

Lignes directrices sur la diffusion de données ouvertes

Portail Données Québec



Lignes directrices sur la diffusion de données ouvertes

Portail Données Québec

Cette publication a été réalisée par le Sous-secrétariat du dirigeant principal de l'information en collaboration avec les partenaires du portail Données Québec et produite par la Direction des communications du Secrétariat du Conseil du trésor.

Vous pouvez obtenir de l'information au sujet du Conseil du trésor et de son Secrétariat en vous adressant à la Direction des communications ou en consultant son site Web.

Direction des communications
Secrétariat du Conseil du trésor
2^e étage, secteur 800
875, Grande Allée Est
Québec (Québec) G1R 5R8

Téléphone : 418 643-1529
Sans frais : 1 866 552-5158

communication@sct.gouv.qc.ca
www.tresor.gouv.qc.ca

Tous droits réservés pour tous les pays.
© Gouvernement du Québec - 2016

Table des matières

Liste des figures	V
Liste des tableaux	V
Liste des sigles et acronymes	VI
Historique des changements	VII
1. Dispositions générales	1
1.1 Objet des lignes directrices	1
1.2 Champ d'application	1
1.3 Mesures transitoires	1
1.4 Schéma synthétique des lignes directrices	1
2. Données tabulaires	2
2.1 Format CSV (fortement recommandé)	2
2.2 Formats XLSX et ODS (à utiliser sous certaines conditions)	5
3. Données hiérarchiques	5
3.1 Format JSON (fortement recommandé)	5
3.2 Format XML (à retenir sous certaines conditions)	7
3.3 Données 3D	8
4. Données en géomatique	9
4.1 Format GeoJSON (fortement recommandé)	9
4.2 Formats secondaires en géomatique (à utiliser sous certaines conditions)	11
4.3 Données d'images géoréférencées (raster)	12
5. Considérations pour les différents types de données	13
5.1 Valeurs inconnues ou non disponibles	13
5.2 Date et heure	13
5.3 Nombre	13
5.4 Booléen	13

6. Considérations générales relatives aux fichiers	14
6.1 Encodage pour les fichiers textes	14
6.2 Nom des fichiers	14
6.3 Format d'image	14
6.4 Compression	15
7. Métadonnées	15
7.1 Métadonnées associées au jeu de données	15
7.2 Métadonnées associées aux fichiers attachés	19
7.3 Métadonnée « Licence »	20
7.4 Métadonnée « Catégorie »	20
7.5 Définition des catégories	22
7.6 Métadonnée « Mots clés »	23
8. Création et mise à jour d'un jeu de données	23
8.1 Fréquence de mise à jour des données ouvertes	24
8.2 Gestion des données historiques	24
Références	25

Liste des figures

Figure 1 Schéma synthétique des lignes directrices	2
Figure 2 Exemple de visualisation d'un CSV sous forme de tableau (jeu de données sur le nombre de titulaires de permis de conduire)	2
Figure 3 Exemple de visualisation d'un CSV sous forme de graphique (jeu de données sur le nombre de titulaires de permis de conduire)	3
Figure 4 Exemple d'un fichier CSV	4
Figure 5 Exemple d'un fichier JSON	6
Figure 6 Balise indiquant l'encodage d'un fichier XML	7
Figure 7 Exemple d'un fichier XML	8
Figure 8 Extrait d'un fichier GeoJSON	9
Figure 9 Visualisation d'un fichier GeoJSON dans l'outil de visualisation de CKAN	10
Figure 10 Visualisation d'un fichier GeoJSON dans IGO-Navigateur	10
Figure 11 Extrait du fichier CSV	11
Figure 12 Visualisation d'un fichier CSV contenant l'information géomatique	12

Liste des tableaux

Tableau 1 Représentation d'un CSV sous forme de tableau	4
Tableau 2 Structure du GeoJSON	9
Tableau 3 Grille de corrélation des métadonnées associées au jeu de données	16
Tableau 4 Métadonnées associées aux fichiers	19
Tableau 5 Grille de corrélation des catégories de données	21

Liste des sigles et acronymes

API	<i>Application Programming Interface</i>
COFOG	<i>Classification of the Functions of Government</i>
CSV	<i>Comma-separated values</i>
CC-BY	<i>Creative Commons Attribution</i>
XML	<i>Extensible Markup Language</i>
GeoJSON	<i>Geographic JavaScript Object Notation</i>
GML	<i>Geography Markup Language</i>
JSON	<i>JavaScript Object Notation</i>
JPEG	<i>Joint Photographic Experts Group</i>
KML	<i>Keyhole Markup Language</i>
XLSX	<i>Office Open XML WorkBook</i>
ODS	<i>Open Document Spreadsheet</i>
ISO	<i>International Organization for Standardization</i>
PNG	<i>Portable Network Graphics</i>
RFC	<i>Requests for Comments</i>
RDF	<i>Resource Description Framework</i>
<i>UTF-8</i>	<i>Universal Character Set Transformation Format – 8 bits</i>
WFS	<i>Web Feature Service</i>
WMS	<i>Web Map Service</i>

Historique des changements

Version	Statut	Modifications
1	Avril 2016	Publication de la première édition

1. Dispositions générales

1.1 Objet des lignes directrices

Ce document vient appuyer les diffuseurs de jeux de données sur le portail Données Québec en présentant les normes techniques à respecter.

1.2 Champ d'application

Les organisations diffusant des données ouvertes sur le portail ont la responsabilité de les respecter et de documenter les dérogations. Les choix retenus ont été déterminés et approuvés par des partenaires du portail.

Il est essentiel de signaler que les lignes directrices identifiées dans le présent document s'appliquent d'abord et avant tout aux nouveaux jeux de données qui seront diffusés pour une première fois après la mise en ligne, en avril 2016, du portail Données Québec.

Au cours de l'année 2016, les diffuseurs de données ouvertes intégreront graduellement, au sein de leur processus de diffusion de données, les lignes directrices qui suivent, de manière à ce que les nouveaux jeux et les mises à jour diffusés après le lancement du portail respectent les présentes lignes directrices.

En ce qui concerne les données diffusées avant 2016, les diffuseurs sont invités à actualiser leurs anciens fichiers de données ouvertes. Considérant toutefois que plusieurs de ces fichiers ont été exportés par différents systèmes, il pourrait être difficile, voire même impossible, de régénérer ces fichiers.

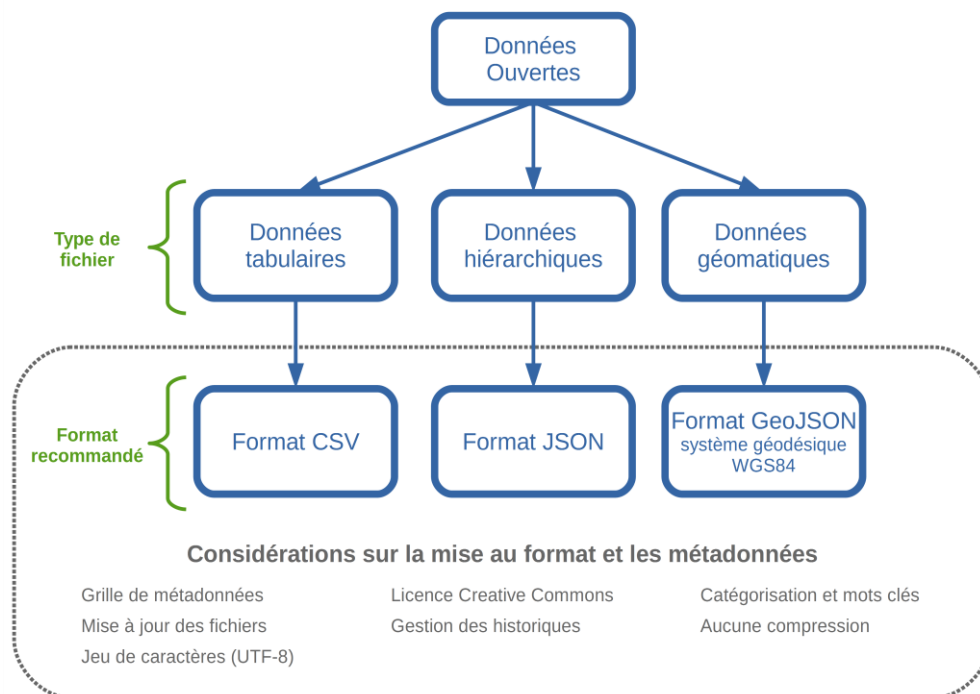
1.3 Mesures transitoires

Les diffuseurs de données ouvertes ont jusqu'au 31 décembre 2016 pour mettre à jour leurs jeux de données en respectant les présentes lignes directrices.

