

# Guide concernant les formats recommandés par BAnQ

- 1 [Introduction](#)
- 3 [Documents textuels](#)
- 4 [Documents de présentation](#)
- 4 [Courriels](#)
- 5 [Images fixes](#)
- 6 [Documents sonores](#)
- 7 [Documents vidéo](#)
- 8 [Données géospatiales](#)
- 9 [Documents de CAO](#)
- 10 [Ensembles de données](#)
- 10 [Collectes de sites Web](#)

Mars 2020

## Introduction

Ce guide désigne les formats que BAnQ considère comme fiables à des fins de préservation. Il n'est pas exhaustif, mais vise à signaler les types de formats courants que BAnQ s'attend à recevoir des organismes publics (ministères, organismes, etc.) ou des donateurs.

Ce document peut également servir de référence aux organismes publics qui conservent leurs propres archives (organismes municipaux, universités, centres de santé et de services sociaux, etc.) ou aux organismes privés (entreprises privées, OBNL, etc.). Il les aidera à prendre des décisions fondées sur les risques d'obsolescence des formats pour des projets de numérisation ou d'autres projets numériques.

BAnQ visera à garantir plus particulièrement la lisibilité de ces formats à long terme.

## La typologie d'un format

Un fichier informatique est un ensemble de données constitué d'une séquence de bits, soit une séquence de 0 et de 1. La convention utilisée pour stocker des données dans un fichier détermine le format de celui-ci.

Il existe des formats dits ouverts ou fermés. Un format est dit ouvert si sa description est publique et qu'aucune disposition légale ne s'oppose à sa libre utilisation. Ce type de format est interopérable. Cela veut dire qu'il sera éventuellement possible d'utiliser différents logiciels pour créer, modifier, lire et imprimer un document enregistré dans ce format. Un format ouvert peut aussi être normalisé lorsqu'il est adopté par un organisme de normalisation comme ISO ou W3C.

Par opposition, un format est dit fermé lorsque sa description n'est pas divulguée ou que des dispositions légales peuvent en restreindre l'utilisation. Souvent, un document enregistré dans un tel format n'est parfaitement lisible qu'avec un logiciel en particulier.

Un format propriétaire est un format dont les spécifications techniques appartiennent à l'entreprise qui les a créés. Il peut être ouvert ou fermé.

Qu'il soit ouvert ou fermé, un format peut devenir un format dit **standard** du fait de sa popularité.

## Format et pérennité

Les évolutions techniques sont tellement rapides qu'aucun format n'est pérenne. En revanche, certains formats, de par leurs caractéristiques, seront plus simples à gérer à long terme que d'autres. Privilégier un format sur la seule prémisses qu'il est ouvert ne suffit pas. En effet, un format ouvert peut être complexe à gérer à long terme, soit parce que sa documentation est pauvre ou mal maintenue, soit parce qu'il est peu populaire et donc qu'il existe peu de logiciels pour le manipuler.

Actuellement, la seule manière de s'assurer qu'un document numérique sera lisible et exploitable par les générations futures est de le créer dans un format répondant à des critères définis dont on suivra les risques d'obsolescence et pour lequel on mettra en place des stratégies de migration vers un autre format advenant qu'il devienne inutilisable.

## Critères de sélections des formats

BAnQ se base sur les critères suivants afin de déterminer la viabilité à long terme des formats :

- **l'accessibilité et l'exhaustivité de la documentation** : un format doit être doté d'une documentation complète et accessible;
- **la fréquence d'utilisation** : un format très utilisé et populaire a plus de chances d'être utilisé longtemps;
- **le degré d'ouverture** : un format ouvert est lisible par plusieurs logiciels; un format propriétaire est souvent lié à un seul logiciel;
- **la reconnaissance du format** : lorsqu'un format est reconnu par une norme, il est adopté plus largement et sa documentation est plus accessible.

## Liste des formats recommandés

BAnQ distingue deux niveaux de préservation de formats basés sur les critères énoncés précédemment.

Les formats **favorisés** sont ceux que BAnQ considère comme les formats présentant un risque très faible de devenir illisibles à long terme. Ils ne demandent pas de gestes immédiats de préservation.

