

Annexe aux délibérations décidant de l'ouverture des données publiques des Parcs Naturels Régionaux

Qu'est-ce que l'Open Data ?

Ce sont les données accessibles en ligne, réutilisables avec un format lisible par tous et une licence qui autorise l'usage libre, même à des fins commerciales. Des données sont ouvertes si tout le monde peut **les utiliser, les réutiliser et les partager librement**. Le pouvoir de l'Open Data est de changer l'interaction entre les gouvernements, les entreprises et la société. Cela nous mène vers de nouvelles applications et vers de nouvelles solutions, en générant donc de la valeur pour la société et l'économie. L'Open Data résout aussi certaines questions environnementales. Mais toute limitation mise en place sur l'usage libre des données en limitera le potentiel d'utilisation.

L'Open Data est **une ressource naturelle de l'ère digitale** où tout le monde peut l'utiliser en même temps. Des bénéfices sont apportés tant dans le secteur privé que public. L'Open Data améliore le service public, développe la confiance des citoyens et renforce l'engagement politique. Il existe de nombreux exemples d'initiatives liées à l'Open Data qui montrent de nombreux avantages sociaux, économiques et environnementaux. Par exemple, l'ouverture de données géographiques et statistiques ont été utilisées par les groupes humanitaires pour les approvisionnements dans les zones sinistrées comme en 2015 en Haïti et dans les Philippines. Les opérateurs d'imagerie spatiale se sont engagés à mettre à disposition des services de secours des photos satellitaires de ces zones sinistrées. Avec la mise à disposition de ces données, OpenStreetMap (OSM) ou encore CrisisMappers ont pu réaliser des cartographies peu de temps après les intempéries. L'Open Data intervient ainsi dans la gestion de crise et dans l'action humanitaire. C'est donc un agent de changement. On rentre dans une logique où les gouvernements, les entreprises et les citoyens sont comme simultanément les fournisseurs et les utilisateurs de données. Mais que faut-il pour que les données soient ouvertes ?

Il faut savoir que **sans licence, les données ne sont pas réellement ouvertes**. Une licence permet d'informer les gens qu'ils ont le droit d'accéder, d'utiliser et de partager les données. S'il n'y a pas de licence qui spécifie ces points, les données ne sont pas réellement ouvertes. La licence donne une certitude et une permission de prendre des risques ainsi qu'une garantie de bénéfices d'Open Data. La règle d'or est la suivante : *plus une licence ouverte est simple, mieux c'est*. En effet, plus facile ce sera pour le réutilisateur de les intégrer à d'autres ensembles de données et de créer de nouveaux produits et services. Il est possible d'inclure des clauses dans tout contrat de licence. Mais il est essentiel de s'assurer que toutes les données impliquées dans les prestations de services importants soient publiées sous une licence ouverte le plus possible. D'après les législations de droits d'auteur ou de bases de données, sans licence, certes, les données sont accessibles au public mais les utilisateurs n'ont pas l'autorisation d'y accéder, de les utiliser et de les partager. Différents types de licence existent :

- **Licences *Creative Commons*** : trois versions sont disponibles :
 - Zero (CCO) ;
 - Paternité (« *Attribution* ») CC-BY v4.0
 - Paternité & Partage à l'identique (« *Attribution and share-alike* ») CC-BY-Sa v4.0

Zero et Paternité sont les deux licences qui offrent plus de souplesse d'utilisation à l'inverse de Partage à l'identique qui émet une limite dans l'utilisation commerciale.

- **Licences sur mesure** : elles peuvent rendre l'utilisation des données plus complexes. Elles peuvent restreindre l'intégration des données etc.