Au 1^{er} janvier 2017, les données qui ne respecteront pas les exigences minimales du présent document seront identifiées par un bandeau dans la fiche de métadonnées, mentionnant leur non-conformité.

1.4 Schéma synthétique des lignes directrices

Le schéma présenté ci-après identifie de manière structurée et synthétisée les lignes directrices à suivre lors de la diffusion de données.

Figure 1 - Schéma synthétique des lignes directrices

2. Données tabulaires

Les jeux de données tabulaires, c'est-à-dire pouvant être représentés sous forme d'un tableau simple comportant des lignes et des colonnes, peuvent être diffusés en données ouvertes sous différents formats de fichiers :

- ✓ le format CSV est fortement recommandé.

Les considérations propres au type des champs (nombre, date, etc.) sont décrites à la section 5 - Considérations pour les différents types de données, alors que les règles s'appliquant à plusieurs formats de fichiers sont décrites à la section 6 - Considérations générales relatives aux fichiers. Le texte qui suit est relatif aux formats proposés pour les données tabulaires.

2.1 Format CSV (fortement recommandé)

Le format *Comma-separated values* (CSV) désigne un format de fichier ouvert représentant des données tabulaires sous forme de valeurs délimitées par des séparateurs de champs. Ce format est celui qui est recommandé car le portail Données Québec offre des fonctionnalités additionnelles pour les jeux de données diffusés en CSV.

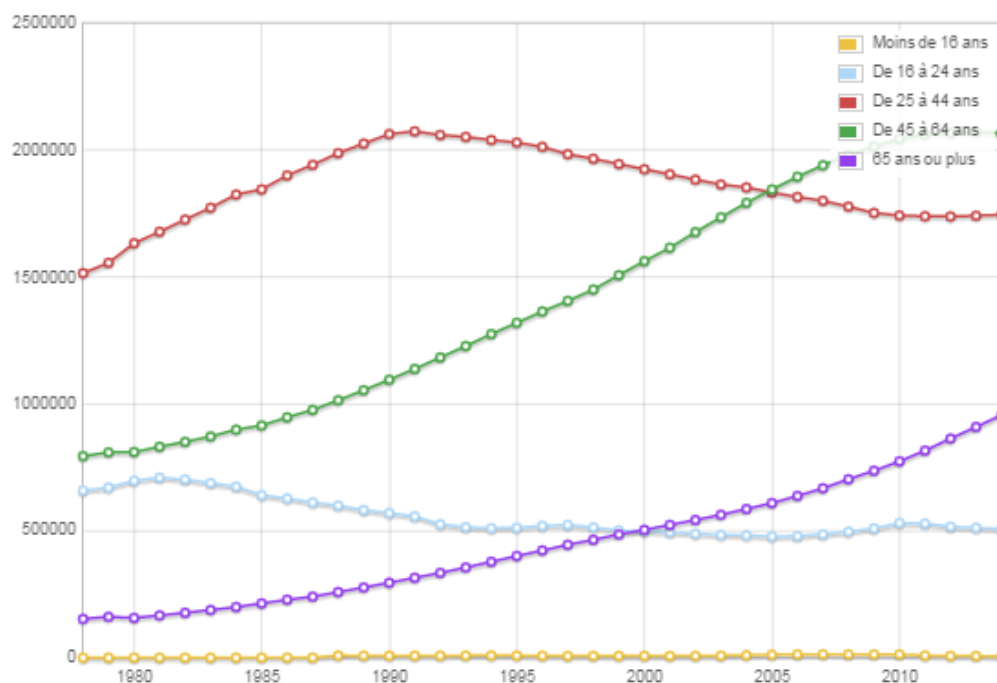
Par exemple, le portail affiche automatiquement le contenu d'un fichier CSV dans un tableau où l'utilisateur du portail peut trier et filtrer les données. Le jeu de données utilisé dans l'exemple ci-dessous est [Nombre de titulaires de permis de conduire de 1978 à 2014, selon l'âge](#), diffusé par la Société de l'assurance automobile du Québec (SAAQ).

Figure 2 - Exemple de visualisation d'un CSV sous forme de tableau (jeu de données sur le nombre de titulaires de permis de conduire)

ANNÉE	Moins de 16 ans	De 16 à 24 ans	De 25 à 44 ans	De 45 à 64 ans	65 ans ou plus
1978	0	658854	1515600	794333	152680
1979	0	670534	1556540	809176	161111
1980	0	697776	1633596	810622	156511
1981	0	709617	1678879	831274	166551
1982	0	702326	1726912	850665	176862
1983	0	688390	1773460	871825	187845
1984	0	674103	1825746	898921	199492
1985	0	640626	1845869	915252	213706
1986	0	627555	1901023	947312	228276
1987	0	611950	1942840	976581	240421
1988	8247	599596	1988929	1014744	258484
1989	7750	581923	2026586	1054125	276570
1990	7839	570001	2064038	1095720	295373
1991	7518	556945	2074856	1138281	314768
1992	7693	525773	2060361	1183299	334100
1993	8094	514097	2053071	1228313	355397
1994	8188	510375	2040697	1275485	377847
1995	8167	512123	2030651	1319812	400594

Pour les jeux de données diffusés en CSV, le portail offre également à l'utilisateur la possibilité de créer une représentation graphique du jeu de données.

Figure 3 - Exemple de visualisation d'un CSV sous forme de graphique (jeu de données sur le nombre de titulaires de permis de conduire)



Le portail offre aussi d'autres fonctionnalités plus avancées pour les fichiers CSV, par exemple l'intégration du jeu de données à une base de données permettant une visualisation plus performante pour l'utilisateur et l'extraction du jeu de données (partiellement ou en totalité) à partir d'un API.

Un jeu de données peut être diffusé en plusieurs formats sur le portail. Considérant les nombreux avantages associés aux CSV, il est fortement recommandé de diffuser les jeux de données tabulaires en format CSV (et optionnellement sous d'autres formats).

Structure du CSV

Le CSV n'a jamais fait l'objet d'une spécification formelle, offrant donc beaucoup de flexibilité au diffuseur. Dans le but de simplifier l'interprétation des CSV et de faciliter l'interopérabilité entre eux, les diffuseurs devront respecter les quelques règles suivantes :

- ✓ la première ligne du fichier désigne l'en-tête des colonnes;
- ✓ les lignes suivantes donnent les valeurs des données;
- ✓ les en-têtes ne comportent pas de caractères accentués ou d'espaces;
- ✓ les éléments sont séparés par des virgules « , »;
- ✓ les guillemets anglais « " " » sont utilisés pour délimiter les champs de texte.

Outil de validation

Il est possible de valider le contenu d'un CSV à l'aide de l'outil CSV Lint : <http://csvlint.io/>.

Exemple de données en CSV

Un jeu de données donnant l'état d'avancement de différents projets pourrait ressembler à ce qui suit en format CSV.

Figure 4 Exemple d'un fichier CSV

```
Id,Nom,Nom_Organisme,Date_Fin_Projetee,Statut,Variation_echeancier
11598,"Refonte ERT - État du réseau routier","Transports",2015-10-30,"Actif",-45.9
13942,"Téléphonie IP","Conseil exécutif",2015-06-30,"Terminé",3.4
13373,"Rehaussement Infrastructure Services","Société d'habitation du Québec",2015-12-31,"Terminé",-25.6
13964,"Refonte du site Internet ministériel","Culture et Communications",2017-03-31,"Suspendu", ,
```

Ce fichier pourrait être représenté dans le tableau suivant.