Les formats **acceptables** sont ceux que BAnQ considère comme des formats qui présentent un risque faible de devenir illisibles à long terme. Ces formats répondent aux critères minimaux de préservation et peuvent faire l'objet d'une intervention pour assurer celle-ci.

La liste des formats recommandés par BAnQ pourra être mise à jour, car les formats sont soumis à une veille technologique par les experts en préservation de BAnQ. Les organismes publics peuvent communiquer avec BAnQ pour signaler d'autres formats de fichiers qui pourraient bonifier cette liste.

Les formats de fichiers sont divisés dans les catégories et les sous-catégories suivantes :

- Documents textuels
- Documents de présentation
- Courriels
  - Messages individuels
  - Boîtes courriels
- Images fixes
  - Photos numériques
  - Textes numérisés
- Documents sonores
- Documents vidéo
  - Cinéma numérique
  - Vidéo numérique
- Données géospatiales
- Documents de conception assistée par ordinateur (CAO)
- Ensembles de données
- Collectes de sites Web

## 1. Documents textuels

| FORMAT                                      | NOTES   | COMMENTAIRES  |
|---|---|---|
| <b>FORMATS FAVORISÉS PAR BAnQ</b>           |   |   |
| TEXTE BRUT (TXT)                            | TEXTE UNICODE<br>(UTF-8, UTF-16)<br><br>ASCII (7 BIT) | Ce type de fichier contient uniquement du texte brut, sans mise en forme.<br>Ce format est public, normalisé et lisible de manière universelle. Son utilisation n'est pas dépendante d'un programme en particulier.<br>Ce format est très utilisé.<br>Les spécifications complètes sont disponibles.  |
| OPENDOCUMENT TEXT<br>FORMAT (ODF)           | OPENDOCUMENT 1.0                                      | C'est un format de fichier ouvert pour les applications bureautiques.<br>Il peut être lu par une variété de logiciels de bureautique.<br>Ce format est très utilisé.<br>Il est reconnu par la norme ISO 26300.  |
| PDF/A-1 (PDF)<br><br>PDF/A-2 (PDF)          | PDF/A-1<br><br>PDF/A-2                                | Ce format est très répandu pour conserver et échanger des documents numériques. C'est un format destiné à une conservation à long terme des documents.<br>Le format PDF/A se distingue du format PDF par le fait qu'il permet d'insérer seulement des images comme contenu multimédia.<br>Ce format est très utilisé.<br>Il est reconnu par la norme ISO 19005.   |
| ELECTRONIC PUBLICATION<br>EPUB 3 (EPUB)     | EPUB 3  | Ce format est utilisé pour les livres électroniques avec un calibrage fluide du texte en fonction du dispositif de lecture et des options d'affichage choisies par l'utilisateur.<br>C'est un format standardisé et ouvert.<br>Il est très utilisé.<br>Les spécifications sont disponibles.   |
| <b>FORMATS ACCEPTABLES</b>                  |   |   |
| PDF 1.7 (PDF)<br><br>PDF 1.0-1.6 (PDF)      | PDF 1.7<br><br>PDF 1.0-1.6                            | Ce format a été développé par Adobe Systems. La spécificité de ce format est de préserver la mise en forme d'un fichier (les polices d'écriture, images, objets graphiques, etc.) telle qu'elle a été définie par son auteur, et cela, quels que soient le logiciel, le système d'exploitation et le matériel utilisés pour l'imprimer ou le visualiser.<br>Ce format est très largement utilisé.<br>Il est reconnu par la norme ISO 32000. |
| MICROSOFT WORD (DOCX)                       | MICROSOFT WORD<br>FOR WINDOWS,<br>VERSION 2007-2010   | Ce format est utilisé par Microsoft Word depuis sa version 2007.<br>Il est basé sur la norme Office Open XML, reconnue norme ISO 29500.<br>Ce format est très largement utilisé.  |
| MICROSOFT WORD (DOC)                        | VERSION 8.0   | Ce format propriétaire de la suite Microsoft Office est très largement utilisé.<br>Ce format reste cependant peu documenté.   |
| ELECTRONIC PUBLICATION<br>EPUB 2.0.1 (EPUB) | EPUB 2.0.1  | Voir EPUB 3.  |

