Submitted on 26 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Définir une politique formats : les neuf critères essentiels.

CELLULE NATIONALE DE VEILLE SUR LES FORMATS

Sous-groupe « connaissance des formats existants et émergents et définition de critères d'obsolescence et de pérennisation »

Avril 2020

Bertrand CARON (Bibliothèque nationale de France)

Marion HUMBERT (Archives départementales de Moselle)

Anne-Flora JOLLY (Ministère des Armées)

Isabelle JOSSE (Ministère de l'Europe et des Affaires étrangères)

Emeline LEVASSEUR (Archives nationales)

Erwann RAMONDENC (Ministère de l'Europe et des Affaires étrangères)

Martine SIN BLIMA-BARRU (Archives nationales)



INTRODUCTION

Les premiers échanges au sein du sous-groupe ont abouti à l'ébauche de deux types de livrables :

- le rapport, par institution, de tous les formats présents dans leurs systèmes dès lors que le nombre de fichiers dépasse une proportion des fonds conservés qui reste à définir. En effet, ce « seuil critique » dépend de l'engagement de l'institution de conservation dans la préservation de certains formats prioritaires et de la stratégie adoptée (connaissance fine de la structuration de l'information, conversion, utilisation de conteneurs, etc.).
- le recensement et l'analyse des pratiques d'autres grandes institutions de conservation en matière de critères d'obsolescence et de pérennisation.

La variété des contextes et des politiques dans les institutions de conservation rend peu pertinent l'établissement d'une liste générale de recommandations relatives aux formats. En effet, « il est fondamental de garder en mémoire que la pérennité n'est pas une caractéristique intrinsèque d'un format mais découle de l'engagement de l'institution qui les utilise à disposer d'outils de lecture, d'analyse, de traitement, et de conversion, et des compétences nécessaires à leur utilisation¹ ».

Les pratiques et les critères de pérennisation propres à cinq grandes institutions de préservation ont été recensés : la Library of Congress², la British Library³, la Harvard Library⁴, la National Archives and Records Administration (NARA)⁵ et la Bibliothèque nationale de France (BnF). Cette sélection n'a pas vocation à être exhaustive, mais cherche à présenter les principaux facteurs retenus pour la conservation sur le long terme des données numériques⁶.

Outre le contexte et la stratégie propre à chacune d'entre elles, l'objectif est ici de tenter de cerner les points communs, à la fois techniques et stratégiques, présents dans leurs travaux, afin de proposer une base de réflexion à toute institution désireuse d'approfondir la question de la pérennité des formats. Une synthèse a été établie et a abouti à la proposition de neuf critères essentiels permettant de qualifier les critères de pérennisation d'un format sur le long terme. Ces informations peuvent être complétées par un historique du format (date de création, dates des dernières mises à jour, identité des créateurs, statuts du format) et tout autre complément utile à l'analyse. Par ailleurs, chacun de ces neuf points entraîne des conséquences en termes de ressources humaines et matérielles qu'il est fondamental d'envisager lors de l'élaboration d'une stratégie de préservation numérique. Certains critères peuvent être pondérés au profit d'autres, en fonction de la stratégie et des moyens de l'institution de conservation.

¹ *Formats de données pour la préservation à long terme : les recommandations BnF*, version de travail du 15 octobre 2019.

² Library of Congress, "Sustainability of Digital Formats: Planning for Library of Congress Collections" [en ligne] URL : <https://www.loc.gov/preservation/digital/formats> (consulté le 23 avril 2020).

³ British Library, "File Formats Assessments" [en ligne] URL : http://wiki.dpconline.org/index.php?title=File_Formats_Assessments (consulté le 23 avril 2020).

⁴ Harvard Library, "Assessment Tools", [en ligne] URL : <https://wiki.harvard.edu/confluence/display/digitalpreservation/Assessment+Tools> (consulté le 4 novembre 2019).

⁵ National Archives and Records Administration, "NARA digital preservation file format risk analysis and preservation plans" [en ligne] URL : <https://github.com/usnationalarchives/digital-preservation> (consulté le 24 avril 2020).

⁶ Voir aussi National Archives of Australia, « Long-term File Formats » [en ligne] URL : <https://www.naa.gov.au/information-management/storing-and-preserving-information/preserving-information/preserving-digital-information/long-term-file-formats> (consulté le 23 avril 2020).

TABLE DES MATIERES

Introduction.....	2
La préservation à long terme des formats en 9 points	4
ANNEXE 1 - Library of Congress.....	6
Sustainability factors	6
Critères pour l'évaluation de la pérennité	6
Format Description.....	6
Description de format	6
ANNEXE 2 - British Library	7
File format assessment factors	7
Critères pour l'évaluation du format de fichier.....	7
Format Preservation Assessment.....	7
Évaluation de l'éligibilité du format à la préservation	7
ANNEXE 3 - Harvard Library	10
Format matrix tool	10
L'outil « matrice de format ».....	10
ANNEXE 4 - National Archives and Records Administration	12
Sustainability factors	12
Critères pour l'évaluation de la pérennité	12
File format risk matrix	12
Matrice des risques à évaluer pour un format de fichier.....	12
ANNEXE 5 : Bibliothèque nationale de France	16
Critères de pérennité : tableau synthétique	16
Critères objectifs	16
Critères subjectifs.....	20

