

Préservation des systèmes d'information géographique (SIG)

Collection de notes d'orientation sur les types de
données

Artefactual Systems et Digital
Preservation Coalition



**Veille technologique de
DPC, Note d'orientation**

Juillet 2021



DigitalPreservationCoalition

Informations sur la traduction française

La traduction française de cette note d'orientation a été réalisée dans le cadre des travaux de la Cellule nationale de veille sur les formats entre janvier et mars 2022. Lancée en 2019, la Cellule nationale de veille sur les formats (CNVF), sous l'égide de l'association française Aristote et de son groupe de travail sur la Pérennisation de l'information numérique (PIN), regroupe à ce jour une douzaine de partenaires. Ses principaux objectifs sont la mutualisation des activités de veille sur les formats, la sensibilisation des professionnels sur le sujet, la contribution ou l'influence sur les outils associés. Elle a pour ambition d'être un interlocuteur francophone reconnu dans les travaux internationaux sur ces sujets.

Le manuel a été traduit en français par : Wilfried Prieur (ministère de la Culture) et Édouard Vasseur (École nationale des Chartes). Une première traduction automatique a été réalisée avec le logiciel DeepL puis a été revue et corrigée par le groupe de travail.

Collection de notes d'orientation sur les types de données

Chaque note d'orientation de la collection « Types de données » est conçue pour fournir une introduction à l'état actuel des connaissances de la communauté sur les types de données couramment rencontrés par ceux qui cherchent à préserver les collections numériques. La préservation numérique consiste à faire en sorte que l'information reste trouvable, utilisable et fiable à long terme. La meilleure approche pour un entrepôt numérique varie en fonction du périmètre et du contenu de ses collections, des ressources dont il dispose et des attentes de ses financeurs et utilisateurs. Il existe cependant de bonnes pratiques largement applicables qui ont été définies à la suite de nombreuses années de recherche, de mise en œuvre concrète et d'élaboration de consensus. Elles sont présentées ici comme un point de départ, et comprennent des ressources complémentaires pour approfondir les recherches.

Cette collection de notes d'orientation sur les types de données a été rédigée par le personnel d'*Artefactual Systems* en collaboration avec la *Digital Preservation Coalition*. Ces notes ont été élaborées en collaboration avec la *UK Nuclear Decommissioning Authority*.

La préservation numérique est un domaine en pleine évolution et qui se développe en réponse à des facteurs externes et des nouveaux défis. De nouveaux formats, normes et exemples de bonnes pratiques apparaîtront au fil du temps ; ainsi, les informations contenues dans ce rapport devront être mises à jour. Nous vous invitons à nous faire part de vos commentaires et de vos réactions à l'adresse suivante : info@dpconline.org.

1. Présentation

Les systèmes d'information géographique (SIG) sont des outils logiciels multi-utilisateurs qui permettent de saisir et d'analyser des données géographiques et spatiales (géospaciales). Ils présentent ces informations sous forme de cartes interactives à multiples facettes. Les données géospaciales représentent des informations sur des lieux (latitude, longitude, altitude, temps) à l'intérieur de limites géographiques définies (étendues). La caractéristique distinctive d'un SIG est qu'il utilise un emplacement spatio-temporel comme clé commune depuis laquelle toutes les autres informations du système sont indexées. Les cartes SIG sont organisées en couches de données, telles que les rues, les restaurants, les rivières, les arbres ou les bouches d'incendie, pouvant être activées ou désactivées pour superposer des informations sur une seule carte. Les données sont horodatées, ce qui permet aux couches d'illustrer les changements dans le temps.

Un SIG peut être constitué de tout ou partie des éléments suivants :

- une ou plusieurs bases de données pour le stockage de l'information (voir aussi : Note d'orientation – Préservation des bases de données (DPC, 2021) ;
- un logiciel middleware pour la prise en compte des règles métier ;
- des cartes sous formes de tuiles ;
- des photographies aériennes/satellites sous forme de tuiles ;
- des fichiers de formes vectorielles ;
- des éléments graphiques de la carte ;
- différentes interfaces graphiques pour la saisie et l'affichage des données.