## 2. Documents de présentation

| FORMAT                                    | NOTES                                       | COMMENTAIRES   |
|---|---|--|
| <b>FORMATS FAVORISÉS PAR BAQ</b>          |   |  |
| OPENDOCUMENT PRESENTATION<br>FORMAT (ODP) | VERSION 1.0                                 | L'extension ODP est utilisée pour les fichiers de type présentation utilisant le format ODF. Voir ODF.   |
| PDF/A-1                                   | PDF/A-1                                     | Voir la section 1.   |
| PDF/A-2                                   | PDF/A-2                                     |  |
| <b>FORMATS ACCEPTABLES</b>                |   |  |
| MICROSOFT POWERPOINT<br>(PPTX)            | MICROSOFT POWERPOINT<br>2007 ET ULTÉRIEURES | Ce format est utilisé par Microsoft PowerPoint depuis sa version 2007. Il est basé sur la norme Office Open XML, reconnue norme ISO 29500. Ce format est très largement utilisé. |
| MICROSOFT POWERPOINT<br>(PPT)             | MICROSOFT POWERPOINT<br>1997 À 2003         | Ce format propriétaire de la suite Microsoft Office est très largement utilisé. Ce format reste cependant peu documenté.   |

## 3. Courriels

### 3.1. Messages individuels

Les fichiers joints doivent être extraits des courriels et fournis dans un format propice à la conservation.

| FORMAT   | NOTES             | COMMENTAIRES   |
|--|-------------------|--|
| <b>FORMATS FAVORISÉS PAR BAQ</b>               |                   |  |
| INTERNET MESSAGE FORMAT<br>(EML)               |                   | Chaque fichier représente un courriel individuel stocké en texte brut. Il contient l'en-tête, le sujet, le corps du courriel ainsi que les pièces jointes. Ce format est très largement utilisé. Les spécifications sont disponibles.              |
| <b>FORMATS ACCEPTABLES</b>                     |                   |  |
| MICROSOFT OUTLOOK ITEM<br>MESSAGE FORMAT (MSG) | MICROSOFT OUTLOOK | Ce format a été créé par Microsoft. Il présente des lacunes pour le visionnement, car il peut y avoir des problèmes de rendu des messages sans le lecteur de Microsoft. Ce format est très largement utilisé. Les spécifications sont disponibles. |

### 3.2. Boîtes courriels

Non accepté pour l'instant.

## 4. Images fixes

### 4.1. Photos numériques

| FORMAT   | NOTES            | COMMENTAIRES  |
|--|------------------|---|
| <b>FORMATS FAVORISÉS PAR BAQ</b>   |                  |   |
| TAGGED IMAGE FILE FORMAT (TIFF), SANS COMPRESSION  | VERSIONS 4, 5, 6 | Le format TIFF 6.0 est un format privilégié pour la préservation.<br>Ce format est largement utilisé.<br>Les spécifications sont disponibles.   |
| <b>FORMATS ACCEPTABLES</b>   |                  |   |
| JPEG FILE INTERCHANGE FORMAT (JFIF) WITH JOINT PHOTOGRAPHIC EXPERTS GROUP (JPEG) COMPRESSION |                  | Ce format est un standard international. C'est le format par défaut des appareils photos numériques.<br>Ce format est très largement utilisé.<br>Il est reconnu par la norme ISO 10918.                                 |
| PORTABLE NETWORK GRAPHICS (PNG)  |                  | Ce format est un standard international, même s'il n'est pas encore adopté largement. Il pourrait remplacer le format JPEG dans les usages à long terme.<br>Il est reconnu par la norme ISO 15948.                      |
| JPEG 2000 (JP2), SANS COMPRESSION  |                  | Ce format est un standard international, même s'il n'est pas adopté largement. Ses performances en compression (avec et sans perte) sont supérieures à celles du format JPEG.<br>Il est reconnu par la norme ISO 15444. |
| DIGITAL NEGATIVE (DNG)   | VERSION 1.4.0.0  | Ce format créé par Adobe Systems est non propriétaire. Il est utilisé par une large variété d'applications et d'appareils qui utilisent des fichiers Raw.<br>Les spécifications sont disponibles.                       |