Tableau 1 - Représentation d'un CSV sous forme de tableau

Id	Nom	Nom_Organisme	Date_Fin_Projetee	Statut	Variation_echeancier
11598	Refonte ERT - État du réseau routier	Transports, Mobilité durable et Électrification des transports	2015-10-30	Actif	-45.9
13942	Téléphonie IP	Conseil exécutif	2015-06-30	Terminé	3.4
13373	Rehaussement Infrastructure Serveurs / Services	Société d'habitation du Québec	2015-12-31	Terminé	-25.6
13964	Refonte du site Internet ministériel	Culture et Communications	2017-03-31	Suspendu	

2.2 Formats XLSX et ODS (à utiliser sous certaines conditions)

Les formats (par exemple, XLSX, ODS) utilisés par les logiciels tableurs courants, tels que Microsoft Excel, ne doivent jamais être privilégiés comme format principal de jeu de données. Ces formats sont à éviter pour les raisons suivantes :

- ✓ dans la majorité des cas, les outils de visualisation du portail ne peuvent pas afficher le contenu de ces fichiers;
- ✓ il est impossible d'accéder au contenu de ces fichiers par l'intermédiaire de l'API du portail;
- ✓ les formats de logiciels tableurs permettent la mise en forme des données, allant à l'encontre de la volonté d'offrir des données brutes aux citoyens et aux entreprises.

Toutefois, lorsque les formats XLSX et ODS doivent être utilisés pour des raisons particulières, on doit prendre en considération les éléments suivants :

- ✓ le fichier ne doit contenir aucun formatage;
- ✓ la première ligne désigne l'en-tête des colonnes et les en-têtes ne comportent pas de caractères accentués ou d'espaces;
- ✓ l'utilisation d'onglets doit être évitée;
- ✓ l'utilisation de fonctions dynamiques doit être évitée;
- ✓ la fusion de cellules est à proscrire.

3. Données hiérarchiques

Les jeux de données hiérarchiques, c'est-à-dire ne pouvant pas être représentés sous forme d'un tableau simple, doivent être diffusés dans des formats permettant de reproduire les relations complexes entre les données. Différents formats de fichiers offrent cette possibilité, mais le format suivant est fortement recommandé :

- ✓ le format JSON.

Malgré le fait que ces formats sont beaucoup plus appropriés pour des données hiérarchiques, ils pourraient aussi être utilisés pour des données tabulaires.

Les considérations propres au type des champs (nombre, date, etc.) sont décrites à la section 5 - Considérations pour les différents types de données, alors que les règles s'appliquant à plusieurs formats de fichiers sont décrites à la section 6 - Considérations générales relatives aux fichiers. Le texte qui suit est relatif aux formats proposés pour les données hiérarchiques.

3.1 Format JSON (fortement recommandé)

Le format *JavaScript Object Notation* (JSON) désigne un format de données ouvertes permettant de représenter des données hiérarchiques. Dans le contexte du portail de données ouvertes, le format JSON est recommandé, puisqu'il est destiné à représenter de l'information brute et structurée, qu'il est simple à interpréter et qu'il est facile à intégrer pour la majorité des langages de programmation.

Outil de validation

Il est possible de valider le contenu d'un format JSON à l'adresse suivante : <http://www.webtoolkitonline.com/json-tester.html>.

Exemple de données en JSON

Un jeu de données répertoriant des activités de loisir pourrait ressembler à ce qui suit en format JSON.

Figure 5 - Exemple d'un fichier JSON

```
{
  "LOISIRS_LIBRES": {
    "LOISIR_LIBRE": [
      {
        "CODE_SESSION": "H2016",
        "DESCRIPTION": "Volleyball - Groupe communautaire",
        "LIEU": "Centre communautaire Lebourgneuf",
        "ARRONDISSEMENT": "Arrondissement des Rivières",
        "SEMAINE_DEB": "2016-01-10",
        "SEMAINE_FIN": "2016-04-02",
        "PLAGES_HORAIRE": {
          "PLAGE_HORAIRE": [
            {
              "JOUR_SEMAINE": "Mercredi",
              "HEURE_DEBUT": "18:15:00",
              "HEURE_FIN": "19:45:00"
            },
            {
              "JOUR_SEMAINE": "Vendredi",
              "HEURE_DEBUT": "19:15:00",
              "HEURE_FIN": "20:45:00"
            }
          ]
        }
      },
      {
        "CODE_SESSION": "H2016",
        "DESCRIPTION": "Musique - Guitare - Enseignement individuel",
        "LIEU": "Centre communautaire Elzéar-Turcotte",
        "ARRONDISSEMENT": "Arrondissement des Rivières",
        "SEMAINE_DEB": "2016-01-17",
        "SEMAINE_FIN": "2016-03-26",
        "PLAGES_HORAIRE": {
          "PLAGE_HORAIRE": [
            {
              "JOUR_SEMAINE": "Jeudi",
              "HEURE_DEBUT": "17:00:00",
              "HEURE_FIN": "20:00:00"
            }
          ]
        }
      }
    ]
  }
}
```


3.2 Format XML (à retenir sous certaines conditions)

Le format *Extensible Markup Language* (XML) est un langage informatique de balisage dont l'objectif est le stockage et l'échange de contenu complexe. Il peut donc être utilisé pour des données tabulaires, mais il est beaucoup plus approprié pour des données hiérarchiques. Par son extensibilité, le format XML est également très flexible.

Structure du format XML

Dans le cadre des données ouvertes, le XML doit respecter les conditions suivantes :

- ✓ le XML est brut et autodescriptif : celui-ci ne contient que les données pertinentes. Il est interdit d'inclure de l'information sur la mise en forme (par exemple, un enregistrement en format XML à partir de l'application Microsoft Excel ne respecte pas ces critères);
- ✓ le XML est valide : l'arborescence est cohérente et logique. Il est possible de vérifier si que le contenu du XML respecte le standard, à l'adresse suivante : http://www.w3schools.com/xml/xml_validator.asp;
- ✓ aucun caractère accentué n'est intégré dans le nom des balises;
- ✓ le XML est encodé en UTF-8 et l'encodage est indiqué à l'aide de la balise ci-dessous.

Figure 6 - Balise indiquant l'encodage d'un fichier XML

```
<?xml version="1.0" encoding="UTF-8"?>
```

Exemple d'un fichier XML

Un jeu de données répertoriant des activités de loisir pourrait ressembler à ce qui suit en format XML.

Figure 7 - Exemple d'un fichier XML

```
<ns1:LOISIRS_LIBRES xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:ns1="http://donnees.ville.abc.qc.ca">
  <LOISIR_LIBRE>
    <CODE_SESSION>H2016</CODE_SESSION>
    <DESCRIPTION>Volleyball - Groupe communautaire</DESCRIPTION>
    <LIEU>Centre communautaire Lebourgneuf</LIEU>
    <ARRONDISSEMENT>Arrondissement des Rivières</ARRONDISSEMENT>
    <SEMAINE_DEB>2016-01-10</SEMAINE_DEB>
    <SEMAINE_FIN>2016-04-02</SEMAINE_FIN>
    <PLAGES_HORAIREES>
      <PLAGE_HORAIRE>
        <JOUR_SEMAINE>Mercredi</JOUR_SEMAINE>
        <HEURE_DEBUT>18:15:00</HEURE_DEBUT>
        <HEURE_FIN>19:45:00</HEURE_FIN>
      </PLAGE_HORAIRE>
      <PLAGE_HORAIRE>
        <JOUR_SEMAINE>Vendredi</JOUR_SEMAINE>
        <HEURE_DEBUT>19:15:00</HEURE_DEBUT>
        <HEURE_FIN>20:45:00</HEURE_FIN>
      </PLAGE_HORAIRE>
    </PLAGES_HORAIREES>
  </LOISIR_LIBRE>
  <LOISIR_LIBRE>
    <CODE_SESSION>H2016</CODE_SESSION>
    <DESCRIPTION>Musique - Guitare - Enseignement individuel</DESCRIPTION>
    <LIEU>Centre communautaire Elzéar-Turcotte</LIEU>
    <ARRONDISSEMENT>Arrondissement des Rivières</ARRONDISSEMENT>
    <SEMAINE_DEB>2016-01-17</SEMAINE_DEB>
    <SEMAINE_FIN>2016-03-26</SEMAINE_FIN>
    <PLAGES_HORAIREES>
      <PLAGE_HORAIRE>
        <JOUR_SEMAINE>Jeudi</JOUR_SEMAINE>
        <HEURE_DEBUT>17:00:00</HEURE_DEBUT>
        <HEURE_FIN>20:00:00</HEURE_FIN>
      </PLAGE_HORAIRE>
    </PLAGES_HORAIREES>
  </LOISIR_LIBRE>
</ns1:LOISIRS_LIBRES>
```

3.3 Données 3D

Le format privilégié pour les données 3D est le format CityGML.

