

Pérennisation de l'information numérique

Evaluation du format TIFF

au regard de son aptitude à être pérennisés

Version du 31 août 2004



Laurent Duploux

BnF



•1. FICHE IDENTITE

- *Nom commun : TIFF*
- *Nom complet : Tagged Image File Format*
- *Version : Revision 6.0 Final du 3 juin 1992*
- *Lien :
<http://partners.adobe.com/public/developer/tiff/index.html>*
- *Codage : par séquence de 8 bits*
- *Taille maximal d'un fichier : 2^{32} octets soit 4 Go*



• 2. DESCRIPTION GENERALE DU FORMAT

- *TIFF est format-enveloppe qui contient :*
 - l'entête du fichier contient,
 - des informations techniques sur l'image : Image File Directory (IFD)
 - les données de l'image selon un mode de représentation prédéfini.



L'entête

- *définit le type d'encodage des bits (BigEndian ou LittleEndian),*
- *la version,*
- *la position (offset) du premier IFD.*



Les IFD ou tags

- *donnent des informations sur :*
 - les dimensions de l'image,
 - le type de compression utilisé,
 - la profondeur de codage,
 - la résolution pour chaque dimension, etc..

Certains tags sont obligatoires. Il est possible de définir des tags privés (pas recommander pour la préservation).

- *Les valeurs possibles des tags sont :*
 - BYTE : Entier non signé sur 8 bits
 - ASCII : Caractère ASCII sur 8 bits terminé par le caractère null
 - SHORT : Entier non signé sur 16 bits
 - LONG : Entier non signé sur 32 bits
 - RATIONAL : Fraction avec 2 valeurs de type LONG, une pour le numérateur, l'autre pour le dénominateur



Les données de l'image

- *peuvent être compressées ou non. Les modes de compression possibles sont :*
 - CCITT Group 3 1-Dimensional modified Huffman RLE
 - Facsimile compatible CCITT Group 3
 - Facsimile compatible CCITT Group 4
 - LZW (Lempel-Ziv & Welch)
 - JPEG
 - PackBits (Macintosh)

Ces modes de compression sont du domaine public (LZW depuis juin 2003 pour UNISYS ou 2006 pour IBM).



Pérennité des images

- *Les formats images en général et le format TIFF en particulier. Six critères principaux sont à considérer eu égard de la pérennité :*
 - *1. la complexité du format*
 - *2. la capacité d'extension*
 - *3. la capacité d'être portable*
 - *4. Réversibilité*
 - *5. le niveau de diffusion des spécifications du format*
 - *6. la diffusion des outils d'exploitation*



3. CRITERE 1

- *APTITUDE DU FORMAT A REPRESENTER TOUTE L'INFORMATION A PERENNISER*
 - Le format TIFF est un très bon pour la pérennisation.
 - possible d'insérer les profils ICC pour la gestion de la couleur.
 - Attention des restrictions suivantes :
 - Longueur maximale de 2^{32} bits ce qui est une limite dans le cadre des images satellites.
 - Type de compression LZW pas encore dans le domaine public.
 - Jusqu'à 24 bits par pixel (à vérifier) ce qui peut représenter une limitation dans certains cas d'utilisation (représentation de œuvres artistiques).



3. CRITERE 2 : FORMAT NORMALISE

- *Format normalisé : Non*
- *Format propriétaire : Oui,*
 - développé par les sociétés Aldus et Microsoft en collaboration avec d'importants manufacturiers de scanner et développeurs de logiciel d'imagerie.
 - La première version a été publiée en 1986 par la société Aldus.
 - Les droits (copyright) détenus depuis 1995 par la société Adobe Systems. La société Adobe Systems a publié une version 6.0.1 faisant état du transfert des droits mais sans incidence sur le contenu strictement technique des spécifications publiées en 1992 par la société Aldus.
- *Format publié : Oui*



3. CRITERE 3

- *POSSIBILITE DE MODIFIER ULTERIEUREMENT LES DONNEES*
 - Aucune restriction particulière est à noter pour ce critère. Il est toujours possible de modifier les images enregistrées au format TIFF avec un outil adéquat.



3. CRITERES COMPLEMENTAIRES

- *DIFFUSION DU FORMAT*

- Le format est très largement répandu.

- Dans le monde de la numérisation aussi bien professionnel que grand public,

- la totalité des numériseurs sont capable de produire des images respectant le format.

- Attention noter que certains constructeurs prennent des libertés avec le format, cas du NewSubfileType.

- *DIFFUSION DES OUTILS D'EXPLOITATION (1)*

- bibliothèque de fonction de TIFF fait partie du domaine public écrite par Sam Leffler de l'Université de la Californie à Berkeley

- est largement diffusée (voir libtiff).

- langage utilisé est le langage ANSI C .



3. CRITERES COMPLEMENTAIRES

- *DIFFUSION DES OUTILS D'EXPLOITATION (2)*
 - logiciels de manipulation d'image sont capables d'exploiter des images au format TIFF. On peut citer : Photoshop (payant), PaintShopPro (shareware), GIMP (freeware, libre)...
 - Attention, ils ne sont pas toujours capables d'utiliser tous les tags de la spécification ni tous les modes de compression (Cas group4 et JPEG).
- *COMPLEXITE/SIMPLICITE DU FORMAT*
 - format est d'une relative simplicité
 - Bien documenté, il est assez simple de créer des outils permettant son exploitation.
 - niveau de complexité concerne le mode de compression utilisé avec ce format.



3. CRITERES COMPLEMENTAIRES

- *CONTROLABILITE*
 - logiciel de contrôle automatique du format peut être être mis en œuvre de façon
- *METADONNEES (1)*
 - La diversité des tags proposés en standard permet de couvrir la plupart des besoins en métadonnées techniques :
 - Nom du document
 - Taille de l'image, résolution, profondeur de codage, orientation
 - Fabricant, modèle utilisé, logiciel utilisé, Description de l'image, Nom de l'artiste, Mention de copyright, Nom de la page
 - Horodatage (format libre sur 20 caractères)



3. CRITERES COMPLEMENTAIRES

- *METADONNEES (2)*
 - données textuelles enregistrées au format ASCII.
 - recherche d'information directement dans le fichier n'est pas une opération habituelle. Peu d'outils exploitent cette possibilité.
 - cas d'utilisation des métadonnées visent plutôt à renseigner un autre format, tel qu'XML (voir le schéma MIX)