

# Open data – Données Libres



De l'utilisation des données en open data et des données  
Libres

Le 19 avril 2021  
Auteur : Benjamin LARCHER

# Droits d'auteurs

## **Licence**

Ce document est sous licence « GNU Free documentation 1.3 », hormis les dessins, ceux-ci n'étant pas la production de Nâga.

L'objet de cette Licence est de rendre tout manuel, livre ou autre document écrit « libre » au sens de la liberté d'utilisation, à savoir : assurer à chacun la liberté effective de le copier ou de le redistribuer, avec ou sans modification, commercialement ou non. En outre, cette Licence garantit à l'auteur et à l'éditeur la reconnaissance de leur travail, sans qu'ils soient pour autant considérés comme responsables des modifications réalisées par des tiers.

Cette démarche est effectuée dans un souci de transparence, de réutilisations et de modifications de données pouvant être sujettes à débat.

# Sommaire

Introduction.....	1
Open data.....	1
<i>Définition.....</i>	<i>1</i>
<i>Utilisation.....</i>	<i>1</i>
Études de cas en open data.....	2
Coté donnée open data.....	2
Coté logiciel.....	2
<i>Fond de carte.....</i>	<i>2</i>
<i>Données.....</i>	<i>3</i>
Rendu.....	3
<i>Appel des données.....</i>	<i>4</i>
Open data / libre données.....	5
4 libertés fondamentales.....	5
Aspect philosophique.....	5
Très concrètement.....	5
Existence de données libres.....	5
<i>Wikipédia.....</i>	<i>5</i>
<i>Open street map.....</i>	<i>6</i>
<i>Wheelmap.....</i>	<i>6</i>
<i>Umap.....</i>	<i>6</i>
Études de cas en libre.....	7
Schéma « idéal ».....	7
<i>Open Street Map.....</i>	<i>8</i>
<i>Sollicitation des structures.....</i>	<i>8</i>
<i>Ajout des données à Open Street Map.....</i>	<i>8</i>
<i>Le site de jeu de données.....</i>	<i>8</i>
<i>Remontée citoyenne ou d'autres institutions.....</i>	<i>8</i>
<i>Centralisation / Décentralisation.....</i>	<i>8</i>
<i>Les peurs habituelles.....</i>	<i>9</i>
Cas particulier de la remontée de données interne.....	9
Site de cartographie.....	10
<i>Centralisation / Décentralisation.....</i>	<i>10</i>
<i>Participation.....</i>	<i>10</i>
<i>Annuaire.....</i>	<i>10</i>

# Introduction

Depuis quelques années l'open data à le vent en poupe, mais à quoi ça sert et comment ça fonctionne ?

Dans le monde du logiciel, on parle d'open source et de logiciel Libre, qu'est ce qui différencie les deux et est-ce que cela peut s'appliquer aux données ?

## Open data

### *Définition*

Wikipédia nous éclaire sur la définition en nous indiquant que les données ouvertes,<sup>1</sup> quelque soit l'organisme qui les produit, sont réutilisables par un tiers, sans restriction.

