

ANALYSER UN FORMAT AU PRISME DE SON POTENTIEL DE PÉRENNISATION L'exemple du TIFF

Webinaire PIN n°5
08/06/2022

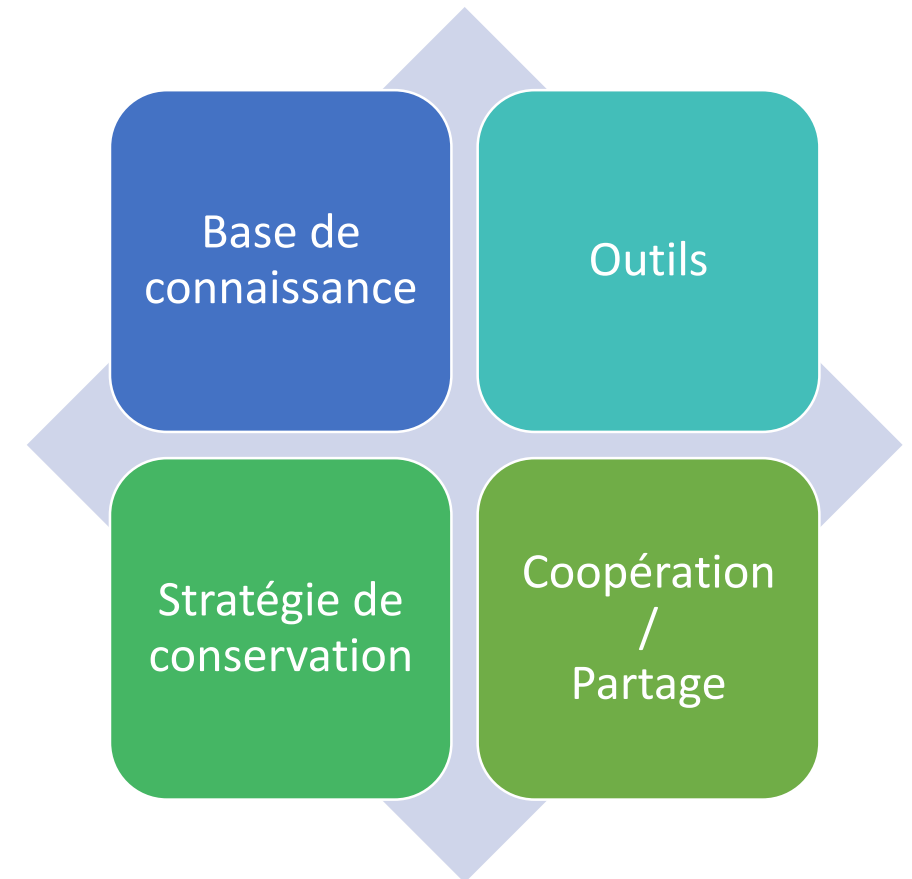
#PINFormats



Cellule nationale de veille sur les formats

➤ Les objectifs

- Mutualiser les activités de veille
- Sensibiliser les professionnels sur le sujet
- Contribuer et influencer sur les outils
- Etre l'interlocuteur francophone dans les travaux internationaux



Les groupes de travail

Réfléchir aux critères de pérennité des formats et constituer une base de connaissance sur les principaux formats de fichier



Connaissance des formats



Expertises

Constituer un annuaire des expertises françaises et internationales sur la problématique des formats de fichier

Proposer à la communauté une sélection de ressources internationales de référence traduites en français
Mettre en valeur les travaux français à l'international



Traduction



Outils & corpus

Recenser et classer les corpus de fichiers de test et les outils disponibles pour leur traitement (identification, validation, conversion, réparation)

Programme de la séance

#PINFormats



01.

Analyser un format – L'exemple du TIFF

*Par Erwann Ramondenc, chef
du pôle Archivage et
pérennisation des données
numériques et Pierre
Marchandin, chef de projet
pérennisation – Archives
diplomatiques (MEAE)*

02.

