

<http://pasq.fr/shapefile-doit-mourir>

Shapefile doit mourir !

Rédigé par Alain 25 mai 2019 [Aucun commentaire](#)



Je vous propose la traduction de <http://switchfromshapefile.org/> car j'adhère totalement à l'idée, surtout que l'on pourrait en dire la même chose de pas mal de formats : .dwg, .xls et .doc, .mp3, .AI....Bref, tous ces formats plus ou moins propriétaires qui vous enferment dans une solution logicielle. Vive les formats "ouverts" et "libres". J'ai commenté avec [NdT].



Se passer de Shapefile

ESRI Shapefile est un format de fichier pour le stockage de données vectorielles géospatiales. Il existe depuis le début des années 1990 et demeure le format d'échange de données vectorielles le plus couramment utilisé. Bien que les Shapefiles aient permis de nombreuses activités réussies au fil des ans, ils ont également un certain nombre de limitations qui compliquent le développement logiciel et réduisent l'efficacité.

Nous, membres de l'industrie des Technologies de l'information géospatiale, **croions qu'il est temps de cesser d'utiliser Shapefiles comme principal format d'échange de données vectorielles** et de les remplacer par un format qui tire parti des énormes progrès qui ont été réalisés depuis l'introduction de Shapefile.

Le bon côté

Shapefile fait beaucoup de choses bien. Voici quelques raisons pour lesquelles Shapefile est si fortement utilisé :

- Shapefile est **de loin le format le plus largement supporté** dans les logiciels existants.
- Bien que le format soit propriétaire, ses **spécifications sont ouvertes**.
- Pour de nombreux cas d'utilisation, **c'est suffisant**.
- Les fichiers d'index (par exemple *.shx) permettent une **bonne performance de lecture**.
- Il est relativement **efficace en termes de taille de fichier**. Le fichier résultant, même décompressé, est relativement petit par rapport à d'autres formats (principalement textuels).

Shapefile est un mauvais format

Pourquoi Shapefile est-il si mauvais ? Voici plusieurs raisons pour lesquelles le Shapefile est un mauvais format et vous devriez éviter son utilisation : [NdT : voir aussi <http://desktop.arcgis.com/fr/arcmap/latest/manage-data/shapefiles/geoprocessing-considerations-for-shapefile-output.htm>]

1. [Pas de définition du système de référence des coordonnées](#).
2. [C'est un format multi-fichier](#).
3. [Les noms d'attributs sont limités à 10 caractères](#).
4. [Seulement 255 attributs](#). Le fichier DBF ne vous permet pas de stocker plus de 255 champs d'attributs.
5. [Types de données limités](#). Les types de données sont limités à flottant, entier, date et texte avec un maximum de 254 caractères.
6. [Jeu de caractères inconnu](#). Il n'y a aucun moyen de spécifier le jeu de caractères utilisé dans la base de données.
7. Il est [limité à 2 Go de taille de fichier](#). Bien que certains outils soient capables de dépasser cette limite, ils ne peuvent jamais dépasser 4 Go de données.
8. [Pas de topologie dans les données](#). Il n'y a aucun moyen de décrire les relations topologiques dans le format.
9. [Type de géométrie unique par fichier](#). Il n'y a aucun moyen de sauvegarder les caractéristiques géométriques mixtes.
10. Des [structures de données plus complexes sont impossibles à sauvegarder](#). C'est un format "flat table".
11. Il n'y a [aucun moyen de stocker des données 3D](#) avec des textures ou des apparences telles que des définitions de matériaux. Il n'y a également aucun moyen de stocker des solides ou des objets paramétriques.
12. [Définition des projections](#). Ils sont incompatibles ou manquants.
13. Le [type de géométrie des lignes et des polygones](#), en une ou plusieurs parties, ne peut pas être déterminé de façon fiable au niveau de la couche, il doit être déterminé au niveau des caractéristiques individuelles.

1. Pas de définition du système de référence des coordonnées

Par défaut, il n'y a pas de définition du système de coordonnées de référence utilisé. Vous pouvez le faire en utilisant par exemple un .prj, mais premièrement, ce n'est pas une partie standard de la spécification et deuxièmement, il y a encore quelques problèmes, voir les problèmes de projection et le [format multi-fichier](#) plus bas.

2. Format Multifile

Le format Shapefile utilise [au moins 3 fichiers \(*.shp, *.dbf, *.prj\) \(wikipédia anglais plus complet\)](#). Les utilisateurs ne peuvent pas partager un seul fichier, vous devez les envoyer tous. Les utilisateurs zippent généralement tous les fichiers dans une archive et les décompressent à l'autre bout de la chaîne de distribution, mais c'est encombrant et **source d'erreurs**.