### 4.2. Textes numérisés

| FORMAT   | NOTES            | COMMENTAIRES         |
|--|------------------|----------------------|
| <b>FORMATS FAVORISÉS PAR BAQ</b>   |                  |                      |
| TAGGED IMAGE FILE FORMAT (TIFF), SANS COMPRESSION  | VERSIONS 4, 5, 6 | Voir la section 4.1. |
| PDF/A-1  | PDF/A-1          | Voir la section 1.   |
| <b>FORMATS ACCEPTABLES</b>   |                  |                      |
| JPEG FILE INTERCHANGE FORMAT (JFIF) WITH JOINT PHOTOGRAPHIC EXPERTS GROUP (JPEG) COMPRESSION |                  | Voir la section 4.1. |
| PORTABLE NETWORK GRAPHICS (PNG)  |                  | Voir la section 4.1. |
| JPEG 2000 (JP2), SANS COMPRESSION  |                  | Voir la section 4.1. |

## 5. Documents sonores

| FORMAT  | NOTES        | COMMENTAIRES  |
|---|--------------|---|
| <b>FORMATS FAVORISÉS PAR BAnQ</b>                             |              |   |
| WAVE WAVEFORM AUDIO FILE FORMAT (WAV) ET BROADCAST WAVE (BWF) |              | <p>Ce format est généralement utilisé pour encoder une représentation de haute qualité audio.</p> <p>Ce format propriétaire est largement utilisé dans l'industrie du son. Les spécifications sont disponibles.</p>   |
| <b>FORMATS ACCEPTABLES</b>                                    |              |   |
| MPEG AUDIO LAYER III (MP3)                                    |              | <p>Ce format permet de produire des fichiers de qualité audio respectable, de petite taille.</p> <p>Ce format est extrêmement utilisé pour diffuser des documents audio sur Internet.</p> <p>Il est reconnu par les normes ISO 11172 et 13818.</p>                                    |
| MPEG-4 AAC ADVANCED AUDIO CODING (AAC)                        |              | <p>Ce format audio est destiné à remplacer le format MP3, car il offre une meilleure qualité pour le même niveau de compression.</p> <p>Ce format voit sa popularité augmenter, car il est largement adopté par l'industrie du son.</p> <p>Il est reconnu par la norme ISO 14496.</p> |
| FREE LOSSLESS AUDIO CODEC (FLAC)                              | VERSION 1.21 | <p>Ce format de compression sans perte produit des fichiers beaucoup plus petits que les fichiers WAV et stocke les métadonnées.</p> <p>Ce format est ouvert et est soutenu par des logiciels libres.</p>   |
| AUDIO INTERCHANGE FORMAT (AIFF)                               | VERSION 1.3  | <p>Ce format est une option alternative d'Apple au WAV. Il prend en charge les métadonnées.</p> <p>Les spécifications sont disponibles.</p>   |

## 6. Documents vidéo

### 6.1. Cinéma numérique

| FORMAT                                       | NOTES         | COMMENTAIRES   |
|--|---------------|--|
| DIGITAL MOVING PICTURE EXCHANGE BITMAP (DPX) | NON COMPRESSÉ | Ce format est utilisé pour une représentation de haute qualité vidéo. Les spécifications sont disponibles. |