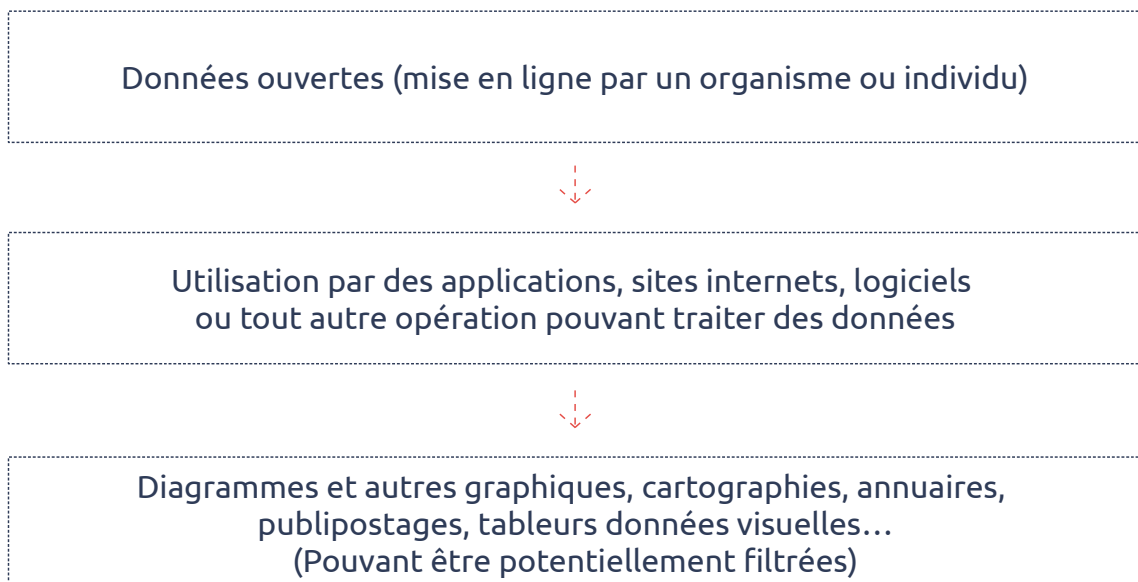
Bien qu'un jeu de données puisse être plus complexe, très souvent, les sites d'open data fournissent des tableurs, ce qui constitue la forme la plus simple dans les bases de données.

### *Utilisation*

À une époque où les données sont omniprésentes, leur réutilisation a un intérêt majeur afin d'alimenter des applications, sites internet, logiciels ou tout autre opération pouvant traiter des données (donc également des machines artistiques...).

Les rendus de l'exploitation de ses données peuvent être très variés, comme des diagrammes et autres graphiques, des cartographies, des annuaires, des publipostages, tableurs données visuelles...

Les données ouvertes permettent de ne pas réinventer la roue lorsque le travail a été précédemment réalisé et partagé.



---

<sup>1</sup>Les données ouvertes ou open data sont des données numériques dont l'accès et l'usage sont laissés libres aux usagers, qui peuvent être d'origine privée mais surtout publique, produites notamment par une collectivité ou un établissement public. Elles sont diffusées de manière structurée selon une méthode et une licence ouverte garantissant leur libre accès et leur réutilisation par tous, sans restriction technique, juridique ou financière.

# Études de cas en open data

## Coté donnée open data

Imaginons des données sous forme de tableur, qui est une des formes les plus classiques (Données non factuelles).

lat	long	nom	description	horaire	site	type
47.1922431	-1.5697547	Mairie de Rezé	Mairie principale	De 10 à 19h, du lundi au samedi	www.reze.fr	institution
47.1872945	-1.5477936	Nâga	Reconditio nneur	De 10 à 18h, du lundi au vendredi	ww.naga44.org	association