- **Licences gouvernementales ouvertes** : certains gouvernements choisissent de développer leurs propres licences. Les meilleurs exemples sont des licences : courtes, compatibles avec d'autres licences utilisées, et faciles à respecter. En France, le [décret n° 2017-638](#) du 27 avril 2017¹ précise les licences d'utilisation des données que les administrations pourront utiliser par défaut pour une démarche Open Data. Si la liste des licences parue dans le décret est vouée à évoluer, les deux licences standards de l'Open Data français sont :
 - o *Licence ouverte* créée par la mission Etalab² (2011) et récemment actualisée (2017) : c'est une licence « permissive » où la totale liberté de réutilisation (y compris à des fins commerciales), des données ou des logiciels, sera accordée aux réutilisateurs ;
 - o *Open Database licence* (ODbL) ou licences « avec obligation de réciprocité » : d'origine anglo-saxonne, elle oblige les réutilisateurs des données ouvertes à ouvrir, à leur tour, le résultat de leur travail. Les réutilisateurs restent libre d'exploiter les données.

Si l'on souhaite établir une licence spécifique, il faudra réaliser une demande motivée auprès de la DINSIC (Direction interministérielle des systèmes d'information et de communication de l'Etat) pour que la licence soit homologuée.

- **Licences et les tiers** : lorsque l'on fait appel à des sous-traitants qui fournissent des services en leur nom, l'engagement quant à la publication de données s'étend aux tiers. Il faut donc l'inclure dans les contrats.

Si ces licences donnent des informations sur l'accès, le partage et la réutilisation des données, il faut également prendre en compte les standards déterminés par la communauté de l'Open Data, ainsi que la durabilité de l'ouverture des données.

Qui dit données ouvertes dit également standards ouverts. Ces derniers définissent les aspects pratiques, légaux, techniques et sociaux d'un Open Data de qualité. L'aspect légal touche ce qui concerne la protection d'informations sensibles ou de données personnelles pour préserver les droits des propriétaires de ces données tout en promouvant l'utilité et l'exploitabilité de ces données. Le côté technique traite du format sous lequel on peut accéder aux données. Le côté social correspond à la création de communautés Open Data, ou encore la création d'outils et de manuels pour travailler ces données. Les « *Open Data Certificates* » de l'*Open Data Institute* (ODI) font office d'auto-évaluation pour les éditeurs. Dans certains cas, ils permettent aux utilisateurs de données ouvertes, d'évaluer et de comparer la facilité d'utilisation d'un ensemble de données. Il existe trois indicateurs permettant de mesurer le succès de l'Open Data :

- *Disponibilité opérationnelle* : soutien des licences ouvertes, politiques et procédures en place, investissements existants, soutien éducatif disponible, engagement au sens large ;
- *Implémentation* : soutien de l'innovation, générer de la valeur, contribuer à la croissance économique ;
- *Impact* : mobilisation en faveur de l'Open Data, croissance économique liée à l'Open Data, avantages sociaux et environnementaux.

Il est intéressant de savoir aussi combien de temps les données seront disponibles. **La durabilité de l'Open Data est importante car elle rassure les utilisateurs.** Pour reconnaître des sources d'Open data durables, il faut se poser les questions suivantes :

- Quelle est la source ?
- Qui finance la publication ?

¹ Décret relatif aux licences de réutilisation à titre gratuit des informations publiques et aux modalités de leur homologation.

² Fait partie de la Direction interministérielle du numérique et du système d'information et de communication de l'Etat (DINSIC). Elle coordonne la politique d'ouverture et de partage des données publiques (« open data »).

- Que dit la licence ?
- Qui utilise déjà les données ?
- Quelle est l'histoire derrière la publication des données ?
- Comment puis-je contribuer ?

Pour qu'une série de données soit durable, elle est appuyée par :

- Un financement sûr pour couvrir les coûts des mises à jour ;
- Un soutien actif des éditeurs pour que les mises à jour soient une priorité ;
- Une large réutilisation des données qui démontrent leur valeur.

Certains choix sont donc décisifs en vue d'un Open Data durable, notamment celui d'une plateforme Internet.