4. Données en géomatique

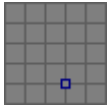
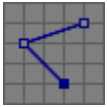

Cette section concerne les jeux de données ouvertes en géomatique. Le format principal à adopter est le format GeoJSON, dans le système géodésique WGS84. Le format GeoJSON est le format ouvert le plus répandu dans le cas de développement d'application Web en données ouvertes, puisqu'il est une adaptation géographique au JSON. Ce format est apprécié des développeurs, puisque celui-ci est supporté par la plupart des systèmes d'information géographique (SIG) sur le marché.

4.1 Format GeoJSON (fortement recommandé)

Le format *Geographic JavaScript Object Notation* (GeoJSON) est un format ouvert d'encodage d'ensemble de données géographiques. Dans le contexte du portail de données ouvertes, le format GeoJSON utilise par défaut le système géodésique WGS84 (EPSG:4326). En utilisant ce système géodésique, lorsque la taille du fichier est inférieure à 5 mégaoctets, il est possible de tirer profit des outils de visualisation du portail CKAN.

Tableau 2 - Structure du GeoJSON

Le format GeoJSON propose trois formes géométriques :

Point	Segment	Polygone
		

Exemple d'un fichier GeoJSON

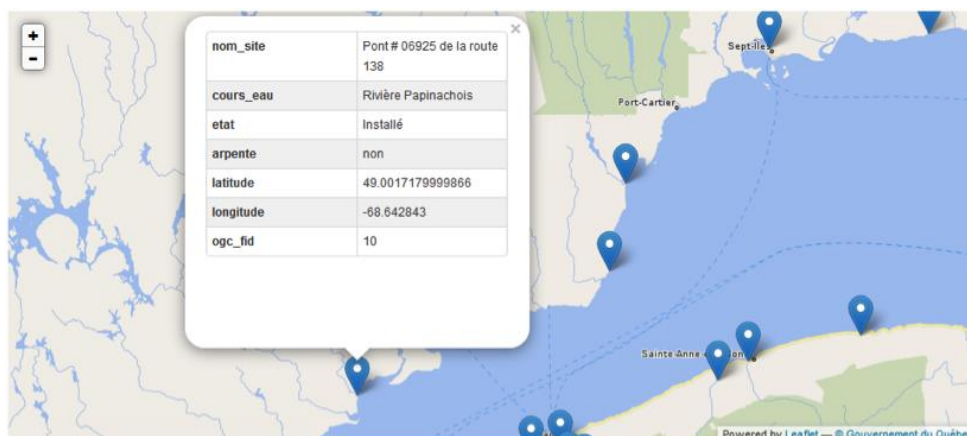
Le jeu de données « [échelle limnimétrique \(publiques\)](#) », publié par le ministère de la Sécurité publique, offre un exemple de fichier GeoJSON. Les deux figures présentées ci-après décrivent, respectivement, un extrait du contenu du fichier de format GeoJSON et son aperçu par l'entremise de l'outil de visualisation CKAN.

Figure 8 - Extrait d'un fichier GeoJSON

```

{
  type: "Feature",
  properties: {
    nom_site: "Pont # 06925 de la route 138",
    cours_eau: "Rivière Papinachois",
    etat: "Installé",
    arpenté: "non",
    latitude: 49.001717999986603,
    longitude: -68.642843,
    ogc_fid: 10
  },
  geometry: {
    type: "Point",
    coordinates: [
      -68.642843,
      49.001717999986646
    ]
  }
},

```

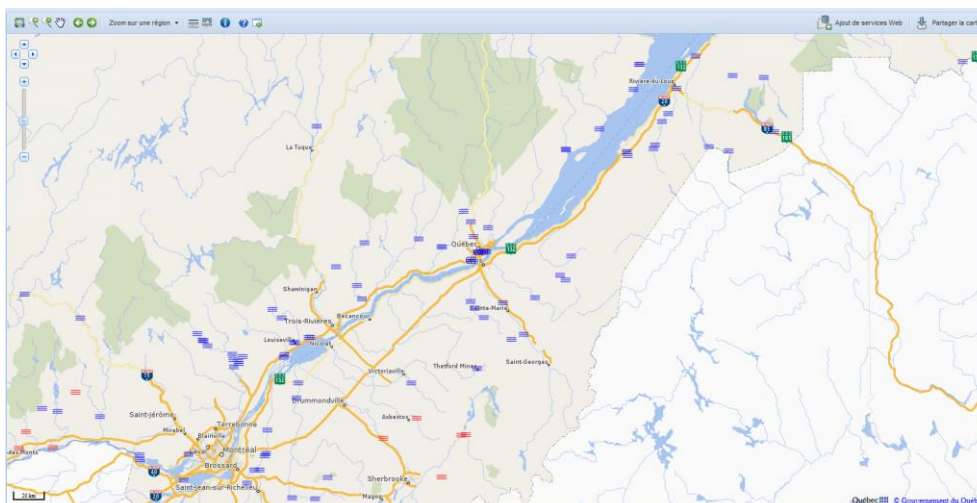
Figure 9 - Visualisation d'un fichier GeoJSON dans l'outil de visualisation de CKAN

Visualisation

Le système de coordonnées géographiques doit être WGS 84 (projection par défaut : EPSG:4326 : <http://spatialreference.org/ref/epsg/4326/>). L'utilisation de celui-ci est obligatoire pour visualiser un jeu de données sur le portail, car le portail ne peut supporter qu'une seule projection. Pour les jeux de données faisant appel à une autre projection que celle utilisée par défaut (4326), le code EPSG de la projection utilisée doit être documenté à la section « information complémentaire » du jeu de données, si la projection utilisée s'applique à toutes les ressources (fichiers) du jeu, ou dans la description de la ressource, lorsque la projection particulière s'applique pour une ressource donnée.

Visualisation enrichie (IGO)

Pour une visualisation plus interactive (symbologie, interroger la couche, connexion à des services Web, superposition avec d'autres jeux de données, etc.) des données géomatiques, l'équipe géomatique du Ministère de la Sécurité publique et ses partenaires, entre autres, le Ministère des Transports, de la Mobilité durable et de l'Électrification des transports, l'Institut de la Santé publique du Québec et le ministère de la Culture et des Communications, permettent, dans le cadre du projet « Infrastructure de géomatique ouverte » l'utilisation d'IGO-Navigateur (<http://igouverte.org/>), qui offre plus de possibilités de visualisation des différents formats de données ouvertes en géomatique. Pour en savoir plus sur l'intégration des jeux de données à IGO, vous pouvez contacter l'équipe d'IGO en utilisant leur formulaire « Contactez-nous ».

Figure 10 - Visualisation d'un fichier GeoJSON dans IGO-Navigateur

4.2 Formats secondaires en géomatique (à utiliser sous certaines conditions)

Lorsque la taille d'un fichier GeoJSON atteint plusieurs dizaines de mégaoctets, ce format n'est pas la meilleure solution pour la visualisation. Dans cette situation, d'autres formats doivent être envisagés, en complémentarité à un standard de services Web cartographiques (Web Map Service, en anglais, ou WMS) :

- ✓ Shapefile : compressé, en complémentarité au Web Map Service (WMS), afin de permettre la visualisation;
- ✓ SpatiaLite : format léger de base de données géomatiques, en complémentarité au Web Map Service (WMS), afin de permettre la visualisation;
- ✓ KML, GML, WFS : en complémentarité au Web Map Service (WMS), afin de permettre la visualisation;
- ✓ CSV :
 - Si l'information géospatiale est un point, le CSV doit contenir un champ « latitude » et un champ « longitude »;
 - Si l'information géospatiale est une ligne ou une région, le CSV doit contenir un champ « well known text ».