LA PRESERVATION A LONG TERME DES FORMATS EN 9 POINTS

Adoption	<i>Adoption</i>	Le terme renvoie à l'utilisation du format pour produire, conserver, diffuser ou échanger l'information. Plus l'adoption est large, plus les outils de création, de lecture, de transformation, d'émulation, etc. sont nombreux et plus le format est maintenu et mis à jour. Pour les services patrimoniaux, l'utilisation d'un format recommandé pour la conservation par d'autres institutions est un point important de l'adoption. L'âge du format et sa diffusion dans plusieurs pays sont notamment des facteurs permettant de mesurer le taux d'adoption.
Existence et disponibilité de la documentation	<i>Disclosure / Software support, Documentation and Guidance</i> ⁷	Le terme renvoie à la complétude, à la précision et à la compréhensibilité des spécifications disponibles pour le format (pouvant provenir d'une source institutionnelle mais pas forcément) et à l'existence d'outils accessibles pour identifier et valider le format. Le facteur peut par exemple être corroboré par l'existence d'outils de validation élaborés par différentes sources, preuve que les spécifications sont accessibles et complètes. Les formats ouverts présentent généralement une mise à disposition de la documentation plus complète que les formats propriétaires mais ce n'est pas forcément systématique.
Transparence	<i>Transparency / Complexity</i>	Le terme renvoie à l'accessibilité du format à l'aide d'outils non spécialisés, comme la possibilité pour un humain de lire et de comprendre le code du format grâce à des outils de type éditeur de texte. Les outils qui servent à produire le format doivent par ailleurs être utilisables sur des machines standard. La transparence suppose également que le format soit non compilé et encodé avec des caractères standardisés. La compression, si elle est inévitable, doit utiliser des algorithmes largement répandus ou précisément documentés par l'utilisateur.
Contenu additionnel embarqué	<i>Self-documentation / Embedded or Attached Content</i>	Capacité d'un format à comprendre au sein de son code les flux de données nécessaires à sa compréhension et son exploitation, à l'inverse d'un format qui obligerait l'utilisateur à stocker ces flux séparément de l'information primaire. Ces flux sont de plusieurs types : représentation (permet de transformer des données en information exploitable), identification, de contexte (lien entre une information et son environnement), d'intégrité (traçabilité des modifications), de provenance (historique de l'information) ⁸ ...
Dépendances externes	<i>External dependencies</i>	Nécessité d'utiliser des logiciels (notamment propriétaires) ou du matériel informatique spécifiques (joystick, microphone, etc.) pour exploiter le format. Ce facteur concerne particulièrement les contenus dynamiques, comme les jeux vidéo ou certaines œuvres d'art numérique ⁹ .
Incidence des brevets	<i>Impact of patents / Legal Issues</i>	Présence de brevets ou de toute contrainte juridique sur un format, qui peut entraîner des coûts de conservation importants et limiter le développement d'outils d'analyse et d'exploitation du format. Toutefois, la présence d'un

⁷ Les termes en italiques sont ceux utilisés par la British Library, cf *supra*.

⁸ Classification OAIS. *Modèle de référence pour un Système ouvert d'archivage*, Consultative Committee for Space Data Systems, octobre 2017 (version française) [en ligne] URL : <https://public.ccsds.org/Pubs/650x0m2%28F%29.pdf> (consulté le 3 janvier 2019), p. 12-13.

⁹ Voir par exemple "Supporting Virtual Reality and 3D in Academic Libraries : Defining Preservation and Curation Challenges", Ipres 2019 [en ligne] URL : <https://ipres2019.org/static/proceedings/IPRES2019.pdf> (consulté le 3 janvier 2020).

Sous-groupe « connaissance de formats existants et émergents et définition de critères d'obsolescence et de pérennisation »

		brevet ou d'une réglementation similaire doit être associée à l'analyse de l'adoption du format, bon indicateur pour savoir si le brevet empêche la large diffusion du format et la création d'outils d'exploitation.
Dispositifs techniques de protection	<i>Technical protection mechanisms</i>	Les dispositifs comme le chiffrement, la protection par mot de passe, éventuellement la présence d'informations de copyright sous forme de watermark, ne permettent pas toujours une libre manipulation de l'information par les services de conservation à des fins de diffusion, transformation, réparation, etc. D'autres dispositifs comme la signature électronique, n'empêchent pas la manipulation du contenu, mais sont conçus pour assurer son authenticité.
Résilience	<i>Resilience</i>	Capacité du format à permettre la lisibilité de son contenu, même en cas d'altération partielle de celui-ci.
Compacité	<i>Compacity</i>	Capacité du format à contenir un grand nombre de données dans un volume réduit. Si la compacité est liée à la compression, s'assurer de la transparence et de la réversibilité de la compression.