## Coté logiciel

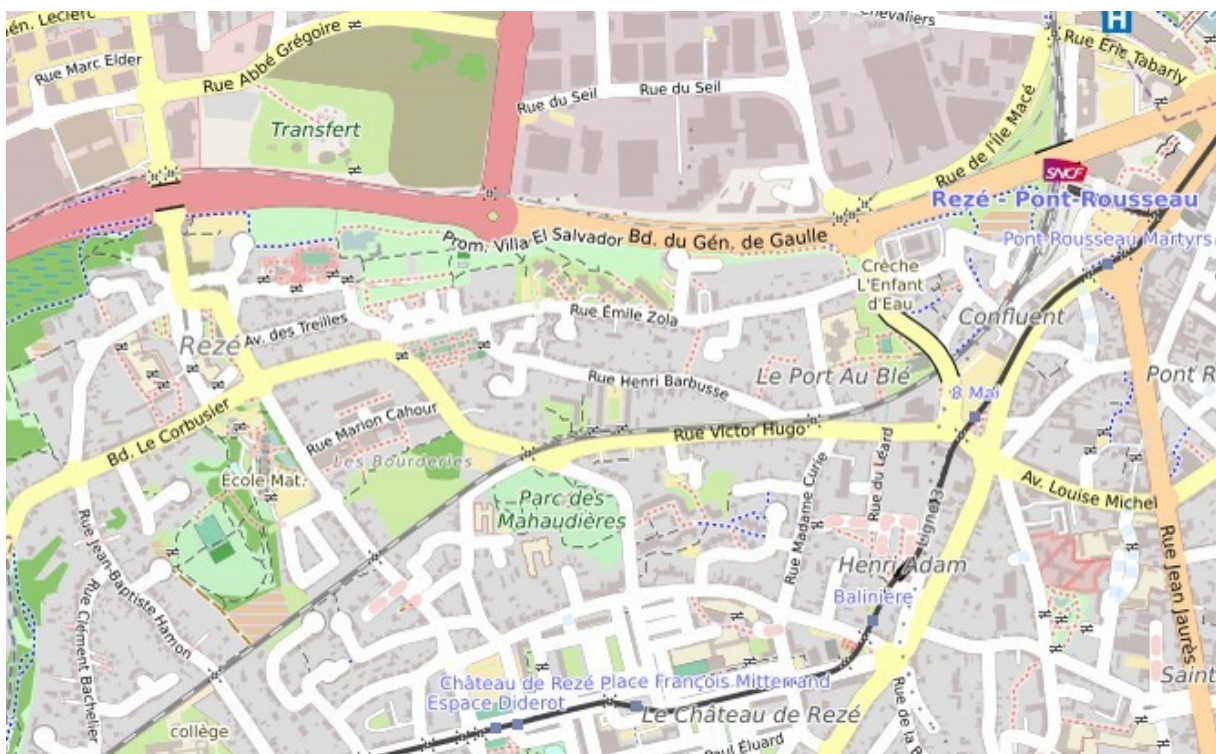
Dans cet exemple, nous voulons afficher les données sur une carte.

Nous affichons ce que l'on nomme un fond de carte. Un fond de carte n'est pas un simple contenant, mais est également constitué de données. Sur ce point de vue, Open Street Map, c'est :

- 7.3 millions d'utilisateurs enregistrés (30/12/2020)
- 7.3 milliards d'objets (30/12/2020)
- 6.6 milliards de nœuds (30/12/2020)
- 730 millions de polygones (30/12/2020)
- 4.5 millions d'éditions par jour (2020-08-18)

Bref, la très grande majorité des données sont les communs qui constituent le fond de carte.

## *Fond de carte*



## Données

Les données utilisées peuvent être renommées ou encore « appelées » dans un ordre différent comme montré ci-dessous en vert. Cela se fait assez aisément.

Pour l'utilisation que nous avons, nous ajoutons une nouvelle colonne : telephone (on évite en général les accents sur les noms de colonnes).

nom	latitude	longitude	description	horaire	site internet	type	telephone
Mairie de Rezé	47.1922431	-1.5697547	Mairie principale	De 10 à 19h, du lundi au samedi	www.reze.fr	institution	02 40 02 02 20
Nâga	47.1872945	-1.5477936	Reconditionneur	De 10 à 18h, du lundi au vendredi	ww.naga44.org	association	02 40 02 02 23

Les données seront dans ce cas utilisées de trois façons :

- par le placement de la position (colonnes latitude et longitude) ;
- via la couleur d'icônes (selon le type : institution ou association. Cela pourrait aussi être un affichage par menu déroulant par case à cocher...) ;
- au moment de l'affichage des autres données du tableur (nom, description, horaire site internet et téléphone) lorsque l'on clique sur une icône.

## Rendu





## *Appel des données*

Dans notre cas, les données sont affichées ainsi, lorsque l'on clique sur une icône :

<:nom:> / <:description:>

Horaire

<:horaire:>

<:site internet:>

<:telephone:>

Ceci n'est pas du vrai code, mais le principe est le même. <:XXX:> appelle des variables que l'on insère au milieu de textes, images, tableaux...