Pour les jeux de données faisant appel à une autre projection que celle utilisée par défaut (4326), le code EPSG de la projection utilisée doit être documenté à la section « Informations complémentaires » du jeu de données, si la projection utilisée s'applique à toutes les ressources (fichiers) du jeu, ou dans la description de la ressource, lorsque la projection particulière s'applique pour une ressource donnée.

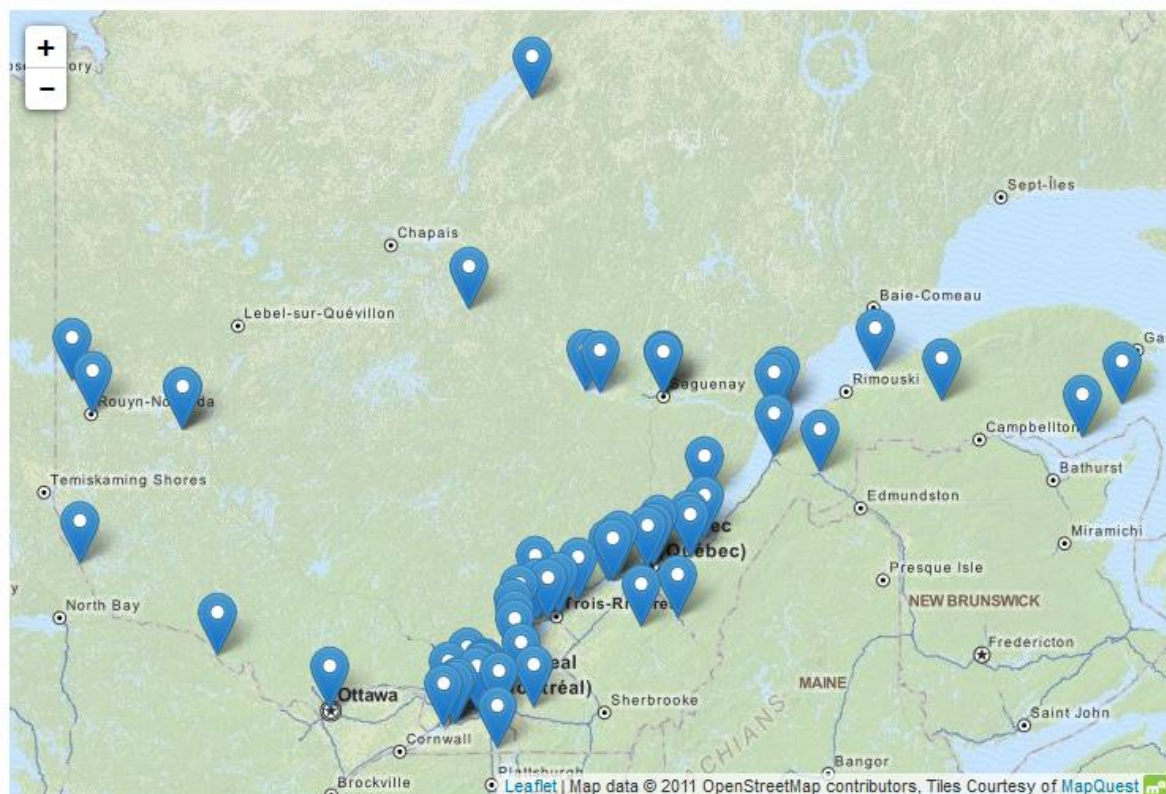
Exemple d'un format CSV contenant de l'information géomatique

À titre d'exemple, le jeu de données « [Sites patrimoniaux cités par les municipalités et les communautés autochtones](#) », diffusé par le ministère de la Culture et des Communications, permet une visualisation par géolocalisation, puisqu'il contient les champs « latitude » et « longitude ». Les deux figures présentées ci-après illustrent, respectivement, un extrait du contenu du fichier de format CSV et la visualisation d'un fichier CSV contenant l'information géomatique.

Figure 11 - Extrait du fichier CSV

```
nom_bien,bien_id,latitude,longitude
Site du patrimoine Flavie-Drapeau,115297,48.611306,-68.229222
Site du patrimoine du canton de Sainte-Edwidge-de-Clifton,157484,45.204075,-71.681717
Ferme Chalifoux-Lockman,165875,45.673778,-75.030306
Maison Marc-Aurèle-Plamondon,166036,46.037442,-71.914692
Maison Fleury,166037,46.037078,-71.915333
Boisé Stein,166038,46.037217,-71.913519
Hameau de Way's Mills,168923,45.099866,-71.975682
```

Figure 12 - Visualisation d'un fichier CSV contenant l'information géomatique



Outils de conversion

Pour la conversion, deux logiciels libres et gratuits sont proposés :

- ✓ Système d'information géographique (SIG) : le logiciel QGIS dispose de nombreuses fonctionnalités, dont un outil de conversion;
- ✓ Application Web : l'outil Web OGRE permet la conversion rapide de fichiers.

Des logiciels propriétaires tels que FME et ArcGIS font également la conversion.

4.3 Données d'images géoréférencées (raster)

Les données d'images géoréférencées sont des représentations numériques d'images du territoire. Pour ce type de données d'images, la projection conique conforme de Lambert (EPSG:32198) est recommandée. Toutefois, il est possible de diffuser les images dans une autre projection. Cette projection doit être documentée par son code EPSG, dans le fichier et dans la métadonnée du jeu de données, à la section « Informations complémentaires », si la projection particulière s'applique à toutes les ressources (fichiers), ou dans la description de la ressource, si la projection est propre à une ressource. Les formats à considérer sont :

- ✓ représentation principale pour des images géoréférencées : GeoTIFF;
- ✓ représentation secondaire : JPEG2000;
- ✓ représentation secondaire : Web Coverage Service WCMS;
- ✓ représentation en complémentarité au GeoTiff, au JPEG2000 ou au WCS : Web Map Service (WMS) pour la visualisation.

Veillez prendre note que les images géoréférencées de taille très volumineuse (plusieurs gigaoctets) ne pourront être hébergées sur le portail. Veuillez communiquer avec le pilote du portail pour discuter d'options alternatives possibles.

5. Considérations pour les différents types de données

Les jeux de données contiennent différents types d'information (texte, nombre, date, etc.). Selon le type d'information, certaines recommandations doivent être respectées pour assurer l'intégrité, simplifier l'interprétation et accroître l'interopérabilité.

5.1 Valeurs inconnues ou non disponibles

Peu importe la nature d'une donnée (date, heure, nombre, booléen, texte), lorsque la valeur n'est pas connue, quelle qu'en soit la raison, il est fortement recommandé de laisser le champ vide plutôt que de mettre une valeur par défaut qui introduirait des erreurs d'interprétation.

5.2 Date et heure

Lorsqu'une donnée représente une date ou une heure, elle doit être fournie selon la norme ISO8601. À cet égard, voici les spécifications techniques :

- ✓ format de la date : AAAA-MM-JJ;
- ✓ format de l'heure : HH:MM:SS;
- ✓ lorsque la date et l'heure sont dans le même champ, on utilise la valeur « T » pour les séparer [AAAA-MM-JJTHH:MM:SS].

5.3 Nombre

Pour ce qui est des nombres, les particularités suivantes sont à prendre en considération :

- ✓ les valeurs décimales sont séparées par un point « . »;
- ✓ les valeurs sont fournies selon les unités de base du système international (par exemple, m, kg);
- ✓ les valeurs monétaires sont en dollars (\$).

5.4 Booléen

En ce qui concerne les booléens, il est proposé d'utiliser des valeurs simples et évidentes à interpréter. Par exemple, les combinaisons suivantes peuvent être retenues :

- ✓ TRUE, FALSE;
- ✓ Vrai, Faux;
- ✓ Oui, Non.

6. Considérations générales relatives aux fichiers

Outre les règles identifiées précédemment et propres aux différents formats, les quelques règles générales suivantes sont à prendre en considération.

6.1 Encodage pour les fichiers textes

Pour tous les formats de fichiers sauvegardés sous forme de texte (y compris, par exemple, les formats CSV, JSON, XML et GEOJSON), l'encodage utilisé doit être « UTF-8 ». Cet encodage de caractères informatiques a été conçu par l'ISO pour respecter l'ensemble du « répertoire universel de caractères codés ». Dans le contexte québécois, UTF-8 permet de respecter les exigences du français intégral.

