

Open Data : l'ouverture des données publiques du Parc naturel régional du Verdon

Alexandra Greco

▶ **To cite this version:**

Alexandra Greco. Open Data : l'ouverture des données publiques du Parc naturel régional du Verdon. Géographie. 2017. dumas-01859109

HAL Id: dumas-01859109

https://dumas.ccsd.cnrs.fr/dumas-01859109

Submitted on 21 Aug 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Université de Toulouse

MASTER 2 GEOMATIQUE

« **Sc**I**e**nces **G**éomatiques en environne**M**ent et **A**ménagement » (**SIGMA**)

http://sigma.univ-toulouse.fr

RAPPORT DE STAGE

OPEN DATA : L'ouverture des données publiques du Parc naturel régional du Verdon





GRECO Alexandra

Parc naturel régional du Verdon

Maître de stage: PETIT Nicolas, chargé de projet géomatique et évaluation au Parc du Verdon

Tuteur-enseignant: HEINTZ Wilfried, ingénieur agronome spécialiste en géomatique et gestion de bases de données à l'INRA/Dynafor

Septembre 2017

*Photos de couverture prises par Alexandra Greco, le 21 mai 2017 à Moustiers-Sainte-Marie.

Contenu sous Licence Ouverte

Auteur : Alexandra GRECO
Contact : alexandra.greco16@gmail.com

RESUME

Le Parc naturel régional du Verdon produit une grande variété de données, et souhaite s'engager dans une démarche d'ouverture de ses données publiques (Open Data). Ce projet suit les grandes lignes de la Charte du Parc et de ses actions. Le but du stage est d'élaborer une démarche Open Data et de mettre en place un catalogue de données. C'est ainsi répondre au contexte juridique actuel sur l'Open Data, et aux obligations de la Directive INSPIRE auxquelles le Parc du Verdon est soumis pour les données géographiques. Inspirée des préconisations de l'association OpenData France (Occitanie), la méthodologie s'est adaptée au contexte d'un Parc naturel régional et à ses enjeux. Plusieurs choix techniques et stratégiques ont été pris pour mener à bout ce projet. La stratégie finale de diffusion des données du Parc s'est portée vers la solution de mutualisation proposée par la Région et le CRIGE PACA, à savoir la plateforme IDGO. Le catalogage de données *via* GeoNetwork a permis, entre autre, de produire des guides techniques disponibles à cette adresse : https://github.com/GeoNetworkTeam. Une délibération de principe pour l'ouverture des données publiques du Parc a été rédigée et votée le 12 juillet 2017. A la suite de ce vote, le Parc a mis à disposition son premier jeu de données ouvert sur la plateforme OpenPACA. Le Parc du Verdon est un territoire pilote en matière d'Open Data pour les autres Parcs de PACA. Une des perspectives pour ces territoires serait d'aller vers le Web de données, où toutes les données des Parcs seraient liées.

Mots-clés : *Open Data, données publiques, métadonnées, droit des données, interopérabilité, GeoNetwork, catalogage, RDF, Web de données*

Géographique : *France, Provence-Alpes-Côte d'Azur, Parc naturel régional du Verdon*

ABSTRACT

The Verdon Regional Nature Park produces a wide range of data, and wishes to engage in a policy of open data. This project is in keeping with the main lines of the Park's charter and actions. The goal of the work placement is to develop an open data approach and set up a catalogue of data. It is thus a way to meet the requirements of today's legal context regarding Open Data and the standards - called INSPIRE - the Verdon Regional Park is subjected to concerning its geographical data. Inspired by the recommendations of the organisation OpenData France Occitanie, the methodology is adapted to a regional nature park's specific context and stakes. Several technical and strategic choices were made to carry out the project. The final strategy for disseminating data turned out to be the pooling solution proposed by the Region and the CRIGE PACA, namely the IDGO platform. Data cataloguing, *via* Geonetwork, has made it possible, among other things, to produce technical guides, available at this address: https://github.com/GeoNetworkTeam. A deliberation in principle was organised in respect of the opening of the Park's public data; a text was drafted and voted on 12th July 2017. Following this vote, the Park made its first dataset available on the OpenPACA platform. The Verdon Nature Park is a pilot territory regarding Open Data for other parks of the PACA region. These territories should keep in sight the prospect of turning to the data web, where all data from parks are to be connected in the future.

Keywords: *Open Data, public data, metadata, data law, interoperability, GeoNetwork, cataloguing, RDF, Web of Data*

Geographic: *France, Provence-Alpes-Côte d'Azur region, Verdon Regional Nature Park*

REMERCIEMENTS

A Nicolas Petit pour m'avoir donné l'opportunité de travailler sur ce projet et de m'avoir fait confiance tout au long du stage. Merci à lui pour ses conseils et ses idées.

A Wilfried Heintz pour avoir accepté d'encadrer mon travail et de m'avoir aidée durant mon stage, notamment sur les éléments techniques. Merci à lui pour son soutien.

A Lucien Duflocq pour avoir travaillé en équipe sur l'installation du catalogue de données GeoNework, ainsi que sur la rédaction de documents techniques.

A Pauline Asselin De Williencourt et à Thomas Bekkers pour m'avoir aiguillée dans la démarche Open Data et m'avoir permis de publier le premier jeu de données du Parc sur la plateforme régionale.

A Mathieu Kazmierski pour m'avoir permis de tester la validité des métadonnées créées *via* le catalogue de données installé en local au Parc du Verdon. Merci à lui et à l'équipe du CRIGE PACA pour leur accueil.

A Mathieu Andro pour avoir partagé ses codes sources sur les outils d'aide à la décision pour l'ouverture des données.

A Julien Dinterich, à Salah Eddine Aitahtman et à Josselin Janisset pour m'avoir donné des pistes en programmation informatique.

A Camille Monchicourt pour avoir partagé des informations sur l'Open Data et son actualité, ainsi que l'intérêt qu'il a porté pour ce projet.

A Fabien Moguen pour ses précieux conseils dans l'élaboration d'une démarche Open Data.

A Julie Delauge pour m'avoir appris les enjeux liés à l'ouverture des données naturalistes, ainsi que le fonctionnement du SILENE.

A Anne Beuzeval, Maxime Sormani et à Kévin Villena pour la relecture de mon rapport de stage.

A la Direction et à l'équipe technique du Parc du Verdon, en particulier :

A Audrey Zorzan pour m'avoir guidé au début de mon stage ;

A Marlène Economidès pour avoir relu et validé le texte publié sur la page Web partenaire de la plateforme régionale OpenPACA ;

A Corinne Gautier pour avoir pris le temps de valider les jeux de données à ouvrir du Pôle Eau ;

A Marjorie Salvarelli pour son jeu de données qui a été le premier à être ouvert au public.

SOMMAIRE

RESUME 1
ABSTRACT 1
REMERCIEMENTS 2
SOMMAIRE 3

INTRODUCTION 5

1. CONTEXTE ET OBJECTIFS DU STAGE 6
1.1. Qu'est-ce qu'un Parc naturel régional ? 6
1.1.1. Le Parc naturel régional du Verdon 6
1.1.2. Les missions d'un Parc naturel régional et les priorités du Verdon 7
1.1.3. Fonctionnement d'un Parc naturel régional 8
1.2. La place de la géomatique dans le Parc du Verdon 11
1.2.1. Historique du Système d'information territorial (SIT) 11
1.2.2. Vers une nouvelle structuration des données et un nouveau SIT 11
1.2.3. Vers l'Open Data interparc 13
1.3. Les besoins exprimés et les moyens 14
1.3.1. Les objectifs du stage 14
1.3.2. Le matériel 15
1.4. Organisation des tâches et répartition du temps de travail 16

2. ADMINISTRATION DES DONNEES : de la donnée brute à l'Open Data dans un Parc naturel régional 18
2.1. Le contexte Open Data 18
2.1.1. Qu'est-ce qu'une donnée ? 18
2.1.2. Qu'est-ce que l'Open Data ? 20
2.1.3. Un cadre juridique en évolution 22
2.2. Quels sont les enjeux de l'Open Data au sein du Parc du Verdon ? 24
2.2.1. Intérêts pour un Parc naturel régional 24
2.2.2. Risques potentiels pour un Parc naturel régional 26
2.2.3. Comment ouvrir les jeux de données ? 29
2.3. Expérimentations de l'Open Data – Parc du Verdon – 34
2.3.1. Evaluation de l'existant 34
2.3.2. Elaboration de la démarche Open Data à suivre 35

3. MISE EN PLACE DE L'OUTIL DE CATALOGAGE DES DONNEES 39
3.1. Interopérabilité des données 39
3.1.1. Qu'est-ce qu'une métadonnée ? 39
3.1.2. Normes et standards sur les données 40
3.1.3. La Directive INSPIRE et la cohérence avec l'Open Data 42
3.2. Solutions envisagées – GeoNetwork – 43
3.2.1. Les outils de catalogage 43
3.2.2. Choix de la solution GeoNetwork 46
3.3. Du catalogage au Web des données : le RDF 47
3.3.1. Installation et configuration de GeoNetwork 47
3.3.2. Développement de la chaîne de traitement pour un Web de données 48

4. BILAN ET PERSPECTIVES 52
4.1. Conclusion 52
4.2. Bilan personnel 53

ACRONYMES 54
BIBLIOGRAPHIE / WEBOGRAPHIE 55
LISTE DES FIGURES 59

INTRODUCTION

Le Parc naturel régional du Verdon produit une grande variété de données qui sont potentiellement réutilisables par les citoyens et les entreprises. La première étape vers cette réutilisation passe par la mise à disposition des données, à savoir l' « ouverture des données publiques » ou l' « Open Data ». C'est une démarche volontaire du Parc que d'aller vers l'ouverture de ses données publiques.

Depuis 2006, le Parc du Verdon s'est engagé dans le Système d'information territorial (SIT) des Parcs naturels régionaux de Provence-Alpes-Côte d'Azur (PACA). Il s'agit d'un dispositif technique et organisationnel qui permet de centraliser et de diffuser les données géographiques disponibles sur les territoires de quatre Parcs de la Région PACA (Luberon, Camargue, Queyras et Verdon). Le SIT a vocation à s'étendre aux autres Parcs de la Région. Son but est d'augmenter l'efficience de l'information produite sur les territoires en question, ainsi que de la porter à connaissance et la diffuser au plus grand nombre.

L'objectif principal du stage est d'aller dans le sens du SIT à savoir : susciter l'intérêt de la mise à disposition des données géographiques. Le but est aussi d'intéresser les utilisateurs aux autres types de données produites par le Parc. En découlent plusieurs enjeux comme l'élaboration de la démarche Open Data dans une réflexion collective interparc ; ou encore la détermination des obligations auxquelles le Parc serait soumis en termes d'accès à l'information géographique, notamment dans le cadre de la directive européenne INSPIRE. Le stage comporte donc deux volets :

1) Elaborer la démarche Open Data avec le Parc du Verdon comme territoire pilote des Parcs de PACA ;
2) Répondre aux normes INSPIRE en matière d'information géographique en mettant en place un catalogue de données.

Décrire et cataloguer les données du Parc sont une façon de répondre à la fois aux attentes de l'Open Data et de la Directive INSPIRE.

Aujourd'hui, de plus en plus de collectivités territoriales ouvrent leurs données publiques par obligation ou volontairement, mais peu nombreux sont les Parcs naturels régionaux qui s'y engagent. Comment mettre en place une démarche Open Data dans une structure productrice de données sensibles à caractère environnemental ? Et comment l'élaborer dans une réflexion collective interparc ? Il est donc important d'analyser le cadre juridique pour mieux connaître les droits et les devoirs en matière de diffusion de la donnée brute. D'un point de vue technique, il faut s'interroger également sur la façon de mettre à disposition les données brutes du Parc. Le choix technique en matière de catalogage de données et de leur diffusion est à prendre en considération pour que la démarche soit pérenne.

Dans un premier temps, le contexte du stage est posé pour mieux comprendre comment s'articule la démarche Open Data au sein du Parc du Verdon. Les besoins exprimés par la structure sont alors mentionnés, et expliquent en grande partie les choix pris pour répondre à la commande de départ. Dans un second temps, il est important de contextualiser l'Open Data du fait que le cadre juridique soit en constante évolution, depuis la loi pour une République numérique de 2016. Le but est de saisir les enjeux en matière d'ouverture des données publiques pour un Parc naturel régional. Ainsi, toute la démarche Open Data, expérimentée lors du stage, est retracée pour mettre en évidence les stratégies prises pour la diffusion des données. D'un point de vue technique, la troisième partie rappelle deux enjeux primordiaux pour répondre à la fois à la Directive INSPIRE et à la démarche Open Data, à savoir : la création des métadonnées et l'interopérabilité des données. Cette troisième partie montre comment répondre à ces deux enjeux, avec notamment la mise en place d'un catalogue de données, GeoNetwork. Enfin, pour aller plus loin dans cette réflexion, une perspective serait d'aboutir à un Web de données.

1. CONTEXTE ET OBJECTIFS DU STAGE

1.1. Qu'est-ce qu'un Parc naturel régional ?

Un Parc naturel régional (PNR)[1] est un territoire rural reconnu par l'Etat, correspondant à une entité géographique qui s'organise autour d'un projet concerté de développement responsable. Le territoire d'un Parc correspond aux limites des communes adhérentes à un document formel appelé la « Charte »[2]. Les collectivités locales et l'Etat se sont engagés dans une politique contractuelle innovante et dynamique, associant protection du patrimoine et développement local. C'est donc en 1967 que sont nés les Parcs naturels régionaux créés par la Délégation à l'Aménagement du Territoire et à l'Action Régionale (DATAR) en tant qu' « outils » d'aménagement du territoire. Aujourd'hui, 51 Parcs sont répartis sur l'ensemble du territoire français (métropole et outre-mer) dont 7 en région Provence-Alpes-Côte d'Azur (PACA).

1.1.1. Le Parc naturel régional du Verdon

Le Parc naturel régional du Verdon compte parmi les 7 Parcs (Alpilles, Baronnies provençales, Camargue, Luberon, Préalpes d'Azur, Queyras, Verdon) et les 2 Parcs en préfiguration (Mont-Ventoux, Sainte-Baume) de la région PACA. Cette continuité géographique, unique en France, qui s'étend sur près du tiers du territoire régional, témoigne de la richesse des patrimoines naturels et culturels de la Région PACA et de la volonté de les préserver.

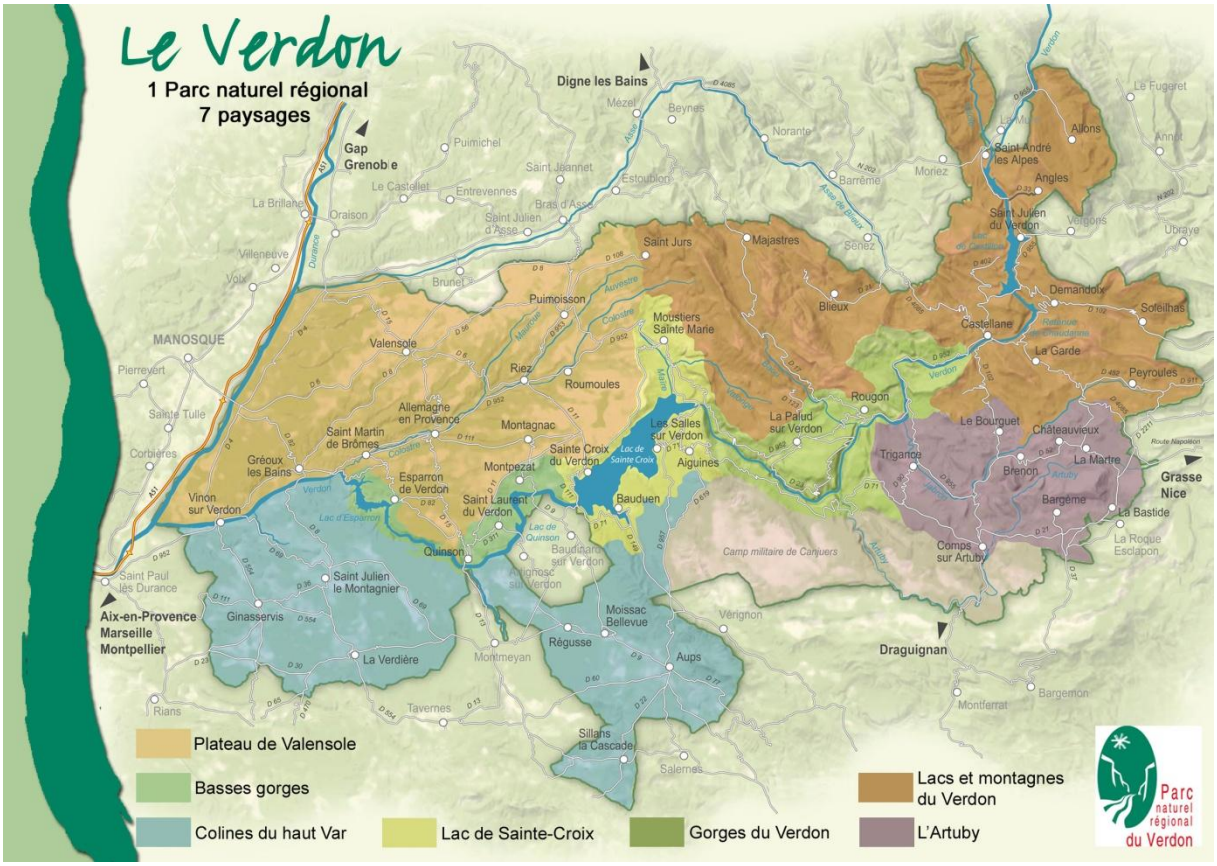


Figure 1 Le territoire du Parc naturel régional du Verdon et ses paysages Source : PNRV, 2014

[1] Dans ce rapport, l'acronyme « PNR » est substitué volontairement par le terme « Parc » (choix préconisé par Marlène Economidès, chargée de communication du Parc du Verdon).

[2] Formalisation des objectifs de préservation et de développement définis par les gestionnaires du territoire.

Créé en 1997, le Parc naturel régional du Verdon s'étend sur une surface de 188 000 hectares. A cheval sur les départements des Alpes-de-Haute-Provence et du Var, il regroupe environ 34 000 habitants répartis sur 46 communes (Figure 1).

Les Parcs de la Région PACA travaillent en réseau depuis des années sur plusieurs sujets tels que le tourisme durable, l'énergie ou l'agriculture. Les Parcs sont des territoires qui allient protection des patrimoines et développement économique et social.

1.1.2. Les missions d'un Parc naturel régional et les priorités du Verdon

Les missions confiées aux Parcs sont les suivantes :

- ✓ Inventorier, mettre en valeur et protéger les patrimoines naturels et culturels, en particulier ceux dont l'équilibre est fragilisé ou menacé et ce par une gestion adaptée ;
- ✓ Contribuer au développement économique et social du territoire en compatibilité avec sa mission de protection ;
- ✓ Assurer un accueil, une information et une éducation de qualité auprès des différents publics : les résidents, les élus, les visiteurs ;
- ✓ Développer des actions innovantes et expérimentales dans tous les domaines afin de faire du Parc un laboratoire de tout instant ;
- ✓ Contribuer de façon originale et consensuelle à l'aménagement du territoire en zone rurale.

Dans la Charte du Parc du Verdon (2008-2020), **trois objectifs prioritaires** y sont définis :

- ✓ **La ressource en eau du Verdon :** la mission consiste à mobiliser de nombreux partenaires concernés par cette ressource à l'échelle du bassin versant pour une gestion exemplaire ;
- ✓ **La transmission des patrimoines :** la mission a pour but de préserver la diversité biologique et l'originalité des paysages pour en faire un atout du cadre de vie ainsi que pour l'activité touristique. La gestion concertée de l'eau sera pleinement intégrée à la politique patrimoniale du Parc ;
- ✓ **La gestion des ressources :** la mission vise un meilleur équilibre et une meilleure répartition entre les activités agricoles, forestières et touristiques dans l'espace et dans la durée. Faire en sorte que ces activités soient plus respectueuses de l'environnement sur lequel elles se développent. Le Parc accorde une importance primordiale à la gestion économe de l'espace, dans un contexte de pressions foncières accrues.

Le Parc souhaite que le plus grand nombre s'approprie les objectifs de la Charte (2008-2020), l'idée étant de faire de la démocratie participative un enjeu central. Pour atteindre ces objectifs, le Parc s'appuie sur son équipe technique et ses partenaires. Saisir le fonctionnement d'un Parc naturel régional est important lorsqu'il s'agit, entre autre, d'administrer ses données.

Que contient la Charte d'un Parc ?

La politique menée par un Parc naturel régional est définie dans sa Charte (renouvelée tous les 12 ans), véritable feuille de route fixant :

- × Les objectifs à atteindre ;
- × Les orientations de protection, de mise en valeur et de développement du Parc ;
- × Les règles du jeu pour mettre en œuvre ce projet ;
- × Un plan d'orientations selon les vocations des différentes zones du Parc ;
- × Les statuts de l'organisme de gestion du Parc ;
- × La marque du Parc, constituée de l'idéogramme et de la dénomination du Parc.

(Source : « A la découverte des Parcs naturels régionaux », Région PACA, juin 2011, p. 68)

1.1.3. Fonctionnement d'un Parc naturel régional

Un Parc naturel régional rassemble un gestionnaire, une équipe technique pluridisciplinaire, des outils d'intervention et un budget de fonctionnement. Pour mettre en œuvre la Charte, un Syndicat Mixte est créé. Ce dernier regroupe, dans le Verdon, les 46 communes adhérentes, les conseils départementaux des Alpes-de-Haute-Provence et du Var et le conseil régional PACA. Un Parc fonctionne comme une collectivité locale selon les principes du service public. L'administration d'un Parc passe par des élus réunis en comité syndical et bureau, qui s'appuient sur des instances de proposition ou consultatives (Figure 2). Les habitants et/ou les visiteurs du territoire sont impliqués dans les actions du Parc. Pour fonctionner, ce dernier mobilise des moyens financiers (subventions, cotisations des membres) et s'appuie sur une équipe technique[3].