Un SIG peut dépendre de logiciels et de plates-formes tiers pour le rendu de ses cartes en ligne. Les deux systèmes les plus couramment utilisés sont Google Earth (Google Earth, 2021) et OpenStreetMap (OpenStreetMap, 2021). Les applications non-SIG utilisent également ces logiciels et plates-formes dans une pratique appelée « geo-tagging » (ou géomarquage) pour intégrer des cartes avec des points de localisation dans tout logiciel accessible sur le Web.

Les SIG sont largement utilisés dans les applications de gestion des limites de propriété, de planification urbaine, d'administration des licences, de services publics, d'intervention d'urgence, de données météorologiques, de gestion des ressources naturelles et du patrimoine archéologique et monumental, etc. Les SIG sont également largement utilisés par les entreprises d'ingénierie, de transport, de logistique, d'assurance et de télécommunications.

Une introduction utile aux données et formats géospaciaux a été produite par la Library of Congress (LoC, 2017).

2. Défis en matière de préservation

Les défis liés à la préservation des SIG sont résumés ci-dessous et comprennent des problèmes fréquemment rencontrés avec des types de données complexes, tels que le volume élevé de fichiers, les relations entre les fichiers et les liens avec des systèmes et des données externes.

2.1. Champ d'application ambigu

Les données des SIG sont souvent utilisées dans des applications qui vont bien au-delà de celles initialement prévues. Leurs informations géolocalisées sous forme de séries chronologiques ont tendance à avoir une grande utilité dans d'autres contextes, ce qui signifie que de nombreuses autres applications sont alimentées par des SIG, telles que les systèmes de gestion de documents, les logiciels de conception assistée par ordinateur (CAO), les systèmes de Building Information Modeling ou BIM (Modélisation des données du bâtiment), les systèmes de planification des ressources de l'entreprise (ERP), les répertoires de fichiers partagés, les applications mobiles ou les services de cartographie en ligne. Cela signifie souvent qu'il n'existe pas de délimitation claire de l'endroit où le SIG commence et finit, ce qui rend l'évaluation et la saisie difficiles.

Vous trouverez de plus amples informations sur les modèles 3D et la CAO dans les Notes d'orientation sur la 3D (Artefactual & DPC, 2021a) et la CAO (Artefactual & DPC, 2021b).

2.2. Volume et volatilité

Un SIG est généralement volumineux en termes de nombre de fichiers, et peut être de grande taille lorsque qu'il contient de nombreux fichiers d'images matricielles. Les informations géospatiales sont fréquemment mises à jour, et de nombreux producteurs de données écrasent les valeurs de données existantes plutôt que d'utiliser une gestion des versions. Les professionnels emploient le terme « archivage » pour désigner certaines fonctionnalités de ces applications (ESRI, 2021), ce qui peut créer des difficultés pour la capture des données du SIG à des fins d'archivage historiques ou de preuve.

2.3. Droits d'auteur sur le contenu

Certains contenus SIG, tels que les photographies aériennes et les relevés LiDAR, peuvent être protégés par des droits d'auteur. Dans ce cas, l'utilisation de ce contenu peut être limitée aux détenteurs de la licence du SIG source.

2.4. Relations complexes entre les données

La plupart des interactions des utilisateurs avec un SIG consistent à rassembler plusieurs fichiers de formats différents pour obtenir une visualisation sous forme cartographique. La gestion des données nécessaires pour recréer des vues avec précision est complexe. La simple préservation d'instantanés des bases de données du SIG avec des copies de leurs tuiles matricielles et de leurs fichiers de forme vectorielle ne suffit souvent pas à fournir un accès adéquat aux informations géospatiales historiques. Un logiciel de SIG est alors nécessaire pour orchestrer les relations et les comportements entre ces différents composants. Cependant, de nombreuses organisations responsables de la préservation numérique n'utilisent pas de logiciel SIG.