ANNEXE 1 - LIBRARY OF CONGRESS

Sustainability factors	Critères pour l'évaluation de la pérennité
------------------------	--

Disclosure	Existence et disponibilité de la documentation
Adoption	Adoption
Transparency	Transparence
Self-documentation	Métadonnées embarquées
External dependencies	Dépendances externes
Impact of patents	Incidence des brevets
Technical protection mechanisms	Dispositifs techniques de protection

Format Description	Description de format
--------------------	-----------------------

<ol style="list-style-type: none"> 1. Identification and description <ol style="list-style-type: none"> 1.1. Full name 1.2. Description 1.3. Production phase 1.4. Relationship to other formats 2. Local use <ol style="list-style-type: none"> 2.1. LC experience or existing holdings 2.2. LC preference 3. Sustainability factors 4. Quality and functionality factors 	<ol style="list-style-type: none"> 1. Identification et description <ol style="list-style-type: none"> 1.1. Nom complet 1.2. Description 1.3. Phase de production¹⁰ 1.4. Relations à d'autres formats 2. Usage au sein de l'institution <ol style="list-style-type: none"> 2.1. Expérience ou exploitations existantes au sein de la Library of Congress 2.2. Préférence de la Library of Congress 3. Critères pour l'évaluation de la pérennité 4. Critères de qualité et de fonctionnalité
--	---

Quality and functionality factors assess a format in relation to functional support for features and aspects that can affect the quality of content.

Les critères de qualité et de fonctionnalité visent à évaluer un format par rapport à la prise en charge fonctionnelle de caractéristiques et d'aspects qui peuvent affecter la qualité du contenu.

<ol style="list-style-type: none"> 5. File type signifiers and format identifiers <p><i>This table documents some of the signifiers that may be used by automated systems to identify a format or the data it contains.</i></p>	<ol style="list-style-type: none"> 5. Indicateur de type de fichier et identifiants de format <p><i>Cette table mentionne certains indicateurs qui peuvent être utilisés par des systèmes automatisés pour identifier un format ou les données qu'il contient¹¹.</i></p>
--	--

6. Notes

6. Notes

¹⁰ Ce critère précise à quel(s) stade(s) le format peut être utilisé : pour créer le document (phase initiale), le réviser (phase intermédiaire), le diffuser (phase finale).

¹¹ Il s'agit ici entre autres de la notion de signature de fichier, ou « magic number », voir [https://fr.wikipedia.org/wiki/Nombre_magique_\(programmation\)](https://fr.wikipedia.org/wiki/Nombre_magique_(programmation)).

ANNEXE 2 - BRITISH LIBRARY

File format assessment factors	Critères pour l'évaluation du format de fichier
--------------------------------	---

Development status	Stade de développement
Adoption and usage	Adoption et usage
Software support	Support logiciel
Documentation and guidance	Documentation et conseils

Complexity	Complexité
Embedded or attached content	Contenu embarqué ou joint

External dependencies	Dépendances externes
Legal issues	Contraintes juridiques
Technical protection mechanisms	Dispositifs techniques de protection
Other preservation risks	Autres risques de préservation

Format Preservation Assessment	Évaluation de l'éligibilité du format à la préservation
--------------------------------	---

1. Introduction	1. Introduction
1.1. Scope	1.1. Périmètre
1.2. Summary	1.2. Résumé
2. Assessment	2. Évaluation
2.1. Development status	2.1. Stade de développement

<i>A summary of the development history of the format and an indication of its current status</i>	<i>Un résumé de l'historique du développement du format et une présentation de son statut actuel</i>
---	--

2.2. Adoption and usage	2.2. Adoption et usage
-------------------------	------------------------

<i>An impression of how widely used the file format is, with reference to use in other memory organisations and their practical experiences of working with the format</i>	<i>Une indication sur le degré d'adoption du format, avec les références à l'utilisation faite dans d'autres organismes de conservation mémorielle et à leurs expériences pratiques de gestion de ce format</i>
--	---

2.3. Software support	2.3. Support logiciel
2.3.1. Rendering Software Support	2.3.1. Support logiciel de restitution

Sous-groupe « connaissance de formats existants et émergents et définition de critères d'obsolescence et de pérennisation »

An overall impression of software support for rendering the format with the reference to typical desktop software; and current support on British Library reading room PCs

Une indication générale sur le support logiciel nécessaire pour restituer le format avec les références aux logiciels de bureau courants et le support actuel existant sur les postes informatiques de la salle de lecture de la British Library

2.3.2. Preservation Software Support

2.3.2. Support logiciel de préservation

An impression of the availability and effectiveness of software for managing and preserving instances of the file format

Une indication sur l'accessibilité et l'efficacité des logiciels pour gérer et préserver les instances du format de fichier

Format identification
Validation and detecting preservation risks

Conformance checking
Metadata extraction
Content extraction
Migration

Identification de format
Validation et analyse des risques en matière de préservation
Vérification de conformité
Extraction de métadonnées
Extraction de contenu
Migration

2.4. Documentation and guidance

2.4. Documentation et recommandations

An indication of the availability of practical documentation or guidance with specific reference to the facilitation of any recommended actions

Une indication sur la disponibilité d'une documentation ou de recommandations pratiques avec les références spécifiques pour faciliter la mise en œuvre des actions recommandées

2.5. Complexity

2.5. Complexité

An impression of the complexity of the format with respect to the impact this is likely to have on the British Library managing or working with content in this format. What level of expertise in the format is required to have confidence in management and preservation?

Une indication sur la complexité du format concernant l'impact qu'il est susceptible d'avoir pour la British Library en termes de gestion et de traitement du contenu de ce format. Quel niveau d'expertise est requis pour avoir confiance dans sa gestion et sa préservation ?

2.6. Embedded or attached content

2.6. Contenu embarqué ou joint

The potential for embedding or attaching files of similar or different formats, and the likely implications of this

La possibilité d'embarquer ou de joindre des fichiers de formats similaires ou différents et ses implications possibles

2.7. External dependencies

2.7. Dépendances externes

An indication of the possibility of content external to an instance of the file format that is complimentary or even essential to the intellectual content of the instance

Une indication sur la possible existence d'un contenu externe à une instance du format de fichier qui est complémentaire ou même essentiel au contenu intellectuel de l'instance.