Actualités de la cellule

*Derniers travaux
& objectifs 2022-2023*

Analyser un format au prisme de son potentiel de pérennisation

L'exemple du TIFF

#PINFormats



Rappel sur les travaux antérieurs

- Construction d'un **tableau de récolement** qui permet de prioriser les formats à analyser
- Définition de **neufs points** essentiels à étudier pour connaître le potentiel de pérennisation d'un format

	A	B	C	D	E	F	G
	DOMAINES (prendre la catégorisation PRONOM)	Type du format	Extension	Type MIME	PUID	Wikidata	Caractéristiques
1	Image (Raster)	Tagged Image File Format	.tiff .tif	image/tiff	fmt/353	https://www.wikidata.org/wiki/Q215106	Tout TIFF 6.0 confondu
2	Image (Raster)	Tagged Image File Format (compression GROUP 4)	.tif	image/tiff	fmt/353	https://www.wikidata.org/wiki/Q215106	TIFF 6.0, compression T6/group 4
3	Image (Raster)	Exchangeable Image File Format (Uncompressed)	.tiff .tif	image/tiff	x-fmt/388 [Exif Uncompress	https://www.wikidata.org/wiki/Q196465	Fichier EXIF non compressé, basé sur le TIFF 6.0

Le choix du TIFF

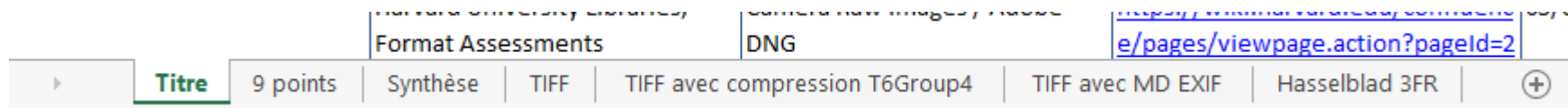
- Nombre important de fichiers collectés au moment où a été fait le récolement
- Fichiers issus de la numérisation des archives papiers
- Analyse liés aux renouvellement des travaux à l'atelier de numérisation

Les objectifs de l'analyse du format

- Répondre aux questions posées par les 9 points à partir de la documentation disponible sur ce format
- Ne pas réétudier les questions auxquelles d'autres institutions ont déjà répondu
- Synthétiser les éléments de réponse pour chaque point
- Traduire ces éléments pour rendre largement accessible les informations sur ce format

Le livrable

- Un tableau excel avec différents onglets :
 - 3 onglets récurrents : page de titre, rappel des 9 points, synthèse des informations
 - Autant d'onglets que de sous-formats identifiés pour l'institution



▶	Titre	9 points	Synthèse	TIFF	TIFF avec compression T6Group4	TIFF avec MD EXIF	Hasselblad 3FR	+
---	--------------	----------	----------	------	--------------------------------	-------------------	----------------	---

Le livrable : la page de titre

- Des informations sur l'identification du format : les PUID associés, les extensions, la famille de format, des identifiants divers, ...
- Une liste des ressources utilisées : avec le lien et la date de mise à jour
- Un résumé de la démarche pour bien comprendre les choix opérés

Le livrable : la page de titre

NOM DU FORMAT		TIFF		
Date de mise à jour de la fiche		08/03/2022		
Rédacteur (institution)		MEAE		
IDENTIFICATION DU FORMAT	Extension	tiff tif		
	Famille / Classification	Image fixe		
	Version décrite	6.0		
	PUID	fmt/353 ; x-fmt/387 ; x-fmt 388 ; fmt/1062		
	ID Library of Congress	fdd000022 ; fdd000024 ; fdd000145		
	ID Wikidata	Q215106 ; Q196465		
	Type MIME	image/tiff		
Source d'information	Titre/nature de la ressource	Lien	Mise à jour de la ressource	
Adobe	TIFF, Revision 6.0, Final	https://developer.adobe.com/content/dam/udp/en/open/standards/tiff/TIFF6.pdf	03/06/1992	
Library of congress	TIFF, Revision 6.0	http://www.loc.gov/preservation/digital/formats/fdd/fdd000022.shtml	18/09/2020	
Harvard University Libraries, Format Assessments	Camera Raw Images / Adobe DNG	https://wiki.harvard.edu/confluence/pages/viewpage.action?pageId=207554565	05/04/2016	