3. Formats de fichiers

Il n'existe pas de format unique parfait pour la préservation et l'utilisation future des SIG. Les décisions concernant les formats de fichiers doivent dépendre des caractéristiques et des fonctionnalités à préserver, ainsi que des cas d'utilisation future à prendre en charge. Il est à noter

que le tableau ci-dessous ne fournit pas une liste exhaustive des formats adaptés à la préservation et à l'accès. Le format le plus approprié pour préserver les caractéristiques et fonctionnalités importantes d'un fichier peut être le format original dans lequel il a été créé. La Library of Congress (2020-2021) recommande de cibler des formats de préservation qui maintiennent l'exhaustivité des données « avec une préférence pour la préservation du format natif et de la projection des données ». Cela inclut l'utilisation de formats de fichiers propriétaires de la version native. Il est donc recommandé d'effectuer des recherches et des analyses minutieuses avant de migrer des fichiers vers un nouveau format.

Format de fichier	Extension	Brève description
Fichier de projet ArcGIS Pro	.aprx	Les fichiers de projet APRX contiennent des cartes, des boîtes à outils, des bases de données, des dossiers et des styles. Ils peuvent également contenir des connexions à des bases de données, des serveurs et des dossiers, et peuvent inclure plusieurs cartes et mises en page dans un seul projet. Ils sont conçus comme un conteneur pour « archiver » toutes les ressources liées à un projet. Cependant, les fichiers APRX constituent un format propriétaire qui nécessite l'application d'ESRI ArcGIS Pro pour les restituer avec leur fonctionnalité intacte.
Format d'échange des données ASPRS LiDAR	.las, .lasd, .laz	Light imaging Detection and Ranging (détection et estimation par la lumière) (LiDAR) désigne une technologie de télédétection optique utilisée dans de nombreux secteurs pour visualiser la surface de la terre et réaliser des cartes haute définition. Les données LiDAR peuvent être importées dans des applications comme ArcGIS pour ajouter de nouvelles couches riches aux cartes. La technologie LiDAR consiste en un réseau dense de coordonnées tridimensionnelles de nuages de points avec des valeurs d'élévation. Le format d'échange des données Lidar (LAS) est un format de fichier binaire qui a remplacé les formats propriétaires et les formats basés sur l'ASCII comme format d'échange standard du secteur pour les données LiDAR (ASPRS, 2021). En utilisant le format de fichier LAS, la compression est appliquée aux fichiers LAS pour économiser un espace de stockage important sans perte d'information. Les ensembles de données LAS (LAS Datasets, LASD) font référence à un ensemble de fichiers LAS pour permettre l'examen des propriétés des nuages de points tridimensionnels, la visualisation des surfaces triangulées et l'analyse statistique.
Base de données	.gdb	La base de données géographiques ESRI désigne la structure de données native des applications ArcGIS (LC, 2020). Il s'agit du stockage physique

géographiques ESRI		<p>des informations géographiques, principalement à l'aide d'un SGBD ou d'un système de fichiers. Les bases de données géographiques ESRI sont les exemples les plus omniprésents de bases de données spatiales, en d'autres termes, des systèmes de gestion de bases de données optimisés pour le stockage et l'interrogation de données qui représentent des objets dans un espace géométrique.</p> <p>ESRI permet aux utilisateurs de créer des « jeux de données » pour organiser et gérer les informations géographiques. Les trois principaux types de jeux de données incluent les classes de caractéristiques, les jeux de données matricielles et les tables d'attributs. Les utilisateurs étendent les jeux de données de leurs bases de données géographiques avec des fonctionnalités plus avancées (en ajoutant des topologies, des réseaux ou des schémas spécifiques à un domaine, par exemple) pour modéliser le comportement du SIG, maintenir l'intégrité des données et travailler avec un ensemble de relations spatiales.</p>
Fichier de formes ESRI	.shp, .dbf, .shx	<p>Le fichier de formes ou « shapefile » désigne de loin le format de fichier géospatial le plus omniprésent. En tant que norme du secteur, ce type est recommandé comme format de préservation pour les données des entités SIG (LC, 2020). Un fichier de formes stocke la géométrie d'une entité sous forme d'un ensemble de coordonnées vectorielles. Il doit comprendre trois sous-fichiers pour être fonctionnel : SHP (géométrie de l'entité), SHX (position de l'indice de forme) et DBF (données d'attribut).</p>
GeoTIFF	.tif .tiff	<p>GeoTIFF (GeoTIFF, 2021) est une extension du format d'image matricielle TIFF qui associe son contenu à une projection cartographique¹ en utilisant un ensemble prédéfini de balises TIFF étendues qui sont intégrées dans l'en-tête du fichier.</p> <p>GeoTIFF est devenu la norme du secteur pour les fichiers d'images SIG et est recommandé comme format de préservation pour les données matricielles (LC, 2009). Il est pris en charge par toutes les principales applications logicielles SIG. La plupart des agences et services qui produisent</p>