Sous-groupe « connaissance de formats existants et émergents et définition de critères d'obsolescence et de pérennisation »

2.8. Legal issues	2.8. Contraintes juridiques
<i>Legal impediments to the use, management or preservation of instances of the file format</i>	<i>Obstacles juridiques à l'utilisation, la gestion ou la préservation des instances du format de fichier</i>
2.9. Technical protection mechanisms	2.9. Dispositifs techniques de protection
<i>Encryption, Digital Rights Management and any other technical mechanisms that might restrict usage, management or preservation of instances of the file format</i>	<i>Chiffrement, gestion des droits numériques et tout autre dispositif technique qui pourrait restreindre l'usage, la gestion ou nuire à la préservation des instances du format de fichier</i>
2.10. Other preservation risks	2.10. Autres risques de préservation
<i>Other evidence based preservation risks, noting that many known preservation risks are format specific and do not easily fit under any of the sustainability factors above</i>	<i>Autres risques de préservation basés sur des faits, il convient de noter que nombre de risques de préservation connus sont propres à un format et se rattachent difficilement à l'un des critères de pérennité cités ci-dessus</i>
2.11. Preservation risk summary	2.11. Résumé des risques en matière de préservation
<i>A summary of preservation risks and recommended actions (where possible)</i>	<i>Un résumé des risques en matière de préservation et des actions recommandées (si possible)</i>
3. Recommendations for action	3. Actions recommandées
<i>Recommended actions in usage and handling of the format. Recommend actions in the support or development of software applications that provide, or have the potential to provide, significant risk mitigation for the format. Note that these recommendations do not take into account other requirements such as those driven by specific British Library collections, or non-preservation issues such as resourcing.</i>	<i>Actions recommandées dans l'usage et le traitement du format. Actions recommandées, en matière de support ou de développement d'applications logicielles, qui fournissent ou peuvent potentiellement fournir une atténuation significative du risque pour le format. Il convient de noter que ces recommandations ne prennent pas en compte les autres exigences telles que celles concernant les collections particulières de la British Library.</i>
Handling recommendations	Recommandations de traitement
Knowledge recommendations	Recommandations en matière de constitution de connaissances
Software recommendations	Recommandations de logiciels
Monitoring recommendations	Recommandations de contrôle
4. References	4. Références

ANNEXE 3 - HARVARD LIBRARY

Format matrix tool	L'outil « matrice de format »
<p>1. Sustainability factors</p> <p>1.1. Adoption</p> <p>1.1.1. Dependency on a single organization or company</p> <p>1.1.2. Widespread use by consumers</p> <p>1.1.3. Widespread use by professionals</p> <p>1.1.4. Archival use</p> <p>1.1.5. Community / 3rd party support</p> <p>1.1.6. Developer / corporate support</p> <p>1.1.7. Geographic spread</p> <p>1.1.8. Lifetime</p> <p>1.1.9. Revision rate</p> <p>1.1.10. Last revision</p> <p>1.1.11. Format expertise available</p> <p>1.1.12. Backward compatibility</p> <p>1.2. Disclosure</p> <p>1.2.1. Availability of specifications</p> <p>1.2.2. Degree to which specification is complete and understandable</p> <p>1.2.3. Standardized</p> <p>1.3. Transparency</p> <p>1.3.1. Level of format complexity</p> <p>1.3.2. Ease of accurate identification</p> <p>1.3.3. Degree to which compression is understood</p> <p>1.3.4. Difficulty of tool development for accurate identification</p> <p>1.3.5. Ease of accurate validation</p> <p>1.4. Self-documentation</p> <p>1.4.1. Error-tolerance</p> <p>1.4.2. Technical metadata support</p>	<p>1. Critères pour l'évaluation de la pérennité</p> <p>1.1. Adoption</p> <p>1.1.1. Dépendance à l'égard d'une seule organisation ou société</p> <p>1.1.2. Utilisation répandue parmi les consommateurs</p> <p>1.1.3. Utilisation répandue parmi les professionnels</p> <p>1.1.4. Usage à des fins archivistiques</p> <p>1.1.5. Soutenu par une communauté ou une tierce partie</p> <p>1.1.6. Soutenu par un développeur ou une entreprise</p> <p>1.1.7. Diffusion géographique</p> <p>1.1.8. Durée de vie</p> <p>1.1.9. Taux de révision</p> <p>1.1.10. Dernière révision</p> <p>1.1.11. Disponibilité de l'expertise sur le format</p> <p>1.1.12. Rétrocompatibilité</p> <p>1.2. Existence et disponibilité de la documentation</p> <p>1.2.4. Disponibilité des spécifications</p> <p>1.2.5. Niveau de complétude et d'intelligibilité des spécifications</p> <p>1.2.6. Standardisé</p> <p>1.3. Transparence</p> <p>1.3.1. Niveau de complexité du format</p> <p>1.3.2. Facilité d'identification exacte</p> <p>1.3.3. Degré de compréhension de la compression</p> <p>1.3.4. Difficulté de développer des outils pour une identification exacte</p> <p>1.3.5. Facilité de validation exacte</p> <p>1.4. Métadonnées embarquées</p> <p>1.4.1. Tolérance aux erreurs¹²</p>

¹² En l'absence de documentation plus précise sur les champs de la matrice, nous pensons que ce point se rapporte à la présence de métadonnées d'intégrité dans les métadonnées embarquées.