### 6.2. Vidéo numérique

| FORMAT                                  | NOTES   | COMMENTAIRES   |
|---|---|--|
| <b>FORMATS FAVORISÉS PAR BAnQ</b>       |   |  |
| QUICKTIME FILE FORMAT (MOV)             | NON COMPRESSÉ 4:2:2   | Ce format a été créé par Apple. C'est un format conteneur qui contient une ou plusieurs pistes pouvant être chacune encodée dans un ou plusieurs codecs différents.<br>Ce format est très largement utilisé.<br>Les spécifications sont disponibles.   |
| QUICKTIME FILE FORMAT (MOV)             | JPEG 2000<br>DV-NTSC<br>AVC/H.264<br>APPLE PRORES<br>AVID DNxHD |  |
| AUDIO VIDEO INTERLEAVED FORMAT (AVI)    | NON COMPRESSÉ 4:2:2   | Ce format a été créé par Microsoft. C'est un format conteneur qui peut utiliser une variété de formats audio et vidéo.<br>Ce format est très largement utilisé. Les spécifications sont disponibles.   |
| MATERIAL EXCHANGE FORMAT (MXF)          | JPEG 2000, SANS PERTE, COMPRESSÉ                                | Ce format a été créé par la SMPTE (Society of Motion Picture and Television Engineers). C'est un format conteneur qui facilite le travail sur les différents composants d'un fichier vidéo.<br>Ce format est largement utilisé. Les spécifications sont disponibles.   |
| <b>FORMATS ACCEPTABLES</b>              |   |  |
| AUDIO VIDEO INTERLEAVED FORMAT (AVI)    | JPEG 2000<br>DV-NTSC<br>AVC/H.264                               | Voir AVI Non compressé 4:2:2.  |
| MPEG-2 VIDEO (MPEG2)                    |   | Ce format est couramment utilisé pour les DVD. Il est aussi utilisé dans la diffusion de télévision numérique par satellite, câble, réseau de télécommunications ou hertzien (TNT).<br>Ce format est un format ouvert et largement utilisé.<br>Il est reconnu par la norme ISO 13818.  |
| MPEG 4                                  | AVC/H.264   | Ce format est utilisé pour les applications multimédias comme le téléchargement et la diffusion en continu sur Internet, le multimédia sur téléphone mobile, la radio numérique, les jeux vidéo, la télévision et les supports de haute définition.<br>Ce format est largement utilisé. Il est reconnu par la norme ISO 14496. |
| WINDOWS MEDIA VIDEO 9 FILE FORMAT (WMV) | VC-1  | Ce format a été créé par Microsoft. C'est un format commun utilisé pour stocker la vidéo sur Windows. Il est largement utilisé.<br>Les spécifications sont disponibles.  |

## 7. Données géospatiales

| FORMAT  | NOTES                             | COMMENTAIRES   |
|---|-----------------------------------|--|
| <b>FORMATS FAVORISÉS PAR BANQ</b>                 |                                   |  |
| GEOSPATIAL TAGGED IMAGE FILE<br>FORMAT (GEO TIFF) |                                   | Ce format ouvert permet d'ajouter des informations de géoréférencement à une image TIFF.   |
| GEOGRAPHIC MARKUP LANGUAGE<br>(GML)               |                                   | Ce format ouvert a été créé par l'Open Geospatial Consortium pour garantir l'interopérabilité des données dans le domaine de l'information géographique et de la géomatique.<br>Il est reconnu par la norme ISO 19136.   |
| KEYHOLE MARKUP LANGUAGE<br>(KML)                  | CE FORMAT EST BASÉ<br>SUR LE XML. | Ce format ouvert a été documenté par l'Open Geospatial Consortium. Il est basé sur le langage XML est utilisé pour afficher des données géographiques dans un navigateur tel que Google Earth. GML et KML sont complémentaires.<br>Ce format est un format ouvert et largement utilisé.          |
| ESRI SHAPEFILE                                    |                                   | Ce format est utilisé par les systèmes d'informations géographiques. Créé au départ par ESRI pour ses logiciels commerciaux, ce format est désormais devenu un standard et est utilisé par un grand nombre de logiciels libres et propriétaires.<br>C'est un format ouvert et largement utilisé. |