¹ NdT : La projection cartographique est un ensemble de techniques géodésiques permettant de représenter une surface non plane (surface de la Terre, d'un autre corps céleste, du ciel, ...) dans son ensemble ou en partie sur la surface plane d'une carte. Il existe de nombreux types de projection, on peut citer par exemple les projections de Cassini, Mercator, Lambert, de Fuller, de Winkel-Tripel.

		<p>des données géospatiales fournissent leurs images au format GeoTIFF (LC, 2009).</p> <p>Notons également le format COG (Cloud Optimized GeoTIFF), une autre variante du format TIFF. Il s'agit d'un fichier GeoTIFF ordinaire qui est hébergé sur un serveur de fichiers HTTP. Il utilise des requêtes HTTP GET pour récupérer uniquement les parties spécifiques d'un fichier GeoTIFF qui sont requises pour un flux de travail basé sur le cloud (Cloud Optimized GeoTIFF, 2021).</p>
Geography Markup Language (GML)	.gml	<p>Le Geography Markup Language (langage de balisage géographique ou GML) est un schéma XML défini par l'Open Geospatial Consortium (OGC) pour exprimer les caractéristiques géographiques (GML, 2021). Les principaux types d'objets géométriques GML incluent les points, les chaînes de lignes et les polygones. Ils définissent des emplacements ou des étendues (régions).</p> <p>GML comprend également des objets caractéristiques pour représenter des entités physiques (par exemple, des bâtiments et des rivières), qui peuvent avoir ou non des aspects géométriques. GML sert de langage de modélisation et de format d'échange ouvert. Il fait l'objet d'une publication croisée sous la référence ISO 19136:2020.</p> <p>GML est adapté à la préservation car son format XML est compatible avec des SIG anciens et disparates. Il permet également l'intégration de multiples formes d'informations géographiques, y compris, par exemple, les objets vectoriels, les couvertures et les données de capteurs. Des communautés d'intérêts spécifiques (par exemple, le secteur du tourisme) ont créé leurs propres schémas d'application XML basés sur le GML pour décrire les types d'objets que leurs applications SIG communautaires doivent pouvoir traiter et exposer.</p>
GeoJSON	.json	<p>GeoJSON est un format d'échange de données géospatiales maintenu en tant que norme ouverte par l'Internet Engineering Task Force comme RFC 7946 (IETF, 2016). GeoJSON représente des entités géographiques et leurs attributs non spatiaux. Il prend en charge les types d'entités suivants : point, chaîne de lignes, polygone et topologies. Étant donné que JSON est plus compact que XML, il s'agit du format idéal pour transférer des données spatiales vers des applications Web et mobiles. GeoJSON est</p>