Encodage UTF-8 et suite bureautique de Microsoft

La suite bureautique de Microsoft Office ne supporte pas la sauvegarde en UTF-8. La façon la plus simple de remédier à cette situation consiste à utiliser le logiciel Bloc-notes de Windows. Voici la procédure à suivre :

1. Ouvrir le fichier avec Bloc-notes;
2. Dans le menu, sélectionner « Fichier », puis « Enregistrer sous »;
3. Dans le menu « encodage », sélectionner UTF-8;
4. Enregistrer le fichier.

6.2 Nom des fichiers

Des noms de fichier structurés et uniformes permettent aux utilisateurs de comprendre ce qu'ils ont téléchargé et limitent les risques d'erreur. Voici la structure de base recommandée pour nommer les fichiers :

préfixe-identifiant-suffixe.extension

(par exemple, gtdo-rapport-2013.csv)

Voici quelques éléments à prendre en considération :

- ✓ tout en minuscules;
- ✓ aucun caractère accentué (par exemple, à, è, î);
- ✓ aucun caractère spécial (par exemple, %, œ, ... , =, ¾);
- ✓ aucun espace;
- ✓ éviter les mots superflus (par exemple, les déterminants);
- ✓ privilégier le trait d'union « - » plutôt que la barre de soulignement « _ ».

6.3 Format d'image

Une image n'est pas considérée comme une donnée ouverte. Toutefois, certains jeux de données peuvent être accompagnés d'images (photos historiques, désastre naturel, etc.). Deux formats de fichiers d'images sont à envisager :

- ✓ représentation principale : PNG;
- ✓ représentation secondaire : JPG.

6.4 Compression

La compression de fichier doit être évitée. En effet, les outils de prévisualisation du portail ne peuvent gérer les formats compressés (.zip, .tar, etc.). Toutefois, voici quelques exceptions à prendre en considération :

- ✓ certains jeux sont accompagnés de plusieurs images. Dans ce cas précis, il est acceptable de compresser les fichiers au format ZIP;
- ✓ certains formats, tels que le Shapefile, sont standardisés sous un format compressé.

7. Métadonnées

Les métadonnées sont les données qui accompagnent les jeux de données afin de décrire leur contenu et leur contexte. Elles sont un complément essentiel aux jeux de données ouvertes, car elles permettent de structurer la définition du jeu de données. Les métadonnées utilisées sur le portail s'appuient sur les standards DCAT et Dublin Core et sont structurées en deux groupes :

- ✓ les métadonnées associées au jeu de données, de manière globale;
- ✓ les métadonnées associées à chacune des ressources (par exemple, les fichiers attachés au jeu de données). Les ressources contiennent généralement les données pour différentes périodes de temps ou pour différentes perspectives, mais les ressources peuvent aussi contenir des compléments d'information décrivant la signification des champs du jeu de données ou toute information pouvant servir à l'utilisateur des données.

Cette section aborde les différents éléments normatifs concernant les métadonnées utilisées sur le portail Données Québec.

7.1 Métadonnées associées au jeu de données

Le tableau suivant présente la corrélation entre les métadonnées québécoises et celles de Dublin Core pour les métadonnées associées globalement au jeu de données. Il décrit également chacune des métadonnées et mentionne si la métadonnée est requise (R) ou facultative (F). La mention A pour une métadonnée indique qu'elle est gérée automatiquement par le portail.

Tableau 3 - Grille de corrélation des métadonnées associées au jeu de données

DUBLIN CORE METADATA ELEMENT SET	MÉTADONNÉE DU PORTAIL	(R) (A) (F)	NOMENCLATURE	COMMENTAIRES
Identifieur	Identifiant unique	A R	Identifiant universel unique ou UUID	Généré automatiquement par CKAN
Title	Titre ¹	R	Texte libre	Afin d'assurer la pérennité de l'information, il est suggéré de ne pas y apporter de modification; on suggère également de ne pas modifier l'URL donnant accès au jeu de données.
Description	Description ¹	R	Texte libre	Permet de décrire le jeu de données de manière générale et de préciser ce qui est commun à la majorité des fichiers associés au jeu de données.
	Information complémentaire ¹	F	Texte libre	Caractéristiques importantes pour soutenir l'utilisation et l'interprétation des données, y compris, notamment, les limites, la fiabilité, les exclusions ou inclusions, l'exhaustivité, la granularité, la précision, le système de coordonnées et tout complément d'information sur une autre métadonnée. Il est important de préciser si les caractéristiques s'appliquent à l'ensemble du jeu de données ou si elles sont propres à un fichier donné.
Subject	Catégorie ¹	R	Menu déroulant Choix multiples possibles parmi les douze catégories prédéfinies (voir les points 7.4 et 7.5 pour le détail)	Le champ « subject » fait aussi référence au thésaurus canadien. Au Québec, le thésaurus québécois sera utilisé pour déterminer la catégorie et les mots clés.

¹ Prenez note que la conception du portail permet de supporter une langue autre que le français. Ainsi, dans une phase de développement ultérieure, nous prévoyons qu'il sera possible d'inscrire un équivalent en anglais pour le titre, la description et les mots clés. Les catégories seront traduites automatiquement.

DUBLIN CORE METADATA ELEMENT SET	MÉTADONNÉE DU PORTAIL	(R) (A) (F)	NOMENCLATURE	COMMENTAIRES
	Mots clés ¹	R	Texte libre avec proposition de saisie (voir le point 7.6)	
Coverage	Portée géographique	R	Portée prédéterminée en fonction du territoire des municipalités et du Québec	Une portée géographique générale est définie par défaut pour les villes participantes et le gouvernement du Québec. Il sera éventuellement possible d'intégrer une portée personnalisée.
	Portée temporelle	F	AAAA-MM-JJ / AAAA-MM-JJ	Période de temps traitée dans le contenu du jeu de données
Date	Diffusion initiale	A	AAAA-MM-JJ HH:mm UTC	La date de diffusion initiale représente la date à laquelle le jeu de données a été diffusé la première fois. Cette première diffusion du jeu de données pourrait être sur le portail donneesquebec.ca ou sur un autre portail. La date est en temps universel coordonné (UTC).
	Mise à jour	A	AAAA-MM-JJ HH:mm UTC	La date de la dernière mise à jour indique la date à laquelle une ressource (fichier de données ou ressource additionnelle) associée au jeu de données a été modifiée ou la date à laquelle la fiche descriptive du jeu de données a été modifiée. La date est en temps universel coordonné (UTC).
	Fréquence de mise à jour	R	Menu déroulant : irrégulier, temps réel, quotidien, hebdomadaire, mensuel, trimestriel, semi-annuel, annuel, bisannuel, triennal	La fréquence de mise à jour informe sur la fréquence à laquelle le jeu de données est actualisé. Généralement, cette fréquence est moins élevée que la fréquence à laquelle les données changent réellement.

DUBLIN CORE METADATA ELEMENT SET	MÉTADONNÉE DU PORTAIL	(R) (A) (F)	NOMENCLATURE	COMMENTAIRES
Rights	Licence	R	Menu déroulant CC0, CC-BY, CC-BY-SA, CC-ND, CC-BY-NC, CC-BY-SA-NC, CC-BY-NC-ND	Voir le point 7.3 pour le détail.
Creator	Organisation participante ²	R	Menu défilant à choix restreint, en fonction de l'organisation principale ²	Ce champ identifie l'organisme diffusant le jeu de données. Le nombre de variantes disponibles pour une organisation principale dépend des préférences de l'organisation principale.
	Responsable	F	Texte libre permettant de préciser l'unité responsable de la création et de la mise à jour	Le niveau de précision peut varier selon l'organisation. À titre d'exemple, au gouvernement du Québec, il est recommandé de limiter la précision au niveau de la direction générale.
Language	Langue	R	Menu déroulant Français, Français et anglais	Par défaut, ce champ sera en français car le portail ne supporte pas l'anglais pour le moment. N. B. : le Canada indique ce champ pour chacune des ressources, mais dans le cas de Données Québec, c'est la langue des métadonnées du jeu de données.
Source	Source URL	F	Possibilité de modification	Lien de l'organisme

Légende : (R) requis, (A) automatique, (F) facultative

² Le concept d'organisation principale existe dans le portail Données Québec pour regrouper les multiples diffuseurs de données ouvertes (organisations participantes) de manière plus globale (par exemple, le gouvernement du Québec, la Ville de Québec, la Ville de Sherbrooke, etc.). Par contre, l'organisation principale à laquelle une organisation participante est rattachée n'est pas disponible par l'entremise des métadonnées, mais générée automatiquement par le portail. En plus d'être observable visuellement sur le portail, l'organisation principale à laquelle une organisation participante est rattachée peut être connue par l'API du portail.