## 8. Documents de conception assistée par ordinateur (CAO)

| FORMAT   | NOTES | COMMENTAIRES   |
|--|-------|--|
| <b>FORMATS FAVORISÉS PAR BAnQ</b>  |       |  |
| AUTODESK'S DRAWING<br>INTERCHANGE FILE<br>FORMAT/DATA EXCHANGE<br>FORMAT (DXF) |       | <p>Ce format a été créé par Autodesk. Il sert à échanger des fichiers entre systèmes CAO n'utilisant pas le même format de fichier natif.</p> <p>Ce format est largement utilisé.</p> <p>Les spécifications sont disponibles.</p>  |
| AUTODESK'S DRAWING FILE<br>(DWG)   |       | <p>Ce format a été créé par Autodesk. C'est le format natif des fichiers de dessins AutoCAD. Il est devenu un standard dans son industrie et est utilisé par de nombreux logiciels.</p> <p>Ce format est très largement utilisé.</p> <p>Les spécifications sont disponibles, car elles ont été documentées par l'Open Design Alliance.</p> |
| <b>FORMATS ACCEPTABLES</b>   |       |  |
| PORTABLE DOCUMENT<br>FORMAT/ENGINEERING (PDF/E)                                |       | <p>Ce format est utilisé pour l'échange de documents d'ingénierie et est basé sur le format PDF.</p> <p>Il est reconnu par la norme ISO 24517.</p>   |
| STANDARD FOR THE EXCHANGE<br>OF PRODUCT MODEL DATA<br>(STEP)                   |       | <p>Ce format est utilisé pour partager des données entre différents systèmes de CAO.</p> <p>Il est reconnu par la norme ISO 10303.</p>   |

## 9. Ensembles de données

Les données structurées doivent respecter les conventions syntaxiques reliées au format et être valides selon les règles définies dans le schéma ou la définition du document (DTD) associé. Les fichiers associés nécessaires pour vérifier la validité des données doivent être fournis. Une documentation exhaustive précisant les noms et la définition des champs devra aussi être fournie.

| FORMAT  | NOTES  | COMMENTAIRES  |
|---|--|---|
| <b>FORMATS FAVORISÉS PAR BAQ</b>                |  |   |
| COMMA SEPARATED VALUE (CSV)                     |  | Ce format permet de représenter des structures de données simples. C'est un format ouvert et très largement utilisé.  |
| ASCII TEXT                                      | 7 BIT  | Voir la section 1.  |
| <b>FORMATS ACCEPTABLES</b>                      |  |   |
| MICROSOFT EXCEL (XLSX)                          | MICROSOFT WORD POUR WINDOWS, VERSION 2007-2010 | Ce format est utilisé par Microsoft Excel depuis sa version 2007. Il est basé sur la norme Office Open XML, reconnue norme ISO 29500. Ce format est très largement utilisé.   |
| MICROSOFT EXCEL 97 BINARY DOCUMENT FORMAT (XLS) | VERSION 8.0                                    | Ce format propriétaire de la suite Microsoft Office est très largement utilisé. Ce format reste cependant peu documenté.  |
| OPENDOCUMENT FORMAT SPREADSHEET (ODS)           |  | L'extension ODS est utilisée pour les fichiers de type tableur utilisant le format ODF. Voir ODF à la section 1.  |
| EXTENSIBLE MARKUP LANGUAGE (XML)                | VERSION 1.1                                    | Ce format est utilisé notamment pour la gestion et l'échange d'informations sur Internet. Le XML est un langage qui permet de structurer des données textuelles. Il est standardisé par le W3C. C'est un format ouvert et très largement utilisé. |
| JAVASCRIPT OBJECT NOTATION (JSON)               |  | Ce format est considéré comme une option alternative légère à XML. Ce format est largement utilisé. Les spécifications sont disponibles.  |

## 10. Collectes de sites Web

| FORMAT                           | NOTES | COMMENTAIRES  |
|----------------------------------|-------|---|
| <b>FORMATS FAVORISÉS PAR BAQ</b> |       |   |
| WEB ARCHIVE FORMAT (WARC)        |       | Ce format est couramment utilisé au niveau international pour effectuer des collectes de sites Web à grande échelle. Il est reconnu par la norme ISO 28500. |
| ARCHIVE FILE FORMAT (ARC)        |       | Ce format est la version précédente du format WARC. Les spécifications sont disponibles.  |