		largement utilisé dans la communauté SIG. ArcGIS et QGIS le prennent en charge. Il est considéré comme un format de préservation acceptable (LC , 2020-2021).
Google Keyhole Markup Language (KML/KMZ)	.kml .kmz	KML a été développé par Keyhole Inc. qui a été racheté par Google (Wikipedia , 2021). KML est un format XML utilisé principalement dans le navigateur Google Earth, ce qui lui assure une large base d'utilisateurs. KML est maintenu par l'Open Geospatial Consortium, Inc. (OGC) depuis 2008. Les fichiers KML sont souvent distribués dans des conteneurs compressés portant l'extension KMZ. Un fichier KMZ inclut un document KML racine unique et des superpositions facultatives, des images, des icônes ou des fichiers de modèle 3D COLLADA. KML peut être utilisé pour rendre du contenu GML, mais diffère considérablement de GML en ce qu'il est avant tout un format de rendu 3D, et non un format d'échange de données. Les créateurs de données doivent donc éviter d'encoder du contenu GML pour le représenter avec KML, car cela entraîne une perte importante et irrécupérable de contenu et de contexte.
PostGIS + PostgreSQL	https://www.postgresql.org/	PostGIS est une application logicielle libre et Open Source qui prend en charge les objets géographiques dans une base de données relationnel-objet PostgreSQL. PostGIS ajoute la géométrie, la géographie, les matrices et d'autres types d'informations géospatiales à la base de données PostgreSQL, ainsi que des fonctions dédiées, des opérateurs et des améliorations d'index qui leur sont applicables. Ces fonctions permettent d'exécuter des requêtes de localisation en SQL. PostGIS suit l'ensemble de normes Simple Features de l'Open Geospatial Consortium qui spécifie un modèle commun de stockage et d'accès pour les caractéristiques géographiques. De nombreux SIG, dont ArcGIS et QGIS, peuvent utiliser PostGIS comme backend de leur base de données. PostGIS + PostgreSQL est recommandé comme format de préservation des données SIG tabulaires.
Fichier de projet QGIS	.qgs, .qgz	Les fichiers de projet QGS stockent les couches cartographiques du projet avec des liens vers les ensembles de données sous-jacents et d'autres propriétés de couche, notamment l'arborescence de la couche cartographique, les étendues, le système de référence spatiale, le système de référence des coordonnées, les styles, les

		<p>moteurs de rendu, les modes de fusion et les valeurs d'opacité. Il contient également des mises en page d'impression, des relations de tableaux, des macros de projet, des paramètres de plugins et des paramètres de serveur QGIS (Association QGIS.org, 2020).</p> <p>Le fichier de projet QGS est enregistré dans un format XML qui peut être modifié en dehors de QGIS. Par défaut, l'application compresse les fichiers de projet dans un format QGZ contenant un fichier QGS et un fichier QGD. Le fichier QGD représente la base de données SQLite associée au projet QGIS qui contient les données auxiliaires du projet. Le fichier QGZ peut être ouvert avec n'importe quel utilitaire ZIP.</p>
PDF géospatial Adobe	.pdf	<p>Le PDF géospatial désigne un ensemble d'extensions géospatiales de la spécification PDF (Portable Document Format) 1.7 qui permet de relier une région de la page du document à une région de l'espace physique (Adobe, 2021). On parle souvent de geo-tagging (« géomarquage » ou « géoréférencement »). Notez que GeoPDF est un format différent qui se réfère spécifiquement aux fichiers produits par les applications TerraGo (TerraGo, 2021).</p> <p>Un PDF géospatial peut contenir des éléments géométriques tels que des points, des lignes et des polygones qui peuvent représenter des caractéristiques telles que des bâtiments, des routes ou des limites de ville. Ces fichiers peuvent être écrits à l'aide d'Adobe Acrobat ou d'applications SIG telles qu'ArcGIS et QGIS. Bien qu'Adobe propose sa propre extension, les normes de l'Open Geospatial Consortium (OGC, 2021) doivent être utilisées pour ajouter des métadonnées géospatiales aux documents PDF. Il s'agit d'un format d'accès acceptable, mais il est recommandé de privilégier le format de fichier natif pour la préservation (LC, 2020-2021).</p>
MapInfo TAB	.tab .dat .id .map .ind	<p>Il s'agit d'un format propriétaire développé pour être utilisé avec le logiciel SIG MapInfo. (LC, 2011). Bien que les fichiers Mapinfo puissent être ouverts dans le logiciel libre et Open Source QGIS, il s'agit d'un format fermé sans spécification complète qui n'est donc pas recommandé pour la préservation à long terme. Il peut toutefois être utilisé comme format d'accès.</p>

4. Standards de métadonnées

Il existe des normes internationales de métadonnées pour les données géospatiales :

- La norme « ISO 19115-1:2014 Information géographique - Métadonnées - Partie 1 : Principes de base » articule le schéma de description des informations et services géographiques ([ISO, 2014](#)).
- La norme « ISO 19115-2:2019 Information géographique - Métadonnées - Partie 2 : Extensions pour l'acquisition et le traitement » étend le schéma ISO 19115-1:2014 pour inclure les métadonnées d'acquisition et de traitement, telles que les méthodes numériques et les procédures de calcul utilisées pour dégager l'information géographique ([ISO, 2019](#)).