Sous-groupe « connaissance de formats existants et émergents et définition de critères d'obsolescence et de pérennisation »

1.4.3.Descriptive metadata support	1.4.2. Support de métadonnées techniques 1.4.3. Support de métadonnées descriptives
1.5. External dependencies 1.5.1.Browser support 1.5.2.Dependencies on particular HW/SW 1.5.3.Quantity and availability of rendering software	1.5. Dépendances externes 1.5.1. Compatibilité aux navigateurs 1.5.2. Dépendances à des Hardware/Software particuliers 1.5.3. Nombre et accessibilité des logiciels de restitution
1.6. Impact of patents 1.6.1.Legal restrictions affecting use now or long-term	1.6. Incidence des brevets 1.6.1.Restrictions légales affectant l'usage à court ou long terme
1.7. Technical protection mechanisms 1.7.1.Support for TPM	1.7. Dispositifs techniques de protection 1.7.2. Prise en charge du Trusted Platform Module (TPM)
2. Cost factors 2.1. Cost to maintain environment for access or processing 2.2. Uncompressed raw storage requirements 2.3. Storage (with maximum compression applied) requirements	2. Critères de coût 2.1. Coût de maintien de l'environnement nécessaire à l'accès et au traitement 2.2. Exigences de stockage brut sans compression 2.3. Exigences de compression (avec application de compression maximum)
3. Quality and functionality factors 3.1. Colorspaces supported 3.2. Max Bits per Color Channel 3.3. Max Channels Supported 3.4. Max Bit Depth Overall 3.5. Max resolution 3.6. Ability to encode in true lossless compression 3.7. Max Chroma subsampling 3.8. Compression ratio	3. Critères de qualité et de fonctionnalité 3.1. Espaces colorimétrique supportés 3.2. Nombre maximum de bits par canal de couleur 3.3. Nombre maximum de canaux supporté 3.4. Profondeur de bits maximale 3.5. Résolution maximale 3.6. Capacité à encoder via une compression fidèle et sans perte 3.7. Sous-échantillonnage de la chrominance maximum 3.8. Taux de compression
4. Additional notes	4. Notes additionnelles

ANNEXE 4 - NATIONAL ARCHIVES AND RECORDS ADMINISTRATION

Sustainability factors	Critères pour l'évaluation de la pérennité
------------------------	--

<p>Disclosure Adoption Transparency Self-documentation External hardware dependencies</p> <p>External software Dependencies Impact of Patents Technical protection mechanisms Format age</p>	<p>Existence et disponibilité de la documentation Adoption Transparence Contenu additionnel embarqué Dépendances externes à des matériels informatiques Dépendances externes à des logiciels Incidence des brevets Dispositifs techniques de protection Âge du format</p>
--	---

File format risk matrix	Matrice des risques à évaluer pour un format de fichier
-------------------------	---

<p>1. Disclosure</p> <p>1.1. Is the format proprietary?</p> <p>1.2. Does the format have a published open specification?</p> <p>1.3. Are there available tools that can validate the technical integrity of a file encoded in this format against the published specification?</p> <p>1.4. Has the specification been approved and published by an internationally recognized standards body?</p> <p>1.5. Is the available specification complete and accurate?</p> <p>2. Adoption</p> <p>2.1. Is the file format commonly used to create or maintain permanent records within the federal government?</p> <p>2.2. Is the file format commonly used outside the federal government?</p>	<p>1. Existence et disponibilité de la documentation</p> <p>1.1. S'agit-il d'un format propriétaire ?</p> <p>1.2. Les spécifications du format sont-elles ouvertes ?</p> <p>1.3. Existe-t-il des outils accessibles qui sont capables de valider la conformité d'un fichier encodé dans ce format au regard des spécifications publiées ?</p> <p>1.4. Les spécifications ont-elles été approuvées et publiées par un organisme de normalisation reconnu internationalement?</p> <p>1.5. Les spécifications du format sont-elles complètes et exactes ?</p> <p>2. Adoption</p> <p>2.1. Le format de fichier est-il utilisé couramment pour créer ou conserver des archives définitives au sein du gouvernement fédéral ?</p> <p>2.2. Le format de fichier est-il utilisé couramment en dehors du gouvernement fédéral ?</p>
---	--

Sous-groupe « connaissance de formats existants et émergents et définition de critères d'obsolescence et de pérennisation »