1.1.3.1. Les instances d'un Parc naturel régional

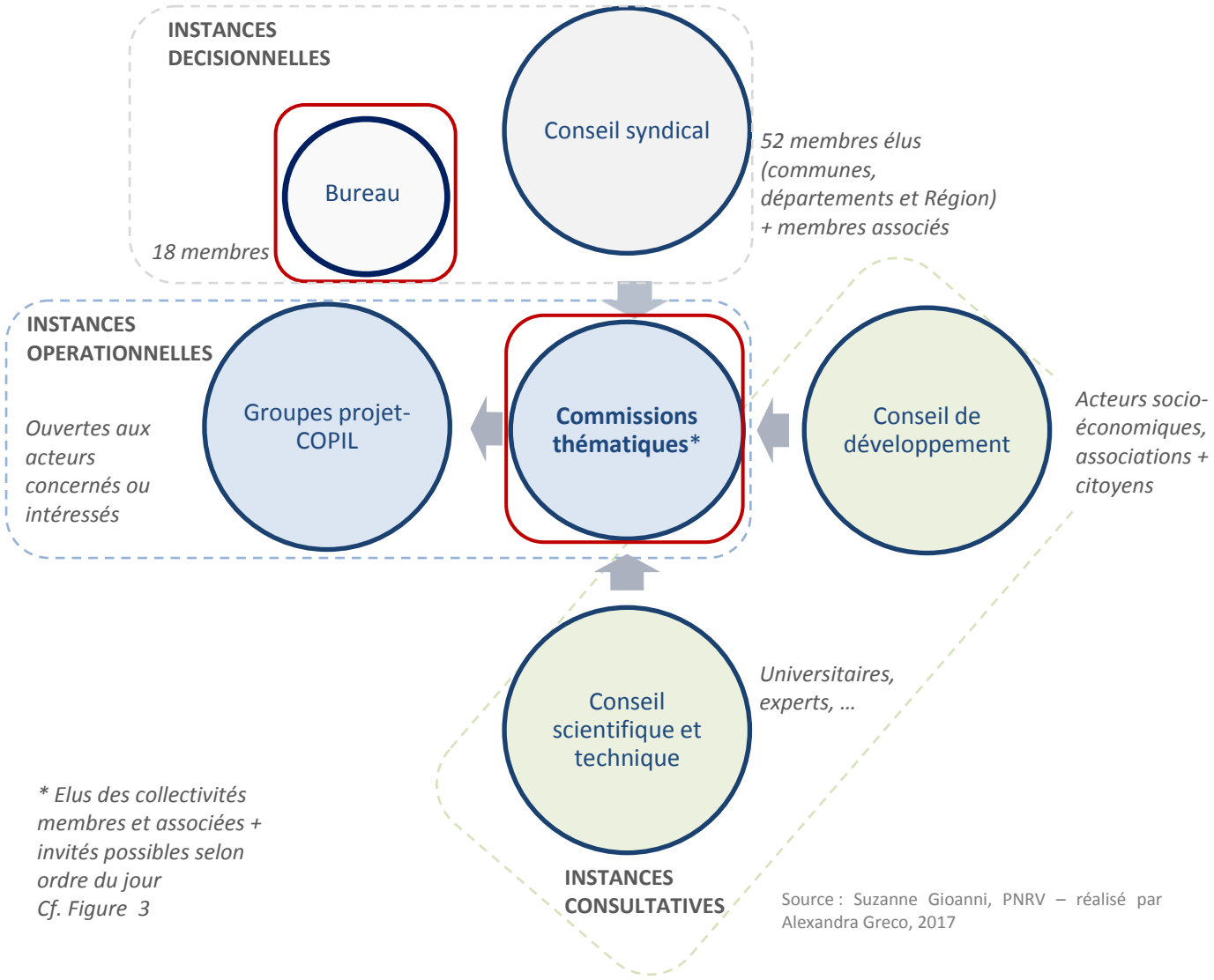
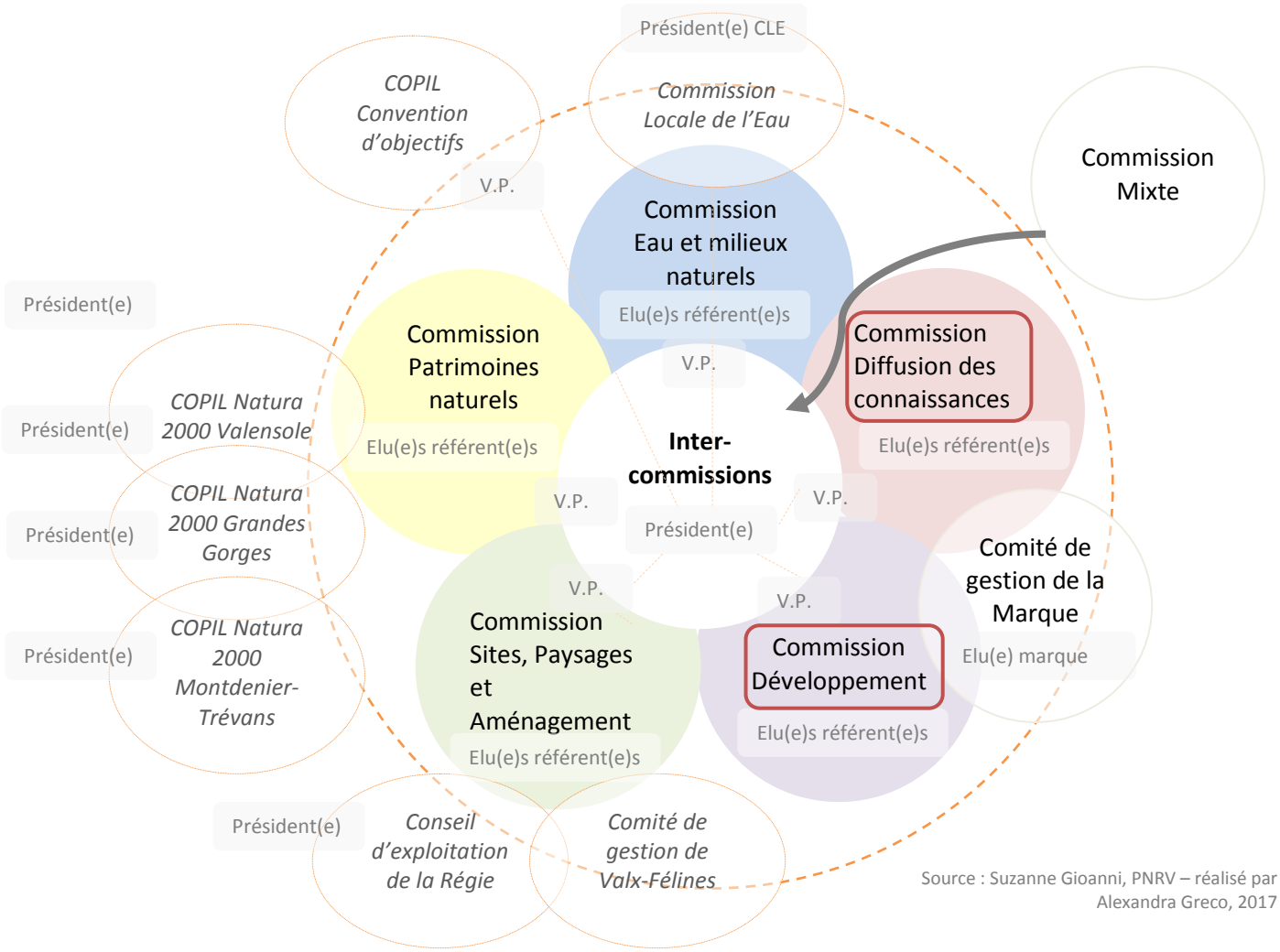


Figure 2 Schéma des instances du Syndicat Mixte de gestion du Parc du Verdon

[3] Cf. organigramme du Parc du Verdon



Source : Suzanne Gioanni, PNRV – réalisé par Alexandra Greco, 2017

Figure 3 Schéma des commissions et autres instances opérationnelles du Syndicat Mixte de gestion du Parc du Verdon

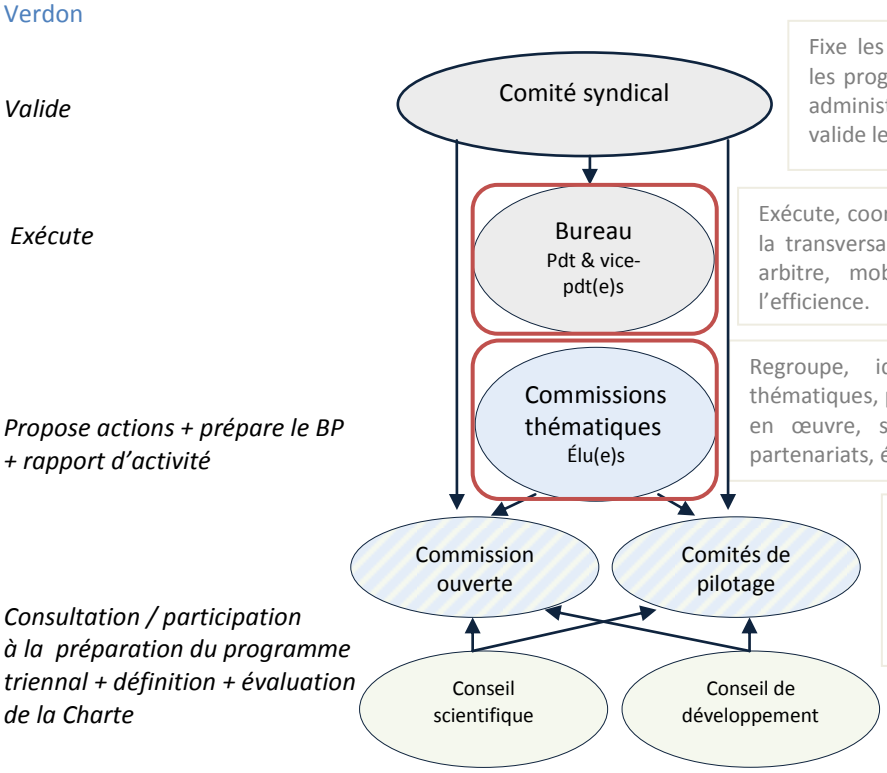


Figure 4 Répartition des tâches selon les instances

Source : Suzanne Gioanni, PNRV – réalisé par Alexandra Greco, 2017

1.1.3.2. La place de l'Open Data au sein des instances du Parc du Verdon

Dans le cadre de mon stage, des réunions de sensibilisation sur l'Open Data ont été organisées avec quasiment chaque Pôle du Parc[4] :

Nom des Pôles	Réunion(s) effectuée(s)	Equipe au complet	Remarques	Compte-rendu produit
Diffusion	✓	✓		✓
Développement	✓	×	Seule une personne manquait	✓
Administration	✓	×	RDV avec la responsable du Pôle	✓
Patrimoine naturel	✓	✓		✓
Sites, paysages et aménagement du territoire	×	×	RDV avec une personne référente sur une donnée précise	×
Eau	✓	✓		✓

Source : Alexandra Greco, 2017

Figure 5 Tableau récapitulatif des réunions de Pôle au Parc du Verdon - sensibilisation Open Data -

Comme il est indiqué dans le tableau ci-dessus, un compte-rendu pour chacune des réunions (Pôles, commissions, …), à l'exception d'une, a été réalisé et transmis au Parc du Verdon pour mettre en évidence certaines perspectives[5] en matière d'ouverture des données selon le Pôle.

Ma participation à un comité de direction (CODIR) a permis également de faire part, auprès des co-directeurs du Parc du Verdon, de la démarche Open Data à suivre. Dans le but de sensibiliser les élu(es) et le Président du Parc du Verdon, j'ai participé à deux instances : opérationnelle (commission Diffusion des connaissances) et décisionnelle (Bureau : vote et signature de la délibération de principe pour l'ouverture des données publiques du Parc du Verdon)[6].

La légende ci-dessous renvoie aux Figure 2 / Figure 3 / Figure 4. L'idée est de montrer comment s'est inscrite ma démarche à travers l'organisation d'un Parc naturel régional.

Place de l'Open Data au niveau des instances du Parc du Verdon

Mon travail étant à présent défini par rapport aux instances du Parc, il s'agit de situer ma démarche dans le contexte géomatique du Parc. En effet, les enjeux de la géomatique dans le Parc du Verdon sont à prendre en compte afin d'articuler au mieux la problématique Open Data avec une réflexion collective interparc.

[4] Les méthodes de travail concernant la mise en place d'une démarche Open Data sont détaillées dans la Partie 2 de ce rapport de stage.

[5] Cf. 2.3

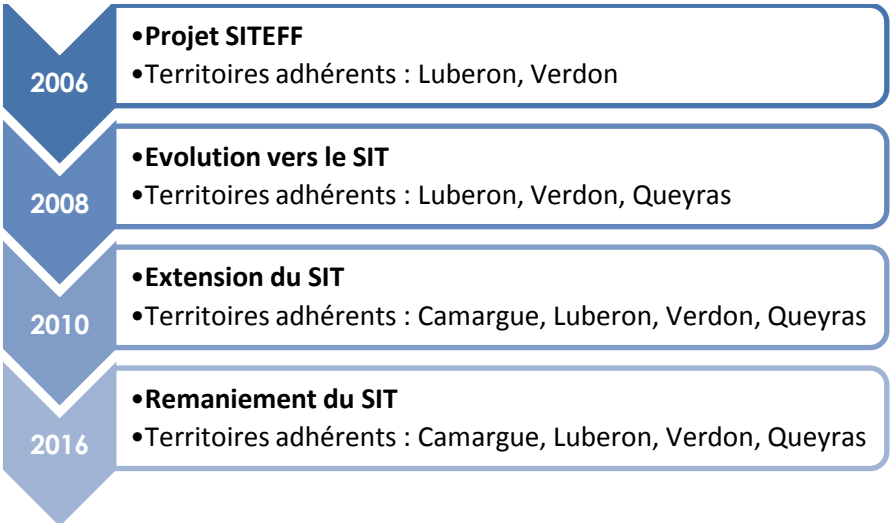
[6] Cf. Figure 10

1.2. La place de la géomatique dans le Parc du Verdon

Les Parcs naturels régionaux de PACA ont une solution commune quant à l'organisation et la mise à disposition des données géographiques : le Système d'information territorial (SIT). Il s'agit d'une plateforme mutualisée pour renforcer la visibilité de l'information géographique produite au niveau des Parcs, de l'adapter au plus grand nombre et de faire du SIT un outil d'aide à la décision. Le SIT a également pour ambition d'entrer en conformité avec les standards établis par l'*Open Geospatial Consortium* (OGC) pour faciliter les échanges.

1.2.1. Historique du Système d'information territorial (SIT)

Le projet de création du SIT a vu le jour en 2006 à l'initiative des Parcs du Luberon et du Verdon. L'intention de départ était de répondre au projet SITEFF[7], à savoir : créer une base de données relative à la faune et à la flore. Dès 2008, le projet a pris de l'ampleur : l'ensemble des données des Parcs est concerné, qu'il s'agisse de données géographiques, alphanumériques ou documentaires (liens vers des sites Internet, documents numériques au format .PDF, .DOC, etc.).



Source : Interparc PACA, 2017
Charte Géomatique SIT-PNR-PACA (Muriel KREBS, PNRL)

Figure 6 Evolution du SIT depuis sa création

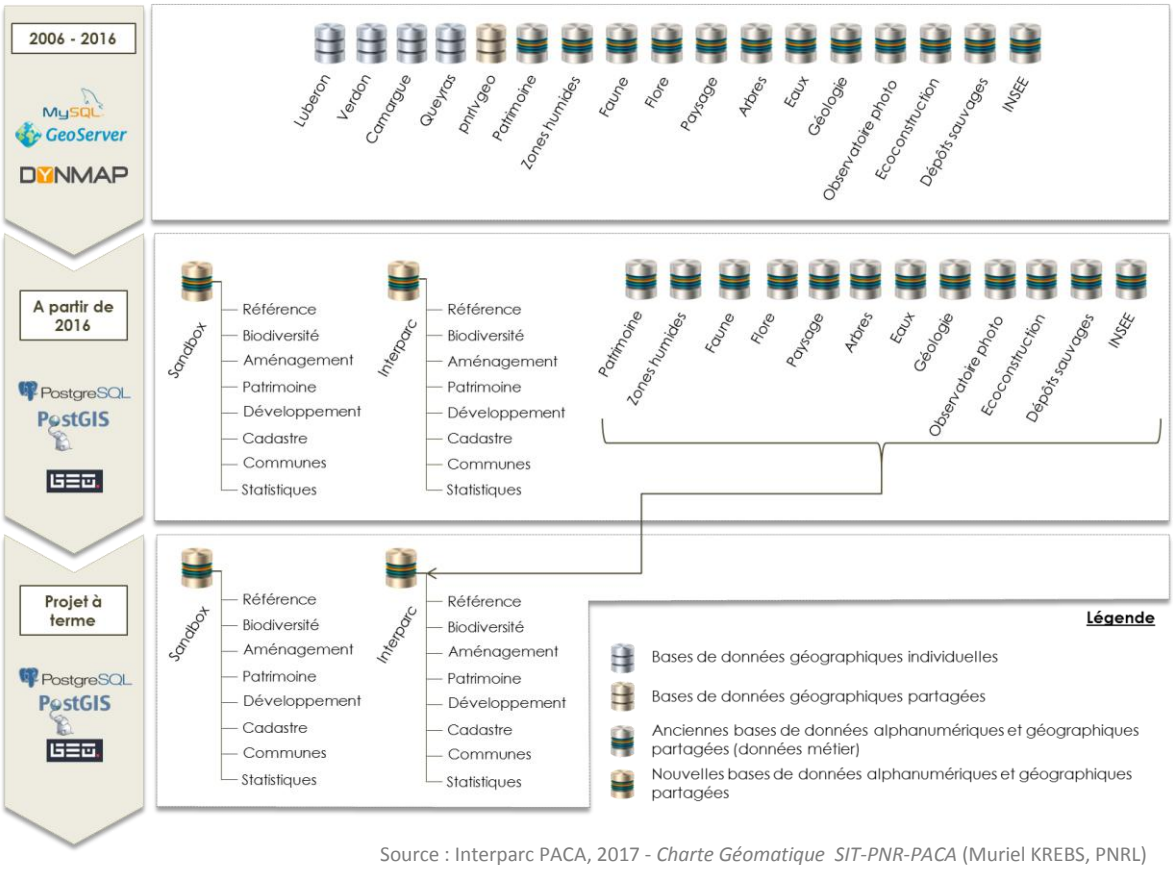
Si le SIT évolue constamment, il se compose désormais d'une infrastructure commune organisée en différentes bases de données partagées (Figure 7). Ses fonctionnalités sont adaptées aux besoins communs ou particuliers des Parcs de PACA. Grâce à une plateforme Internet, le contenu du SIT est diffusé, ce qui permet à chacun d'en prendre connaissance ou de l'enrichir en fonction des niveaux d'accès différenciés (Intranet, Extranet, Internet). Ce SIT permet aux Parcs de répondre à des exigences liées à la transposition de la directive européenne 2007/2/CE du 14 mars 2007, dite Directive INSPIRE[8], et à la mise en place de ses services. En basculant leur base de données du SIT vers PostGRES/PostGIS, les Parcs de PACA entre ainsi en conformité avec les normes OGC.

1.2.2. Vers une nouvelle structuration des données et un nouveau SIT

Depuis près d'un an, l'infrastructure de données spatiales (IDS) évolue. Toutes les données communes aux membres de l'Interparc sont rassemblées dans une seule base de données « interparc » (Figure 7). Cette dernière peut contenir des données spécifiques à certains Parcs

[7] Système d'Information Territorial pour l'Etude de la Faune et de la Flore
[8] *Infrastructure for Spatial Information in the European Community* – Cf. 3.1.3

adhérents lorsque cela relève d'un besoin particulier. Par exemple : les indicateurs de suivi sur les niveaux d'eau pour le Parc de Camargue. Mutualiser toutes ces données interparc a pour but de fournir une information géographique la plus homogène et cohérente possible sur l'ensemble des territoires des Parcs. Dans cette optique, une Charte Géomatique Interparc a été rédigée afin de poser les différentes règles communes que doivent adopter les Parcs adhérents au projet SIT (règles de nommage, nomenclature des tables de la base de données, formats des données etc.).



Source : Interparc PACA, 2017 - *Charte Géomatique SIT-PNR-PACA* (Muriel KREBS, PNRL)

Figure 7 Evolution de l'infrastructure de données spatiales

Les données du SIT sont hébergées sur un serveur appartenant aux quatre Parcs membres de l'Interparc (Figure 8). Il faut savoir aussi que l'outil de gestion de base de données MySQL a été remplacé par PostgreSQL/PostGIS (PgAdminIII). Quant à l'affichage des données et à la cartothèque, ils sont gérés par l'outil Web SIG[9] GEO (Figure 8). Cet outil est né de la fusion entre DynMap et le logiciel Aigle. De plus, en 2015, le CRIGE[10] PACA a été sollicité par les Parcs pour que leurs métadonnées[11] soient moissonnées par le catalogue de données du CRIGE. Mais la bascule DynMap vers GEO a rendu la situation complexe du fait que l'infrastructure ait changé. Le catalogue de données étant un service optionnel (payant) proposé par GEO, les Parcs se sont organisés pour installer un GéoSource[12] sur le serveur interparc. A cause du dysfonctionnement de ce catalogue de données, les objectifs du stage ont quelque peu évolués (Cf. 1.3.1). Le schéma ci-dessous représente l'architecture du Système d'information (SI) du Parc du Verdon tel qu'il est aujourd'hui.

[9] Système d'information géographique
[10] Centre régional de l'information géographique.
[11] Littéralement, cela correspond aux données sur les données (fiche descriptive de la donnée).
[12] Version française de GeoNetwork (catalogue de données en *open source*).

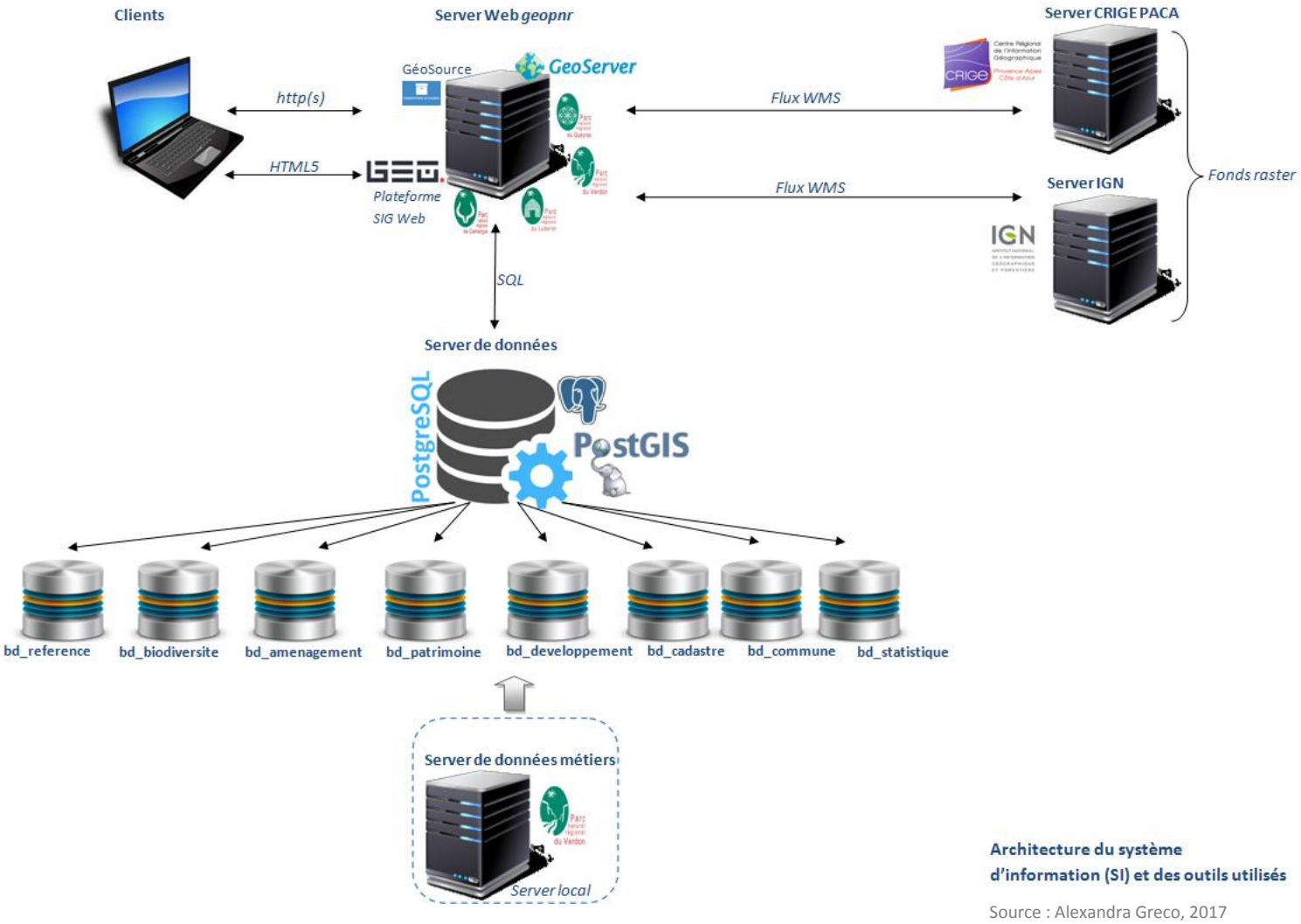


Figure 8 Architecture du système d'information et des outils utilisés par le Parc du Verdon

1.2.3. Vers l'Open Data interparc

De la Charte (2008-2020) du Parc du Verdon à la Charte Géomatique Interparc en passant par le SIT, l'objectif commun est d'augmenter l'efficience de l'information produite sur les territoires des Parcs, de la porter à connaissance et de la diffuser au plus grand nombre. Cette idée rejoint celle d'une réflexion collective interparc sur l'ouverture des données publiques (Open Data) des Parcs de PACA. Avec la nouvelle structuration des données interparc, la refonte du SIT et la volonté de répondre à la Directive INSPIRE, la mise à disposition des données géographiques est en cours. Le Parc du Verdon devient un territoire d'expérimentations dans le cadre d'une démarche Open Data pour tous types de données.