Le livrable : la page de titre

<p style="text-align: center;">DEMARCHES</p>	<p>Le recensement des formats de fichiers conservés au MEAE, effectué en février 2020, avait montré l'existence de 4 sous-catégories de TIFF dans les collections des Archives diplomatiques. Ces catégories avaient été construites en fonction de certaines caractéristiques (PUID, extension, existence ou non d'une compression...) pouvant justifier des décisions différentes en matière de prise en charge :</p> <ul style="list-style-type: none">* Les fichiers de format TIFF 6.0, comprenant l'ensemble des fichiers TIFF sans distinction, sauf les fichiers TIFF avec compression T6 de groupe 4 ;* Les fichiers TIFF avec compression T6 groupe 4 ;* Les fichiers TIFF avec métadonnées EXIF ;* Les fichiers Hasselblad 3FR. <p>Le MEAE a fait le choix de fonder son analyse des fichiers TIFF au regard des 9 critères de préservation en reprenant ces sous-catégories de TIFF, dans la mesure où le but de cette analyse est de compléter le travail de recensement des formats de fichiers déjà effectué. La non congruence entre les catégories de fichiers recensés et les catégories de fichiers analysées rendrait en effet plus difficile la prise de décision quant à la prise en charge de ces formats par l'institution.</p> <p>Constitution du tableau : Pour chaque feuillet d'analyse d'une sous-catégorie de TIFF, les informations contenues dans chaque source sont ventilées dans une ligne. Lorsque la source ne contient pas de réponse à la question posée, la case est laissée blanche.</p> <p>Les cases des lignes de synthèses sont toutes complétées de quatre manières :</p> <ul style="list-style-type: none">(i) Lorsque une ou plusieurs des ressources consultées permettent de répondre à une question, les cases offrent une synthèse des informations présentes dans les sources(ii) NA est indiqué lorsque la question ne se pose pas pour le type de format considéré(ii) Les déductions faites par les concepteurs du tableau en l'absence d'éléments de réponse aux questions dans les sources sont indiquées en italiques.
--	---

Le livrable : les 9 points

- Les critères sont rappelés, ainsi que leur description
- Chaque critère fait l'objet de plusieurs questions auxquelles on tente de répondre dans les onglets suivants

Le livrable : les 9 points

Critère	Présentation	Équivalences anglo-saxonnes
Adoption	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.	Adoption / Software support
Existence et mise à disposition de spécifications	Le terme renvoie à la complétude, à la précision et à l'intelligibilité 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 a minima des spécifications accessibles, tandis que les formats propriétaires se limitent plus souvent à une documentation commerciale ou élaborée a posteriori par les utilisateurs.	Disclosure / Documentation and Guidance
Transparence	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. Par exemple, le code source d'un logiciel est plus transparent que son code objet. De même, un texte est plus transparent si le codage de ses caractères est standard (UTF-8, en particulier). La compression, si elle est inévitable, doit utiliser des algorithmes largement répandus	Transparency / Complexity
Compacité	Capacité du format à contenir un grand nombre de données dans un volume réduit. Si la compacité est liée à la compression, la transparence et la réversibilité de la compression sont des critères supplémentaires à évaluer.	Compactness
Contenu additionnel embarqué	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	Self-documentation / Embedded or Attached Content

Le livrable : les analyses par format

- Regroupement subjectif de l'ensemble des formats identifiés qui semble ressortir d'un même format
- Traduction des éléments de réponses apportés par les institutions (souvent depuis l'anglais)