Le choix de la plateforme Open Data est également important car c'est **une aide à l'utilisateur dans la découverte, l'exploration, le filtrage et la visualisation des données pour une utilisation directe**. Plusieurs solutions possibles :

- Cataloguer les données qui existent déjà à plusieurs endroits ;
- Une approche plus directe de gestion de données : on télécharge la donnée *via* la plateforme (visualisation et exploration des données sur la plateforme).

Il peut être nécessaire de combiner ces deux approches. Les caractéristiques clés des plateformes Open Data sont des données faciles à trouver, une cohérence et une uniformité pour l'utilisateur et pour l'accès à ces données.

	Plateforme catalogue	Plateforme gestion de données
Tableau récapitulatif des 2 types de plateformes Open data	Recherches fédérées dans un vaste domaine	Des sources fiables de données
	Mécanismes de découverte	Une qualité de données améliorée
	Métadonnées ³ clés liées à la licence et à la mise à jour des données	La possibilité de rechercher et de manipuler les données

Grâce à ces plateformes, l'Internaute peut visualiser les données disponibles, les filtrer, les partager et notamment les exploiter. Pour se faire, le choix du format est important pour qu'un utilisateur puisse y accéder et la réutiliser.

Le bon format des données est crucial pour maintenir l'exploitabilité et l'accès aux données. Il est conseillé de choisir le format le plus réutilisable, qui soit déjà en place pour gérer et utiliser les données. Mais le format doit aussi s'adapter au contexte d'utilisation des données, ce qui signifie qu'il ne soit pas forcément le plus répandu ou le plus largement compris. Dans ce type de cas, il faudra peut-être publier à la fois sous le format .CSV en plus du format d'origine. Il faut penser également au choix de la structure (hiérarchique ou en réseau) et au moyen de distribution. Si le format le plus utilisé pour l'Open Data est le .CSV, les formats de données géospatiales sont souvent plus complexes : geoJSON et le KML sont les plus largement exploitables.

Il existe un moyen d'évaluer à quel point les données sont utilisables : **les 5 étoiles des Données Ouvertes & Liées (LOD)⁴**. Il s'agit de directives mesurant la qualité d'intégration des données sur le Web. Ces étoiles examinent l'accessibilité et la facilité d'utilisation technique d'une série de données allant de 1 étoile (disponibilité en ligne) à 5 étoiles (appartenance à un Web de données). Ces recommandations ont été développées par Sir Tim Berners-Lee.

³ « Métadonnée » signifie littéralement la donnée de la donnée.

⁴ <http://5stardata.info/fr/>

En plus d'un bon format des données, il est important que les données à ouvrir soient « propres » afin de les travailler et de générer de la nouvelle valeur. On entend par « données propres » des données cohérentes et uniformes sans doublons et prêtes à être consommées par des machines. Ces données « propres » facilitent la combinaison de séries de données et elles permettent une compréhension plus approfondie. Des outils du nettoyage de données existent : *Open Refine*, un logiciel conçu pour traiter des données brutes, ou encore des programmes de feuilles de calcul pour traiter les fautes d'orthographe, les données redondantes etc.

Un des objectifs ultimes de l'Open Data est d'aboutir à un Web de données ouvertes et liées. Ce Web en question fournit des panneaux indicateurs en exigeant que toute donnée soit connectée à d'autres *via* des liens. L'idée est de naviguer dans le Web en utilisant les données au lieu de pages web. L'intérêt de ce Web de données ouvertes et liées est qu'il évite de répéter les données puisque l'on partage une seule ressource parmi un nombre infini de séries de données. Le but n'est pas ici d'exposer la structure du Web de données liées mais de faire comprendre que c'est un des objectifs vers lequel tend l'Open Data.

L'Open data et les Parcs naturels régionaux

L'Open Data s'inscrit dans une dynamique interparc dans le sens d'une valorisation d'un patrimoine numérique commun aux PNR de PACA pour servir l'intérêt général. La donnée publique est avant tout un bien commun.