# Open data / libre données

## 4 libertés fondamentales

D'un point de vue technique, rien ne différencie l'open source du Libre dans le monde du logiciel.

Dans les deux cas, il y a quatre libertés fondamentales :

- la liberté d'utiliser le logiciel
- la liberté de copier le logiciel
- la liberté d'étudier le logiciel
- la liberté de modifier le logiciel et de redistribuer les versions modifiées

On retrouve d'ailleurs dans l'open source la plupart de ces caractéristiques. La liberté d'étudier a effectivement moins de sens pour les données, celles-ci n'ayant pas de code à examiner.

## Aspect philosophique

À l'origine le libre est une production de hackers. Contrairement à la pensée répandue, les hackers ne sont pas des pirates d'internet (cela correspond aux crackers ou black hat).

La grande majorité des hackers produisent du commun ou de la sécurité (white hat). Un autre communauté (les grey hat) attaquent des services, afin de proposer des solutions de sécurité.

Bref, l'éthique des hackers suit six principes :

- L'accès aux ordinateurs — ainsi que tout ce qui peut permettre de comprendre comment le monde fonctionne — doit être universel et sans limitations. Il ne faut pas hésiter à se retrousser les manches pour surmonter les difficultés.
- Toute information doit être libre.
- Se méfier de l'autorité — encourager la décentralisation.
- Les hackers doivent être jugés selon leurs hacks, et non selon de faux critères comme les diplômes, l'âge, l'origine ethnique ou le rang social.
- On peut créer l'art et le beau à l'aide d'un ordinateur.
- Les ordinateurs peuvent améliorer notre vie.

Dans l'éthique hacker, les individus sont des acteurs du bien commun.

## Très concrètement

Concrètement on peut considérer que l'open data partage des données produites.

Le libre partage les données produites et permet des corrections, des remontées de bugs, des propositions d'amélioration (ajout de colonne par exemple), le partage dans les communs...

Le libre, dans l'esprit d'accès au plus grand nombre, implique souvent le fait de documenter, afin d'être réutilisable par le plus grand nombre.

## Existence de données libres

La dimension libre ouvre des possibles bien plus larges qui ne sont tout simplement pas existants en open source. Voici quelques exemples.

### *Wikipédia*

Parmi les modèles utilisant ce type de système, on peut citer Wikipédia, considérée à présent comme l'encyclopédie la plus fournie et parmi les plus qualitatives au monde.