La démarche Open Data à mettre en place au sein du Parc du Verdon doit être reproductible pour les autres Parcs de PACA, notamment pour les adhérents du SIT qui partagent déjà des règles communes de mutualisation et d'homogénéisation des données. La mission phare du stage est donc d'élaborer cette démarche d'ouverture des données publiques de façon globale. Les besoins exprimés par le Parc du Verdon sont explicites concernant ce projet.

1.3. Les besoins exprimés et les moyens

Si un dispositif technique de diffusion et de partage en ligne des données géographiques existe à travers le SIT, le Parc du Verdon souhaite initier une démarche d'ouverture de ses données publiques au sens large (Open Data), à savoir : la mise à disposition et la réutilisation possible de tout type de données. L'élaboration d'une démarche Open Data dans une réflexion interparc constitue le cœur du stage. La dimension « INSPIRE » pour les données géographiques, avec la création de métadonnées et la mise en place d'un catalogue de données, fait partie intégrante de cette démarche Open Data. En effet, lorsqu'on sait qu'une étape cruciale de l'Open Data est de faire en sorte que les données soient diffusables et réutilisables, cette dimension « INSPIRE » est exemplaire.

1.3.1. Les objectifs du stage

Les objectifs du stage au sein du Parc du Verdon sont de :

1) Réaliser un état des lieux des sources de données disponibles (gisement de données) dans les archives numériques, sur la plateforme SIT et les outils connexes ;
2) Déterminer les risques potentiels de l'ouverture des données publiques (notamment les données « sensibles » à caractère environnemental) ;
3) Formaliser les problématiques auxquelles le Parc est confronté et doit répondre pour aider les élus du Parc à faire les choix stratégiques pour la démarche Open Data ;
4) Élaborer des recommandations pour citer, présenter et diffuser les données : licence, cadre juridique… ;
5) A partir de cet état des lieux, définir un répertoire d'informations publiques du Parc à défaut son architecture ;
6) Faire un point sur les formats et standards existants et émergents ;
7) Définir les possibilités de diffusion des données (notamment sur les plateformes de données publiques régionales, nationales, …) ;
8) Participer aux développements informatiques nécessaires à la mise en place des bases de données et outils pour l'ouverture de données - celles à composante géographique - en installant GéoSource sur le serveur interparc ;
9) Spécifier les besoins pour l'ouverture des données pertinentes en Open Data, statiques ou dynamiques en coordination avec le chargé de projet géomatique et évaluation.

A l'entrée de mon stage, l'objectif n°8 était accompli à moitié puisqu'un catalogue GéoSource avait déjà été installé sur le serveur interparc. Mon but était alors de paramétrer GéoSource pour cataloguer les premières métadonnées. Le problème relatif aux droits d'accès/d'administration se posait déjà. Mi-avril, l'objectif n°8 a changé à cause d'un dysfonctionnement rencontré avec GéoSource. N'ayant pas les droits d'administration, le problème n'a pas pu être résolu directement *via* le serveur. Il a fallu trouver une nouvelle solution pour répondre à cet objectif qui est de mettre en place un catalogue de données pleinement opérationnel. Plusieurs solutions ont été envisagées en fonction du matériel mis à disposition lors du stage (Cf. 3.2).

1.3.2. Le matériel

Initialement je disposais d'un ordinateur sur lequel aucun droit d'administration n'était accordé :

Nom du périphérique : EDUC-CZC0367WYN
Nom du domaine : pnrverdon.loc
Edition Windows: Windows 7 Professionnel
Mémoire installée (RAM): 3, 00 Go (2, 87 Go utilisable)
Processeur : Intel® Core(TM) i5 CPU 650@ 3.20GHz
Système d'exploitation : 64 bits
Disque : 78 Go (6,30 Go de libre)

Cet ordinateur m'a permis exclusivement de réaliser un travail de bureautique.

A la suite du changement d'objectif (n°8), un deuxième ordinateur (système d'exploitation : *Windows Vista*) m'a été confié. L'idée était de pouvoir réaliser des tests d'installation de catalogage de données en local, avec le même environnement que sur le serveur interparc, à savoir sous Debian 9. Pour réaliser ce travail, je l'ai donc entièrement réinstallé[13], à deux reprises, à cause d'une surchauffe du processeur (44°C/111°F) : problème lié à la température ambiante. Ci-dessous la description de cet ordinateur :

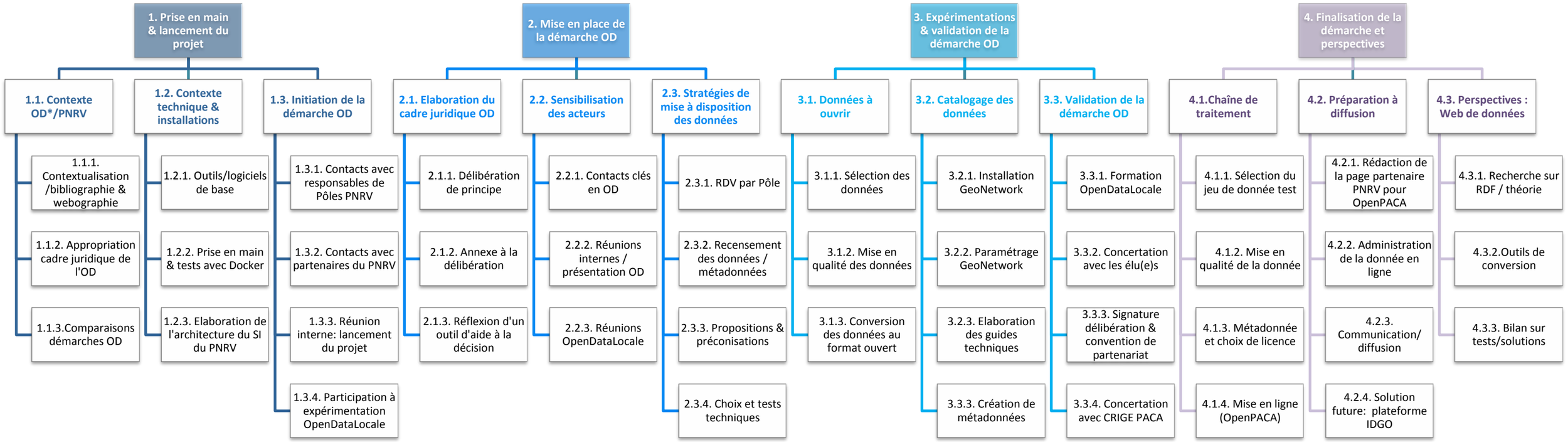
GNOME version 3.22.2
Nom du périphérique : sig
Mémoire : 2,00 Go
Processeur : Intel Pentium (R) 4 CPU 3.00Ghz × 2
Carte graphique : Intel 945G ×86/MMX/SSE2
Distribution : Debian GNU/Linux 9 (stretch) 32 bits
Système d'exploitation : Debian 9 « Stretch », i3861
Disque : 76,1 Go

La performance du matériel et l'absence de droits d'administration ont constitué un frein pour la mise en place d'un catalogue de données. A deux reprises, le système d'exploitation Debian 9 et le catalogue de données GeoNetwork ont été réinstallés suite à des problèmes de mémoire vive et de surchauffe du processeur. Cette perte de temps a eu des conséquences sur l'expérimentation de conversion des données au format RDF (Cf. 3.3), notamment pour aller plus loin dans la démarche Open Data.

Largement autonome dans mon travail, l'organisation des tâches à effectuer pendant les 6 mois de stage était importante pour garantir une cohérence dans la mise en place de la démarche Open Data.

[13] Cf. p. 34 du Guide d'installation de GeoNetwork

1.4. Organisation des tâches et répartition du temps de travail



*OD = Open Data

Figure 9 Organigramme des tâches

Source : Alexandra Greco, 2017

	Date de début	Date de fin
1. Prise en main	13/03/17	17/05/17
Contextualisation, appropriation de la bibliographie	13/03/17	13/04/17
Appropriation du cadre juridique de l'Open data	13/03/17	31/03/17
Comparaison des démarches Open data	13/03/17	13/04/17
Prise de contacts	20/03/17	13/04/17
Installation des logiciels "basiques"	20/03/17	20/03/17
Réunion interne de lancement du projet	22/03/17	22/03/17
Recherche de solutions techniques	24/03/17	13/04/17
Consultation des données	27/03/17	13/04/17
Réunion d'information et de sensibilisation par OpenPACA (OpenDataLocale)	05/04/17	05/04/17
Contacts avec les partenaires	05/04/17	12/04/17
Prises de RDV avec les responsables des pôles	06/04/17	17/05/17
RDV Pôle Diffusion (1)	13/04/17	13/04/17
Tests d'installation sur Docker	10/04/17	28/04/17
2. Mise en place de la démarche	14/04/17	09/06/17
Rédaction de la délibération de principes et validation	14/04/17	04/05/17
Rédaction de l'annexe à la délibération	05/05/17	31/05/17
Définition des stratégies de mise à disposition des données	17/04/17	31/05/17
RDV Pôle Diffusion (2)	20/04/17	20/04/17
Finalisation de l'état de l'art Open data	19/04/17	21/04/17
Préparation de la réunion de direction (CODIR)	24/04/17	24/04/17
CODIR : sensibilisation auprès des responsables des Pôles du PNRV	25/04/17	25/04/17
Recensement des données publiques à ouvrir	26/04/17	09/06/17
RDV avec Fabien Moguen - OpenDataLab -	26/04/17	26/04/17
Installation de Debian 9	02/05/17	02/05/17
RDV Pôle Patrimoine naturel (1)	04/05/17	04/05/17
Réunion au Conseil Départemental (04) - SDTSUN -	11/05/17	11/05/17
1er essai d'installation de GeoNetwork	03/05/17	08/05/17
Réunion OpenDataLocale (04)	19/05/17	19/05/17
RDV Pôle Eau (1)	22/05/17	22/05/17
Réparation - Réinstallation de Debian 9	23/05/17	24/05/17
RDV avec Julie Delauge - CEN PACA -	24/05/17	24/05/17
RDV Pôle Administration (1)	30/05/17	30/05/17
RDV Pôle Développement (1)	30/05/17	30/05/17
3. Préparation des jeux de données et des solutions techniques	01/06/17	10/08/17
Sélection des jeux de données à ouvrir	01/06/17	13/07/17
Mise en qualité et en format des données	06/06/17	14/07/17
2ème essai d'installation de GeoNetwork	07/06/17	07/06/17
Formation OpenDataLocale (13)	09/06/17	09/06/17
Paramétrage de GeoNetwork	12/06/17	05/07/17
Commission Diffusion des connaissances	22/06/17	22/06/17
Rédaction d'un guide d'installation pour GeoNetwork	13/06/17	11/07/17
Paramétrage et création de métadonnées	23/06/17	10/08/17
Rédaction d'un guide de paramétrage GeoNetwork	23/06/17	12/07/17
Conception du MCD GeoNetwork	03/07/17	03/07/17
Synthèse de l'état d'avancement	04/07/17	06/07/17
Réunion du Pôle Diffusion	04/07/17	04/07/17
Réunion PNR (SIT) au CRIGE PACA	07/07/17	07/07/17
Bureau du Parc - délibération sur l'Open Data	12/07/17	12/07/17
Validation de la délibération + convention de partenariat avec la Région	19/07/17	19/07/17
4. Mise à disposition des jeux de données	13/07/17	29/08/17
Rédaction et validation du descriptif du PNRV pour OpenPaca	13/07/17	13/07/17
Chaîne de traitement pour ouvrir les données	13/07/17	26/07/17
Tests de diffusion sur la plateforme choisie	14/07/17	29/08/17
Conception d'un outil d'aide à la décision	17/07/17	04/08/17
Tests de conversion des données en RDF	31/07/17	29/08/17
Rédaction du rapport	13/07/17	31/08/17
Soutenances	13/09/17	15/09/17

Figure 10 Diagramme de Gantt

Source : Alexandra Greco, 2017, Gantt Project

2. ADMINISTRATION DES DONNEES : de la donnée brute à l'Open Data dans un Parc naturel régional

2.1. Le contexte Open Data

Pour répondre aux objectifs du stage, il m'a fallu cerner ce qu'est l'Open Data, le contexte actuel et son cadre juridique, notamment par rapport à un Parc naturel régional. Nombreuses sont les collectivités territoriales qui ont adoptée une démarche Open Data, mais peu sont concernées par des données sensibles à caractère environnemental. Il est donc important de présenter ce contexte et les enjeux pour comprendre ma démarche dans la mise en place de l'Open Data au sein du Parc du Verdon. Les expérimentations concernant cette démarche sont ensuite exposées et révèlent ainsi le management d'un tel projet, avec ses atouts et ses limites.

2.1.1. Qu'est-ce qu'une donnée ?

Une confusion est souvent faite dans le langage courant entre les « données publiques » et un «ensemble des données accessibles en ligne ». Mais la politique d'ouverture et de partage des données publiques est bien fondée sur la loi sur l'accès aux documents administratifs (CADA) et sur la directive européenne sur les informations du secteur public (PSI). Afin de poser au mieux le contexte Open Data, revenons sur la définition d'une donnée[14].

2.1.1.1. Donnée brute

Une donnée brute est une donnée non analysée, qui n'a subi aucun traitement au préalable. La donnée brute correspond à la matière première à partir de laquelle on peut retirer des connaissances. Lorsque la donnée est contextualisée, il s'agit d'une information. La collecte et la présentation des données constituent l'information. Et quand cette dernière est interprétée et utilisée, il s'agit d'une connaissance. Acquérir des connaissances c'est transformer l'information en choix. Dans le cas de l'Open Data, il s'agit bien de données numériques, à ne pas confondre avec une information numérisée. Cette dernière peut être copiée/collée mais elle n'est pas exploitable directement dans un logiciel, tandis qu'une donnée numérique peut être traitée automatiquement. Mais qu'est-ce qui rend les données « ouvertes » ?

2.1.1.2. Donnée ouverte

Une donnée ouverte est la traduction littérale de l' « open data ». Les données sont ouvertes si tout le monde peut y accéder, les utiliser et les partager. Il faut limiter au maximum les restrictions quant à l'utilisation des données pour qu'elles soient ouvertes. N'importe quel utilisateur doit être libre de modifier, de combiner et de partager les données, même à des fins commerciales. Si l'on souhaite restreindre cet usage commercial, il est possible *via* une licence apposée à un jeu de données de cadrer l'utilisation de la donnée en y spécifiant certaines clauses. La licence fixe les conditions dans lesquelles les données peuvent être utilisées, partagées etc. (Cf.2.2.3.2). Techniquement, la donnée doit être livrée dans un format ouvert c'est-à-dire qui peut être lu par tout logiciel standard. Il s'agit souvent de formats associés à des logiciels open source[15] (.CSV, .JSON,

[14] Pour plus de détails concernant la définition d'une donnée, retrouvez un état de l'art en annexe (n°1).
[15] « Un logiciel *open source* correspond à un programme informatique dont le code source est distribué sous une licence. Cette dernière permet à quiconque de lire, modifier ou redistribuer ce logiciel ».
(Source : https://fr.wikipedia.org/wiki/Open_Source_Definition).

.RDF etc.). Produite par des acteurs publics ou par des acteurs privés dans le cadre de gestion de service public, la donnée ouverte respecte le principe de la loi CADA[16] (Cf. 2.1.3).

2.1.1.3. Donnée publique

La donnée publique est une information contenue dans tout document produit ou reçu par une personne agissant *« dans le cadre [d'une] mission de service public ».* (Etat, collectivité territoriale, établissement public administratif ou personne privée (Loi CADA, art. 1). La loi CADA ne parle pas de données publiques mais de « documents administratifs » ou d'« informations publiques » (Loi n°78-753 du 17/07/1978). Etant donné qu'il n'y a pas de définition de « donnée publique » de façon juridique, il faut se référer donc au terme de « document administratif ». A partir du moment où une personne agit dans le cadre d'une mission de service public, alors tous les documents qu'elle produit ou qu'elle reçoit sont publics. D'après l'article 10 de la loi CADA, étant donné qu'un document administratif contient des informations (ici dans le sens de données), les documents administratifs contiennent donc tous des données publiques. Leur format, leur support ou leur conservation importent peu. Donc si l'on considère les données publiques comme ouvertes, cela revient à une définition de données produites ou détenues par une administration qui entrent dans le champ des données publiques et qui doivent être partagées, gratuitement et librement réutilisables. Ceci étant, l'ouverture des données publiques ne s'applique pas à tous les documents administratifs comme l'expose le schéma ci-dessous :

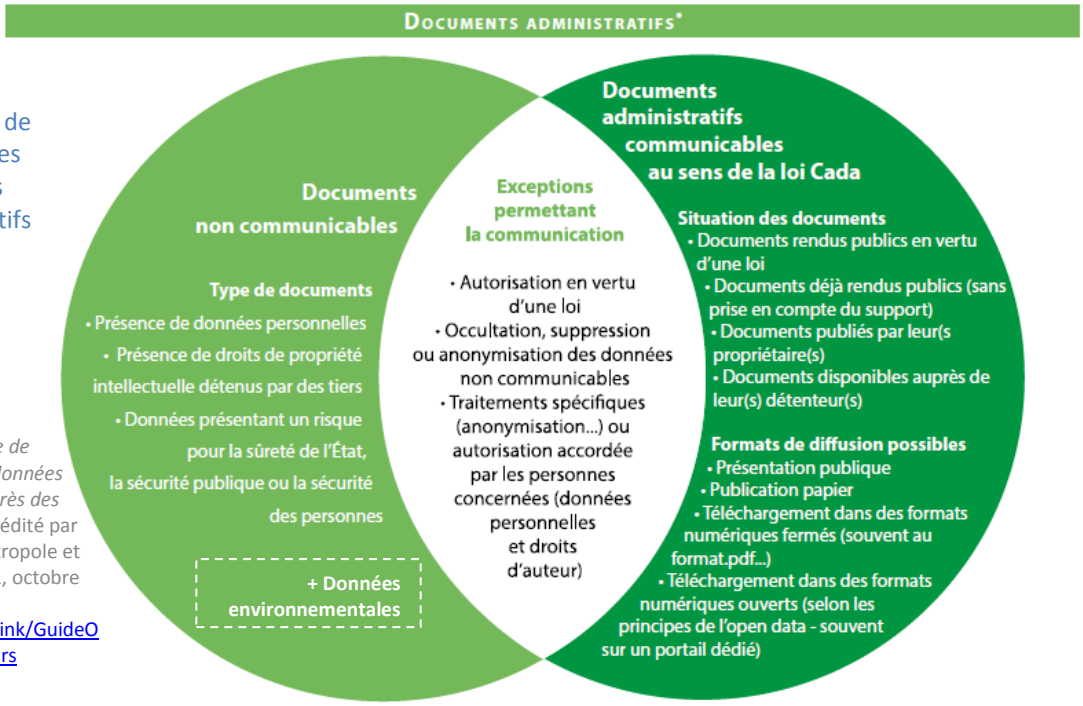


Figure 11
Conditions de diffusion des documents administratifs

Source : *Guide de demande de données publiques auprès des collectivités »* édité par Bordeaux métropole et AEC, version 1, octobre 2014, page 7, http://frama.link/GuideOD_reutilisateurs

Les données sensibles ou personnelles concernant l'intimité d'une personne physique sont protégées par la loi Informatiques et Libertés (LIL) (Figure 11). Elle prend en compte les données sur les origines raciales ou ethniques, opinions politiques, philosophiques ou religieuses etc. (OpenData France, version 1, juin 2016). Par conséquent, tout n'est pas à « ouvrir » (Cf. 2.2.2.1).

[16] Loi Commission d'accès aux documents administratifs.

2.1.1.4. Donnée liée

Pour aller plus loin dans la démarche Open Data, il faudrait faire en sorte que les données soient liées. Les données liées correspondent à la publication et à la connexion de données structurées sur le Web. Grâce à des technologies Web standards, des connexions par ordinateur permettent aux données de différentes sources d'être connectées et interrogées. Le lien peut se faire par les métadonnées ou les en-têtes de colonne des bases de données. 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. Il s'agirait ainsi de données identifiées par des URL[17] (avec date de mise à jour) afin que l'on puisse « pointer » un lien vers elles. Pour concrétiser cette idée, il faudrait convertir les données en triplets RDF[18], langage de base du Web sémantique. 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.

2.1.2. Qu'est-ce que l'Open Data ?

Littéralement, « open data » est une donnée « ouverte » ou « libre ». Cette dernière se rapporte au mouvement autour de la culture du libre : la donnée est mise à disposition de tous et peut être librement réutilisée. Cela correspond parfaitement à l'*Open Definition: « that is, it can be freely used, reused and redistributed by anyone* ». L'Open Data correspond donc aux données accessibles en ligne, réutilisables avec un format lisible par tous et une licence qui autorise l'usage libre, même à des fins commerciales.

2.1.2.1. Un vecteur de changement

L'Open Data[19] change l'interaction entre les gouvernements, les entreprises et la société. On rentre dans une logique où les gouvernements, les entreprises et les citoyens sont simultanément les fournisseurs et les utilisateurs de données. 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. En effet, la réutilisation des données environnementales peut être à l'origine de nouveaux services et usages, le tout ancré dans une démarche collaborative et respectueuse de l'environnement. Par exemple, en 2011, la ville de Montpellier a reçu le prix de la croissance verte, notamment grâce à son application mobile BioZoom (Cf. Annexe n°4). L'Open Data améliore la diffusion d'information, 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 a permis d'aider les Organisations non gouvernementales (ONG) par l'utilisation des données pour les approvisionnements dans les zones sinistrées, comme en 2015 en Haïti et dans les Philippines (Cf. Annexe n°4). L'Open Data intervient ainsi dans la gestion de crise et dans l'action humanitaire. C'est donc un vecteur de changement.

---

[17] *Uniform Resource Locator*

[18] *Resource Description Framework :* modèle de graphe destiné à décrire les ressources Web et leurs métadonnées pour permettre le traitement automatique de telles descriptions. Le RDF est aussi de l'interopérabilité syntaxique, c'est-à-dire un format d'échange internationalement reconnu (Cf. 3.3.2.2).