- | | |
|--|---|
| <p>2.3. Is the format actively maintained and updated by an organization, individual, or community?</p> <p>2.4. Are multiple renderers available?</p> <p>2.5. Have the archives or library communities identified the format as one they prefer for creation and transfer of permanent materials?</p> | <p>2.3. Le format est-il maintenu et mis à jour activement par une organisation, une personne physique ou une communauté ?</p> <p>2.4. Des outils de restitution multiples sont-ils disponibles ?</p> <p>2.5. Les communautés des services d'archives et des bibliothèques ont-elles identifié le format comme l'un des formats qu'elles utilisent de préférence pour la création et le transfert des matériaux permanents ?</p> |
| <p>3. Transparency</p> | |
| <p>3.1. Is the format human readable and can it be opened with a text editor?</p> <p>3.2. Does the available specification provide enough detail to allow basic analysis of the file format with hex editors or other inspection tools?</p> <p>3.3. Does the format rely on standard character or other encoding methods such as IEEE notations?</p> <p>3.4. Is the source code of the software used to create the format available for little or no cost?</p> <p>3.5. Is the software used to create the format supported by current computing environments?</p> <p>3.6. Does the format require the use of compression?</p> <p>3.7. Does the format support user-definable compression levels or other quality settings that affect essential characteristics?</p> | <p>3.1. Le format est-il lisible pour une personne et peut-il être ouvert avec un éditeur de texte ?</p> <p>3.2. Les spécifications disponibles fournissent-elles assez de détail pour permettre une analyse de base du format de fichier avec des éditeurs hexadécimaux ou d'autres outils de contrôle ?</p> <p>3.3. Le format repose-t-il sur des caractères standards ou des méthodes d'encodage telles que les notations IEEE ?</p> <p>3.4. Le code source du logiciel utilisé pour créer le format est-il disponible à un coût réduit ou gratuitement ?</p> <p>3.5. Le logiciel utilisé pour créer le format est-il supporté par des environnements informatiques courants ?</p> <p>3.6. Le format nécessite-t-il le recours à la compression ?</p> <p>3.7. Le format supporte-t-il des niveaux de compression paramétrables par l'utilisateur ou d'autres paramètres de qualité qui affectent des caractéristiques essentielles ?</p> |
| <p>4. Self-documentation</p> | |
| <p>4.1. Does the format support self-contained (embedded) descriptive metadata?</p> <p>4.2. Does the format support self-contained technical metadata?</p> <p>4.3. Does the format support self-contained administrative metadata?</p> <p>4.4. Does the format metadata adhere to an international standard?</p> <p>4.5. Is the format metadata robust enough for an accurate file analysis?</p> | <p>4. Contenu additionnel embarqué</p> <p>4.1. Le format contient-il des métadonnées descriptives portées par lui-même (embarquées) ?</p> <p>4.2. Le format contient-il des métadonnées techniques embarquées ?</p> <p>4.3. Le format contient-il des métadonnées administratives embarquées ?</p> <p>4.4. Les métadonnées du format se conforment-elles à un standard international ?</p> <p>4.5. Les métadonnées du format sont-elles assez étoffées pour une analyse précise du fichier ?</p> |

Sous-groupe « connaissance de formats existants et émergents et définition de critères d'obsolescence et de pérennisation »

- | | |
|--|---|
| 5. External hardware dependencies | 5. Dépendances externes à des matériels informatiques |
| 5.1. Does the format require a specific hardware environment, such as a specific graphics card, chipset, or memory requirements, to process or interact with it? | 5.1. Le format nécessite-t-il un environnement matériel informatique spécifique comme une carte graphique spécifique, un chipset, ou des besoins en mémoire pour traiter le format et interagir avec lui ? |
| 5.2. Does the format require specific playback hardware (e.g., Blu-Ray, Audio CD, etc) to transfer the format to the NARA environment? | 5.2. Le format nécessite-t-il un matériel de lecture spécifique (par exemple un lecteur Blu-Ray ou de CD audio, etc.) pour transférer le format dans l'environnement de la National Archives and Records Administration ? |
| 6. External software dependencies | 6. Dépendances externes à des logiciels |
| 6.1. Does the format rely on proprietary software to render or view files? | 6.1. Le format s'appuie-t-il sur des logiciels propriétaires pour restituer ou permettre la visualisation des fichiers ? |
| 6.2. Does the format rely on plug-ins, scripts, etc. to render or view files? | 6.2. Le format s'appuie-t-il sur des plug-ins, des scripts, etc., pour restituer ou permettre la visualisation des fichiers ? |
| 6.3. Does the format rely on specific computing operating system(s) to render or view files? | 6.3. Le format s'appuie-t-il sur un ou des système(s) d'exploitation informatique(s) spécifique(s) pour restituer ou permettre la visualisation des fichiers ? |
| 6.4. Does the format rely on special tools to render or view files ? | 6.4. Le format repose-t-il sur des outils spécifiques pour restituer ou permettre la visualisation des fichiers ? |
| 7. Impact of patents | 7. Incidence des brevets |
| 7.1. Is the format subject to patent claims that may impede the development of open source tools for opening and managing the files? | 7.1. Le format est-il soumis à des revendications de brevets qui peuvent empêcher le développement d'outils <i>open source</i> pour ouvrir et gérer les fichiers ? |
| 7.2. Have the patent claims expired? | 7.2. Les droits en revendications de brevet ont-ils expiré ? |
| 7.3. Are there fees associated with the format as a result of patent claims? | 7.3. Des frais sont-ils associés à l'usage du format du fait de revendications de brevet ? |
| 7.4. Does the format have open source license terms? | 7.4. Le format bénéficie-t-il des conditions d'une licence <i>open source</i> ? |
| 8. Technical protection mechanisms | 8. Dispositifs techniques de protection |
| 8.1. Does the format have capability to encrypt all or part of the resulting file? | 8.1. Le format a-t-il la capacité à chiffrer tout ou partie du fichier de sortie? |
| 8.2. Does the format require the use of encryption? | 8.2. Le format nécessite-t-il le recours au chiffrement ? |