<https://fr.wikipedia.org/>

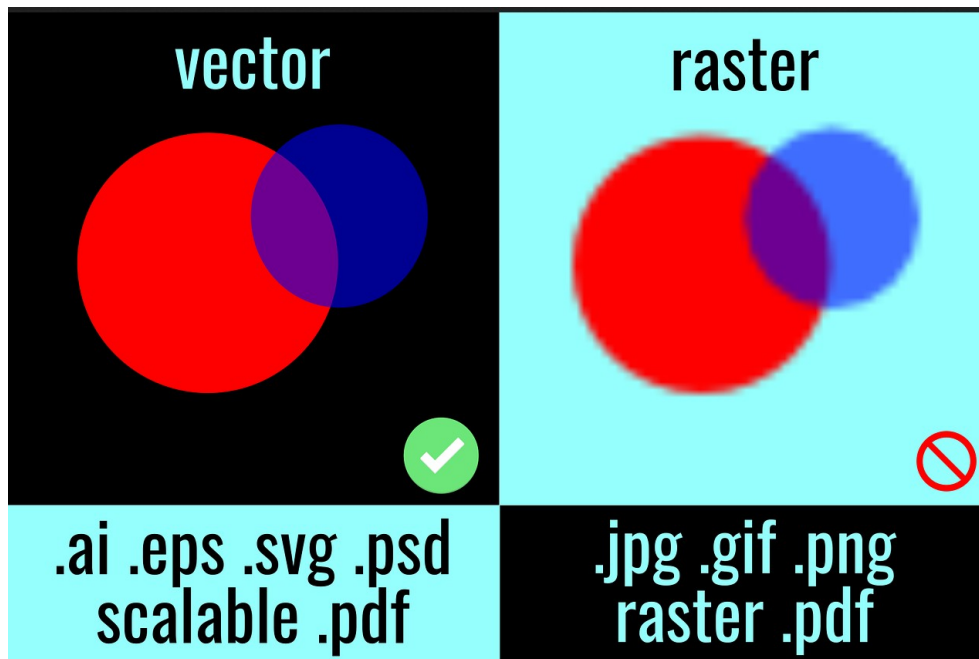


## Open street map

Open Street Map (OSM) est un projet collaboratif de cartographie en ligne qui vise à constituer une base de données géographiques libre du monde. <https://www.openstreetmap.org>

Depuis le 19 janvier 2009, la Direction Générale des Finances Publiques (DGFIP) en charge du cadastre en ligne a autorisé Open Street Map à utiliser ces données. De nombreux bénévoles ont pu vectoriser des cartes parfois très pixelisées par les nombreuses photocopies.

Les vecteurs sont plus précis, plus légers et n'ont pas « l'effet escalier » (pas de pixellisation) quand nous zoomons dessus.



Crédit image : Rfind / Réutilisation autorisée

<https://pixabay.com/fr/vectors/raster-vecteur-vs-%C5%93uvres-d-art-5171211/>

L'IGN autorise également les contributeurs OSM à utiliser ses images aériennes.

## Wheelmap

Grace aux nombreux contributeurs d'Open Street Map, des projets comme Wheelmap peuvent avec des millions de données, indiquer les lieux accessibles aux personnes à mobilité réduite.

Lorsque l'on renseigne une donnée sur Wheelmap, elle s'ajoute à Open Street Map si elles n'étaient pas présentes, profitant ainsi même à d'autres logiciels / site internet / application / base de données...