[19] Cf. *Open Data Index* : indicateur qui liste des pays selon l'ouverture de leurs données publiques : https://index.okfn.org/

2.1.2.2. Les conditions pour faire de l'Open Data

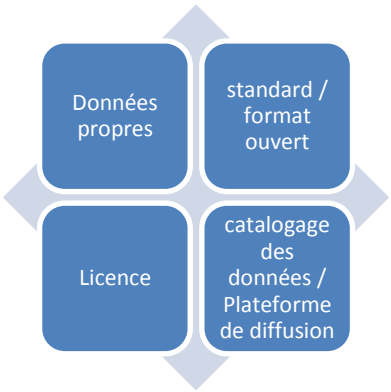


Figure 12 Conditions pour faire de l'Open Data

Comme il a été expliqué dans la définition d'une donnée ouverte (Cf. 2.1.1.2), sans licence, il n'y a pas réellement d'Open Data. **Une licence** permet d'informer les gens qu'ils ont le droit d'accéder, d'utiliser et de partager les données. La règle d'or est la suivante : *plus une licence ouverte est simple, mieux c'est* (*European Data Portal*). En effet, plus facile ce sera pour l'utilisateur 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 la plus ouverte 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 explicite d'y accéder, de les utiliser et de les partager. Différents types de licence existent (Cf. Annexes n°7 et 8) : *Creative Commons* (CC0 / CC-BY / CC-BY-Sa), des licences sur mesure, des licences gouvernementales ouvertes comme la Licence Ouverte d'Etalab[20] (versions 1.0 & 2.0), ou encore l'*Open Database licence* (ODbL).

La libération d'une donnée implique également une standardisation. Les **standards ouverts** définissent les aspects pratiques, légaux, techniques (format de fichier, projection,…) 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 des 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 à la création d'outils et de manuels pour travailler ces données. **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 un format ouvert en plus du format d'origine. Il faut penser également au choix de la structure (hiérarchique ou en réseau) et au moyen de distribution. Par exemple, 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 est intéressant de savoir également 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 ?
- 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 doit être 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.

[20] 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 »).

Certains choix sont donc décisifs en vue d'un Open Data durable, notamment le choix porté sur une plateforme Internet. **Le choix de la plateforme** Open Data est 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 (cela sous-entend de créer des métadonnées (informations sur la donnée);
- Une approche plus directe de gestion de données : on télécharge la donnée *via* une plateforme (visualisation et exploration des données sur la plateforme).

Il peut être nécessaire de combiner ces deux approches pour qu'il y ait une cohérence et une uniformité pour l'utilisateur et pour l'accès aux données. Grâce à ces plateformes, l'utilisateur 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. Mais comment évaluer cette exploitabilité des données ?

2.1.2.3. Evaluer la qualité des données ouvertes

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*[21], 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. Il faut savoir également qu'un des objectifs ultimes de l'Open Data est d'aboutir à un Web de données ouvertes et liées. Il existe notamment un moyen d'évaluer à quel point les données sont utilisables : **les 5 étoiles des Données Ouvertes & Liées (LOD)**[22]. 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 (inventeur du WWW)[23], en 2010. Quant aux **« *Open Data Certificates* »** de l'*Open Data Institute* (ODI), ils 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.

Le contexte Open Data évolue très rapidement, ce qui est notamment le cas d'un point de vue juridique.

2.1.3. Un cadre juridique en évolution

Une des missions de mon stage était de cerner le cadre juridique de l'Open Data. Tout un travail de recherche et de veille informatique a donc été réalisé pour comprendre les enjeux au sein du Parc

---

[21] Formation sur l'utilisation d'*OpenRefine* suivie dans le cadre de l'OpenDataLocale, à Marseille, le 09/06/2017.
[22] LOD
[23] *World Wide Web*

du Verdon. Vous trouverez donc en annexe (Cf. Annexe n°2), une fiche détaillée des différentes lois et décisions relatives à l'ouverture des données.

2.1.3.1. Les lois clés de l'Open Data

C'est en 1789 qu'est apparu le concept de l'ouverture des données, avec l'article 15 de la Déclaration des Droits de l'Homme et du Citoyen : *« la Société a le droit de demander compte à tout Agent public de son administration ».* En 1978, la loi « Informatique et libertés » et la loi CADA donnent le droit d'accès à l'information publique. En 2010 seulement, la Ville de Rennes est la première collectivité à mettre en ligne un catalogue de données ouvertes. En 2016, la loi pour une République numérique a marqué une nouvelle étape avec le principe de l'ouverture des données par défaut et gratuite, ainsi que l'obligation d'une publication des standards ouverts (pour les collectivités de plus de 3500 habitants et de plus de 50 agents). Dans une démarche Open Data, le cadre juridique doit être posé pour cerner les enjeux actuels de la politique numérique : répondre à la loi et faire preuve de transparence envers les citoyens. Les points suivants sont à prendre en compte :

- Les frontières entre les données publiques et personnelles ;
- Les données à anonymiser ;
- La gouvernance publique des données ;
- La confidentialité ;
- La réutilisation ;
- Les licences et le cadre réglementaire.

L'objectif ici est de revenir sur les lois clés relatives à la démarche Open Data[24].

- **La loi CADA** sur la Communication des Documents Administratifs –impose de fournir l'information à qui la demande. Cela ne concerne pas uniquement les données numériques. Le but de l'ouverture des données publiques est de les rendre accessibles au plus grand nombre en diminuant les contraintes administratives. Si les réutilisations des données sont libres, il existe des exceptions mentionnées dans la loi CADA. Ex : Article 13 : *« La réutilisation d'informations publiques comportant des données à caractères personnel est subordonnée au respect des dispositions de la loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés ».*

- **Loi pour une République numérique ou loi Lemaire** : l'obligation pour les organisations publiques (pour les collectivités de plus de 3500 habitants) de publier sur Internet leurs bases de données, sous réserve notamment d'anonymisation et de protection de la propriété intellectuelle et du secret industriel et commercial. Le décret d'application du 28 décembre 2016 fixe à 50 agents le seuil minimal à partir duquel les organisations publiques sont dans l'obligation d'ouvrir leurs données.

- 14 mars 2017, le **décret n°2017-331** relatif au service public de mise à disposition des données de référence précise la mission de service public relevant de l'Etat. Toutes les administrations mentionnées dans l'article L300-2 concourent à cette mission. Ces données de référence sont des informations publiques mentionnées dans l'article L 321-1, à savoir des informations dans des documents communiqués ou publiés pouvant être utilisés par toute personne qui le souhaite à d'autres fins que celles de la mission de service public.

- Le 14 mars 2017, un autre décret voit le jour : **décret n° 2017-330** relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique. L'administration sera dans l'obligation de divulguer les informations mentionnées à l'adresse suivante : https://www.legifrance.gouv.fr/eli/decret/2017/3/14/PRMJ1632786D/jo/texte

[24] Les dates clés de la démarche Open Data (source : Alexandra Greco, 2017)

- **Directive européenne INSPIRE** –impose aux organisations publiques de publier leurs données environnementales géographiques et leurs métadonnées. Cette démarche de cataloguer les données géographiques en créant des métadonnées rentre dans le cadre de l'Open Data. En effet, cela rejoint l'objectif de l'interopérabilité des données.

- Le 27 avril 2017, **le décret n°2017-638** fixe une liste de licences de réutilisations autorisées pour les administrations qui souhaitent soumettre la réutilisation à titre gratuit leurs informations publiques à une licence. Ce décret fixe les conditions d'homologation des licences dans le cas où les administrations désirent recourir à des licences qui ne figureraient pas dans cette liste. A l'issue de la publication de ce décret et d'une concertation, la version 2.0 de la Licence Ouverte est définitivement adoptée le 28 avril 2017 et remplace la Licence Ouverte Administrations publiée le 20 mars 2017.

Ce travail m'a permis de rédiger une délibération de principe pour l'ouverture des données publiques du Parc du Verdon, ainsi que son annexe (Cf. Annexes n°3 et 4). Ainsi, j'ai pu cerner les droits et devoirs du Parc du Verdon en matière d'Open Data.

2.1.3.2. Quelles sont les obligations du Parc du Verdon ?

En pratique, sur le vote Open Data, la loi pour une République Numérique du 07/10/2016 crée l'obligation, pour les collectivités territoriales de plus de 3500 habitants, de publier sur Internet leurs bases de données, sous réserve notamment d'anonymisation et de protection de la propriété intellectuelle, du secret industriel et commercial. Ces données pourront ainsi être exploitées et réutilisées facilement par chacun, particulier comme entreprise. Le décret d'application du 28/12/16 fixe à 50 agents, ou salariés équivalents à temps plein, le seuil minimal en-deçà duquel les collectivités territoriales ne sont pas concernées par l'obligation d'ouverture. Avec ce décret, le Parc du Verdon n'est donc pas dans l'obligation d'ouvrir ses données. En revanche, le Parc naturel régional du Luberon (52 agents en août 2017) est dans l'obligation d'ouvrir ses données. Quant au Parc de Camargue avec ses 48 agents en 2017, il est à la limite du seuil fixé par le décret du 28/12/16. Si le Parc du Verdon est une collectivité territoriale de 3500 habitants mais de moins de 50 agents (37 agents en février 2017), il exerce avant tout des missions de service public. Et pour différentes raisons, notamment pour respecter la loi pour une République numérique, le Parc prend l'initiative d'ouvrir ses données.

Avec les récentes lois en matière d'ouverture des données publiques, les collectivités territoriales sont concernées par une démarche Open Data qui est de plus en plus cadrée, notamment d'un point de vue méthodologique. Mais quels sont les enjeux à prendre en compte en tant que Parc naturel régional avant de lancer une démarche Open Data ?

2.2. Quels sont les enjeux de l'Open Data au sein du Parc du Verdon ?

Après avoir contextualisé l'Open Data, il est important de connaître la position du Parc du Verdon dans une telle démarche en répondant à différentes interrogations : quels sont les intérêts et les risques potentiels ? Comment ouvrir ses données ? Quels choix sont à prendre en matière d'Open Data ?

2.2.1. Intérêts pour un Parc naturel régional

En grande partie, les intérêts du Parc du Verdon à engager une démarche Open Data sont inscrits dans sa Charte. Cet engagement suit la logique de la loi pour une République numérique où l'on passe du paradigme de la demande à celui de l'offre : une logique où l'on doit tout ouvrir par défaut. Les risques sont à cerner quant à la diffusion des données publiques du Parc, notamment pour les palier. Cerner ces différents enjeux permet de mettre en place au mieux une démarche Open Data adaptée à la situation d'un Parc naturel régional.

2.2.1.1. Mise en cohérence avec la Charte du Parc et la dynamique territoriale

S'engager dans une démarche Open Data rentre parfaitement en cohérence avec les actions du Parc du Verdon ainsi que les axes de sa Charte 2008-2020 dont l'objectif principal est de : capitaliser, valoriser et informer des connaissances auprès du grand public. Le tableau ci-dessous reprend les orientations de chaque axe de la Charte à partir desquelles on peut faire le lien avec l'Open Data :

Axes	Orientations en lien avec l'Open Data
A	**Mettre en place une stratégie de protection et de gestion du patrimoine naturel :** Développer les échanges de données entre détenteurs des sources et gestionnaires. Recenser l'ensemble des données naturalistes relatives au territoire du Parc sur le SIT.
	Rôle de l'organisme gestionnaire du Parc : organiser le partenariat des associations pour un échange systématique de données naturalistes. Concevoir des outils de communication pour la transmission des connaissances scientifiques acquises sur le territoire (site Internet, expositions, éditions, courrier scientifique du Parc, bases de données cartographiques…). Organiser l'accès à distance des habitants du Parc aux données consultables du SIG du Parc (cf. mesure B.2.2).
	Implication des partenaires : L'Etat met à disposition du Parc les données relatives au territoire du Parc dont ses services favorisent le partenariat pour la valorisation de ces connaissances entre les Parcs et l'Education nationale par le biais des inspections académiques.
B	**Favoriser la réappropriation, l'expression et la pratique culturelle** Développer une conscience citoyenne pour l'éducation : Assurer la mutualisation des ressources et des données en matière d'éducation à l'environnement.
	Rôle de l'organisme gestionnaire du Parc : Animer un observatoire du territoire qui recense les données relatives aux patrimoines préhistoriques, archéologiques et historiques, en relation avec les administrations et les services en charge de leur protection (cf. mesure D1.4). Diffuser les données et les travaux scientifiques réalisés par des médias appropriés, destinés prioritairement à la population du Verdon et ses partenaires.
C	**Contribuer au développement d'une agriculture diversifiée économiquement viable, écologiquement responsable** Capitaliser les données existantes sur les productions au travers d'un centre de ressources (valorisation études, recherches ou expérimentations).
	Promouvoir une gestion multifonctionnelle des espaces forestiers. Dans le cadre d'échange réciproque des données, l'Office nationale des Forêts (ONF) informe le Parc du lancement des études d'aménagements forestiers et sollicite son avis technique en amont sur les enjeux écologiques particuliers à prendre en compte.
	Implication des partenaires : l'Etat transmet des données relatives à l'activité agricole au Parc. Les ADASEA[25] participent à la mise en place et à l'alimentation des données de l'observatoire des exploitations agricoles…
D	**Promouvoir une qualité d'aménagement** **Rôle de l'organisme gestionnaire du Parc :** Négocier des conventions de partenariat avec les collectivités et les organismes publics pour la fourniture et la publication de données pertinentes pour le suivi de l'action du Parc et des évolutions du territoire ; Mettre à disposition des données communicables et des analyses issues de l'Observatoire ; Réaliser des traitements de données pour le compte de collectivités adhérentes, notamment pour le besoin de porter à connaissance ; Mise à disposition de ses compétences techniques pour conseiller les associations sur la manière de recueillir des données utiles pour renseigner les bases de données du patrimoine.

[25] Association ou Organisme Départemental pour l'Aménagement des Structures des Exploitations Agricoles

D	**Implication des partenaires :** Communes et intercommunalités compétentes alimentent l'observatoire à partir des données dont elles disposent. La Région PACA et les départements des Alpes de Haute-Provence et du Var définissent des protocoles d'accord avec le Parc pour la mise à disposition des données relatives au territoire du Parc, réunies à l'occasion d'études ou collectées par leurs services. Associations, établissements d'enseignements, conseil scientifique du Parc, maisons thématiques, musées, acteurs socio-économiques valorisent les données produites par l'observatoire dans le cadre de leurs actions d'information, d'éducation et d'animation.

Source : Charte 2008-2020 PNRV – Alexandra Greco, 2017

Figure 13 Orientations des axes de la Charte du Parc du Verdon en lien avec l'Open Data

Le Parc du Verdon, comme tout autre Parc naturel régional, mène des études sur des thématiques diverses : patrimoine, biodiversité, aménagement, développement, ... Différents supports et formats sont également utilisés : couches de données géographiques, études/rapports/thèses, images, vidéos, documents administratifs, données scientifiques etc. Cette diversité à la fois au niveau des thématiques et des supports de travail a tout intérêt à être valorisée auprès du grand public : porter à connaissance les données produites par le Parc, mais aussi innover et développer de nouveaux usages et services numériques. De plus, adopter une démarche Open Data permet d'aller plus loin dans la réflexion collective interparc, à savoir de centraliser et diffuser les informations géographiques disponibles sur les territoires des Parcs de PACA. C'est également un moyen de relever le défi de la transition numérique : se moderniser, digitaliser et transformer le service public en ouvrant les données. L'Open Data permet aux structures qui s'y engagent de faire preuve de transparence auprès des utilisateurs, tout comme il génère de nouvelles interactions avec ces derniers. En dehors de cette question de transparence/ interaction avec les utilisateurs, il existe un réel enrichissement du service public grâce à l'Open Data. Enfin, les Parcs de PACA ont l'opportunité de pouvoir suivre la même démarche Open Data entreprise par la Région et de se mettre en cohérence avec les dynamiques du territoire. En effet, la Région PACA s'est engagée dans une démarche d'ouverture des données publiques par délibération n°11-1631 du 16 décembre 2011. Un des autres atouts de l'adhésion à cette démarche est que le Parc est ainsi garant de la fiabilité des données qu'il diffuse.

2.2.1.2. Garantie de la fiabilité des données

En ouvrant ses données lui-même, le Parc du Verdon se porte garant de leur fiabilité. A travers cette ouverture des données publiques, le Parc diffuse aussi une expertise des données acquise depuis de nombreuses années. L'information est ainsi garantie tout comme la qualité d'acquisition des données brutes. Par exemple : l'accès au sentier des Pêcheurs sur la commune de La Palud-sur-Verdon est interdit depuis le 31 mars 2017. Ce sentier est très largement fréquenté en saison estivale et il est recommandé dans de nombreux guides de randonnée, ou autres (sites Internet etc.). Si le Parc transmet en Open Data ses données sur les itinéraires de randonnée aménagés autour des Gorges du Verdon, il permet ainsi la diffusion d'une donnée fiable et récemment actualisée. D'autres utilisateurs de cette donnée diffuseront à leur tour une information juste qui sera sans doute réutilisée par d'autres, ainsi de suite. Dans notre exemple, diffuser ce type de données permet concrètement d'éviter qu'une information obsolète sur l'accès à ce sentier soit diffusée largement auprès du grand public. Si les avantages et les intérêts d'ouvrir les données publiques du Parc ont été largement exposés, il faut tenir compte également des éventuels risques.

2.2.2. Risques potentiels pour un Parc naturel régional

Toutes les données ne peuvent pas être ouvertes y compris lorsqu'il s'agit de données publiques. Certaines conditions font que des documents administratifs ne sont pas communicables (Cf. Figure 11). C'est également le cas pour des données sensibles à caractère environnemental.

2.2.2.1. *Restrictions dans l'ouverture des données*

L'article 6 de la loi n°78-753 du 17/07/1978, modifié en 2013, expose les limites à la transmission des données dans certaines situations relatives :

a) Au secret des délibérations du Gouvernement et des autorités responsables relevant du pouvoir exécutif ;
b) Au secret de la défense nationale ;
c) A la conduite de la politique extérieure de la France ;
d) A la sûreté de l'Etat, à la sécurité publique ou à la sécurité des personnes ;
e) A la monnaie et au crédit public ;
f) Au déroulement des procédures engagées devant les juridictions ou d'opérations préliminaires à de telles procédures, sauf autorisation donnée par l'autorité compétente ;
g) A la recherche, par les services compétents, des infractions fiscales et douanières ;
h) Ou, sous réserve de l'article L. 124-4 du code de l'environnement, aux autres secrets protégés par la loi ;

Ne sont communicables qu'à l'intéressé les documents administratifs :

- dont la communication porterait atteinte à la protection de la vie privée, au secret médical et au secret en matière commerciale et industrielle ;
- portant une appréciation ou un jugement de valeur sur une personne physique, nommément désignée ou facilement identifiable ;
- faisant apparaître le comportement d'une personne, dès lors que la divulgation de ce comportement pourrait lui porter préjudice.

Compte-tenu du fonctionnement et des données produites par le Parc, les contraintes de publication des données à retenir sont les suivantes :

- Les données portant atteinte à la vie privée, portant appréciation sur une personne physique ou divulguant des comportements : *« Ces données sont assez faciles à identifier, mais peuvent se nicher dans des jeux de données complexes et il convient d'être attentif lors de chaque publication. Notamment à propos des données géographiques comme par exemple l'occupation des sols, des zonages divers, ou des données statistiques sur la population dont les jeux de données pourraient contenir des noms ou données pouvant formellement identifier des personnes. Il faudra alors avant publication anonymiser les jeux de données en supprimant les colonnes permettant l'identification des personnes. »* (PNR du Vercors, 2014, p. 6).

- Les données réalisées dans le cadre d'une prestation dont la cession des droits d'exploitation de diffusion et de reproduction n'a pas fait objet d'un contrat de cession : il est préconisé d'ajouter une clause autorisant l'ouverture des données dans les conventions ou contrats de projets (Cf.2.3.2.2);

A SAVOIR :

Les œuvres réalisées par les agents autres que les chercheurs et enseignants-chercheurs appartiennent à l'établissement public. C'est le cas également des œuvres commandées par l'établissement et réalisées selon des directives précises données par la structure. Sans accord préalable des auteurs, la structure peut diffuser les données dès lors que cette diffusion est strictement nécessaire à l'accomplissement d'une mission de service public (Article L.131-3-1 Code de la Propriété Intellectuelle).

Concernant les données contenues dans des « documents » sur lesquels des tiers détiennent des droits de propriété intellectuelle, leur diffusion est soumise à l'accord préalable de ces tiers. C'est notamment le cas pour les dessins et modèles, les marques, les photos, les brevets ou encore le droit des producteurs de bases de données.

- Les données pouvant porter atteinte à la protection de l'environnement :
  L'article L.124-4 du code de l'environnement apporte la précision suivante : la communication des données est rejetée lorsqu'elle porte atteinte à la protection à l'environnement auquel elle se rapporte. La loi explique que lorsqu'un jeu de données à transmettre comprend des données non communicables, il est possible d'occulter ou de disjoindre ces dernières pour que l'on puisse transmettre le jeu de données en question.

La diffusion des données sensibles à caractère environnemental constitue un cas particulier dans le cadre de l'Open Data au sein d'un Parc naturel régional.

Pour plus de détails relatifs à la communication des données sous condition, je vous renvoie au guide d'analyse du cadre juridique en France pour l'ouverture des données de la recherche : http://prodinra.inra.fr/record/382263

2.2.2.2. Cas particulier : données sensibles à caractère environnemental

Entrée en vigueur en 2002, la convention d'Aarhus instaure un droit d'accès aux données environnementales. Ce droit a été complété en 2003 par la directive sur la réutilisation des données publiques, puis en 2007 par la Directive INSPIRE. La convention d'Aarhus a eu pour but de faciliter la prise en compte de l'environnement dans les décisions publiques et par les citoyens. Les lois Valter[26] et Lemaire complètent ce mouvement. C'est ensuite aux services publics comme les Parcs de mettre en place cette dynamique progressivement. Concernant les données sensibles à caractère environnemental, l'argument des naturalistes réticents au partage de ces données est le suivant : *« des personnes mal intentionnées pourraient détruire ou piller les éléments remarquables qui seraient divulgués »* (Espaces naturels n°59, juillet-septembre 2017, p.33). Le cas des données sensibles est défini dans le protocole du Système d'Information sur la Nature et les Paysages (SINP) : ce sont des données que l'on ne doit pas diffuser largement pour éviter de porter atteinte aux éléments qu'elles concernent. D'après un groupe de travail partenarial, coordonnée par le Muséum national d'histoire naturelle (MNHN), une espèce est définie comme potentiellement sensible selon les critères suivants :

Risque d'atteinte volontaire	Sensibilité intrinsèque de l'espèce	Effet probable de la diffusion	Aspects de contexte
Est-ce que l'espèce est sujette à prélèvement pour la consommation, la pharmacie, la collection, l'horticulture… à destruction volontaire ? Est-ce que sa présence risque d'attirer de nombreux curieux ou naturalistes ?	L'espèce est-elle menacée au sens d'une liste rouge ? Ou très localisée dans la région ?	Est-ce que la diffusion de la donnée augmenterait le risque d'impact ? (dans le cas où des informations précises existent déjà dans la bibliographie et sur Internet)	Ancienneté de l'observation et statut biologique. Ex : l'endroit précis de nidification de la Cigogne noire est une donnée sensible, alors qu'une observation de l'espèce en migration n'en est pas une.