7.2 Métadonnées associées aux fichiers attachés

Le tableau suivant dresse la liste des métadonnées du portail associées à chacun des fichiers attachés au jeu de données. Il décrit également chacune des métadonnées et mentionne si la métadonnée est requise (R) ou facultative (F). La mention A pour une métadonnée indique qu'elle est gérée automatiquement par le portail.

Tableau 4 - Métadonnées associées aux fichiers

MÉTADONNÉE DU PORTAIL	R/F/A	NOMENCLATURE
Identifiant unique du fichier	A	
Titre	R	Texte libre
Description	R	Texte libre permettant de décrire ce qui est propre au fichier
Taille	F	Saisie manuelle : Ce champ est particulièrement important si le jeu occupe plusieurs centaines de mégaoctets
Format	R	Saisie manuelle : (XML, CSV, JSON, etc.)
Lien URL	A	
Diffusion initiale	A	Date de la première diffusion de fichier ou du lien vers le fichier sur le portail donneesquebec.ca. La date est en temps universel coordonné (UTC) au format suivant : AAAA-MM-JJ HH:mm UTC.
Respect du standard	R	Booléen : Si le jeu de données ne satisfait pas à une exigence précise des lignes directrices, l'organisation peut préciser celle-ci.
Mise à jour (fichier ou lien)	A	La date de la dernière mise à jour indique la date à laquelle la ressource (fichier de données ou lien vers ce fichier dans le cas d'un fichier hébergé à l'externe) a été modifiée. Pour les fichiers qui ne sont pas hébergés sur donneesquebec.ca, la date de la dernière mise à jour (du lien vers le fichier) pourrait être différée de la date de mise à jour réelle du fichier référé. La date est en temps universel coordonné (UTC), au format suivant : AAAA-MM-JJ HH:mm UTC.

Légende : (R) requis, (A) automatique, (F) facultative

7.3 Métadonnée « Licence »

La métadonnée « licence » fait partie des différentes métadonnées du jeu de données. La « Creative Commons 4.0 » est la licence adoptée pour les jeux de données du portail. Plus précisément, la variante **CC-BY** est assignée par défaut aux données du portail. Toutefois, en fonction du contexte des exigences présentes au sein de votre organisme, une variante différente peut être retenue. Pour identifier quelle variante de la licence CC peut convenir à un jeu de données, Creative Commons a créé un [outil d'aide à la décision](#).

Considérations particulières

- ✓ La **licence est irrévocable**. Ainsi, une fois accordée, elle ne peut être révoquée. Vous devez donc vous assurer que votre organisme est le titulaire des droits d'auteur des données qu'il diffuse. Bien sûr, vous pouvez en tout temps cesser de diffuser un jeu de données. Toutefois, la licence continuera de s'appliquer aux données qui auront été antérieurement diffusées.
- ✓ L'organisme détient les **droits d'auteurs** (patrimoniaux et moraux) relatifs aux données qu'il diffuse, ou il les a obtenus sous une licence compatible à la CC4.
- ✓ L'organisme adhère au **respect de la vie privée**. Il est donc important de :
 - ne pas diffuser de données contenant des renseignements personnels non publics ou qui seraient contraires aux lois en la matière;
 - prendre soin que l'identification d'une personne ne puisse pas se faire par croisement de données.

7.4 Métadonnée « Catégorie »

Les jeux de données du portail sont classés selon différentes catégories, précisées par l'entremise de métadonnées du jeu de données. Les catégories du portail de données ouvertes s'appuient sur la Classification des fonctions des administrations publiques (COFOG) et sur le thésaurus de l'activité gouvernementale. Le fait d'assurer une corrélation avec ces deux nomenclatures favorise une compréhension mutuelle, tant sur le plan national que sur le plan international, des termes utilisés lors de la classification de l'information. Le thésaurus, disponible à l'adresse <http://www.thesaurus.gouv.qc.ca>, peut être utilisé pour identifier la bonne catégorie pour un jeu de données.

Le tableau suivant présente la corrélation entre le COFOG, le thésaurus et les catégories offertes pour classer les jeux de données du portail.

Tableau 5 - Grille de corrélation des catégories de données

COFOG	THÉSAURUS QUÉBÉCOIS	CATÉGORIE CORRESPONDANTE SUR LE PORTAL
Protection de l'environnement	Ressources naturelles, agriculture, environnement	Environnement, ressources naturelles et énergie
Affaires économiques	Ressources naturelles, agriculture, environnement	Agriculture et alimentation
	Économie, finances et industrie	Économie et entreprises
		Transport
		Infrastructure
Ordre et sécurité publique	Loi, justice et droit	Loi, justice et sécurité publique
Services généraux des administrations publiques	Gouvernance, politique et administration publique	Gouvernement et finances
Enseignement	Éducation, emploi et sciences	Éducation et recherche
Loisir, culture et culte	Information, culture, communications	Société et culture
	Tourisme et loisirs	Tourisme, sports et loisirs
Santé	Santé	Santé
Protection sociale	Soutien aux personnes, familles et communauté	Politiques sociales

7.5 Définition des catégories

En s'inspirant du thésaurus de l'activité gouvernementale du gouvernement du Québec, les catégories permettant de classer les jeux de données du portail ont été définies de la façon suivante.

Environnement, ressources naturelles et énergie

Domaine d'affaires correspondant à la mise en valeur et l'utilisation optimale du territoire et des ressources énergétiques, forestières et minérales, et ce, dans une perspective de développement durable. Cela inclut les forêts, les gisements de pétrole, de gaz naturel ou de minerai, les ressources hydro-électriques et autres biens de même nature qui ont une valeur économique certaine. Ce domaine concerne aussi la protection des écosystèmes et de la biodiversité, la prévention, la réduction et la suppression de la pollution dans l'environnement ainsi que toute activité climatique pouvant avoir des impacts sur la vie des citoyens.

Agriculture et alimentation

Domaine d'affaires correspondant à la mise en valeur, l'utilisation optimale de tout ce qui touche l'agriculture, les pêcheries et l'alimentation : production, transformation, distribution et commercialisation des produits agricoles et des pêches.

Économie et entreprises

Domaine d'affaires correspondant au développement de la compétitivité des entreprises et des marchés dans le secteur industriel, de la technologie ou toute autre entreprise à but lucratif.

Transport

Ensemble des moyens publics et privés permettant de déplacer, à l'aide d'un équipement de transport, des personnes ou des marchandises sur une distance relativement importante.

Infrastructure

Ensemble des équipements collectifs de base nécessaires à la vie économique de la nation : routes, ponts, voies ferrées, canaux, ports, réseaux de télécommunication, d'énergie, gestion de l'eau.

Loi, justice et sécurité publique

Domaine d'affaires qui correspond à tout ce qui a trait à la législation et son application, tout en reconnaissant les droits des citoyens et leur accès au système de justice. En ce sens, cette catégorie inclut le domaine législatif : préparation des lois et règlements (civil et pénal), administration de la justice, protection des personnes, prévention de la criminalité, encadrement des activités policières et administration des décisions des tribunaux. Celle-ci concerne aussi l'ensemble des mesures instaurées par l'État dans le but de garantir l'ordre public et la sécurité des citoyens.

Gouvernement et finances

Domaine d'affaires correspondant à la manière dont le pouvoir est exercé par le gouvernement sur les citoyens et l'appareil d'État. Cette catégorie inclut la gestion des services publics et de son administration, le contrôle du processus législatif et démocratique ainsi que le maintien et le développement des relations entretenues par le gouvernement, que ce soit du point de vue municipal ou international.