Sous-groupe « connaissance de formats existants et émergents et définition de critères d'obsolescence et de pérennisation »

- | | |
|---|--|
| 8.3. Does the format support robust encryption? | 8.3. Le format supporte-t-il un chiffrement robuste ? |
| 8.4. Can technical protection measures (e.g. digital rights management) be applied? | 8.4. Des mesures techniques de protection peuvent-elles être appliquées (par exemple en termes de gestion des droits numériques) ? |
| 8.5. Does the format allow for embedded information such as watermarking? | 8.5. Le format permet-il la prise en charge d'informations embarquées comme un filigrane ? |
| 9. Format age | 9. Âge du format |
| 9.1. When was the Format Specification First Created? | 9.1. Quand les premières spécifications du format ont-elles été créées ? |
| 9.2. When was the Format Specification last updated? | 9.2. A quelle date a eu lieu la dernière mise à jour des spécifications du format ? |

ANNEXE 5 : BIBLIOTHEQUE NATIONALE DE FRANCE

Critères de pérennité : tableau synthétique

Critères objectifs

Intitulé	Définition	Justification
Communauté d'utilisateurs / Sociabilité	Le format est-il largement utilisé dans sa communauté cible ? Par le grand public ? Par les institutions de conservation ?	<p>L'utilisation d'un format au sein de sa communauté est un indice de son adaptation aux besoins spécifiques de cette communauté. Un format également utilisé au-delà des institutions de conservation fournit des garanties supplémentaires, car les moyens de telles institutions pour maintenir un format sont limités comparés à ceux des industries culturelles.</p> <p>Ce critère est lié à celui des outils disponibles : plus la communauté d'utilisateurs est conséquente, plus elle est susceptible d'avoir développé ou fait développer des outils adaptés.</p>
Documentation	Les spécifications du format sont-elles publiées ? Si oui, sont-elles maintenues par un organisme de normalisation reconnu ? Quel est leur coût ?	<p>Si les spécifications du format sont librement accessibles, il est possible à tout un chacun d'en comprendre la structure et, si le format est également libre, de développer des outils qui le prennent en charge.</p> <p>La documentation peut être partielle : ainsi certains industriels publient-ils des documents décrivant seulement la structure générale de leur format. Un processus de normalisation garantit que l'on dispose de spécifications décrivant l'intégralité des caractéristiques significatives d'un format.</p> <p>Les formats maintenus par des organismes de normalisation nationaux (AFNOR) ou internationaux (ISO, W3C, IETF, etc.) offrent de meilleures garanties de maintenance et de disponibilité des spécifications, mais peuvent faire payer leurs normes.</p> <p>On parle de standards <i>de facto</i> lorsqu'une spécification technique produite par une organisation est majoritairement adoptée dans</p>

Sous-groupe « connaissance de formats existants et émergents et définition de critères d'obsolescence et de pérennisation »

		une communauté bien qu'elle n'ait pas fait l'objet d'un processus de normalisation officiel.
Liberté d'utilisation	Existe-t-il des obstacles juridiques à l'utilisation du format ?	Si un format peut être totalement ouvert (documenté et utilisable par quiconque), il arrive également que des limitations d'usage pèsent sur des formats documentés, notamment en raison de brevets accordant des droits de propriété industrielle déposés au profit d'une organisation donnée. Ces brevets peuvent limiter ou interdire le développement d'outils prenant en charge le format.
Indépendance / autonomie	L'utilisation du format requiert-elle d'autres formats, encodages, environnement logiciels ou matériels ?	La consultation et l'utilisation d'un fichier numérique étant médiates, elles sont systématiquement dépendantes d'un environnement technique. Outre la dépendance à un environnement logiciel qui peut être propriétaire, abordée dans le critère « Liberté d'utilisation », l'utilisation de certains formats est dépendante d'environnements matériels, de bibliothèques logicielles, ou d'éléments habituellement non embarqués dans le fichier (par exemple, la dépendance de la plupart des PDF aux polices installées sur le poste de l'utilisateur).
Résilience / robustesse / tolérance à l'erreur	En cas d'altération d'une partie de son contenu, les parties intactes sont-elles toujours lisibles ?	<p>Ce critère évalue la résistance des fichiers de ce format à l'altération. Cette altération peut provenir d'une dégradation du support ou d'une erreur du matériel de lecture, mais elle est plus souvent encore le fait d'un transfert interrompu, notamment en raison d'une défaillance du réseau ou de la connectique. De ce fait, les formats de fichiers destinés à être échangés sur le réseau par le biais du <i>streaming</i> sont souvent conçus pour être robustes.</p> <p>La robustesse dépend donc de la structure du format. Elle peut être renforcée par la présence d'empreintes numériques caractérisant chacune des zones d'un fichier, ce qui permet à un outil de validation d'identifier précisément la zone corrompue.</p> <p>On notera que certaines méthodes de compression, particulièrement celles s'appliquant à l'ensemble du fichier et non à chacune de ses parties, peuvent avoir un effet négatif sur la</p>