<https://wheelmap.org>

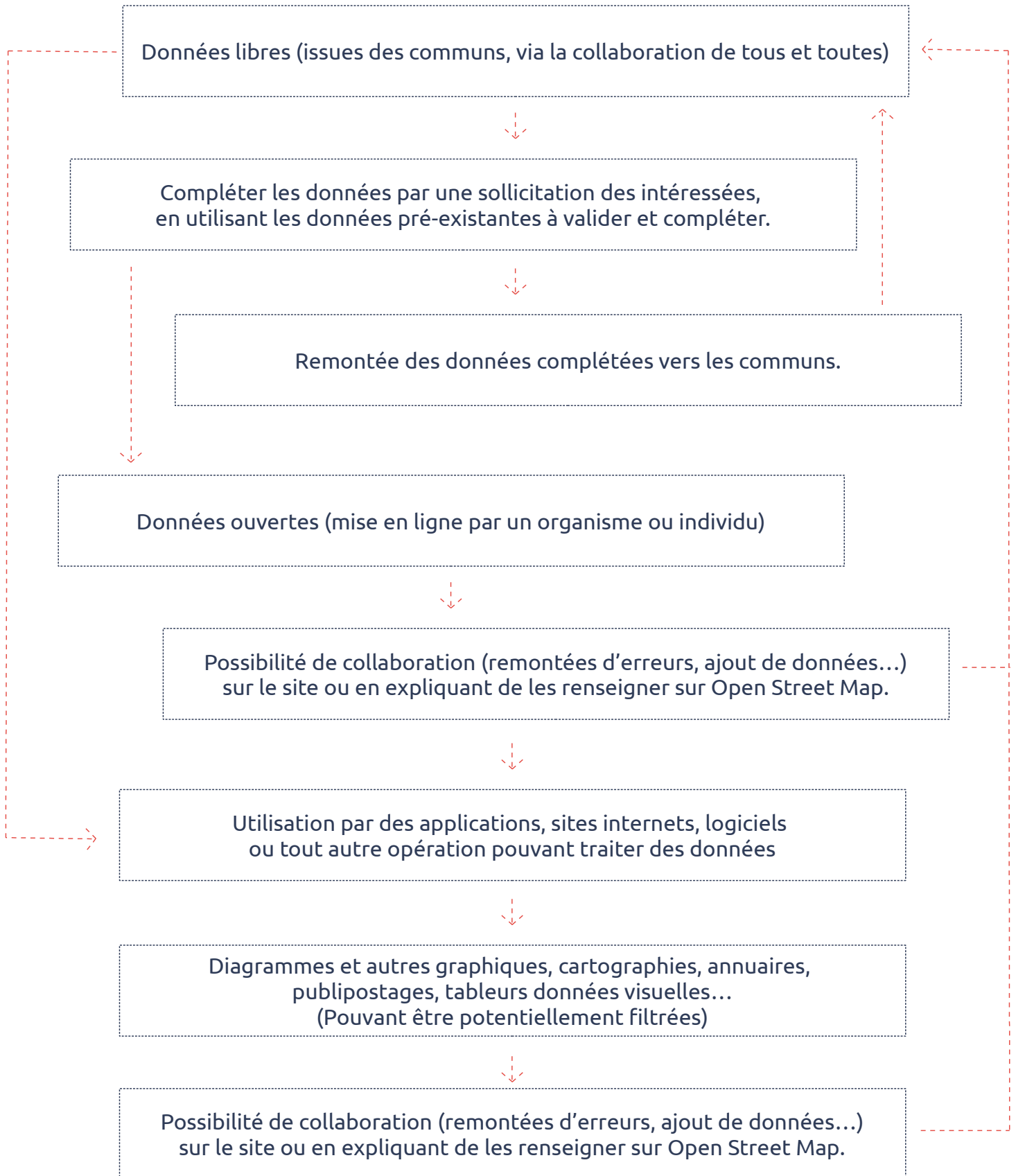
## Umap

Basé sur Open Street Map, Umap permet de créer des cartes avec des données personnalisées ou temporaires qui n'ont pas forcément de raison de se trouver sur Open Street Map (étant donné la non-existence de certains tags sur OSM ou car les données sont temporaires, par exemple emplacements de stands sur un week-end).

<http://umap.openstreetmap.fr/>

# Études de cas en libre

## Schéma « idéal »



Dans un schéma idéal, on aurait tendance à constituer sa base de données en respectant un certain ordre.

Dans le cas d'une données cartographique, on commencerait par voir les couches libres existantes sur Open Street Map.

## *Open Street Map*

Dans le cas de l'exemple que nous avons pris cela donnerait les données suivantes.

Donnée	nom	description	horaire	site	type	telephone
cadastre-dgi-fr source : Direction Générale des Impôts - Cadastre. Mise à jour : 2010	Mairie de Rezé	Mairie principale			institution	
Mieux trier à Nantes;CRESS Pays de la Loire	Association Nâga	Reconditionne ment et don d'ordinateurs sous Linux. Sur rendez-vous uniquement		<a href="http://www.naga44.org">www.naga44.org</a>		02 85 52 31 22

Ainsi, il est possible de mettre d'appeler déjà pas mal de données, sans avoir à envoyer des formulaires et demander aux structures (qui sont sollicitées plusieurs fois par an !) de remplir leurs données.

## *Sollicitation des structures*

Plutôt que le classique formulaire à remplir une énième fois, il serait alors possible de demander de vérifier les données envoyées et de compléter celles manquantes.

## *Ajout des données à Open Street Map*

Les données récupérées peuvent alors être ajoutées à Open Street Map avant de créer son jeu de donnée personnel.

## *Le site de jeu de données*

Le site de jeu de données (comme <https://data.nantesmetropole.fr>) peut alors afficher les données récoltées, en ajoutant ce qui lui est propre et ne se trouve pas sur Open Street Map.