Éducation et recherche

Domaine d'affaires correspondant au développement et au maintien du système éducatif ainsi que l'accès à celui-ci. Cette catégorie inclut également tout ce qui a trait à l'innovation et à la recherche industrielle et scientifique.

Société et culture

Domaine d'affaires correspondant à la mission du gouvernement en matière de culture, notamment au niveau de son développement et de sa diffusion. Le terme « culture et société » réfère à un ensemble des usages, des coutumes, des structures sociales et des manifestations artistiques, culturelles, religieuses et intellectuelles qui définissent un groupe ou une société par rapport à une autre.

Tourisme, sports et loisirs

Domaine d'affaires correspondant à l'action gouvernementale en matière de loisir, de sport et de tourisme. Ce domaine se traduit par le développement de l'offre touristique, de la mise en marché du Québec et de ses acquis touristiques et toute l'infrastructure visant l'accueil des visiteurs.

Santé

Domaine d'affaires correspondant à la régulation du système de santé, c'est-à-dire la définition des règles de fonctionnement et de leur application ainsi que toutes les spécificités reliées aux domaines médical et social.

Politiques sociales

Domaine d'affaires correspondant à la mission sociale du gouvernement. Il touche entre autres les relations avec les citoyens (jeunes, famille, aînés, femmes), l'immigration, le développement communautaire et l'intégration des individus. Ce domaine inclut également le soutien socio-économique par le biais des divers programmes gouvernementaux, des personnes en difficulté ou caractérisées comme vulnérables.

7.6 Métadonnée « Mots clés »

Le thésaurus contient des milliers de termes liés hiérarchiquement aux grands domaines de l'activité gouvernementale. Il permet donc d'identifier, non seulement la bonne catégorie pour un jeu de données, mais aussi les mots clés recommandés. Le thésaurus est disponible à l'adresse suivante : <http://www.thesaurus.gouv.qc.ca>.

Exemple de classification

Par exemple, pour classer le jeu fictif « liste des parcs nationaux du Québec », une recherche de « Parc national » dans le thésaurus propose la suite hiérarchique suivante : « parc → attrait touristique → tourisme → tourisme et loisir ». La catégorie équivalente sur le portail est donc « tourisme, sport et loisir ». Les mots clés à considérer sont les suivants : parc, attrait touristique, ainsi qu'aire protégée proposée comme terme associé par le thésaurus.

8. Création et mise à jour d'un jeu de données

La création ainsi que la mise à jour des jeux de données et des métadonnées associées à ceux-ci peuvent se faire manuellement, en utilisant le formulaire offert sur le portail de données ouvertes, ou de manière automatisée, en utilisant l'API.

L'API permet d'actualiser un fichier de données. Lorsqu'il est utilisé en combinaison avec la base de données interne du portail, l'API peut servir à remplacer par une information plus actuelle des valeurs

dans un jeu de données ou à insérer une mise à jour de la valeur tout en conservant les anciennes valeurs. Cette fonctionnalité est seulement disponible avec les fichiers de format CSV.

L'utilisation de l'API et de la base de données du portail est documentée, en anglais, aux adresses suivantes :

- ✓ API : <http://docs.ckan.org/en/ckan-2.3/api/index.html>
- ✓ Base de données : <http://docs.ckan.org/en/latest/maintaining/datastore.html>

8.1 Fréquence de mise à jour des données ouvertes

En fonction de la fréquence à laquelle la valeur d'une donnée change, il est proposé de diffuser une mise à jour en données ouvertes selon les options suivantes :

- ✓ **Données mesurées de manière horaire ou quotidienne** : il est suggéré d'effectuer un instantané (*snapshot*) à une fréquence quotidienne et de rendre cette valeur disponible en données ouvertes;
- ✓ **Données mesurées de manière hebdomadaire ou mensuelle** : il est suggéré d'effectuer un instantané à une fréquence mensuelle et de rendre cette valeur disponible en données ouvertes;
- ✓ **Données mesurées de manière mensuelle, annuelle ou occasionnelle** : il est suggéré d'effectuer un instantané à une fréquence trimestrielle ou annuelle et de rendre cette valeur disponible en données ouvertes.

Au lieu d'un instantané, les valeurs publiées en données ouvertes pourraient également être une somme ou une moyenne, dépendamment de l'information diffusée. Le diffuseur du jeu de données doit donc inscrire dans les métadonnées toute information pertinente à propos de la méthodologie utilisée pour colliger et diffuser les données.

8.2 Gestion des données historiques

Peu importe à quelle fréquence la valeur d'une donnée est mesurée et peu importe la fréquence à laquelle elle est rendue disponible en données ouvertes, il est souvent intéressant de conserver un historique des valeurs pour une certaine période de temps. Différentes manières existent pour conserver un historique des valeurs antérieures, notamment :

- ✓ en diffusant un fichier unique cumulant toutes les valeurs des données;
- ✓ en diffusant un nouveau fichier pour chaque mise à jour. Le fichier peut comporter uniquement les nouvelles valeurs ou contenir une certaine période (par exemple, les cinq dernières années);
- ✓ en actualisant la base de données interne du portail pour insérer les dernières valeurs. La base de données peut être utilisée pour stocker toutes les valeurs diffusées dans le passé ou une procédure peut effectuer les suppressions nécessaires pour conserver uniquement une certaine période. L'utilisation de la base de données a l'avantage pour les utilisateurs des données de pouvoir recourir à l'API pour extraire les valeurs pour la période qu'ils désirent. Cette fonctionnalité est seulement disponible avec les fichiers de format CSV.

La manière retenue pour conserver l'historique dépend généralement des besoins des utilisateurs pour ces données et de la stabilité dans le temps des données antérieures (certaines données historiques peuvent nécessiter d'être revues pour faciliter leur comparaison avec la dernière mise à jour). Peu importe la manière choisie par le diffuseur, ce dernier doit bien décrire comment il gère l'historique des données par l'entremise des métadonnées.

Références

CREATIVE COMMONS. CC-BY 4.0, [en ligne]. <https://creativecommons.org/licenses/by/4.0/> (2013)

DUBLIN CORE. *Dublin Core Metadata Element Set, Version 1.1*, [en ligne].

<http://dublincore.org/documents/dces/> (2012)

<https://creativecommons.org/licenses/by/4.0/deed.fr>

GEOJSON. *The GeoJSON Format Specification*, [en ligne]. <http://geojson.org/geojson-spec.html> (2008)

GOUVERNEMENT DU QUÉBEC. *Thésaurus de l'activité gouvernementale*, [en ligne].

<http://www.thesaurus.gouv.qc.ca>

Gouvernement du Québec, *Profils de métadonnées gouvernementaux* [En ligne]

http://www.banq.qc.ca/archives/archivistique_gestion/ressources/publications/gestion_integree/profils_metadonnees/ (2009)

MAPBOX. *GeoJSON.io*, [en ligne]. <http://geojson.io/#map=2/20.0/0.0>

OPEN DATA INSTITUTE. *Check your CSV files with CSVLint*, [en ligne]. <http://csvlint.io/> (2013)

ORGANISATION DES NATIONS UNIES (ONU). *Classification des fonctions des administrations publiques (COFOG)*, [en ligne]. <http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=4&Lg=2&Top=1> (1999)

ORGANISATION INTERNATIONALE DE NORMALISATION. *ISO 8601 Éléments de données et formats d'échange -- Échange d'information -- Représentation de la date et de l'heure*, [en ligne].

<http://www.iso.org/iso/fr/home/standards/iso8601.htm> (2004)

W3C. *Data Catalog Vocabulary (DCAT)*, [en ligne]. <http://www.w3.org/TR/vocab-dcat/> (2014)

W3C. *Extensible Markup Language (XML) 1.0*, cinquième édition, [en ligne].

<http://www.w3.org/TR/2008/REC-xml-20081126/> (2008)

W3C. *Model for Tabular Data and Metadata on the Web*, [en ligne].

<http://www.w3.org/TR/2015/PR-tabular-data-model-20151117/#empty-and-quoted-cells> (novembre 2015)

W3C SCHOOLS. *XML validator*, [en ligne]. http://www.w3schools.com/xml/xml_validator.asp

WEBTOOLKITONLINE. *json tester*, [en ligne]. <http://www.webtoolkitonline.com/json-tester.html> (2014)