Figure 14 Critères de définition d'une espèce sensible

Source : Espaces naturels n°59 – retravaillé par Alexandra Greco, 2017

Parmi les atteintes possibles provenant de la publication de ces données sensibles, il a été fait état de la publication de données géographiques trop précises : risque d'atteinte volontaire aux éléments concernés par la donnée diffusée (Cf. Figure 14). Pour dépasser cette réticence à publier ce type de données, une diffusion à géométrie variable peut être une solution :

[26] En 2015, traduction en loi française de la directive européenne PSI 2013/37/UE du 26/06/2013.

- ✓ **Pour les citoyens** - l'accès et la visualisation des données sensibles est possible mais leur localisation est dégradée soit : à la commune, à la maille de 10 x 10 km par exemple, au département (pour les données les plus sensibles). Une demande nominative et motivée peut être envoyée auprès de la structure, pour accéder à la donnée précise, sous réserve de validation.
- ✓ **Pour les acteurs habilités** – l'accès restreint (avec mot de passe) pour obtenir la donnée précise.

Ces deux suggestions de diffusion différenciée de la donnée sensible à caractère environnemental sont proposées par les plateformes du SINP[27] et sont largement recommandées par les professionnels du milieu des espaces naturels. Mais il ne faut pas perdre d'esprit qu'en dehors de ces données sensibles, il faut publier des données les plus brutes possibles si l'on souhaite faire réellement de l'Open Data. En effet, traiter des données avant diffusion ferait perdre un des grands intérêts de l'Open Data. Une fois que l'on a pris connaissance des atouts et des risques dans la diffusion des données publiques, il faut savoir comment les ouvrir concrètement.

2.2.3. Comment ouvrir les jeux de données ?

La partie Le contexte Open Data rappelle les conditions à suivre pour faire de l'Open Data, notamment nettoyer/vérifier ses données, choisir un format/ standard ouvert, apposer une licence de diffusion aux données, et choisir un moyen de publication. Il est question ici de montrer concrètement :

- quels sont les formats de fichier à préconiser lorsque l'on souhaite ouvrir des données ;
- les licences de diffusion ;
- quelles sont les solutions possibles pour la publication des données.

Répondre *a minima* à ces trois éléments après avoir pris connaissance des restrictions, c'est pouvoir ouvrir des jeux de données.

2.2.3.1. Choix des formats

Il existe de nombreux formats de fichiers spécifiques à des types de fichiers. Il est possible de distinguer les formats dits ouverts des formats dits fermés (Cf. Annexe n°9). Un fichier est ouvert à condition que :

- Son encodage soit transparent ;
- Sa recette de fabrication fasse partie du domaine public ;
- Tous les logiciels destinés à traiter le type du fichier (image, texte, audio, etc.) puissent le créer, le lire et le modifier, c'est-à-dire qu'il soit interopérable.

Ces formats sont souvent des standards ouverts, comme le HTML qui sert à construire des pages Web.

On a tendance à employer le terme « standard » pour se référer à tout format ou protocole. Or, utiliser un standard a pour but de créer un référentiel commun (Cf. 3.3.2.1). Deux principales catégories de standard existent :

- **Un standard de fait :** *« un référentiel s'imposant de manière spontanée dans un marché donné. Il n'est pas nécessairement ouvert, ni libre »* (Goffinet, L., 2010). Par exemple, les formats de la suite *Microsoft office* est un exemple de standard de fait ;

[27] Cf. 2.2.3.3

- **Un standard de droit :** *« également appelé « norme », est un référentiel établi par une autorité de normalisation reconnue – publique ou privée, nationale ou internationale – comme l'ISO, l'IETF, le W3C ou la CEN… »* (Goffinet, L., 2010).

A l'inverse, les formats fermés n'appartiennent pas au domaine public. On est le plus souvent contraint d'avoir le logiciel adéquat pour pouvoir lire et modifier un fichier convenablement. Par exemple, pour lire un fichier au format .PSD, on doit posséder le logiciel Photoshop. De plus, il s'agit le plus souvent de formats propriétaires (Cf. exemple de Photoshop appartenant à Adobe). Mais, tous les formats propriétaires ne sont pas fermés. En effet, pour reprendre le cas d'Adobe, son format PDF est devenu un standard (norme ISO). Tout le monde peut ouvrir un fichier PDF.

L'utilisation de formats ouverts est à privilégier pour ne pas dépendre d'un logiciel en particulier, et pour faire en sorte que les données soient lisibles et réutilisables dans le temps. Mais il est possible de recourir à des formats propriétaires standards (comme le PDF).

Microsoft a essayé d'ouvrir les formats de sa suite *Microsoft Office*. Toutes ses extensions terminant par –x (.DOCX, .XLSX, .PPTX…) sont des formats créés dans le but de répondre à l'interopérabilité des données. Toutefois, leur format ouvert fait encore débat. Par exemple, sur la plateforme régionale OpenPACA, les formats .DOC, .XLS, .PPT… sont préconisés.

Choisir une licence de diffusion à apposer à un jeu de données à ouvrir est une étape importante dans l'Open Data, notamment dans un but de réutilisation et de partage de la donnée.

2.2.3.2. Choix des licences

La loi pour une République numérique indique que la licence doit être choisie sur une liste fixée par décret ou, à défaut, selon une licence homologuée par l'Etat. Le décret n° 2017-638 du 27 avril 2017 fixe ces licences et la procédure d'homologation. L'article L. 323-2 du Code des relations entre le public et l'administration (CRPA) a pour seule vocation de limiter le nombre de licences que les administrations peuvent proposer aux réutilisateurs, et ne doit pas avoir pour effet de limiter ou de faire obstacle aux projets auxquels une administration peut participer. Le décret couvre à la fois la réutilisation des données publiques et des logiciels. C'est dans ce contexte que pour chacune des catégories de document administratif (données ou logiciels), deux types de licences sont proposés :

- **Licence Ouverte** ou licences « permissives » : où il y a une totale liberté de réutilisation, y compris à des fins commerciales. La dernière Licence Ouverte d'avril 2017 est compatible avec toute licence libre, notamment avec sa version antérieure 1.0 ;

- **Open data base licence (ODbL)** ou licences « avec obligation de réciprocité » : où le réutilisateur est libre d'exploiter :

    - Une base de données à condition de maintenir la licence sur la base de données dérivées ou du moins qu'elles soient licenciées sous l'une des licences compatibles ;

    - Un logiciel avec garanti de liberté de : l'exécuter, étudier son fonctionnement et l'adapter à ses besoins, redistribuer des copies, améliorer le logiciel et publier ces améliorations. Les œuvres dérivées de ce logiciel offriront aussi ces libertés.

Pour établir une licence particulière, il faut faire une demande d'homologation à l'Etat après avoir adressé une demande motivée à la DINSIC. Mais au préalable, l'administration doit mener une concertation avec les principaux réutilisateurs et en présenter la synthèse. Cette décision d'homologation sera prononcée par décision du Premier ministre. Si de nombreuses demandes portent sur la même licence, il est possible qu'elle rejoigne les licences listées dans le décret.

Le décret dresse une 1ère liste des licences les plus employées actuellement et compatibles avec les critères mentionnés ci-dessus. Cette liste « v0 » est amenée à évoluer (au moins tous les 5 ans comme fixé par la loi).

Vous trouverez en annexe un tableau comparatif des licences CC4.0 et CC0 ainsi qu'un tableau comparatif des licences gratuites ODbL et Licence Ouvertes mis à jour (Cf. Annexes n° 7 et 8).

Vous pouvez également vous référer aux outils d'aide à la décision pour l'ouverture des données publiques à l'adresse suivante[28] : http://193.55.175.126/sigmaetu10/PNRV/
Ces outils sont basés sur les codes sources développés par Mathieu Andro de l'INRA[29] (Délégation à l'information scientifique et technique). Ses codes en PHP ne fonctionnant pas sur mon serveur test (de la formation SIGMA), j'ai dû les retravailler en utilisant du JavaScript. Ces codes sont en cours de développement mais cela m'a permis de monter en compétence en programmation informatique, et d'apprendre ainsi un nouveau langage de programmation qu'est le JavaScript.

2.2.3.3. Solutions de publication des données

Après avoir choisi les formats des données et les licences de diffusion, il faut savoir où publier les données. Dans le cas du Parc du Verdon, il existe plusieurs vecteurs de diffusion potentiels gratuits :

- ✓ Data.gouv.fr : permet d'envoyer le jeu de données directement sur leur site, ou bien de n'y indiquer que le lien permettant de le télécharger. Cette souplesse est très appréciable car elle rend possible la publication de données même à des personnes ou des institutions ne disposant pas d'un serveur en ligne capable de les héberger. Il est à noter que l'inscription et l'hébergement des données sur data.gouv.fr sont gratuits.

  Autre critère intéressant pour ce lieu de dépôt : la fréquence de mise à jour des fichiers de données. Les fichiers fréquemment mis à jour seront avantageusement publiés sur le site du Parc. Des fichiers mis à jour peu souvent peuvent être envoyés sur le site de data.gouv.fr.

- ✓ Metaclic : un catalogue de données ouvertes libre et gratuit. Il permet d'intégrer dans un site Internet un catalogue des données diffusées par la plateforme data.gouv.fr. C'est une API[30] catalogue développée dans le cadre du projet OpenDataLab Occitanie. Si ce projet a pris fin le 20 mai 2017, il est possible d'utiliser encore cette solution. Techniquement, il s'agit d'un code en javascript (JS) que l'on peut implémenter dans le code source de son site Internet. Il est également possible de configurer et de personnaliser cette solution[31].

[28] Page Web développée avec le framework Bootstrap
[29] Institut national de recherche agronomique
[30] *Application Programming Interface*
[31] https://github.com/datakode/metaclic

- ✓ OpenPACA : La Région PACA a mis en ligne le portail partenarial régional des données publiques, celui-ci est ouvert à tous les partenaires et acteurs de Provence-Alpes-Côte d'Azur. La plateforme actuelle ne permet pas le moissonnage[32] des données entre data.gouv.fr et OpenPACA. Si l'on met de la donnée sur OpenPACA, l'actualisation de cette donnée sur data.gouv.fr ne se fait pas automatiquement. Cette plateforme OpenPACA va évoluer début 2018, et elle sera alors basée sur l'outil CKAN[33].

- ✓ IDGO (Infrastructure de données géographiques ouvertes régionale) sera la nouvelle IDS mutualisée entre la Région et le CRIGE PACA. Elle permettra d'indexer les données à un seul endroit. En octobre 2017, il sera possible d'expérimenter cette solution de mise à disposition de la donnée et du moissonnage des métadonnées.

D'autres solutions existent mais sont pour la plupart payantes comme OpenDataSoft[34]. Il s'agit d'une solution qui fonctionne bien pour la publication des données, mais pour l'instant peu de plugins sont développés pour les données géographiques, notamment lorsque le poids de ce type de fichiers est relativement important. Cette solution reste également onéreuse. D'autres solutions de diffusion des données du Parc du Verdon, notamment internes voire interparc sont envisageables :

- ✓ Site Internet du Parc ;
- ✓ Centre de ressources du site Internet du Parc : pour les rapports / études / dossiers / comptes-rendus ;
- ✓ SIT pour l'information géolocalisée (soumise à la Directive INSPIRE) ;
- ✓ Une plateforme OpenData interparc pourrait être à envisager (questions de concertation, de coût et de maintenance sont à prévoir).

Concernant les données sensibles à caractère environnemental, une solution déjà mise en place au sein du Parc du Verdon reste à préconiser[35] :

- ✓ SILENE (Système d'information et de localisation des espèces natives et envahissantes) PACA: il s'agit de la plateforme régionale du SINP qui repose sur trois principes :
  - *« un outil public et professionnel ;*
  - *un pilotage partenarial ;*
  - *une garantie scientifique des données »* (Espaces naturels n°59, 2017).

[32] 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 même procédé de « codage technique ». Géosource (version française de GeoNetwork) permet de découvrir quelles sont les métadonnées ayant été ajoutées, supprimées ou mise à jour dans le nœud distant.

[33] Logiciel libre, développé et maintenu par l'Open Knowledge Foundation qui permet de créer des catalogues de jeux de données. CKAN conserve un historique complet des modifications dans le même principe que Wikipédia (Cf 3.3).

[34] *« Plateforme en ligne complète conçue pour la transformation rapide et économique de tous types de données en services innovants (APIs, visualisations de données, monitoring en temps réel) sans aucune limite de volumétrie. Sa mission : faciliter la publication, le partage et la réutilisation de données et d'APIs par les utilisateurs métiers ».* https://www.opendatasoft.fr/societe/

[35] Les enjeux soulevés auparavant ainsi qu'un entretien avec Julie Delauge réconforte l'idée selon laquelle l'Open Data pour les données naturalistes doit être régulé à l'échelle régionale a minima, et par des experts en la matière.

Des échanges opérationnels et effectifs se font avec l'Inventaire national du patrimoine naturel (INPN). C'est le Conservatoire botanique national méditerranéen (CBNMed) qui a mis au point la plateforme SILENE. Cette dernière permet un premier niveau d'expertise pour les données naturalistes regroupant les données flore et faune régionales. Ce projet repose sur un travail de coordination des CBNMED et du Conservatoire des espaces naturels (Cen PACA) avec la DREAL PACA et le Conseil régional (Cf. Figure 15). Une charte de référence fixe les droits et les devoirs des partenaires et des utilisateurs de la plateforme. Ainsi, des cadres techniques sont posés pour faciliter l'intégration des données :

- un format standard incitatif ;
- une assistance technique personnalisée ;
- une validation formelle et scientifique avant publication en ligne.

L'intérêt d'utiliser cet outil pour l'accès à l'information naturaliste est que la validation des données repose sur la méthodologie du SINP. Cette validation passe par les administrateurs des données de la plateforme et des experts thématiques.

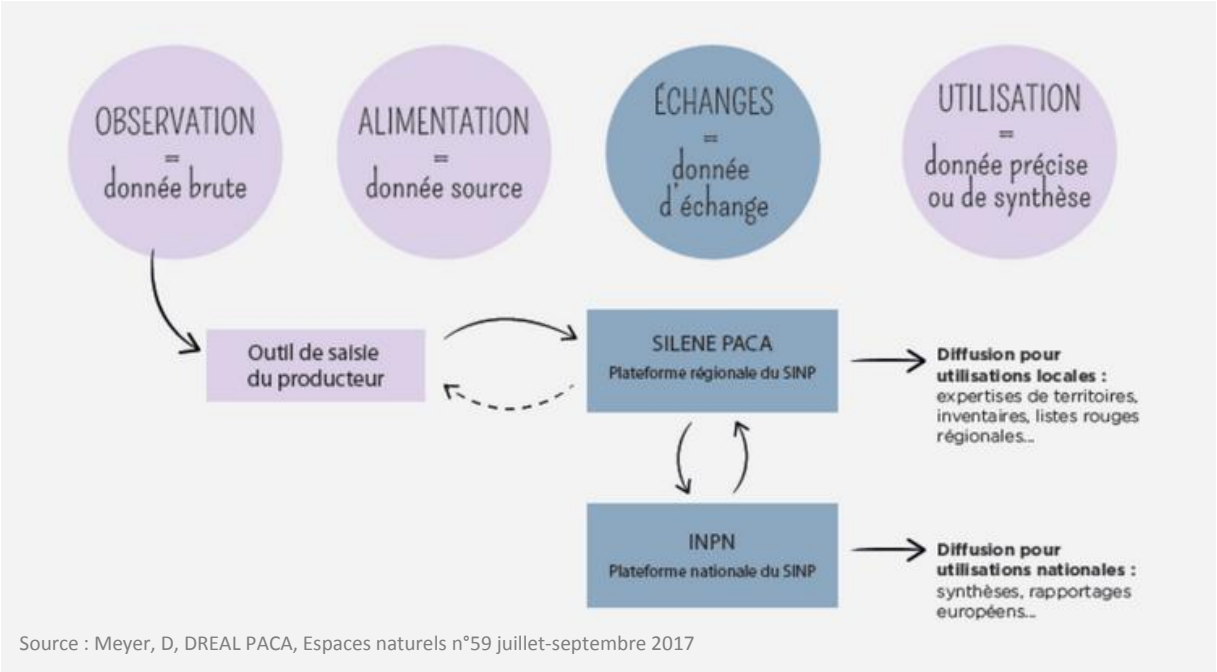
Fonctionnalités de SILENE

« Permet 2 grands types de requêtes :

- **L'interrogation par espèce** (ou groupe d'espèces) pour connaître sa répartition (représentation cartographique) et les informations associées (statut réglementaire, fiche taxon, détail des relevés…) ;
- **L'interrogation par territoire** pour connaître toutes les espèces présentes inventoriées. Les espèces protégées sont identifiées.

Un premier niveau d'information est accessible sans inscription. Les autres adhérents et les autres utilisateurs identifiés (sur demande ponctuelle) peuvent accéder à l'information la plus précise, affiner la recherche par des tris, dessiner ou importer une couche SIG, faire des exports… »

(Source : Espaces naturels n°59 juillet-septembre 2017)



Source : Meyer, D, DREAL PACA, Espaces naturels n°59 juillet-septembre 2017

Figure 15 De l'observation à la connaissance partagée : le chemin de la donnée naturaliste

Une fois que le contexte actuel de l'Open Data a bien été cerné et que tous les enjeux ont été pris en compte, des choix ont été pris pour élaborer la démarche Open Data adpatée à la situation du Parc du Verdon, comme territoire pilote. Il s'agit donc de voir à présent quelles décisions j'ai dû prendre en tant que chargée de mission Open Data et administratrice de données dans le cadre de mon stage.

2.3. Expérimentations de l'Open Data – Parc du Verdon –

Il est question ici de retracer toute la démarche Open Data mise en place au sein du Parc du Verdon pendant la durée de mon stage.

2.3.1. Evaluation de l'existant

Il est intéressant de voir où en sont les autres Parcs naturels régionaux en matière d'Open Data. Une petite enquête a été menée à l'échelle nationale, des réponses sont encore en attente. Ce rapport de stage étant sous Licence Ouverte, il sera possible de le modifier par la suite pour le mettre à jour (sous respect des conditions mentionnées).

PNR	Engagement dans l'Open Data	Remarques
Camargue	**En cours**	A venir suite à l'expérimentation avec le Parc du Verdon. Souhaite mener une démarche Open Data que pour les données principales à référence géographique.
Golfe du Morbihan	✓	Versement des données (très modéré faute de temps) sur la plateforme infrarégionale : GéoBretagne. Evite un maximum les conventions par rapport à la production des données. Les données géographiques relevant de la Directive INSPIRE sont publiques sur demande.
Gâtinais français	**En cours**	2017 : mise en place d'un catalogue de données dans le but d'ouvrir leurs données publiques dans une démarche interparc.
Haut-Languedoc	×	
Livradois-Forez	**En reprise**	Les données géographiques sont mises à disposition sur l'IDG[36] régional (CRAIG). Démarche Open Data lancée auparavant avec les EPCI[37] mais l'expérimentation a été stoppée. Récemment, création d'un poste de technicien développement numérique dont les missions sont fléchées Open Data (mise à disposition du code source des applications).
Luberon	**En cours**	La délibération de principe est en cours de signature pour l'ouverture de tout type de données.
Narbonnaise en Méditerranée	×	Pas de démarche Open Data faute de moyen humain.
Queyras	×	Pas de démarche Open Data à proprement parler qui soit en cours. Mise à disposition de leurs données *via* le SIT et SILENE (GeoNature).
Vercors	✓	Démarche lancée depuis début 2012. Rédaction d'un guide pratique sur l'ouverture des données.

Figure 16 Engagement des PNR dans la démarche Open Data Source : Alexandra Greco, 2017

Sur neuf réponses, deux Parcs se sont officiellement engagés dans une démarche Open Data, sans compter le Parc du Verdon. Trois Parcs vont entamer une telle démarche. Même sur ce petit échantillon de réponses, on voit bien que la dynamique Open Data se met en place timidement au sein des Parcs naturels régionaux.

[36] Infrastructure de données géographiques
[37] Etablissement public de coopération intercommunale

2.3.2. Elaboration de la démarche Open Data à suivre

Concernant l'élaboration de la démarche Open Data, je me suis largement inspirée des recommandations du projet OpenDataLab Occitanie, tout en m'appuyant sur leurs outils méthodologiques[38] mis à disposition. Le travail de recherche et de veille informatique sur l'Open Data et son cadre juridique a permis également de lancer la démarche.

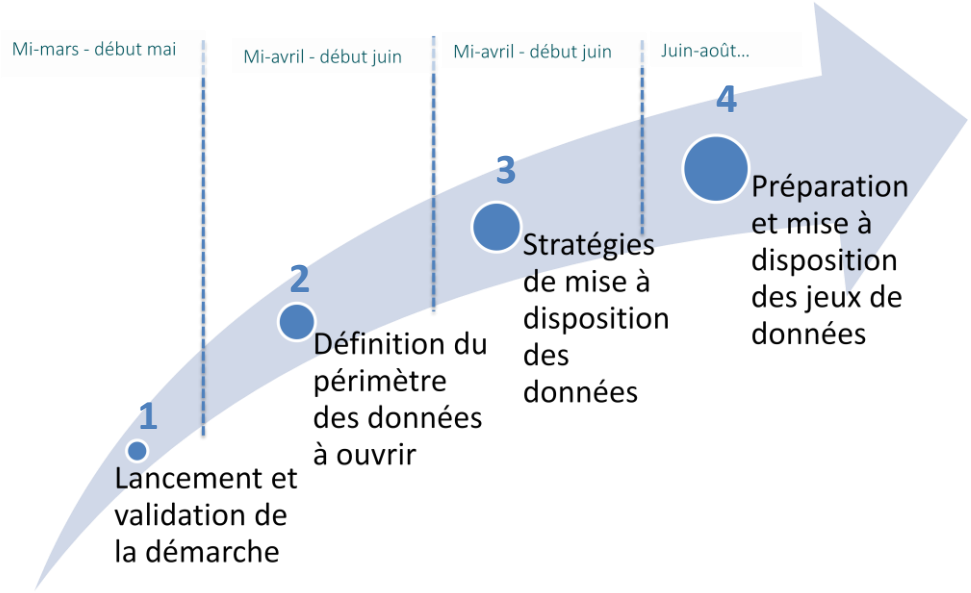


Figure 17 Etapes clés de la démarche Open Data

Source : Alexandra Greco, 2017

2.3.2.1. Lancement et validation de la démarche

Pour cette première phase, l'appropriation du contexte Open Data et de son cadre juridique est primordiale, notamment pour présenter l'initiative à la direction du Parc ainsi qu'à l'équipe technique. Passer en revue les recommandations nationales et analyser le cadre juridique ont été un moyen de valider l'éligibilité à la publication des données. Dans un premier temps, une brève sensibilisation sur l'Open Data auprès de l'équipe du Parc a été faite par mail. J'ai réalisé une frise chronologique en reprenant les dates clés des événements, notamment juridiques, liés à l'Open Data. J'ai fait le point sur ce qui se faisait déjà en matière d'Open Data en France mais aussi à l'étranger. Cela m'a permis de montrer des exemples concrets lors de mes réunions de sensibilisation réalisées auprès de chaque Pôle du Parc (Cf. Figure 5 et Figure 10). En parallèle de ces réunions, j'ai rédigé une délibération de principe pour l'ouverture des données publiques du Parc du Verdon ainsi qu'une annexe (Cf. Annexes n°3 et 4). Cette dernière a pour but d'expliquer ce qu'est l'Open Data et la démarche à suivre, tout en mettant en évidence les intérêts du Parc à ouvrir ses données publiques. En faisant de la veille informatique, j'ai rapidement identifié le projet d'expérimentation régionale OpenDataLocale auquel j'ai participé par la suite. Cette expérimentation est un programme national d'accompagnement des collectivités à l'ouverture des données publiques. Ce programme est coordonné par l'association OpenData France. Ainsi, la Région PACA et ses partenaires[39] sensibilisent, forment et accompagnent toutes les collectivités territoriales qui s'engagent dans une démarche Open Data. Avec mon tuteur de stage, nous avons donc participé à la première réunion de sensibilisation sur l'Open Data. Nous sommes rentrés ainsi en contact direct avec des acteurs clés,

[38] http://www.opendatalab.fr/laboratoire : check-list des prérequis, tutoriels, guides…

[39] Le CRIGE PACA, le Syndicat Intercommunal des Collectivités Territoriales Informatisées Alpes Méditerranée (SICTIAM), l'association OpenData France, les Conseils Départementaux, la Métropole Nice Côte d'Azur et la Fondation Internet Nouvelle Génération (FING).

notamment Mathieu Kazmierski[40] du CRIGE PACA, Pauline Asselin de Williencourt[41] et Thomas Bekkers[42] de la Région PACA. Après plusieurs relectures et modifications, la délibération de principe a été validée et sa signature a été planifiée pour passer en Bureau du Parc le 12 juillet 2017. A cette même date, la convention de partenariat avec la Région PACA pour la diffusion des données ouvertes et intelligentes a été également soumise à la signature (Cf. Annexe n°5). Cette date du Bureau marque la validation officielle de la démarche, même si une délibération de principe n'est pas un document obligatoire lorsque l'on s'engage dans une démarche Open Data. La loi pour une République numérique ne l'exige pas. Cependant, c'est un moyen de communication et d'officialisation de la procédure. En s'inscrivant à l'expérimentation OpenDataLocale, j'ai pu appliquer au fur et à mesure les étapes de la mise en place de l'Open Data au sein du Parc, et les valider.