## *Remontée citoyenne ou d'autres institutions*

Permettre et faciliter les remontées de données offre de nombreux avantages comme plus de réactivité sur des erreurs, plus de données, une mise à jour constante...

## *Centralisation / Décentralisation*

Tous les X temps des programmes peuvent examiner si les données sont toujours identiques à celle d'Open Street Map. Si cela n'est pas le cas, voir si la nouvelle donnée ajoutée à Open Street Map est la bonne ou si elle est à corriger.

Ainsi, si certaines données sont centralisées sur Open Street Map, ce qui enrichit les applications / logiciels / site les utilisant, cela permet aussi une bien meilleure mise à jour des données en décentralisant les données sur les divers sites les ayant exportées.

## Les peurs habituelles

Les craintes habituelles sont souvent les mêmes. Et si quelqu'un écrit n'importe quoi ?

Cette crainte états aussi à l'origine de Wikipédia et Open Street Map. Divers couches de sécurité ont été mise en place. Par exemple avec des modérateurs, des vérifications de robots, des demandes d'inscription...

De plus, les jeux de données ne se mettent pas à jour automatiquement avec les données d'Open Street Map, mais des programmes peuvent examiner les données qui sont « en conflit » uniquement entre les données d'Open Street Map et le site d'open data. Ainsi, cela élimine tout risque d'affichage d'âneries potentielles.

## Cas particulier de la remontée de données interne

Bien qu'il soit plus simple et plus efficace de remonter des données via Open Street Map, il serait aussi possible de les remonter à partir du site d'open data.

Reprenons les mêmes données (les colonnes latitude et longitude sont enlevées pour améliorer la lecture).

Donnée	nom	description	horaire	site	type
Originale	Mairie de Rezé	Mairie principale	De 10 à 19h, du lundi au samedi	www.reze.fr	institution
Originale	Nâga	Reconditionneur	De 10 à 18h, du lundi au vendredi	ww.naga44.org	association

Si une personne voit une erreur elle peut alors corriger. Cela ne retire pas l'ancienne donnée, mais stocke une nouvelle données.

Donnée	nom	description	horaire	site	type
Originale	Mairie de Rezé	Mairie principale	De 10 à 19h, du lundi au samedi	www.reze.fr	institution
Originale	Nâga	Reconditionneur	De 10 à 18h, du lundi au vendredi	ww.naga44.org	association
<b>Proposition 1</b>	Nâga		<b>De 10 à 18h, du lundi au jeudi</b>		

Ainsi, pour les utilisateurs, les choix sont plus nombreux. Certaines données sont conflictuelles, ce qui est une bonne chose, car à présent on peu choisir comme préférence :

- les derniers ajouts ;
- les confirmés par le propriétaire ;
- les plus approuvés ;
- les remontées citoyennes ;
- les données originales ;
- ou encore montrer les conflit et sélectionner ses préférences pour chaque conflit...

Le site peut aussi inciter à entrer les données directement sur des sites libres comme Open Street Map en indiquant qu'il met à jour ces données à partir de là de temps en temps. Ainsi, il peut se délester de la partie remontée de données, ce qui est plus simple et efficace.

# Site de cartographie

Un site internet intéressant comme modèle est la carte ouverte.

<https://rennes.carte-ouverte.org/>

## *Centralisation / Décentralisation*

À priori, il s'agit d'une carte un peu fourre-tout.

Cependant, lorsque l'on sélectionne des catégories, uniquement celles-ci s'affichent si on copie le lien créé par la touche « permalien ».

Ceci peut facilement être mis sur un autre site avec idéalement uniquement les catégories sélectionnées d'affichées.

## *Participation*

Il est possible de participer au site !

## *Annuaire*

Une fonction annuaire existe.

Il est possible d'imaginer un annuaire imprimable suite à la sélection de données. Ce qui pour des travailleurs sociaux / association / public / institutions (...) aurait un intérêt pour distribuer à leur public.

Des fonctions comme l'impression d'une seule donnée placée sur une carte, avec l'adresse, les horaires... pourrait aussi être pertinent.