Sous-groupe « connaissance de formats existants et émergents et définition de critères d'obsolescence et de pérennisation »

		robustesse des données. En raison de la réduction de la redondance que ces méthodes impliquent, une altération pourra affecter simultanément plusieurs zones du fichier.
Compacité	Le format exprime-t-il une quantité d'information conséquente dans un espace contraint ? Si cette compacité est liée à une méthode de compression, celle-ci est-elle réversible (sans perte d'information) ou non ?	<p>Un des risques majeurs pesant sur la pérennité des données est le risque budgétaire. Si les budgets que l'utilisateur peut allouer à l'achat d'espace de stockage sont limités, le critère de compacité peut devenir décisif.</p> <p>La compacité peut être liée à la structure d'encodage des données ou, le cas échéant, à la méthode de compression. Une compression sans perte, également dite « réversible », permet généralement une réduction significative du poids du fichier tout en garantissant la possibilité, grâce à la même méthode, de décompresser le fichier et d'obtenir une copie exacte, au bit près, du fichier source.</p>
Disponibilité d'outils de traitement	Existe-t-il des outils de restitution, de validation, d'analyse, de migration ? L'organisme de maintenance du format en développe-t-il officiellement ?	La disponibilité d'outils de traitement est liée à la documentation (plus un format est documenté, plus les chances sont élevées que des outils de traitement l'exploitent avec précision), à la sociabilité (plus sa communauté d'utilisateurs est étendue, plus elle a de chances d'avoir développé des outils de traitement nombreux et efficaces).
Contenu additionnel embarqué	Le format permet-il d'embarquer des flux complémentaires nécessaires à l'utilisation, l'identification et la gestion du fichier (métadonnées, documentation, visuels associés, etc.) ?	<p>Outre le contenu principal du fichier, le format peut être conçu pour permettre d'embarquer du contenu additionnel nécessaires à son utilisation. En l'absence de telles dispositions, l'utilisateur sera amené à transmettre séparément ce contenu, avec les risques de perte que cela implique.</p> <p>Le contenu additionnel peut être constitué de métadonnées permettant l'identification précise du contenu du fichier, des agents ayant contribué à sa création, des droits associés, etc. Il peut également s'agir de flux spécifiques comme un visuel pour un fichier audio MP3, des sources pour un document PDF/A-3 résultant d'une conversion, etc.</p>

Sous-groupe « connaissance de formats existants et émergents et définition de critères d'obsolescence et de pérennisation »

Mécanismes de protection	Le format dispose-t-il de mécanismes de protection de son contenu ?	Ce critère est ambivalent. Certains mécanismes de protection visant à interdire l'accès ou l'utilisation d'une ou plusieurs des fonctions du fichier, tels que les DRM, peuvent empêcher l'utilisateur de mener à bien des opérations à but de conservation. Ceux à l'inverse qui n'ont pas pour but de limiter l'accès ou l'utilisation, tels les signatures électroniques embarquées comme XMLDSig pour XML, sont un atout pour garantir l'intégrité et l'authenticité du fichier.
Simplicité	Le format a-t-il une structure simple ou complexe ?	Maintenir une compétence et des outils sur un format complexe demandera nécessairement un investissement plus lourd que sur un format simple. Une méthode de compression ajoute un niveau de complexité supplémentaire. Selon les méthodes, ce niveau de complexité peut être conséquent (ex. : JPEG 2000) ou plus limité (ex. : MP3).
Stabilité / évolutivité	Le format connaît-il une évolution soutenue et des versions qui se succèdent à une fréquence élevée ?	Suivre l'évolution d'un format fréquemment mis à jour peut s'avérer complexe et coûteux en investissement ; à l'inverse, un format qui ne connaît plus d'évolutions depuis plusieurs années supposera un effort moindre d'adaptation à son évolution. Plus généralement, ce critère interroge le stade de développement du format : est-il dans sa phase d'expansion initiale ou a-t-il atteint sa maturité, voire est-il toujours maintenu ?
Transparence	Le format est-il aisément lisible et compréhensible ou sa structure est-elle opaque ?	En l'absence d'outils spécifiques, un format relativement transparent pourra être plus facilement interprété et compris par un humain à l'aide d'outils génériques tels que des éditeurs de texte, XML ou hexadécimaux. La mise en œuvre d'une compression limite généralement le niveau de transparence d'un format.

Sous-groupe « connaissance de formats existants et émergents et définition de critères d'obsolescence et de pérennisation »

Critères subjectifs

Ces critères reposent sur une évaluation directement liée aux besoins et aux moyens de l'utilisateur réalisant l'évaluation.

Intitulé	Définition	Justification
Expressivité	Le format permet-il d'encoder toute l'information que le producteur souhaite exprimer ?	<p>Lorsqu'on évalue un format candidat pour une création ou une conversion, on considérera sa fidélité à l'intention originelle : toute l'information exprimée par le créateur ou contenue dans le fichier source peut-elle être adéquatement transformée dans les structures de données du fichier cible ?</p> <p>Une expressivité excessive au regard du besoin n'est pas non plus souhaitable car elle révèle généralement une inadéquation du format, trop riche pour l'emploi que l'utilisateur souhaite en faire, qui peut devenir contre-productive à long terme.</p>
Maîtrise	L'utilisateur dispose-t-il déjà des compétences et des outils nécessaires à une exploitation précise du format ?	<p>L'enregistrement de fichiers dans un format de données différent de celui d'origine équivaut à une conversion. Si celle-ci n'est pas réalisée en connaissance de cause, elle peut avoir un impact négatif sur la fidélité de la copie à son original.</p> <p>Dans le cas de créateurs confiant leurs données à une institution de conservation, il est important que le format soit maîtrisé par les deux parties.</p>