2.3.2.2. Définition du périmètre des données à ouvrir

Une deuxième étape a été de recenser les données disponibles au sein du Parc du Verdon. Deux méthodes ont été employées :

1) Faire l'état des lieux du gisement des données pour chaque Pôle du Parc, c'est-à-dire : accéder aux dossiers de chaque Pôle préalablement sélectionnés par le Responsable et recenser les données « propres ». Ce recensement s'est fait sous forme de tableur Excel dans lequel je renseigne entre autre si la donnée est brute, quel est son format, si le Parc est producteur de la donnée, quelle licence à apposer etc. (Cf. Annexe n°6).

- Cette démarche a été très laborieuse, du fait de la quantité des données et du manque, voire de l'absence des métadonnées. N'étant pas experte dans chacune des thématiques, il m'était impossible de remplir des fiches de métadonnées à partir des données dont je disposais. Cela nécessitait un retour systématique vers le Responsable du Pôle. Faute de temps et d'efficacité, une seconde méthode a été mise au point.

2) Chaque Pôle sélectionne des jeux de données « propres » dont il est le producteur/ propriétaire (minimum 3 jeux de données par Pôle). Pour chaque donnée sélectionnée, les agents doivent créer sa métadonnée selon deux modèles (Cf. 2.3.2.3) :

- Modèle « SIT PNR PACA »[43] (modèle INSPIRE proposé par le CRIGE PACA) et/ou à l'avenir le modèle « OpenData » (norme ISO 19115) sous GeoNetwork (Cf. 2.3.2.2) ;

- Modèle « OpenPACA »[44] (réalisé à partir de la plateforme régionale de catalogage de données).

Une règle de nommage des fichiers de métadonnées a été instaurée à destination de chaque Pôle, à savoir : nommer la fiche de métadonnée « meta_nomdufichier », et l'enregistrer au format .XLSX (Excel) ou .ODS (LibreOffice) sous le même dossier où se trouve la donnée brute.
Cette méthodologie a été explicitée dans la Charte Géomatique Interparc en y ajoutant une partie « Métadonnées » (Cf. Annexe n°11). Avec cette deuxième méthode, après la réception des métadonnées, j'y inclus également une concertation avec l'équipe pour définir la licence à apposer à la donnée ainsi que le format des données (Cf. 2.2.3.1).

[40] Responsable INSPIRE et Open Data
[41] Editrice des données publiques – Open Data –
[42] Chef de projet Open et Smart Data – Service Smart Région -
[43] Cf. Annexe n°12
[44] Cf. Annexes n°13 et 14

2.3.2.3. Stratégie de mise à disposition des données

Les échanges avec différents acteurs clés en matière d'Open Data, ou du moins de mise à disposition des données, m'ont permis de travailler une stratégie pour l'ouverture des données publiques du Parc du Verdon. Par exemple, mon échange avec Julie Delauge[45] du CEN PACA a été nécessaire pour comprendre le cas particulier de la mise à disposition des données naturalistes. D'après Julie Delauge, *« s'il n'existe pas officiellement de droit d'auteur pour les données naturalistes, c'est le cas philosophiquement* » (entretien du 24/05/17). Une observation naturaliste a une valeur scientifique, et elle ne vaut rien sans son observateur. L'anonymisation de ces données serait donc inenvisageable du fait de l'envergure de la propriété intellectuelle dans le monde scientifique. Ce besoin de reconnaissance de la part du naturaliste et la volonté de connaître le devenir de réutilisation de sa donnée constitue un frein dans une démarche Open Data où l'on se doit d'ouvrir toutes les données par défaut. C'est sans compter l'exception relative aux données sensibles à caractère environnemental. Par conséquent, les données naturalistes continueront d'être mises en ligne *via* la plateforme SILENE (Cf. 2.2.3.3) afin de bénéficier d'une expertise et d'une validation scientifique.

Concernant les autres données non naturalistes, elles seront mises à disposition *via* la plateforme régionale OpenPACA gratuite pour ses partenaires[46]. Le Parc du Verdon bénéficie ainsi d'un moyen de diffusion pour tout type de données, sans coût de maintenance. En termes de visibilité du Parc dans la démarche Open Data, cela montre également une cohérence avec la dynamique régionale et une homogénéisation de mise à disposition des données par rapport aux autres partenaires d'Open PACA. En 2018, avec la nouvelle plateforme IDGO, le Parc bénéficiera de nouveaux services, notamment pour les données géographiques. En effet, deux niveaux de catalogage de données seront proposés par cette plateforme (Cf. 2.2.3.3) : un catalogue de données géographiques et un catalogue pour tout autre type de données. Ce premier catalogue génère des métadonnées respectant la norme ISO 19115 suggérée par la Directive INSPIRE, pour faciliter l'échange des informations relatives à l'environnement.

Si une des commandes du stage était d'installer un catalogue de données, notamment afin de l'installer sur le serveur interparc, le faire en local a permis de créer des métadonnées respectant la Directive INSPIRE et de réaliser des tests avec le CRIGE PACA (Cf. Partie 3). Cette stratégie a donc évolué lorsqu'une autre solution énoncée auparavant nous a été proposée par la Région et le CRIGE PACA. Ceci-étant, l'idée d'installer un catalogue de données (GeoNetwork), sur le serveur local du Parc du Verdon, est toujours envisageable afin que chaque agent puisse remplir ses métadonnées *via* le catalogue directement.

D'autres éléments ont contribué à une stratégie de mise à disposition des données, notamment le développement d'un outil d'aide à la décision pour l'ouverture des données publiques, adressé aux agents du Parc. Cet outil a pour unique objectif de guider les professionnels du Parc à mettre en conformité leurs données avant de les diffuser, ainsi qu'à les aider à choisir la licence qui conviendrait le mieux à leurs données[47]. Enfin, une autre stratégie a été de rédiger des clauses à intégrer dans les marchés publics (CCAP/CCTP[48]) relatives à la mise à disposition des données en Open Data (Cf. Annexe n°10). Une fois l'élaboration d'une stratégie de mise à disposition des données, il a fallu préparer les jeux de données pour les diffuser.

---

[45] Adjointe à la Direction – Responsable Connaissance et Programmes & Responsable du Pôle Biodiversité régionale
[46] J'ai rédigé la page Web partenaire du Parc qui a été validée par la chargée de communication, Marlène Economidès.
[47] Lien vers la plateforme d'aide à la décision dans l'ouverture des données publiques pour les PNR.
[48] Cahier des Clauses Administratives Particulières marchés publics / Cahier des Clauses Techniques Particulières

2.3.2.4. *Préparation et mise à disposition des jeux de données*

Cette dernière étape consiste à vérifier la qualité et le format des données, les préparer afin de les mettre à disposition sur la ou les plateforme(s) définie(s). Cette étape s'est concrétisée à partir d'un jeu de données test : [les panneaux du patrimoine bâti sur le territoire du Parc du Verdon]. Le fichier d'origine est un *shapefile*, une couche de points. Le contenu de ce jeu de données du Pôle Sites, paysages et aménagement du territoire a été vérifié au préalable par les producteurs de la donnée. J'ai ensuite vérifié, par exemple, s'il n'y avait pas de problèmes d'affichage des caractères pour les champs contenant du texte (encodage en Utf8) etc. Le fichier d'origine a été converti au format .CSV en y intégrant les champs X et Y de géolocalisation des points, ainsi qu'au format geoJSON. La Licence Ouverte a été apposée à ce jeu de données, qui a été ensuite mis en ligne sur la plateforme OpenPACA, à la suite de la création de sa fiche métadonnée *via* le modèle « OpenPACA » (Cf. Annexes n°13 et 14). L'étape ultime de la mise à disposition des données *via* la future plateforme IDGO est remplie en grande partie par la Région et le CRIGE PACA, à savoir :

- **L'évolution des modes de fonctionnement :** formaliser des évolutions à apporter aux processus pour assurer la production, la mise en qualité et la mise à disposition des données ;
- **L'animation de l'écosystème :** avec le suivi des téléchargements et des réutilisations, le développement de liens entre les jeux de données ou autres pour enrichir le contexte. La sensibilisation des acteurs locaux pouvant utiliser ces données est un point auquel le Parc devra porter attention, notamment grâce à la communication *via* les *Newsletters*, la page *Facebook* du Parc et autres réunions de sensibilisation internes et interparc.

Le schéma ci-dessous récapitule la chaîne de traitement à adopter par les Parcs de PACA qui souhaitent intégrer la dynamique Open Data, depuis la donnée brute jusqu'à sa mise à disposition :

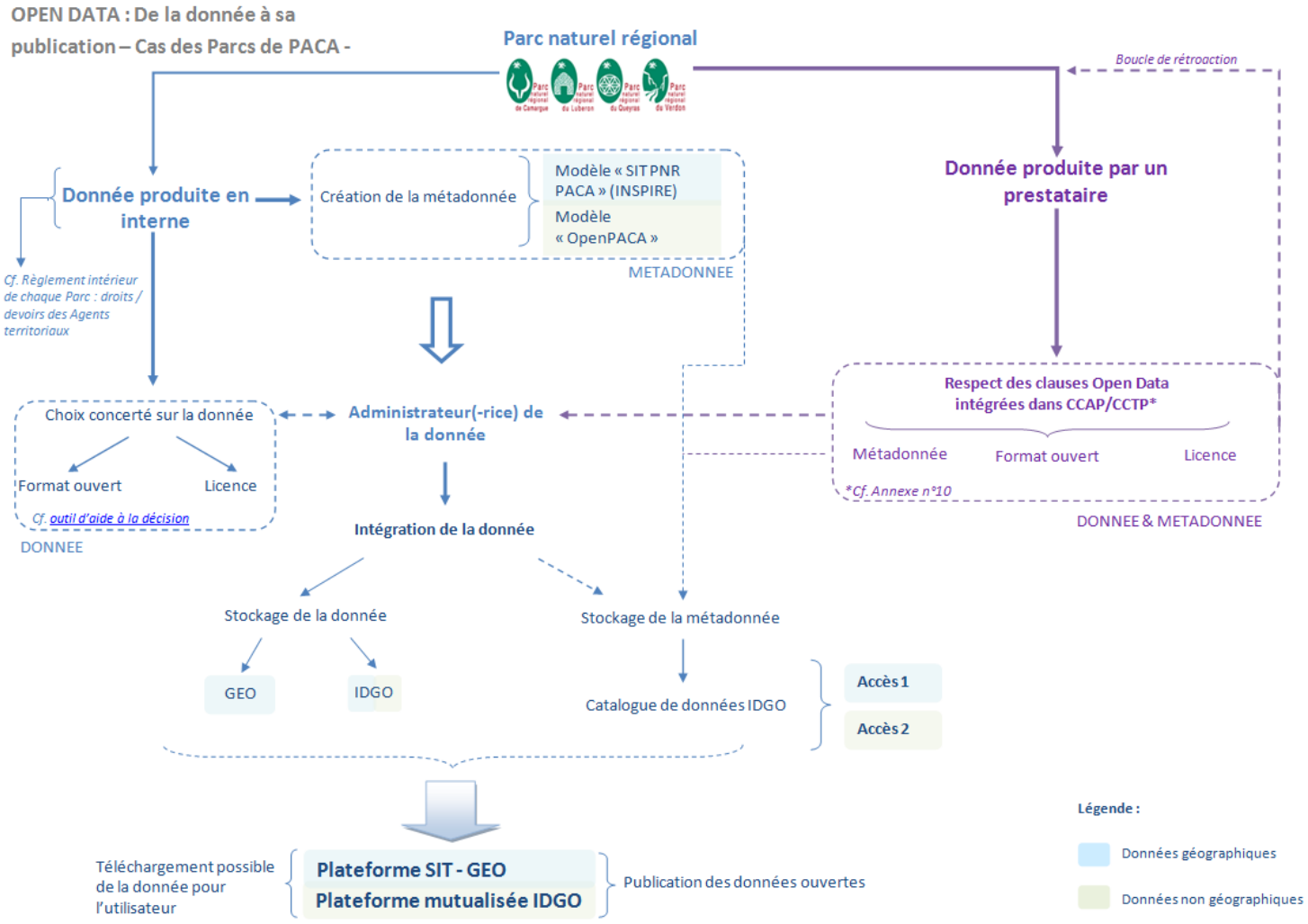


Figure 18 Chaîne de traitement – De la donnée à sa publication Source : Alexandra Greco, 2017

Toute la démarche Open Data mise en place au sein du Parc du Verdon a été retracée, mais un de mes objectifs durant ce stage a été de mettre en place également un catalogue de données afin de répondre à l'interopérabilité des données.

3. MISE EN PLACE DE L'OUTIL DE CATALOGAGE DES DONNEES

3.1. Interopérabilité des données

Rendre les données du Parc interopérables, c'est faire en sorte qu'elles soient réutilisables sans problème d'exploitation. Il s'agit de renseigner les informations sur la donnée concernée et les normaliser, répondant ainsi à la Directive INSPIRE, notamment pour les données géographiques. Il existe plusieurs niveaux d'interopérabilité[49] :

- **Juridique :** respecter le cadre légal et les accords contractuels ;
- **Organisationnel :** définir les rôles et les responsabilités, les processus d'échanges ;
- **Sémantique :** s'entendre sur la signification des données échangées ;
- **Technique :**
    - *Interopérabilité de transport :* définir des protocoles d'échanges de données (exemple du protocole de communication client-serveur HTTP) ;
    - *Interopérabilité syntaxique :* choisir les formats des fichiers de données (exemple de syntaxe standardisée et normalisée : le XML (Cf. 3.3). Ou vers l'utilisation du triplet RDF pour encoder les données d'une ressource).

La Directive INSPIRE est cohérente avec la démarche Open Data lorsque l'on élargit la question de l'interopérabilité à tout type de données.

3.1.1. Qu'est-ce qu'une métadonnée ?

Toutes les données peuvent être enrichies d'informations de contextes portées à connaissance dans un fichier texte, à savoir des données contextualisées, des « métadonnées ». Littéralement, les **métadonnées** sont les données sur les données. La métadonnée structure l'information, la décrit, l'explique, la localise etc. Elle répond à : quoi, pourquoi, quand, comment, où … La métadonnée est importante pour trouver et accéder à l'ensemble des données, mais aussi pour apprendre à les utiliser correctement, et respecter leur interopérabilité. L'*Open Data Support* dans son « Introduction à la gestion des données » définit les métadonnées comme fournissant *« des informations permettant de comprendre les données (documents, images, bases de données), des concepts (par exemple, les systèmes de classification) et des entités du monde réel (par exemple, les personnes, les organisations, les lieux, peintures, produits ».* Les métadonnées peuvent être descriptives, structurelles ou encore administratives. En résumé, dans le contexte actuel de l'Open Data et de la Directive INSPIRE, ces métadonnées sont importantes pour :

- ✓ l'interopérabilité (une meilleure identification des données) et une description universelle des données ;

[49] Cf. notes des cours (décembre 2016) de Wilfried Heintz et d'Emilie Lerigoleur dans le cadre la formation SIGMA 2016-2017

- ✓ la réutilisation (contextualisation des données, conditions d'utilisation, vers la *Data science*[50]) ;
- ✓ la pérennité (sauvegarde caractérisée par une diffusion facilitée, et donc une meilleure portée des connaissances) ;
- ✓ le respect de la loi avec la Directive INSPIRE qui l'exige.

Il existe de nombreuses normalisations pour les métadonnées (Dcat, W3C, ISO, OGC : consortium…)[51]. En effet, pour créer des métadonnées, il faut respecter des normes, avoir des logiciels appropriés et y consacrer du temps. Une des logiques à respecter est que la personne qui a collecté la donnée se doit de créer sa métadonnée. Concrètement, pour créer sa métadonnée, on peut utiliser un outil "maison" ou un logiciel existant, soit un outil international (FAO[52]) comme GeoNetwork qui implémente tous les standards des formats (OGC) des logiciels. Les métadonnées peuvent être donc des fichiers écrits en XML, JS ou encore des fichiers au format .XLS(X). Dans la même idée de partage que les métadonnées, les données liées contribuent également à une meilleure analyse.

3.1.2. Normes et standards sur les données

Pour mettre en place ces différentes interopérabilités, il faut normaliser et standardiser les données. **La norme** fait référence surtout à l'*International Organization for Standardization* (ISO) et à ses instances nationales comme l'Association française de normalisation (AFNOR), ou encore le Comité européen de normalisation (CEN), avec des processus de validation assez lourds. D'après la définition officielle ISO, une norme est un *« document établi par un consensus et approuvé par un organisme reconnu, qui fournit, pour des usages communs et repérés, des règles, des lignes directrices ou des caractéristiques, pour des activités ou leurs résultats, garantissant un niveau d'ordre optimal dans un contexte donné».* **Le standard** est plus assimilé à un processus réactif de consensus du monde économique ou du monde technique. Un standard est un *« ensemble de recommandations développées et préconisées par un groupe représentatif d'utilisateurs »* (Eduscol). Par exemple, pour des organismes comme le W3C, l'enjeu est un accord consensuel pour le développement rapide du commerce. Pour réutiliser du vocabulaire existant pour créer des métadonnées, on a donc différentes normes :

Normes internationales	Normes standards
Référentiel consensuel approuvé par l'organisme **officiel**	Publiées par un groupe **non officiel** mais représentatif d'utilisateurs.
ISO, AFNOR, CEN, …	Organismes privés publiant des standards ouverts comme W3C, OGC (reconnus par INSPIRE)

Figure 19 Comparaison entre les principales normes internationales et standards

Les principaux standards à retenir sont les suivants :

- **Pour tout type de données :** depuis mars 1995, le *Dublin Core* est un standard de métadonnées consensuel, établi par des professionnels de différents domaines (informatique, communauté muséologique etc.). *Dublin Core Metadata Initiative* (DCMI) ou *DCMI Core Metadata Element Set*

[50] Principes de la *Data Science* : fouiller, extraire, raffiner et transformer. On part de données variées (issues de sources hétérogènes), donc orphelines, pour les lier, grouper, croiser et les compléter etc. Un des objectifs est donc de structurer le *Big Data* (ou données massives). Ce *Big Data* désignent « des ensembles de données devenus si volumineux qu'ils dépassent l'intuition et les capacités humaines d'analyse et même celles des outils informatiques classiques de gestion de base de données ou de l'information » (Wikipédia).

[51] Cf. 3.1

[52] Fabrication assistée par ordinateur

(ISO Standard 15836 :2009 et ses 15 champs). Le DCMI *Metadaterms* est validé par le W3C. Cette norme est intéressante car elle est minimaliste en termes de nombre de champs descriptifs de la donnée.

- **Pour les données « biodiversité » :** *Darwin Core* (DWC) produit par *Taxonomic Databases Working Group* (TDWG). Utilisé par le GBIF (Système mondial d'information sur la biodiversité). Ou encore l'*Ecological metadata language* (EML).

- **Pour les données élémentaires d'échange** (DEE) : «occurrences de taxons» (OccTax). Cf. l'*Extract-transform-Load* (ETL) : faire un *mapping* avec une correspondance entre les champs.

- **Pour les données géographiques :** la plus utilisée pour les métadonnées géographiques et recommandée par la Directive INSPIRE est l'ISO 19115. C'est une norme conceptuelle et internationale depuis 2003.

- **Pour les métadonnées de services** : ISO 19119 (complète l'ISO 19115).

- **Pour la transcription XML :** ISO 19139 (complète l'ISO 19115).

Pour partager des données, sous le standard *Dublin Core* :

1) Titre
2) Créateur (auteur de la ressource)
3) Sujet
4) Description
5) Editeur (en charge de la publication)
6) Contributeur (personnes impliquées)
7) Date (dimension temporelle de la ressource) : elle peut se décliner en sous-champs : date début, date création, date de mise à jour...
8) Type (nature ou genre du contenu de la ressource)
9) Format (physique de la donnée)
10) Identifiant (URI/UUID/ID/DOI = ID unique) : problème d'un id unique pour le partage à l'échelle de l'Europe.
11) Source (généalogie de la source)
12) Langue (langage de ressource)
13) Relation (données liées)
14) couverture (étendue géographique)
15) Droits (d'utilisation)

[Une URI (*Uniform Resource Indicator*) peut être une URL (Uniform Resource Locator). Un DOI (Digital Object Identifier est à la fois une URI et une URL].

Il faut savoir que le W3C prend en charge la normalisation de l'ensemble des protocoles Internet :

- « standards de base : HTTP, HTML, DOM, XML, XSL…
- standards autour de l'interopérabilité et des services Web : SOAP, WDSL et Web Services
- standards concernant les documents et le multimédia : HTML, XML, CSS, SMIL, VML, MathML, SVG
- standards concernant l'accessibilité : WAI
- standards liés à la sémantique et à la description de ressources : XML Schema, RDF (norme développée dans le cadre du W3C *Metadata activity* pour faciliter le traitement des métadonnées), langages d'ontologies OWL, et tout ce qui tourne autour du Web sémantique » (source : Dominique Cavet / SCEREN[53]).

De fortes recommandations et obligations sont faites par la Directive INSPIRE en termes de données géographiques. Cette dernière propose des règles dans le domaine sémantique grâce à des modèles de données par thème.

