

## BiJoe - Development #13667

### Ajouter une partie Open Data à BiJoe

19 octobre 2016 10:11 - Brice Mallet

<b>Statut:</b>	Nouveau	<b>Début:</b>	19 octobre 2016
<b>Priorité:</b>	Normal	<b>Echéance:</b>	
<b>Assigné à:</b>		<b>% réalisé:</b>	0%
<b>Catégorie:</b>		<b>Temps estimé:</b>	0:00 heure
<b>Version cible:</b>		<b>Planning:</b>	Non
<b>Patch proposed:</b>	Non		
<b>Description</b>			
Pour :			
- exporter automatiquement des jeux de données			
- proposer une API ?			
Intérêts :			
- devancer futures demandes de nos clients ?			
- <del>se faire mousser auprès du cabinet d'Axelle</del> [trop tard ;-]			
"Le rapport propose de travailler avec les principaux éditeurs de logiciels et progiciels accompagnant les collectivités locales, pour intégrer dans leurs produits une fonctionnalité d'export pour l'open data. Ainsi, le logiciel de gestion de cantines pourrait par exemple exporter de manière automatisée les informations d'horaires, de menus, de prix."			
<a href="http://www.localtis.info/cs/ContentServer?pagename=Localtis/LOActu/ArticleActualite&amp;jid=1250271626823&amp;cid=1250271620708&amp;n=1">http://www.localtis.info/cs/ContentServer?pagename=Localtis/LOActu/ArticleActualite&amp;jid=1250271626823&amp;cid=1250271620708&amp;n=1</a>			
<b>Demandes liées:</b>			
Lié à Publik - Development #19760: Open Data		<b>Rejeté</b>	<b>29 octobre 2017</b>

### Historique

#### #1 - 19 octobre 2016 10:13 - Brice Mallet

- Description mis à jour

#### #2 - 19 octobre 2016 10:23 - Benjamin Dauvergne

Déjà est-ce qu'il existe un format de catalogue de donnée opendata qu'on pourrait publier pour reprise dans n'importe quel portail open-data ? Ça nous éviterait d'avoir à faire des interfaces publics. On pourrait ainsi se concentrer sur les aspects backoffice de curation: choisir les données, sous quel nom et avec quels colonnes (on n'est pas sûr que les données dans nos exports soient toujours "anonymisé" ce serait mieux d'avoir un dernier garde-fou au niveau de bijoe que d'exporter bêtement tout).

#### #3 - 19 octobre 2016 10:51 - Brice Mallet

Pas très détaillée mais page de référence à mon avis : <https://www.data.gouv.fr/fr/faq/producer/>

#### #4 - 19 octobre 2016 11:01 - Benjamin Dauvergne

C'est plus clair ici: on a le choix soit de se faire moissonner en protocole CJAN ou OpenDataSoft (à voir si ce dernier est documenté) soit de pousser via l'API de data.gouv.fr.

#### #5 - 19 octobre 2016 11:33 - Benjamin Dauvergne

Je n'ai pas donné l'URL ou soit disant ce serait plus clair: <https://www.data.gouv.fr/fr/faq/developer/#api>

#### #6 - 01 avril 2020 10:46 - Brice Mallet

Je n'ai pas donné l'URL ou soit disant ce serait plus clair: <https://www.data.gouv.fr/fr/faq/developer/#api>

Cette page (enfin je suppose que c'était celle-ci) est maintenant accessible ici :

<https://doc.data.gouv.fr/jeux-de-donnees/demander-a-datagouvfr-de-moissonner-votre-site/>

#### #7 - 01 avril 2020 10:48 - Brice Mallet

- Lié à Development #19760: Open Data ajouté

## #8 - 01 avril 2020 10:50 - Brice Mallet

- Sujet changé de Ajouter une partie public/open-data à BiJoe à Ajouter une partie Open Data à BiJoe

## #10 - 01 avril 2020 11:03 - Brice Mallet

- Description mis à jour

- Echéance 15 décembre 2016 supprimé

- Assigné à Benjamin Dauvergne supprimé

## #11 - 01 avril 2020 11:15 - Brice Mallet

Déjà est-ce qu'il existe un format de catalogue de donnée opendata qu'on pourrait publier pour reprise dans n'importe quel portail open-data ?

Si existait, pourrait être ici : <https://scdl.opendatafrance.net/docs/> (mais existe pê qqchse à l'échelon européen ?)

Ce serait donc à nous de créer notre "spécification du modèle de données relatif aux demandes en lignes", puis la proposer pour standardisation à <http://www.opendatafrance.net/>

## #12 - 06 avril 2020 07:59 - Nicolas Roche

J'ai regardé du côté de <https://doc.data.gouv.fr/>.