Le Système d'Information Territorial (SIT) des Parcs naturels régionaux de Provence-Alpes-Côte d'Azur (PACA) est un dispositif technique et organisationnel qui permet de centraliser et de diffuser les informations géographiques disponibles sur les territoires de quatre PNR (Luberon, Camargue, Queyras, Verdon) de la région PACA. C'est donc dans une réflexion collective interparc que les Parcs initie une démarche Open Data. Il s'agit de mettre à disposition des citoyens, des entreprises, des chercheurs, des étudiants, des associations, des autres acteurs publics, les jeux de données numériques que le Parc produit à travers ses actions. Ces données numériques pourront être réutilisées, par exemple, pour développer des applications numériques ou pour appuyer une étude, ou encore pour réaliser une étude comparative à l'échelle nationale, européenne voire internationale.

L'ouverture des données publiques des Parcs *via* Internet peut contribuer au développement d'innovations, de nouveaux usages et de services numériques mettant en valeur son patrimoine au sens large. Le Parc relève ainsi le défi de la transition numérique pour diffuser et partager les données géographiques relatives à la protection environnementale et du savoir. Cette libération des données publiques des Parcs est ainsi un moyen de valoriser la pluralité du travail quotidien des acteurs des Parcs et de sensibiliser ainsi un public plus large.

Dans les chartes des Parcs, les axes font référence au partage des données publiques par le développement de la connaissance et des partenariats. Dans le cadre de la démarche Open Data, les Parcs proposeront dans un premier temps des jeux de données géographiques normalisés, avec une priorité sur les thématiques de l'environnement et du patrimoine. Dans l'optique du SIT, 8 bases de données communes aux quatre PNR de PACA sont alimentées et uniformisées suivant la charte géomatique interparc. L'objectif est de rendre ces données accessibles à la fois *via* le SIT, sur les plateformes de l'OpenPACA, du CRIGE PACA et du gouvernement data.gouv.fr. Dans le cadre de la refonte du portail SIT, le but est de mettre en place un catalogue de données ouvertes avec un accès aux informations sur ces données, la possibilité d'explorer les jeux de données disponibles et de les télécharger. Dans le cadre de la directive européenne INSPIRE⁵, un projet a débuté avec le CRIGE PACA pour le moissonnage des métadonnées⁶. Si la directive INSPIRE impose ces

⁵ Cf. Délibération de principes relative à l'ouverture des données publiques des PNR du Luberon et du Verdon.

⁶ Le moissonnage consiste, pour un fournisseur de données, à récolter (moissonner) des données figurant dans les bases de données. Pour qu'il puisse se faire, il faut que le service qui récolte les données et la base de données recourent au

métadonnées, cela rentre également dans une démarche Open Data qui est de faire en sorte que les données soient utilisables et durables. Pour aller plus loin, une de nos ambitions techniques serait d'atteindre la 4^e étoile du LOD, à savoir des données identifiées par des URL⁷ (avec date de mise à jour) afin que l'on puisse « pointer » un lien vers elles. Pour concrétiser cette idée, il s'agirait de convertir les données en triplets RDF⁸, langage de base du Web sémantique⁹. Ceci étant, une des étapes primordiales est de commencer par normaliser les premiers jeux de données à ouvrir.