[53] http://eurolibnet.eu/files/REPOSITORY/20061013020036_050908metadata.pdf

3.1.3. La Directive INSPIRE et la cohérence avec l'Open Data

Les différentes interopérabilités sont importantes dans plusieurs domaines, notamment juridique. La Directive INSPIRE a constitué un élan sur l'interopérabilité avec comme but d'*«établir une infrastructure géographique permettant la diffusion et le partage de données géographiques dans la Communauté européenne pour favoriser la protection de l'environnement »* (Journal officiel de l'Union européenne, 2007)[54]. Les autorités publiques sont dans l'obligation de mettre à disposition et partager entre elles des données géographiques. Ces dernières sont définies à travers 34 thèmes[55]. La Directive INSPIRE a donc pour but d'encourager les pays à créer une IDG à l'échelle de la Communauté européenne. Pour publier les données en conformité à INSPIRE, il faut donc :

a) **Créer et gérer les métadonnées :** norme internationale ISO 19115 relative aux métadonnées géographiques est suggérée par la Directive ;

b) **Publier ces métadonnées sur le Géocatalogue national** (réunit toutes les métadonnées des données INSPIRE). Lorsqu'une autorité publique utilise déjà son propre catalogue de données, un moissonnage peut être mis en place par le Géocatalogue national pour récupérer automatiquement des métadonnées. C'est notamment le cas du Géocatalogue du CRIGE PACA ;

c) **Publier les données :** INSPIRE exige que les données en question soient téléchargeables (le Géocatalogue national centralise également les liens de téléchargement par moissonnage) ;

d) **Proposer des services de consultation des données :** les catalogues de données moissonnées par le Géocatalogue national doivent proposer les données *via* un service WMS.

Ces différents éléments montrent que la Directive INSPIRE va dans le sens d'une plus grande ouverture des données publiques, notamment environnementales, et en particulier géographiques. Si cette directive ne concerne que l'information géographique, elle va au-delà de la communication sur demande en imposant d'une façon générale la publication sur Internet. De plus, l'idée d'utiliser la norme suggérée par INSPIRE pour les données géographiques dans le cadre d'une démarche Open Data s'est concrétisée. L'AFIGEO[56] et Etalab ont réfléchi conjointement à la valorisation des données géographiques ouvertes sur le portail national data.gouv.fr. En effet, en avril 2014, ce groupe de travail « OGC Open Data » (AFIGEO et ETALAB) constate que *« la grande majorité des données géographiques soumises à la directive INSPIRE sont ouvertes ou devraient l'être, et elles sont de plus en plus nombreuses à être documentées, référencées et téléchargeables à l'aide de protocoles communs »* (Le blog d'Etalab, 2015)[57]. La passerelle INSPIRE et Open Data est officiellement lancée avec le site inspire.data.gouv.fr. Si le Parc du Verdon, producteur de données au « format Inspire », peut partager ses données sur data.gouv.fr, le but était de trouver la stratégie la plus adaptée en termes d'Open Data pour l'ensemble de ses données (y compris non géographiques), et de l'intégrer dans une réflexion collective interparc.

L'enjeu de cataloguer les données du Parc est évident en termes d'interopérabilité. C'est pourquoi à présent je vais exposer toutes les solutions, notamment techniques, qui ont été envisagées pour le catalogage de données.

---

[54] http://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:32007L0002

[55] Une des difficultés est de savoir si une donnée traitée relève d'un des 34 thèmes INSPIRE, plusieurs documents sont disponibles en ligne pour aider cette détermination.

[56] Association française pour l'information géographique.

[57] https://www.etalab.gouv.fr/data-gouv-fr-invite-le-standard-inspire-dans-lopen-data

3.2. Solutions envisagées – GeoNetwork –

Plusieurs catalogues de données existent, mais en tenant compte de certains facteurs, seuls trois outils avaient retenu mon attention. Les propriétés de ces outils seront détaillées brièvement, ainsi que mon choix d'utiliser le catalogue GeoNetwork.

3.2.1. Les outils de catalogage

Des logiciels de gestion de métadonnées existent spécifiquement pour les données géographiques. Le but est que toute personne concernée puisse avoir connaissance de la description du contenu du stock de données. L'échange et la diffusion de l'information de manière standard sont possibles grâce à la prise en compte par ces logiciels des normes ISO 19115 et 19139. Cela évite la dispersion des fiches de métadonnées, et les rend plus accessibles aux utilisateurs. En ce qui concerne les gestionnaires de métadonnées, on peut trouver notamment GeoNetwork et Géosource. Mais d'autres solutions de catalogage existent qu'elles soient libres ou non, ou en développement. L'idée ici est de voir les solutions les plus intéressantes dans une démarche interparc concernant en premier lieu le Parc du Verdon.

3.2.1.1. GEO

Il s'agit d'une solution payante. GEO est la nouvelle plateforme géomatique de *Business Geografic*. Le Parc était jusqu'à présent sur Dynmap, plateforme développée également par le prestataire SIG *Business Geografic*. GEO a recours à des techniques et des solutions pour générer, diffuser et partager les données en mode HTML5. Ses atouts résident donc dans son interopérabilité, ses fonctionnalités et sa puissance. Son utilisation garantit aux utilisateurs un hébergement Haute Sécurité en mode SaaS grâce aux datacenters SynAApS certifiés ISO 27001 : 2013 et agréés Sabté de *Business Geofabric* et Ciril GROUP[58]. Ce GEO est l'IDS des 4 Parcs de PACA participant au SIT. L'option catalogage de données existe sous GEO, mais il s'agit d'un service payant.

3.2.1.2. Géosource

C'est une solution *Open Source* de gestion des métadonnées, correspondant à la version française de GeoNetwork, et mise à disposition par le Bureau de Recherches Géologiques et Minières (BRGM). Cet outil fonctionne avec une application n-tiers Web et un serveur d'application Java :

- Serveur Web : Apache (nécessaire pour publication des pages Web) ;
- Serveur d'application Java : Tomcat, Jetty … (permet d'interpréter les pages .JSP) ;
- Un SGBD (Système de gestion de bases de données) : MySQL, Postgres …

Le Géosource peut être installé de 2 façons différentes :

- « **Installation monoposte** :

Déploiement sur une machine physique et pour un usage restreint. Une fois téléchargé, il suffit de lancer le fichier .jar (vérifier au préalable que l'ordinateur possède Java) et suivre les instructions. Par défaut sont installés : Jetty (serveur Web et d'application Java) et McKoi (serveur de base de données où sont stockées les fiches de métadonnées). Si on n'apporte aucune modification aux paramètres, Géosource sera utilisable rapidement.

- **Installation serveur** :

Prévue pour un déploiement en mode Intranet/Internet. L'installation serveur n'utilise pas tous les composants de base. En l'occurrence, le couple Jetty/McKOI sont les composants de base utilisés. Géosource peut faire plusieurs combinaisons possibles : Oracle/Jetty, Postgres/Tomcat,

[58] Source : https://www.business-geografic.com/fr/geo-software/geo-solutions.html

McKOI/Tomcat etc. L'idée est donc de faire des liens entre le SGBD et le serveur d'application Java dans les fichiers adéquats. » (François, J., 2014).
Ce Géosource a été installé sur le serveur interparc. Suite aux dysfonctionnements rencontrés, la question a été de savoir s'il ne serait pas préférable d'installer directement GeoNetwork.

3.2.1.3. GeoNetwork

C'est une autre solution de catalogage *Open Source* basée sur de nombreuses normes internationales pour le stockage et l'échange de métadonnées (ISO, OGC, etc.). Il s'agit d'une application en fort développement. Ce type d'infrastructure vise une meilleure connaissance des données géospatiales afin d'en améliorer leur gestion, consultation et diffusion, sans corrompre la propriété de la donnée ou ses droits d'accès et/ou droits d'usage. Ces fonctionnalités sont les suivantes :

- « Recherche par mots clés, par date ou par localisation spatiale de métadonnées, présentes sur des catalogues locaux et/ou distribués.
- Les métadonnées récemment mises à jour sont accessibles par flux RSS/GeoRSS.
- Récupération, au format XML, PDF ou texte, d'une sélection de métadonnées.
- Gestion des groupes et des utilisateurs avec un contrôle des accès (aux données, aux métadonnées, aux Web Services, etc.) totalement paramétrable.
- Gestion des catégories de métadonnées et gestion des thesaurii importés au format SKOS[59].
- Rédaction en ligne de métadonnées en lien avec un système de modèles par défaut (compatibles ISO19115/ISO19119/ISO19139/ISO19110 et ISO Profiles, FGDC[60] et *Dublin Core*) ou entièrement défini par l'administrateur. La saisie d'information sur les données attributaires est gérée au sein de GeoNetwork.
- Création possible de fiches de métadonnées.
- Chargement et téléchargement de données et de documents associés.
- Visualisation interactive de données raster et vectorielles à partir de *Web Map Services* (WMS) de serveurs distribués dans le monde entier (*NASA Earth Observations*, Demis, etc.).
- Moissonnage possible entre géocatalogues distribués, répondant aux standards en vigueur (c'est-à-dire de type GeoNetwork, CSW, Z39.50, OGC WxS, WebDav, Thredds, Local filesystem, OAI-PMH).
- Ajout du logo de la structure afin d'identifier d'un coup d'œil les métadonnées propres à un établissement.
- Récupération d'une métadonnée au format XML ou d'un ensemble de métadonnées dans un fichier ZIP.
- Import/Export des métadonnées avec l'outil GAST afin de créer une sauvegarde complète du jeu de métadonnées ou afin d'implémenter un nouveau géorépertoire.
- Interface utilisateur multi-langues (anglais, catalan, japonais, allemand, espagnol, français, néerlandais, portugais, grec, turc).» (François, J., 2014).

GeoNetwork est basé sur les standards internationaux ouverts pour les services et protocoles comme ISO-TC211 et les spécifications de l'OGC, ainsi que les standards internationaux pour la

[59] *Simple knowledge organization system*
[60] *Federal geographic data committee*

documentation (*Dublin Core*). Pour la visualisation des cartes sur Internet, GeoNetwork utilise aussi des standards (OGS WMS).

Logiciels nécessaires à installer :

- Par défaut : *Java Runtime Environment* (JRE), container Java (Jetty par défaut ou Tomcat), base de données compatible JDBC[61] (McKoi par défaut ou MySQL, PostreSQL, PostGIS, Oracle, SQLServer) ;
- GeoServer. Mais il est possible d'utiliser d'autres serveurs WMS/WFS, voire des solutions propriétaires (MapServer, ArcGIS Server etc.) ;

Logiciels en complément :

- Luke pour visualiser le contenu de l'index Lucene ;
- Druid pour inspecter la base de données.

La pérennité du projet GeoNetwork repose sur différents éléments :

- Le soutien d'organisations reconnues sur la scène internationale ;
- La communauté de développeurs et d'utilisateurs est en plein essor.

3.2.1.4. geOrchestra

L'entreprise *Camptocamp* est porteuse du projet. Il s'agit d'un projet vivant et libre (licence GNU GPL)[62] donc réutilisable gratuitement. Plusieurs collectivités ont déjà mis en place un geOrchestra. Dans le cadre de la démarche interparc, ce geOrchestra a été envisagé mais la question était de savoir s'il allait remplacer GEO, déjà installé sur le serveur, ou s'il serait une solution complémentaire.
C'est un ensemble de logiciels qui forment une IDS. Du fait de sa structure en brique, on peut utiliser ou non chaque logiciel sans nuire au fonctionnement des autres. Les fonctionnalités principales sont les suivantes :
- « Catalogue de données (GeoNetwork) ;
- Extracteur de données (Extractorapp) ;
- Visualiseur cartographique (Mapfishapp) ;
- Serveur cartographique (Geoserver) ;
- Logiciel de gestion du cache (Geowebcache) ;
- Gestion des droits d'utilisateur (Idapadmin) ;
- Service d'authentification (CAS) ;
- Proxy interne basé sur Spring security (ROOT) ;
- Statistiques de consultation et de téléchargement (Analytics & Downloadform).» (François, J., 2014).

Les sources de geOrchestra sont accessibles *via* un dépôt GitHub. De plus, la distribution Debian de Linux est préconisée pour installer geOrchestra. Si geOrchestra semble être l'outil *open source* le plus complet parmi ces solutions envisageables, plusieurs raisons justifient notre choix d'installer GeoNetwork.

[61] *Java database connectivity*
[62] *General public licence*

3.2.2. Choix de la solution GeoNetwork

3.2.2.1. En lien avec les besoins et les moyens de la structure

Mon choix s'est tourné vers l'installation de GeoNetwork pour différentes raisons : d'un point de vue matériel, je n'avais accès à aucun droit d'administration sur la machine en local ni sur le serveur interparc. De plus, installer une machine virtuelle m'a été refusé pour différentes raisons justifiées. J'ai donc utilisé une machine test (en local), sous le même environnement d'exploitation informatique que le serveur interparc, à savoir sous Debian 9. L'idée était de tester l'installation et le paramétrage du catalogue de données avant qu'il puisse être installé sur le serveur interparc. A la suite de ma formation, j'avais acquis certaines notions informatiques, notamment sous Linux. Le fait de passer sous un environnement Debian m'a permis de monter en compétence dans ce domaine. Enfin, le fait que le Parc dispose déjà d'une IDS, avec GEO, a été décisif dans le choix de l'outil de catalogage à installer. En effet, le Parc bénéficie de nombreux services de GEO à l'exception du catalogue de données. geOrchestra est une solution très intéressante mais suite aux recherches et aux conseils reçus, son installation s'avérait complexe et chronophage par rapport à la durée de mon stage. De plus, cette dernière solution engendrait une perte de temps et de bénéfices pour le Parc qui bénéficie déjà d'une IDS et de sa maintenance. Cette solution ne répondait pas non plus à la commande initiale du stage. Au lieu de partir de zéro, et après concertation avec mon tuteur de stage, le choix s'est donc porté sur l'installation du catalogue de données GeoNetwork, qui complète ainsi l'IDS déjà en place.

3.2.2.2. Bilan sur les alternatives de déploiement de l'outil : Docker – en local – sur le serveur

Les moyens matériels dont je disposais m'ont donc obligée à envisager l'installation du catalogue de données sous différents angles. Au début de mon stage, je me suis concentrée sur l'installation de GeoNetwork *via* Docker[63] à partir de l'ordinateur personnel de mon tuteur de stage.

Docker est un outil libre permettant la création d'un environnement (appelé conteneurs) pour isoler des applications pour ne pas en gêner d'autres. Des fonctionnalités natives au noyau Linux sont utilisées par Docker. Ce dernier permet de simplifier l'utilisation de ces outils présents dans le noyau Linux comme les *namespaces* et le *cgroups.* Contrairement aux machines virtuelles traditionnelles, un conteneur n'inclut pas de système d'exploitation, s'appuyant sur les fonctionnalités du système d'exploitation fournies par l'infrastructure sous-jacente. En comparant avec les machines virtuelles, la finalité est pratiquement la même mais le fonctionnement de Docker est totalement différent. Pour les conteneurs, on installe un *rootfs* qui est appelé image, et non pas une OS (*Operating system* ou système d'exploitation) à proprement parler. Cette image contient les librairies et les binaires nécessaires. Quant au noyau, il est partagé avec le système hôte.

Déploiement	Développement	Installation d'applications
Créer un conteneur pour sa propre application et la répartir où bon nous semble. Un conteneur fonctionne sur une machine avec une distribution X, et fonctionnera sur une autre machine de distribution Y.	Facile d'avoir le même environnement de développement qu'en production. Possibilité sur la même machine de tester avec plusieurs versions d'un même logiciel.	Installer rapidement une application. Le plus souvent une seule ligne de commande suffit.

Figure 20 Cas d'utilisation possibles de Docker

Un mois complet a été consacré à l'apprentissage et à la familiarisation de Docker mais deux éléments rendaient la situation laborieuse : accès au débit Internet extrêmement faible[64] et travail

[63] Lien pour installer Docker sur une machine Debian.

sur l'ordinateur personnel de mon tuteur de stage. Des fiches sur la manipulation et l'usage de Docker ont été rédigées à titre personnel[65]. Une fois un ordinateur réquisitionné, je l'ai recyclé en le rebootant complètement afin de travailler sous un système d'exploitation Debian 9 comme prévu.

Travailler sur une machine en local m'a permis d'établir toute la chaîne de traitement de la mise en place d'un catalogue de données et de son paramétrage. Plusieurs tests ont été nécessaires avant d'aboutir à l'installation finale de GeoNetwork.

L'installation de GeoNetwork sur le serveur interparc n'est finalement pas prévue suite à la solution adoptée dans le cadre de la démarche Open Data interparc, à savoir utiliser la future plateforme IDGO. Ceci étant, son installation sur un serveur a été testée par le chargé de mission SIG du Parc naturel régional du Vexin français (Ile-de-France), Eric Marot (Cf. 3.3.1).

Rechercher toutes les solutions envisageables, s'adapter aux besoins et à l'environnement de travail ont été une étape cruciale avant de pouvoir mettre en place un catalogue de données.

3.3. Du catalogage au Web des données : le RDF

Cette dernière partie expose les résultats et les perspectives techniques en termes d'ouverture des données du Parc du Verdon. Après avoir installé et paramétré le catalogue de données GeoNetwork, j'ai réalisé des tests de création de métadonnées respectant la norme INSPIRE (ISO 19115). Si l'idée d'installer GeoNetwork sur le serveur interparc a été abandonnée récemment suite au choix de cataloguer les données du Parc *via* la solution IDGO, cela a permis de tester la validité des métadonnées (INSPIRE) créées à partir du catalogue de données, installé en local. Cette validation a été faite avec le CRIGE PACA afin de vérifier qu'ils puissent lire et intégrer les métadonnées du Parc au format XML dans leur propre catalogue de données. Ainsi toutes les métadonnées créées pour les premiers jeux de données à ouvrir du Parc ont été réalisées en respectant la norme INSPIRE pour les données géographiques, et en s'appuyant sur le modèle de métadonnées de la Région PACA pour les autres types de données. Officiellement, un seul jeu de données brutes du Parc est déjà mis à disposition sur la plateforme OpenPACA sous Licence Ouverte. Cinq autres jeux de données du Pôle Diffusion et six autres du Pôle Eau sont en prévision. Le Pôle Développement du Parc a des données brutes à ouvrir qui sont en cours également de validation. Prochainement, ces données seront mises à disposition également sur la plateforme OpenPACA, après vérification.

D'un point de vue technique, la démarche Open Data du Parc est également cadrée. Cela dit, il est intéressant de montrer aussi les perspectives en matière d'ouverture des données, notamment l'idée d'aller vers la publication de données liées.

3.3.1. Installation et configuration de GeoNetwork

A la suite de problèmes techniques, à deux reprises GeoNetwork a été installé sur la machine en local. Du fait du débit Internet faible, l'installation du paquet GeoNetwork a mis 30 minutes lors du premier essai. En moyenne, tous les traitements réalisés dans le cadre de l'installation de GeoNetwork sur une machine Debian en local (avec un faible débit Interne) prenaient au minimum 10 minutes. L'installation et la configuration du catalogue de données ont donc pris plusieurs semaines à cause de ces imprévus matériels. Pour palier à ces problèmes et rattraper le temps perdu,

---

[64] Par exemple, évaluation du débit le 17/08/17 : débit descendant = 0,17 Mbps – débit ascendant = 0,54 Mbps. Le Parc a deux lignes spécialisées dédiées chez Orange de 2 x 2 Mo partagées entre la téléphonie, la messagerie avec le serveur exchange du Parc et Internet.

[65] Cf. Annexe n° 15 + https://memo-linux.com/installer-docker-sur-gnulinux/

j'ai travaillé en équipe avec un de mes collègues de formation en géomatique, Lucien Duflocq, chargé de mettre en place un catalogue de données au sein du Parc naturel régional du Gâtinais français.

3.3.1.1. Guides d'installation et de paramétrage du catalogue de données

Deux guides d'installation et de paramétrage de GeoNetwork ont été ainsi rédigés avec la collaboration de Lucien Duflocq (Cf. Annexes n°16 et 17). Un Modèle conceptuel de données (MCD) du catalogue de données est également disponible (Cf. Annexe n°18). Le contenu de ces documents est mis sous Licence Ouverte. Vous pouvez accéder à toute la documentation à l'adresse suivante : https://github.com/GeoNetworkTeam

La documentation a été également partagée aux communautés de GeoNetwork et de GeoRezo afin d'avoir des retours des utilisateurs et qu'elle puisse être ainsi complétée et/ou améliorée.

3.3.2. Développement de la chaîne de traitement pour un Web de données

La conversion des données ouvertes du Parc au format RDF *(**R**esource **D**escription **F**ramework)* serait un pas vers le Web de données. Ce dernier est une des finalités optimale vers laquelle tend l'Open Data, notamment en termes d'interopérabilité.

3.3.2.1. Qu'est-ce que le Web de données ?

Dans le Web actuel, les liens entre tous les documents doivent être faits par des personnes physiques tandis que dans le Web des données, on a des données liées ou *linked data* utilisables par des machines[66]. Les technologies du Web sémantique et du Web actuel sont le fondement des données liées. L'idée est de partager des données structurées à une échelle globale selon les principes suivants :

1) Donner un nom, un identifiant universel non ambiguë, à savoir un URI[67] ;
2) Rechercher et/ou consulter ce nom en utilisant des URI HTTP ;
3) Utiliser des standards (RDF, SPARQL) pour fournir des informations ;
4) Inclure des hyperliens (liens vers d'autres URI pour découvrir d'autres éléments).

Les données liées se fondent sur l'architecture du Web actuel construit sur un ensemble de standards simples :

- URI ;
- HTTP (*HyperText Transfer Protocol*) : protocole de transfert hypertexte ;
- HTML (*HyperText Markup Language) :* langage de balisage hypertexte.

Le troisième principe des données liées repose sur un modèle de données unique pour publier des données structurées : le RDF.

3.3.2.2. Qu'est-ce que le RDF ?

« C'est un modèle de données simple fondé sur une structure en graphe conçue spécifiquement pour le contexte du Web » (Bizer C., Health T., 2012). Il s'agit d'un modèle qui permet de lier les données et il est notamment puissant pour le partage des métadonnées. Son élément de base est le triplet, il se compose donc d'un :

- **Sujet** : une ressource à décrire -> **Parc du Verdon** ;
- **Prédicat** : un type de propriété applicable à cette ressource -> **est créé** ;

---

[66] Pour aller plus loin : Gautier Poupeau, 2010

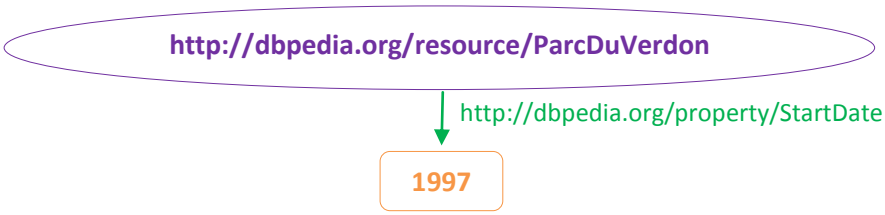
[67] C'est un système d'identification qui étend le principe d'URL et qui sert à identifier une ressource accessible sur Internet ou non.

- **Objet** : une valeur de la propriété -> **1997**.

Ci-dessous la représentation *via* des graphes du triplet pris en exemple :



Si on se base sur DBPedia, la version des données liées de Wikipédia, cela donnerait le graphe suivant :



Il est possible d'améliorer ce graphe[68], mais l'objectif ici est de voir comment se modélise les données si on souhaite les publier sur le Web de données.