Il y a à présent **3 formats** : DCAT, CKAN et OpenDataSoft. Le code du moissonneur CKAN utilisé par doc.data.gouv.fr est publié [ici](#) alors que celui de DCAT est intégré au projet udata.

cf la [doc technique du logiciel udata](#) (utilisé pour doc.data.gouv.fr) : DCAT is now the recommended way to harvest remote portals and repositories (and so to expose opendata metadata for any portal and repository).

DCAT est décrit par le W3C :

- Il s'agit d'un [modèle de métadonnées](#) (la [version européenne](#) fait plus peur à voir).
- [Plusieurs langages](#) peuvent être utilisés pour la description du catalogue. udata accepte les suivants :
  - RDF XML
  - JSON-LD
  - Turtle
  - N3
  - NT
  - Trig

cf google, CKAN is the most mature opensource solution

CKAN propose un [plugin pour importer/exporter suivant le format DCAT](#) qui fourni de la doc avec des exemple dont un en JSON-LD : <https://opendata.swiss/en/dataset/verbreitung-der-steinbockkolonien.jsonld> qui au final serait plus pratique que de l'XML.

En python, on peut jouer avec le format RDF XML via la [rdflib](#). Elle permet de construire des [jeux de données](#) et fourni (entre autres) un [serializer RDF-XML](#)

Sur <https://www.data.gouv.fr/fr/admin> où l'on peut se créer un compte, on ne peut pas jouer directement avec les catalogues, mais on peut créer un moissonneur (via le '+') qui devra ensuite être validé par un humain. Les moissonneurs sont composés par 2 attributs :

- Implémentation : parmi DCAT, ODS, MAAF, CKAN et DKAN
- URL : source base url

En regardant le [catalogue de Tours](#) on peut exporter le catalogue en [DRE](#). Il utilise <https://www.w3.org/ns/dcat.rdf> mentionné dans la [doc de doc.data.gouv.fr](#). Étrangement, la doc ne donne aucune indication sur la structure des données : [à priori](#) il faut juste donner un type mime.

- Dataset noeud de type dcat:Dataset en RDF : une collection de données, publiée ou élaborée par une seule source, et disponible pour accès ou téléchargement dans un ou plusieurs formats.
- Resource noeud de type dcat:Distribution en RDF : représente une forme spécifique d'un jeu de données : CSV...

extrait (Tours publie via opendatasoft) :

```
<rdf:Description rdf:about="https://data.tours-metropole.fr/api/v2/catalog/exports/rdf?lang=fr">
  <dcat:dataset rdf:resource="https://data.tours-metropole.fr/api/v2/catalog/datasets/sections-sur-le-territoire-de-tours-metropole-val-de-loire"/>
</rdf:Description>
```

```
<rdf:Description rdf:about="https://data.tours-metropole.fr/api/v2/catalog/datasets/sections-sur-le-territoire-de-tours-metropole-val-de-loire-csv">
  <rdf:type rdf:resource="http://www.w3.org/ns/dcat#Distribution"/>
  <dcat:accessURL rdf:resource="https://data.tours-metropole.fr/api/v2/catalog/datasets/sections-sur-le-territoire-de-tours-metropole-val-de-loire/exports/csv"/>
```

```
<dct:description>csv export of https://data.tours-metropole.fr/api/v2/catalog/datasets/sections-sur-le-territoire-de-tours-metropole-val-de-loire</dct:description>
<dct:format>csv</dct:format>
<dcat:mediaType>text/csv</dcat:mediaType>
</rdf:Description>
```

Ce qui (d'après le catalogue de vidéos produit l'exemple donné sur la rdflib) devrait aussi pouvoir s'écrire comme ça :

```
<rdf:Description rdf:about="url de ce fichier">
  <dcat:dataset rdf:resource="url de ce fichier#ID"/>
</rdf:Description>
```

```
<rdf:Description rdf:nodeID="ID">
  <rdf:type rdf:resource="http://www.w3.org/ns/dcat#Distribution"/>
  <dcat:accessURL rdf:resource="url du jeux de donnée"/>
  <dct:format>json</dct:format>
</rdf:Description>
```

Tout ça pour dire que la question du format de catalogue de donnée opendata concerne uniquement les métadonnées et ne répond pas à la question du format des données, pour qu'elles soient ensuite reprises dans n'importe quel portail open-data.

A cette seconde question, d'après ce que je lis dans la doc d'[opendatasoft](#) je dirais du GeoJSON ou un [JSON du type](#) :

```
{
  "data": [
    {"id1": {"key1": "val1", "key2": "val2", ...}},
    {"id2": {"key1": "val1", "key2": "val2", ...}},
    ...]
}
```