D'autres normes nationales ou régionales peuvent également être pertinentes. Elles sont souvent basées sur la famille de normes ISO 19115 mais sont adaptées à des régions et/ou des cas d'utilisation spécifiques.

- [INSPIRE](#) est une directive de la Commission européenne visant à créer une infrastructure de données spatiales de l'Union européenne pour permettre le partage d'informations spatiales environnementales entre les organisations du secteur public, faciliter l'accès aux données spatiales et aider à l'élaboration de politiques au-delà des frontières.
- [UK GEMINI](#)² est la norme britannique sur les métadonnées géographiques, qui fournit des conseils sur la manière de publier des métadonnées géographiques conformément aux directives du gouvernement britannique et aux normes ISO pertinentes.

5. Conseils pour les créateurs

- Les créateurs doivent documenter leur architecture et leurs processus SIG. Pour les archivistes, il est essentiel de comprendre le contexte de création et d'utilisation des documents. Cependant, la plupart des informations sur l'architecture du SIG et les processus métier sont implicites. Si des instructions explicites pour l'utilisateur du SIG sont préservées et maintenues de manière proactive, cela aidera grandement les archivistes à rendre les copies historiques du SIG plus compréhensibles pour les futurs utilisateurs.
- Dans la mesure du possible, documentez le contenu et les métadonnées des couches spécifiques du SIG. Si le SIG a été déployé à l'aide d'une architecture formelle et d'une méthodologie de processus métier, la préservation de cette documentation sera d'une grande aide pour comprendre et restituer les interfaces et les informations du SIG original. Il est important pour les futurs utilisateurs que les modifications apportées au contenu du SIG au fil du temps puissent être suivies ; par conséquent, les producteurs de données SIG doivent activer les fonctions de gestion des versions offertes par les plates-formes SIG pour garantir que l'historique des données n'est pas écrasé.
- Utilisez des normes ouvertes ou sectorielles. Les créateurs doivent s'assurer que les outils qu'ils utilisent permettent d'enregistrer ou d'exporter le contenu dans des formats ouverts ou standard tels que *Shapefile*, *GeoJSON* et *GeoTIFF*. L'utilisation de normes ouvertes augmentera la probabilité que les données SIG restent accessibles et utilisables à long terme dans les applications actuelles ou futures.

² NdT : Pour la France, on pourra se reporter sur les guides élaborés par le Centre national de l'information géographique disponible à cette adresse, http://cnig.gouv.fr/?page_id=9237

6. Conseils pour les archivistes

6.1. Conseils généraux³

Les ressources suivantes fournissent des conseils sur la préservation et l'accès aux SIG :

- [Geospatial Multistate Archive and Preservation Partnership's](#) (2011) *Key Findings and Best Practices*.
- [Library of Congress's](#) (2020-2021) *Recommended formats statement: GIS, Geospatial and Non-GIS Cartographic*
- [Rapport final du North Carolina Geospatial Data Archiving Project](#) (2010).

6.2. Aide communautaire

Faites appel à la communauté pour obtenir des conseils et un soutien. Les communautés professionnelles peuvent fournir une mine de connaissances spécialisées sur le sujet, et proposent souvent des publications utiles et des moyens de s'impliquer :

- [L'Open Geospatial Consortium](#) est un consortium international qui compte plus de 500 membres désireux de rendre les informations et services géospatiaux trouvables, accessibles, interopérables et réutilisables (FAIR ou « Findable, Accessible, Interoperable and Reusable »).
- Le [Geospatial Interest Group](#) de la *Research Data Alliance* est un groupe d'intérêt axé sur un domaine particulier, qui s'intéresse à l'interopérabilité des données et à la quantification de l'incertitude dans les ensembles de données.