De plus, d'autres progiciels géospatiaux ajoutent régulièrement leurs propres extensions pour tenter de surmonter les limites de Shapefile. **Les ajouts personnalisés ne sont pas pris en charge par d'autres outils et limitent l'interopérabilité.** ([cf. liste wikipédia](#))

3. 10 caractères par noms d'attributs

Les **noms d'attributs sont limités à 10 caractères maximum**. Les noms plus longs sont généralement automatiquement raccourcis. Cela conduit à des noms d'attributs abrégés et/ou cryptiques qui ne sont pas intuitifs pour le destinataire des données.

4. 255 champs d'attributs

Il ne peut y avoir que **255 champs d'attributs** dans le fichier de base de données. Pour certaines applications, cela est limitatif, en particulier en combinaison avec la structure de la table plate.

5. Faible prise en charge des types de données d'attributs

Les types de données **flottant, entier, date et chaîne de caractères** sont supportés. Les nombres à virgule flottante peuvent être stockés sous forme de texte, mais il n'y a pas de support pour les grands nombres entiers (donc le format n'est pas utilisable, vous avez des données avec de grands nombres entiers, comme les cartes cadastrales) et le texte est limité à 254 caractères seulement.

Il n'y a pas de support pour les champs de données plus avancés tels que les blobs, les images ou les tableaux. [NdT : il ne peuvent pas stocker les heures dans les formats date aussi]

6. Jeu de caractères inconnu

Il n'y a aucun moyen de spécifier le **jeu de caractères utilisé dans la base de données**. De nombreuses applications utilisent les anciens encodages de données Windows-* ou ISO-*, alors qu'aujourd'hui nous avons tendance à utiliser UTF-8 davantage. Il n'y a toujours pas moyen de le spécifier dans l'en-tête du fichier. Le support des caractères Unicode est également très limité.

7. 2 Go Limite de taille

La taille des fichiers de composants.shp et.dbf ne peut dépasser 2 Go. GDAL Shapefile dépasse cette limite, mais le pilote

Le format Shapefile utilise explicitement des décalages de 32 bits et ne peut donc pas dépasser 8 Go (il utilise en fait des décalages de 32 bits en mots de 16 bits), mais l'implémentation OGR shapefile est limitée à 4 Go.

Pour des raisons de compatibilité avec d'autres implémentations logicielles, **il n'est pas recommandé d'utiliser une taille de fichier supérieure à 2 Go** pour les fichiers .SHP et .DBF

Ainsi, 4 Go est tout ce que vous pouvez avoir dans un seul fichier Shapefile. Cela semble suffisant, mais pas dans tous les cas.

8. Format non topologique

Shapefile est un format à fonctions simples. Il n'y a **aucun moyen de stocker des relations géométriques plus complexes**.

9. Pas de géométrie mixte

Chaque fichier ne peut être **qu'un des formats géométriques supportés** (Point, Ligne, Polygone et autres). Les caractéristiques géométriques mixtes ne sont pas possibles.

10 .Structure de données plate

La structure des données est limitée aux tables "plates" sans hiérarchies, relations ou arborescence.

11. Support 3D très limité

Shapefile ne peut pas stocker les définitions de matériaux ni les textures (images avec coordonnées de texture). Les modèles 3D sont stockés sous la forme d'une soupe triangulaire ou polygonale, sans qu'aucun modèle étanche *[NdT : il est possible que les modèles présentent des lacunes ou des trous. C'est pourquoi nous nous référons au terme "étanche à l'eau". Cela signifie essentiellement que vos modèles doivent être scellés ou fermés. Un bon moyen de visualiser ceci est de demander si cela flotterait ou coulerait.]* ou géométrie paramétrique ne soit pris en charge.

12. Incohérences dans la définition des projections

Par défaut, Shapefile ne contient **aucune information sur le système de référence des coordonnées**. Mais certains logiciels acceptent les fichiers *.prj, qui peuvent contenir une description CRS.

Il utilise les définitions Esri WKT, qui sont souvent incompatibles avec les définitions standard de l'EPSG ou d'autres sources concernant des aspects tels que l'ordre des axes ou les définitions des unités. De plus, ils manquent souvent les paramètres nécessaires à la reprojection ("Missing [Bursa Wolf Parameters](#)", n'importe qui ?)