Le livrable : les analyses par format

Sources d'informations	Liens	Mise à jour de la ressource	Description	Historique du format					Par qui ? (professionnels du secteur, grand public, institutions patrimoniales)	Le fo recon l'insti
				Editeur historique	Date de la première version	Version la plus récente et date de publication	Propriétaire actuel	Historique de l'évolution		
Spécifications de 1992 publiées par ADOBE	https://developer.adobe.com/content/dam/udp/sn/open/standards/tiff/TIFF6.pdf	03/06/1992	Type : format conteneur. Le Tiff permet de conserver des images dans une qualité très élevée et prend en charge les calques. Dans sa version compressée, il permet de choisir parmi de nombreux systèmes de compression (Zip, LZW), même celui de Jpeg. La production de fichiers très volumineux le destine surtout à un usage professionnel et à la préparation des images pour l'impression et exclut tout usage pour une publication	Aldus et Microsoft Corporations. Créé par Aldus, racheté ensuite par Adobe	La première version de la spécification TIFF a été publiée par Aldus Corporations à l'automne 1986, après une série de rencontres avec différents fabricants de scanners et développeurs de logiciels.	1992	Aldus (depuis absorbé par Adobe Corporation), propriétaire du format	On peut distinguer 4 révisions majeures (nom de révision ont été donnés a posteriori) : Révision 3.0, la révision 4.0 qui contient surtout des améliorations mineures, publiée en avril 1987. La révision 5 publié en octobre 1988 qui ajoute la palette des couleurs (support for palette colorimages) et la compression LZW. La version la plus récente est la 6.0, publiée en 1992. Les descriptions dans la spécification de 1992 sont		
Library of Congress	http://www.loc.gov/preservation/digital/formats/tiff/tiff000022.shtml		Un format basé sur des tags pour stocker et échanger des images matricielles. Le TIFF sert de container pour différents types d'encodage de flux binaire pour images matricielles. Les différents encodages peuvent représenter différents schémas de compression et différents schéma de représentation des couleur (interprétation photométrique).	Aldus et Microsoft.	1986	Tiff 6.0 (1992)	Adobe Systems	Le format a été créé à l'origine par Aldus, avec Microsoft, pour l'impression PostScript. La première version a été publiée en 1986, elle est connue comme la version 3. La spécification a été créée pour encourager les vendeurs de scanners de bureau du milieu des années 1980 à se mettre d'accord sur un format commun d'image numérisées, au lieu que chaque entreprise ne publie son propre format propriétaire. au départ, le TIFF n'était qu'un format image	Certains sous-types sont largement répandus comme formats master pour des images numérisées, Un des les im géo de la L (dans : Form précis recom	

Le livrable : les analyses par format

Sources d'informations	Liens	Mise à jour de la ressource	Description	Adoption				Existence et mise à di		
				Par qui ? (professionnels du secteur, grand public, institutions patrimoniales)	Le format est-il recommandé par l'institution ?	Le format est-il largement implémenté dans des outils de création, d'édition, de lecture, de diffusion ou de transformation ?	Le format est-il stable ?	Des spécifications complètes et intelligibles sont-elles disponibles ?	Le format est-il normalisé ?	Le forma maintenu organisational normalis-reconnue secteur d
Spécifications de 1992 publiées par ADOBE	https://developer.adobe.com/content/dam/udpt/en/ops/standards/tiff/TIFF6.pdf	03/06/1992	Type : format conteneur. Le Tiff permet de conserver des images dans une qualité très élevée et prend en charge les calques. Dans sa version compressée, il permet de choisir parmi de nombreux systèmes de compression (Zip, LZW), même celui de Jpeg. La production de fichiers très volumineux le destine surtout à un usage professionnel et à la préparation des images pour l'impression et exclut tout usage pour une publication						Non	Maintenu dans les fai jour depuis à jour ont été publication
Library of Congress	http://www.loc.gov/preservation/digital/formats/tiff/tiff000022.c.html		Un format basé sur des tags pour stocker et échanger des images matricielles. Le TIFF sert de container pour différents types d'encodage de flux binaire pour images matricielles. Les différents encodages peuvent représenter différents schémas de compression et différents schéma de representation des couleur (internation	Certains sous-types sont largement répandus comme formats master pour des images numérisées,	Un des formats préférés pour les images matricielles et géoréférencées. Le référentiel de la Library of Congress (dans son <i>Recommended Formats Statement</i>) ne précise pas la version du TIFF recommandée.	Utilisable par de nombreux outils de manipulation d'image, de numérisation, d'OCR, de visionneuse	La dernière version a été publiée en 1992	Le format est propriétaire, les spécifications sont détenues par Adobe qui les met à disposition (elles sont également disponibles sur Internet Archives). Complétude partielle des spécifications : le tags peuvent être étendus : la liste des extensions est maintenue par Adobe mais il ne les fournit pas. Cependant la LOC en fournit la liste.		Maintenu p

Le livrable : la synthèse

- Permet en un coup d'œil de se rendre compte du potentiel de pérennisation d'un format
- Les mentions en italiques sont des déductions en l'absence de réponse dans la documentation consultée