6.3. Accès

- Envisagez d'installer une copie du logiciel libre et Open Source *QGIS* pour accéder aux données du SIG. *QGIS* peut ouvrir les fichiers ESRI Shapefiles, PostGIS, SpatiaLite, Oracle Spatial, MSSQL Spatial, AutoCAD DXF, et de nombreux autres formats propriétaires ou ouverts. *QGIS* prend en charge les couches vectorielles sous forme de points, de lignes ou de polygones, ainsi que de nombreux formats d'images matricielles. Le logiciel peut également géo-référencer des fichiers image et s'interfacer avec des services de cartographie Web, notamment l'API de géocodage de Google.
- *QGIS* doit être installé sur un poste de travail de bureau dans la salle de référence, car il n'existe pas de version complète accessible par Internet. L'utilisation d'une machine virtuelle ou d'un logiciel de virtualisation comme *VirtualBox* ou d'une plate-forme de conteneurs comme *Docker* peut simplifier le déploiement d'une instance *QGIS*.

6.4. Collecte et évaluation

- Effectuez une analyse des fonctions opérationnelles de l'organisme créateur afin de déterminer où et quand les documents SIG doivent être collectés, à moins que cette information ne soit déjà disponible dans les calendriers de préservation des documents de l'organisme. Une telle analyse devrait permettre de collecter la documentation relative aux utilisateurs, la documentation relative aux processus métier, ainsi que les informations sur les types et la fréquence des mises à jour des données.

³ NdT : Parmi les données disponibles en français, on pourra notamment [consulter le manuel sur l'archivage des géodonnées de la Confédération](#) élaboré par les Archives Fédérales Suisse et l'Office fédéral de topographie (swisstopo) ou encore Laetitia Le Clech, « Archives et géographie : typologie, caractéristiques et perspectives », Archives, Volume 47, numéro 1, 2017, p. 59–83 [https://id.erudit.org/iderudit/1041826ar]

- Des critères suggérés pour l'évaluation et la sélection des données géospatiales sont énumérés dans un rapport de la *National Digital Stewardship Alliance* ([NDSA, 2013](#)).
- Identifiez les rapports ou les captures d'écran qui peuvent être collectés à partir du SIG au format TIFF ou PDF pour compléter les données et la documentation du SIG.
- Pour capturer des données SIG en vue de les intégrer dans un référentiel numérique, utilisez des utilitaires qui exportent des données ou des interfaces de programmation d'applications (API) qui mettent les données SQL, les formes vectorielles et les tuiles matricielles à la disposition d'applications externes.
- [L'Open Geospatial Consortium](#) (OGC) a publié une famille de normes API pour permettre un accès automatisé uniforme aux données géospatiales dans les applications et les services. Les créateurs de données peuvent être encouragés à rendre leurs systèmes compatibles avec ces normes.

6.5. Caractérisation des formats

La caractérisation peut être utile pour identifier les formats de fichiers, extraire les métadonnées, identifier les contenus non fonctionnels ou cryptés, ou encore vérifier la conformité aux profils ou aux normes. Le périmètre d'utilisation et l'efficacité des outils peuvent varier considérablement selon les formats de fichiers.

- Identifiez les formats de fichiers à l'aide d'un outil tel que [DROID](#), [FIDO](#) ou [Siegfried](#) qui utilise le registre de formats de fichiers [PRONOM](#).
- Les outils de validation pour les formats GIS ne sont pas disponibles pour le moment.

6.6. Métadonnées

- Soyez conscient des différents types de métadonnées nécessaires. Tous les SIG contiennent un noyau commun de données spatio-temporelles sur les lieux. Les métadonnées de base peuvent être complétées par des données spécifiques au secteur. Ce large éventail peut rendre difficile l'importation, la normalisation, le traitement ou l'accès aux métadonnées SIG de manière cohérente.

7. Références

Artefactual & DPC (2021a) Preserving 3D. Disponible à l'adresse suivante : <http://doi.org/10.7207/twgn21-14>

Artefactual & DPC (2021b) Preserving CAD. Disponible à l'adresse suivante : <http://doi.org/10.7207/twgn21-15>

ASPRS (2021) LASer (LAS) File Format Exchange Activities. Disponible à l'adresse suivante : <https://web.archive.org/web/20200922151120/https://www.asprs.org/divisions-committees/lidar-division/laser-las-file-format-exchange-activities>

ESRI (2021) What is archiving? Disponible à l'adresse suivante : <https://web.archive.org/web/20210202215457/https://pro.arcgis.com/en/pro-app/latest/help/data/geodatabases/overview/what-is-archiving-.htm>