Dans ce contexte, pour chaque jeu de données, il s'agira d'y associer une licence qui apportera un cadre juridique dans le but d'une réutilisation libre et gratuite des données, permettant l'échange et la construction de nouveaux services qu'une diffusion libre est en mesure d'offrir. Le tout en respectant les différentes lois relatives à la protection des données personnelles. Dans le cadre des Parcs, les jeux de données portant sur des espèces sensibles seront retravaillés avant publication, pour assurer la protection de l'environnement auxquelles elles font référence (art. L124-4 de la loi CADA modifié par l'ordonnance n°2010-1233 du 21 octobre 2010). La licence *Open Database Licence* (ODbL) permet par exemple le libre accès, la libre reproduction, redistribution, et la libre réutilisation (y compris à des fins commerciales) des données, avec, comme seule éventuelle condition de devoir redistribuer sous une licence analogue les bases de données dérivées, dans leur totalité. La licence ouverte d'Etalab (version 2.0) semble également intéressante du fait qu'il est stipulé que le producteur de données « *ne peut être tenu pour responsable de toute perte, préjudice ou dommage de quelque sorte causé à des tiers du fait de la réutilisation* » (p. de la licence ouverte).

A long terme l'ensemble des PNR de PACA, pourraient envisager de lancer un appel à projets innovants en matière de réutilisation de ces données publiques, en collaboration avec des partenaires publics et privés. De nombreuses démarches Open Data ont été initiées un peu partout dans le monde et notamment en France, classée troisième au classement mondial de l'*Open Data Index* en 2014 (source : <https://www.etalab.gouv.fr/la-france-3eme-au-classement-mondial-de-lopen-data-index>).

A titre indicatif :

Lien vers une carte interactive de l'Open data en France : <http://opendata.brocas.fr/content/la-carte-de-lopen-data-en-france>

Lien vers la cartographie interactive des portails Open data dans le monde : <https://opendatainception.io/>

Document réalisé à partir des sources suivantes :

Commission d'accès aux documents administratifs (Cada) : <http://www.cada.fr/informations-relatives-a-l-environnement,6086.html>

Charte 2008-2020 du Parc naturel régional du Luberon : http://parcduluberon.fr/sites/default/files/pnrLuberon/pdf/2008_charte.pdf

Direction du numérique de Bordeaux métropole et Cédric Favre, Version 1 – octobre 2014 – « Guide de demande de données publiques auprès des collectivités »

même procédé de « codage technique ». Géosource permet de découvrir quelles sont les métadonnées ayant été ajoutées, supprimées ou mise à jour dans le nœud distant.

⁷ *Uniform Resource Locator*

⁸ *Resource Description Framework* : modèle de graphe destiné à décrire les ressources Web et leurs métadonnées pour permettre le traitement automatisé de telles descriptions. Le RDF est aussi de l'interopérabilité syntaxique, c'est-à-dire un format d'échange internationalement reconnu.

⁹ Correspond aux langages et aux concepts spécifiques définis par le W3C (consortium du Web : *World Wide Web*).

Etalab.gouv.fr, septembre 2013, « Vade-mecum sur l'ouverture et le partage des données publiques »

Grotius International – *Géopolitiques de l'humanitaire* – Gaël Musquet, 28/09/2011 : <http://www.grotius.fr/open-data-gestion-de-crise-et-action-humanitaire-le-%C2%AB-web-humanitaire-%C2%BB-est-deja-une-realite/>

Le kit collectivités d'OpenDataLab : <http://www.opendatalab.fr/laboratoire/le-kit-collectivites>

Legifrance.gouv.fr : [Décret n° 2017-638](#) du 27/04/2017 relatif aux licences de réutilisation à titre gratuit des informations publiques et aux modalités de leur homologation

Participation à la réunion d'information et de sensibilisation à l'Open Data à la région PACA dans le cadre de l'expérimentation Open Data Locale (05/04/2017) : <http://opendata.regionpaca.fr/opendatalocale.html>

Programme e-Learning du Portail Européen de Données : <https://www.europeandataportal.eu/elearning/fr/#/id/co-01>

Star data : <http://5stardata.info/fr/>

Wikipédia - *Resource Description Framework* https://fr.wikipedia.org/wiki/Resource_Description_Framework

Référent de l'Open data du SIT des Parcs :

Nicolas PETIT, Chargé de mission SIT, observatoire du territoire, évaluation – npetit@parcduluberon.fr