Le livrable : la synthèse

Nom	Version décrite le cas échéant	Sources d'information	Description	Historique du format				
				Editeur historique	Date de la première version	Version la plus récente et date de publication	Propriétaire actuel	Historique de l'évolution
Tagged Image File Format (TIFF)	6.0	Spécifications publiées en 1992 par Adobe, File Format Info (The Digital Rosetta Stone) Libraries and Archives Canada National Archives and Records Administration Bibliothèque nationale de France Centre de coordination pour l'archivage à long terme de documents électroniques (CECO, Suisse) CINES PRONOM Fileinfo.com	Le TIFF est un format conteneur qui permet de conserver des images dans une qualité très élevée et prend en charge les calques. Dans sa version compressée, il permet de choisir parmi de nombreux systèmes de compression (Zip, LZW), même celui de Jpeg. La production de fichiers très volumineux le destine surtout à un usage professionnel et à la préparation des images pour l'impression et exclut tout usage pour une publication Web	Aldus and Microsoft Corporation. Créé par Aldus, racheté ensuite par Adobe.	Automne 1986	Tiff 6.0 (1992)	Adobe Systems Inc	On peut distinguer 4 révisions majeures (nom de révision ont été donnés a posteriori) : *Révision 3.0. *Révision 4.0 qui contient surtout des améliorations mineures, publiée en avril 1987. *Révision 5, publiée en octobre 1988 qui ajoute la palette des couleurs (support for palette colorimages) et la compression LZW. *La version la plus récente est la 6.0, publiée en 1992. Les descriptions dans la spécification de 1992 sont pertinentes pour les versions antérieures du standard. Ces évolutions sont dues à l'évolution du matériel informatique : le TIFF a été créé pour que les constructeurs de scanners (alors en noir et blanc), se
TIFF avec compression T6/Group 4		Library of Congress Harvard University Library	Un format basé sur des balises pour stocker et échanger des images raster. Ce sous-type de Tiff renferme une image raster compressée selon la recommandation UIT-T T.6 (Schémas de codage et fonctions de commande de codage de la télécopie pour les télécopieurs de groupe 4). Il est par conséquent limité aux	Cf. TIFF 6.0	Cf. TIFF 6.0	Cf. TIFF 6.0	Cf. TIFF 6.0	Cf. TIFF 6.0

Le livrable : la synthèse

Nom	Version décrite le cas échéant	Sources d'information	Description	Incidences des brevets	Dispositifs de	Résilience	
				Existe-il un ou des brevet(s) susceptible(s) d'empêcher une large diffusion ou utilisation de ce format ?	Existe-t-il des mécanismes de protection qui pourraient empêcher l'exploitation du format ?	Le format permet-il sa lecture même en cas d'altération ? (exemple : marqueur de reprise)	Le format embarque-t-il des moyens internes de contrôle d'intégrité ?
Tagged Image File Format (TIFF)	6.0	Spécifications publiées en 1992 par Adobe, File Format Info (The Digital Rosetta Stone) Libraries and Archives Canada National Archives and Records Administration Bibliothèque nationale de France Centre de coordination pour l'archivage à long terme de documents électroniques (CECO, Suisse) CINES PRONOM Fileinfo.com	Le TIFF est un format conteneur qui permet de conserver des images dans une qualité très élevée et prend en charge les calques. Dans sa version compressée, il permet de choisir parmi de nombreux systèmes de compression (Zip, LZW), même celui de Jpeg. La production de fichiers très volumineux le destine surtout à un usage professionnel et à la préparation des images pour l'impression et exclut tout usage pour une publication Web	Non exploité (l'a été pour la compression LZW expiré en 2003 aux US et 2004 au Japon et UE). Le TIFF est maintenant suffisamment ancien pour que les brevets essentiels aient expiré.	Non	<i>Non</i>	<i>Pas d'élément de réponse dans les ressources consultées</i>
TIFF avec compression T6/Group 4		Library of Congress Harvard University Library	Un format basé sur des balises pour stocker et échanger des images raster. Ce sous-type de Tiff renferme une image raster compressée selon la recommandation UIT-T T.6 (Schémas de codage et fonctions de commande de codage de la télécopie pour les télécopieurs de groupe 4). Il est par conséquent limité aux	Non	Non	<i>Pas d'élément de réponse dans les ressources consultées</i>	<i>Pas d'élément de réponse dans les ressources consultées</i>

Actualités de la cellule

Derniers travaux

#PINFormats



Merci de votre attention !

#PINFormats



Retrouvez nos précédents webinaires

Présentation et vidéos disponibles sur le site d'Aristote

[Webinaires de présentation de la cellule](#)

Novembre 2020

[Digital Preservation Handbook](#)

Juin 2021

[Stratégie de préservation](#)

Novembre 2021