Geospatial Multistate Archive and Preservation Partnership (2011) Key Findings and Best Practices. Disponible à l'adresse suivante :

<https://web.archive.org/web/20200502204535/https://files.nc.gov/ncdit/documents/files/GeoMAP-P-Best-Practices-2011-12-31.pdf>

GISGeography (2020) The Ultimate List of GIS Formats and Geospatial File Extensions. Disponible à l'adresse suivante : <https://gisgeography.com/gis-formats/>

Internet Engineering Task Force (2016) The GeoJSON Format. Disponible à l'adresse suivante : <https://web.archive.org/web/20210104100025/https://tools.ietf.org/html/rfc7946>

ISO (2019) ISO 19115-2:2019 Geographic information — Metadata — Part 2: Extensions for acquisition and processing. Disponible à l'adresse suivante : <https://web.archive.org/web/20201206022231/https://www.iso.org/standard/67039.html>

ISO (2014) ISO 19115-1:2014 Geographic information — Metadata — Part 1: Fundamentals. Disponible à l'adresse suivante : <https://web.archive.org/web/20210320110509/https://www.iso.org/standard/53798.html>

Library of Congress (2020-2021). Recommended formats statement: GIS, Geospatial and Non-GIS Cartographic. Disponible à l'adresse suivante : <https://web.archive.org/web/20210320174757/https://www.loc.gov/preservation/resources/rfs/geo-carto.html>

Library of Congress (2020) ESRI Shapefile. Disponible à l'adresse suivante : <https://web.archive.org/web/20201117143240/https://www.loc.gov/preservation/digital/formats/fdd/fdd000280.shtml>

Library of Congress (2017) Introduction to Geospatial Resources and Formats. Disponible à l'adresse suivante : https://www.loc.gov/preservation/digital/formats/content/gis_intro.shtml

Library of Congress (2011) MapInfo Dataset format. Disponible à l'adresse suivante : <https://web.archive.org/web/20200223233237/https://www.loc.gov/preservation/digital/formats/fdd/fdd000300.shtml>

Library of Congress (2009) GeoTIFF, Revision 1.0. Disponible à l'adresse suivante : <https://web.archive.org/web/20201030114515/https://www.loc.gov/preservation/digital/formats/fdd/fdd000279.shtml>

National Digital Stewardship Alliance (2013) Issues in the Appraisal and Selection of Geospatial Data. Disponible à l'adresse suivante : https://web.archive.org/web/20170102034153/https://www.digitalpreservation.gov/documents/NDSA_AppraisalSelection_report_final102413.pdf?loclr=blogsig

North Carolina Geographic Information Coordinating Council, Archival and Long Term Access Ad Hoc Committee (2008) Final Report. Disponible à l'adresse suivante : <https://web.archive.org/web/20200706080529/https://files.nc.gov/ncdit/documents/files/GICC-Archival-Long-Term-Access-Report-11-08.pdf>

North Carolina Geospatial Data Archiving Project (2010) Final Report. Disponible à l'adresse suivante : https://web.archive.org/web/20130130005923/https://www.digitalpreservation.gov/partners/documents/ncgdap_final_report.pdf

PostgreSQL Global Development Group (2021) PostgreSQL: The World's Most Advanced Open Source Relational Database. Disponible à l'adresse suivante :

<https://web.archive.org/web/20210201221045/https://www.postgresql.org/>

QGIS.org Association (2020) Working with Project Files. Disponible à l'adresse suivante :

https://web.archive.org/web/20201129005713/https://docs.qgis.org/3.16/en/docs/user_manual/introduction/project_files.html

TerraGo (2021) TerraGo. Disponible à l'adresse suivante :

<https://web.archive.org/web/20201127085818/https://terragotech.com/>

Wikipedia (2021) Keyhole Markup Language. Disponible à l'adresse suivante :

https://web.archive.org/web/20210308184851/https://en.wikipedia.org/wiki/Keyhole_Markup_Language

Wikipedia (2020) Geographic Information System. Disponible à l'adresse suivante :

https://web.archive.org/web/20201205232727/https://en.wikipedia.org/wiki/Geographic_information_system