On modélise donc des connaissances grâce au Web sémantique. Et ce Web repose sur deux standards du W3C pour la description respective des vocabulaires et des ontologies : RDFS (*Resources Description Framework Schema*) et OWL (*Web Ontology Language*). L'ontologie fixe le vocabulaire du domaine ainsi que les relations sémantiques entre les éléments du vocabulaire. La représentation des connaissances se compose de la façon suivante :

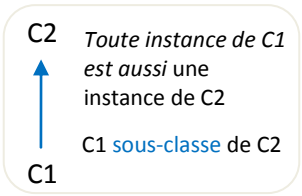


Figure 21 Instances d'ontologie

Concept/classe : catégories d'entités qui partagent certaines caractéristiques ;
Instance d'une classe : membre particulier de cette classe ;

Spécialisation : relation fondamentale entre classes (sorte de, sous classe).

Et il existe plusieurs propriétés des relations d'instance (Cf. Figure 21), comme par exemple : *hasVariety / hasPart / isPartOf / includes / isIncluded*

Les ontologies permettent l'inférence c'est-à-dire l'opération logique de déduction qui consiste, à partir d'indices présents dans le texte, à rendre explicite une information.

Quand une URI est déférençable par une URL, c'est qu'elle est accessible et donc référencée (attention ne pas se fier à la traduction de « déférençable » qui laisserait supposer qu'elle ne soit pas référencée, or, c'est bien le contraire qui est correcte).

Le Web sémantique pose ainsi le cadre d'interopérabilité pour mettre à disposition, consulter, lier et partager des données sur un réseau de machines connectées. En résumé, pour aller vers le Web de données, il faut :

Installer Datalift sous Linux

- Se mettre en mode root : su
- Cloner le dépôt Git existant de Datalift ou récupérer le dossier d'installation compressé (ici)
- Décompresser le dossier dans le répertoire souhaité
- Et lancer la commande suivante : sh datalift.sh ou sh./datalift.sh

Démonstration de l'utilisation de la plateforme Datalift

[68] Pour aller plus loin
& https://www.canal-u.tv/video/inria/introduction_aux_ontologies.20327

1) Convertir sa donnée en RDF grâce par exemple à des outils *open source* proposés par Datalift : c'est une plateforme qui publie et interconnecte les jeux de données sur le Web de données. De plus, Datalift propose un ensemble d'outils pour faciliter le processus de publication des jeux de données comme :
   - **Sélectionner** des ontologies pour décrire les données ;
   - **Convertir** les données en RDF en rapport avec la ou les ontologies sélectionnées ;
   - **Publier** les données sur le Web de données ;
   - **Interconnecter** les données avec d'autres jeux de données.

CKAN (*Comprehensive Knowledge Archive Network*) contribue également au Web de données ouvertes. Accessible *via* une plateforme publique sur le site ckan.org, CKAN permet de rechercher des données à partir de mots-clés. C'est un système Web pour le stockage et la distribution de données. Par ailleurs, la future plateforme IDGO sera sous CKAN.

2) Créer une ontologie[69] à partir des standards du Web sémantique : il existe des catalogues d'ontologies comme LOV (*Linked Open Vocabularies*). L'idée est d'utiliser des vocabulaires ouverts liées pour des données ouvertes liées. *« L'initiative LOV a pour objectif de favoriser et faciliter la réutilisation de vocabulaires pour la description de jeux de données sémantiques liés »* (Vandenbussche P.-Y., Vatant B., Charlet J., 2012). Grâce à ce projet, on a un jeu de données en RDF qui inventorie les vocabulaires descriptifs des jeux de données du Web sémantique, ainsi que les liens entre ces vocabulaires. Il existe d'autres catalogues d'inventaire d'ontologies comme *Open Metadata Registry* avec des vocabulaires au format SKOS, ou encore Schemapedia. ;

3) Interroger la donnée grâce à un langage de requêtes comme le SPARQL (*Simple Protocol and RDF Query Language*): le SPARQL permet de rechercher, modifier, ajouter, supprimer des données RDF (Cf. Figure 22). Il s'agit d'un langage normalisé et ouvert, développé et maintenu par le W3C. Les principes du SPARQL sont les suivants :
   - Appariement de graphe ;
   - Ecriture d'un graphe de requête ;
   - Recherche de tous les sous-graphes qui correspondent au patron de la requête dans le graphe cible.

Ce langage est proche du SQL en base de données, et il existe plusieurs types de requêtes : SELECT, WHERE, ASK, DESCRIBE, CONSTRUCT … Ci-dessous l'exemple de l'anatomie d'une requête :

```
1.  PREFIX p1 :<…>                    } Déclaration des espaces de nommage (optionnel)
2.  PREFIX p2 :<…>                    }
3.  …                                 } Identification des variables dont on veut récupérer les résultats
4.  SELECT …                          }
5.  FROM [NAMED]…                     } Sélection du jeu de données (optionnel)
6.  WHERE {                           }
7.  …                                 } Graphe recherché
8.  …                                 }
9.  }
10. OFFSET …                          } Limitation des résultats et tris (optionnel)
11. LIMIT …                           }
12. ORDER BY                          }
```

Figure 22 Anatomie d'une requête en SPARQL — Source : Neveu P., Symeonidou D., Tireau A., 2017

[69] Pour aller plus loin : http://www.iro.umontreal.ca/~lapalme/ift6282/OWL/EtapesCreationOntologie.html

Avec le RDF et le SPARQL, il est donc possible d'interroger l'information structurée contenue dans les métadonnées. Cette interrogation peut passer par l'interface Web SPARQL Endpoint.

« L'intérêt de RDF réside dans le fait qu'il est possible d'exploiter des triplets RDF quelque soit le vocabulaire utilisé sans conversion, à l'inverse de XML pour lequel il est nécessaire de convertir les données si elles n'utilisent pas le même schéma. Ainsi, il n'impose pas aux différents producteurs de se mettre d'accord strictement sur une structure de métadonnées, comme c'est le cas dans le protocole OAI-PMH avec le profil d'applications OAI_DC (Dublin Core simple), ou de se limiter à un plus petit dénominateur commun pour assurer l'interopérabilité »[70]. Lier les données du Parc à d'autres données pour mieux les contextualiser c'est atteindre la 5ème étoile du LOD. Ci-dessous est présentée l'architecture du SI du Parc du Verdon telle que l'on peut la concevoir pour aboutir au Web de données.

3.3.2.3. Chaîne de traitement : perspectives

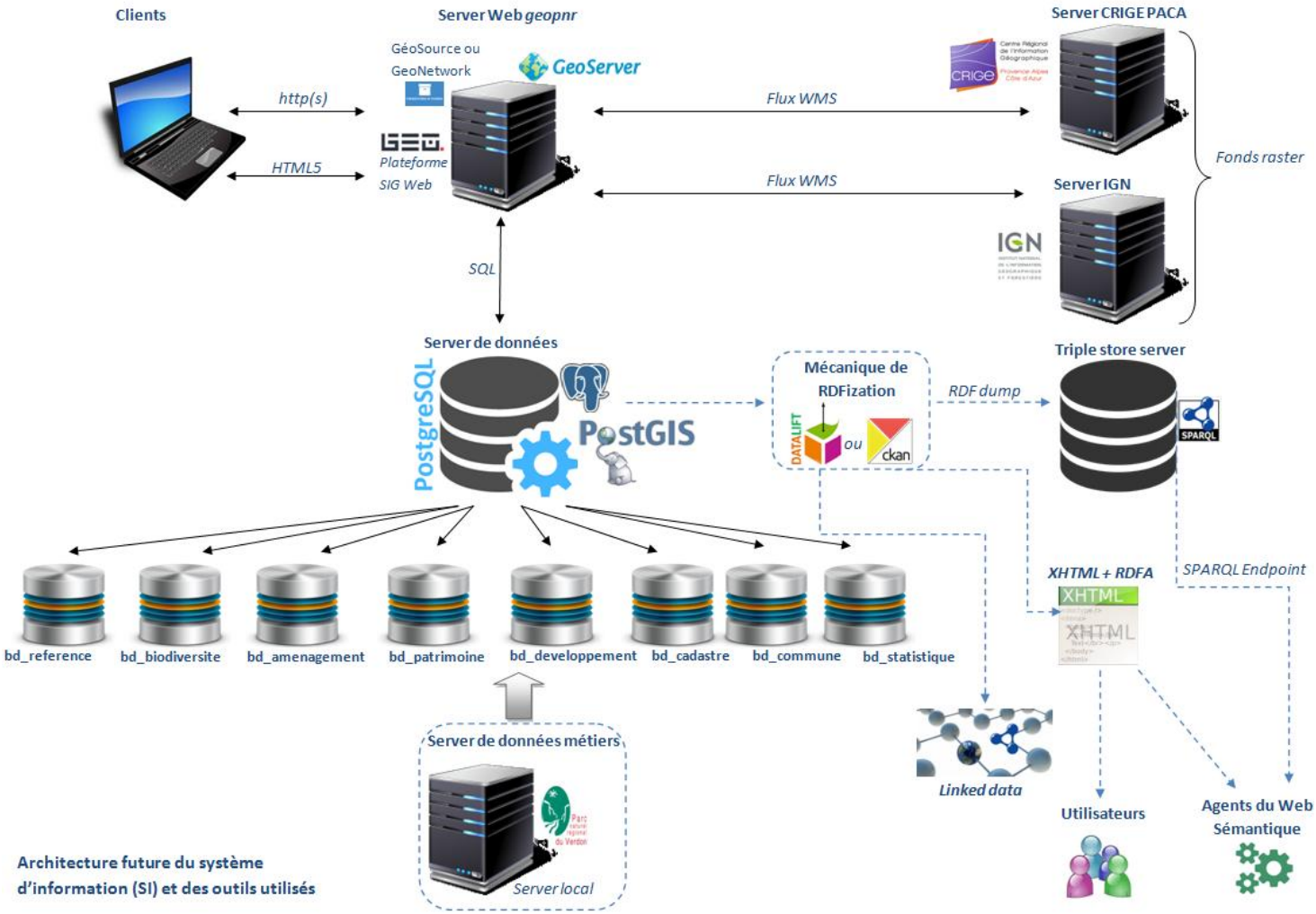


Figure 23 Perspectives pour le SI du Parc du Verdon : vers le Web de données

Source : Alexandra Greco, 2017

Des préconisations ont été faites au sein du Parc du Verdon pour « faire de l'Open Data », tout comme des suggestions pour aller plus loin dans cette démarche, si à l'avenir se présentent des opportunités, notamment en termes de valorisation des données.

[70] Source : http://www.rechercheisidore.fr/sqe/

4. BILAN ET PERSPECTIVES

4.1. Conclusion

Les deux objectifs du stage ont été atteints : engager une démarche Open Data au sein du Parc du Verdon dans une réflexion collective interparc et répondre à la Directive INSPIRE en termes d'accès aux données géographiques.

Pour atteindre ces objectifs, la méthodologie a évolué et s'est améliorée au cours du stage grâce aux rencontres avec des acteurs clés et à des formations en matière d'Open Data. Si la méthodologie de base s'est appuyée sur les consignes et les préconisations développées par l'association OpenData France (Occitanie), elles ont été adaptées au contexte d'un Parc naturel régional. Ainsi, la démarche mise en place tient compte des données sensibles à caractère environnemental. Concernant le catalogage de données, la méthode a consisté à installer et paramétrer GeoNetwork *via* une machine locale, dont l'environnement informatique a été reproduit à l'identique du serveur interparc sur lequel devait être installé le catalogue de données. Ce travail a fait l'objet de la rédaction de deux documentations techniques coécrites avec Lucien Duflocq. Un tableau récapitulatif des livrables restitués auprès du Parc du Verdon est disponible en annexe de ce rapport (Cf. Annexe n°19).

Deux difficultés principales pour venir à bout des attentes ont été de trouver des alternatives pour l'installation d'un catalogue de données. En effet, la performance du matériel mis à disposition et du réseau Internet extrêmement faible ont été un frein dans la mise en place d'un catalogue de données. La deuxième difficulté a été de sensibiliser au mieux l'équipe du Parc du Verdon, tous Pôles confondus. L'enjeu était que les agents du Parc prennent conscience de l'intérêt d'ouvrir leurs données publiques afin qu'elles puissent être réutilisées sous certaines conditions (choix des licences de diffusion, choix des formats de données, etc.). Certes, des risques existent quant à l'ouverture de certaines données, mais lorsqu'ils sont pris en compte et écartés, les atouts, ne serait-ce qu'au sein du Parc, restent nombreux.

A la suite d'une réunion interparc en présence de professionnels du CRIGE et de la Région PACA, la stratégie adoptée par le Parc du Verdon est finalement d'utiliser le catalogue de données de la future plateforme régionale IDGO. Actuellement, les données ouvertes du Parc et leurs métadonnées sont disponibles sur la plateforme régionale OpenPACA. Quant aux métadonnées des données géographiques, elles sont cataloguées par le CRIGE PACA. Ces données géographiques sont disponibles sur OpenPACA. A l'avenir, ces dernières seront téléchargeables *via* la plateforme interparc GEO dont les métadonnées seront stockées dans le catalogue de données IDGO. Pour les autres données, elles seront stockées et donc téléchargeables *via* la plateforme IDGO.

Deux perspectives pour le Parc sont envisageables. La première consisterait à installer le catalogue de données GeoNetwork sur le serveur local du Parc du Verdon. Chaque agent pourrait ainsi remplir ses métadonnées *via* deux modèles que j'ai configurés, dont un respectant la norme INSPIRE. Les métadonnées au format XML seraient alors moissonnées directement par le CRIGE PACA. La deuxième perspective serait de pouvoir aboutir à des données liées, en convertissant les données ouvertes au format RDF *via* par exemple un des outils de Datalift. Les données seraient ainsi interrogeables grâce à un langage de requêtes tel que le SPARQL. La démarche Open Data du Parc atteindrait ainsi la 5ème étoile des Données ouvertes et liées (LOD) définie par Sir Tim Berners-Lee.

4.2. Bilan personnel

D'un point de vue professionnel, ce stage était très complet. Chargée de l'ouverture des données publiques du Parc du Verdon, j'ai découvert le métier de chef de projet Open Data et d'administratrice de données. En effet, j'ai piloté la mise en place d'une démarche Open Data intégrant le catalogage des données, l'administration, la diffusion et le contrôle qualité. Dans le cadre de l'ouverture des données du Parc, j'ai participé au développement de partenariats et de mutualisations avec la Région et le CRIGE PACA. Si en tant que géomaticienne, je suis rattachée au Système d'information géographique (SIG), j'ai dû me familiariser avec les Systèmes d'informations (SI) en général, avec donc toutes ses données, géographiques ou non. Cela sous-entend de prendre en compte tous les enjeux du projet, analyser sa valeur et ses risques, notamment juridiques.

En complément de ma formation en géomatique, ce stage m'a permis d'acquérir de nouvelles compétences. J'ai eu la chance de pouvoir piloter un projet de A à Z en évaluant les besoins de la structure et ses moyens. J'ai dû faire de nombreux choix stratégiques et techniques pour mettre en place une démarche Open Data dans une réflexion collective interparc. Durant ce stage, des échanges fréquents ont eu lieu avec l'équipe et la direction du Parc, ainsi qu'avec la Région et le CRIGE PACA. D'autres échanges avec des acteurs clés de l'Open Data (OpenData France, CEN PACA…) m'ont permis aussi d'améliorer et valider la démarche par rapport au contexte d'un Parc naturel régional. Fortement impliquée dans les problématiques d'échange/partage de données, j'ai à présent des connaissances solides sur le cadre juridique lié à l'Open Data et à la Directive INSPIRE. J'ai pu définir ainsi des clauses techniques et juridiques, et rédiger une délibération de principe.

D'un point de vue technique, j'ai acquis des compétences notamment en informatique : tester Docker, installer un système d'exploitation Debian 9, installer et paramétrer GeoNetwork, coder en PHP et Javascript. J'aurais aimé aller plus loin dans la démarche Open Data en aboutissant à des données liées (installation de Datalift, conversion des données en RDF…). J'ai pris connaissance des solutions techniques en termes de Web sémantique, et je peux à présent évaluer les nouveaux enjeux de l'Open Data. Réaliser ce stage dans un Parc naturel régional m'a permis également de comprendre l'organisation du fonctionnement d'une telle structure publique.

D'un point de vue humain, j'ai tissé de nombreuses relations dans le cadre de mon travail. Si j'ai piloté la démarche Open Data du Parc du Verdon, c'est dans une réflexion collective et grâce à la concertation avec différents professionnels. Il a fallu motiver, intéresser et avertir tous mes collègues pour que le projet puisse fonctionner. J'espère que ceux qui sont convaincus de l'intérêt d'une telle démarche pourront convaincre à leur tour les plus sceptiques. La démarche mise en place est engagée mais pour qu'elle soit durable, il faut adopter des réflexes et certaines préconisations pour faciliter la diffusion et la réutilisation des données. Un de mes objectifs personnels a été de donner tous les moyens techniques et méthodologiques au Parc du Verdon pour qu'il puisse devenir un territoire pilote en termes d'Open Data.

ACRONYMES

ADASEA : Association départementale pour l'aménagement des structures des exploitations agricoles
AFIGEO: Association française pour l'information géographique
API: *Application programming interface*
BRGM : Bureau de recherches géologiques et minières
CADA : Commission d'accès aux documents administratifs
CBNMed : Conservatoire botanique national méditerranéen
CC0: *Creative commons zero*
CCAP: Cahier des clauses administratives particulières
CCTP : Cahier des clauses techniques particulières
CEN (PACA) : Conservatoire des espaces naturels
CEN : Comité européen de normalisation
CKAN: *Comprehensive knowledge archive network*
CODIR : Comité de direction
COPIL : Comité de pilotage
CRIGE : Centre régional de l'information géographique
CRPA : Code des relations entre le public et l'administration
CSW: *Catalog service for harvesting the web*
DATAR: Délégation à l'aménagement du territoire et à l'action régionale
DCMI: *Dublin Core Metadata Initiativ*
DEE: Données élémentaires d'échange
DINSIC : Direction interministérielle du numérique et du système d'information et de communication de l'Etat
DOI : *Digital object identifier*
DOM : *Document object model*
DREAL: Direction régionale de l'environnement de l'aménagement et du logement
DWC: *Darwin core*
EML: *Ecological metadata language*
EPCI: Etablissement public de coopération intercommunale
ETL: *Extract-transform-load*
FAO: Fabrication assistée par ordinateur
FGDC: *Federal geographic data committee*
FING: Fondation Internet nouvelle génération
GBIF: *Global Biodiversity Information Facility*
GPL: *General public licence*
HTML: *HyperText markup*
HTTP: *HyperText transfer protocol*
IDG : Infrastructure de données géographiques
IDGO : Infrastructure de données géographiques ouvertes (régionale)
IDS : Infrastructures de données spatiales
IETF: *Internet engineering task force*
INPN: Inventaire national du patrimoine naturel
INRA: Institut national de recherche automatique
INSPIRE: *Infrastructure for spatial information in the European community*
ISO: *International organization for standardization*
JDC: *Java database connectivity*
JRE: *Java runtime environment*
LOD : *Données ouvertes et liées*
LOV : *Linked open vocabularies*
MCD : Modèle conceptuel de données
MNHN: Muséum national d'histoire naturelle
OAI-PMH: *Open archives-initiative-protocol for metadata*
ODbL: *Open database licence*
ODI: *Open data institute*
OGC: *Open geospatial consortium*
OWL: *Web ontology language*
PACA : Provence-Alpes-Côte d'Azur
PHP : *Hypertext PreProcessor*
PNR: Parc naturel régional
PSI : Informations du secteur public
RDF : *Resource description framework*
RDFS: *Resource description framework schema*
RSS: *Rich Site Summary* ou *Really Simple Syndication*
SGBD: Système de gestion de base de données
SICTIAM : Syndicat intercommunal des collectivités territoriales informatisées Alpes Méditerranée
SIG: Système d'information géographique
SILENE : Système d'information et localisation des espèces natives et envahissantes
SINP : Système d'information sur la nature et les paysages
SIT: Système d'information territorial
SITEFF : Système d'information territorial pour l'étude de la faune et flore
SKOS: *Simple knowledge organization system*
SMIL: *Synchronized multimedia integration language*
SOAP: *Simple object access protocol*
SPARQL: *Simple protocol and RDF query language*
SVG: *Scalable vector graphics*
TDWG: *Taxonomic databases working group*
URI: *Uniform resource indicator*
URL: *Uniform resource locator*
VML: *Vector markup language*
WAI: *Web accessibility initiative*
W3C: *World Wide Web consortium*
WDSL: *Web services description language*
WMS: *WebMap service*
XML: *Extensible markup language*
XSL: *eXtensible stylesheet language*

BIBLIOGRAPHIE / WEBOGRAPHIE

Becard, N., Castets-Renard, C., Chassang, G., Courtois, M.-A., Dantant, M., Gandon, N., Martin, C., Martelletti, A., Mendoza-Caminade, A., Morcrette, N., Neirac, C., 2016, *Ouverture des données de la recherche. Guide d'analyse du cadre juridique en France*. DOI : 10.15454/1.481273124091092E12 http://prodinra.inra.fr/record/382263

Bizer, C., Health, T., 2012, *Web de données : méthodes et outils pour les données liées*, Pearson, p. 136

Bordeaux Métropole et AEC, octobre 2014 (version 1), *Guide de demande de données publiques auprès des collectivités*, p.7

Cavet, D., 2003, SCEREN, p. 46:
http://eurolibnet.eu/files/REPOSITORY/20061013020036_050908metadata.pdf

Centre de ressources du Parc naturel régional du Verdon :
http://centre-ressources.parcduverdon.fr/opac/

CKAN: https://ckan.org/

Data.gouv.fr: http://www.data.gouv.fr/fr/

Datalift: http://www.datalift.org/index.php/development/download/

Developpez.com, 2013 (mis à jour en 2016), "Une introduction aux données liées et à RDF » : http://jplu.developpez.com/tutoriels/web-semantique/introduction-donnees-liees-et-rdf/

Docker (installation): https://memo-linux.com/installer-docker-sur-gnulinux/

Données liées et ouvertes : http://5stardata.info/fr/

Duflocq, L., Greco, A., 2017, *Installing GeoNetwork on Debian virtual machine*, p. 35:
https://github.com/GeoNetworkTeam/GeoNetwork_doc/blob/master/install_Geonetwork_dv.pdf

Duflocq, L., Greco, A., 2017, *Configuring the catalog –GeoNetwork-,* p. 47:
https://github.com/GeoNetworkTeam/GeoNetwork_doc/blob/master/Parametrage_GeoNetwork_dv.pdf

Duflocq, L., Greco, A., 2017, *Conceptual Data Model of GeoNetwork*, p. 21:
https://github.com/GeoNetworkTeam/GeoNetwork_doc/blob/master/MCD_Geonetwork_dv.pdf

Eduscol: http://eduscol.education.fr/numerique/dossier/archives/metadata/normes-et-standards

Etalab, 2015, « Data.gouv.fr invite le standard Inspire dans l'Open Data » :
https://www.etalab.gouv.fr/data-gouv-fr-invite-le-standard-inspire-dans-lopen-data

Etalab, 2017, « Nouvelle licence pour la réutilisation des informations publiques : éléments de clarification » :
https://www.etalab.gouv.fr/nouvelle-licence-pour-la-reutilisation-des-informations-publiques-elements-de-clarification

European Data Portal: https://www.europeandataportal.eu/elearning/fr/module4/#/id/co-01

François, J., 2014, *Mise en place d'une infrastructure de données géographiques dans la communauté d'agglomération du Pays Viennois*, rapport de stage Master SIG St-Etienne, p.107

Gandon, F., Faron Zucker, C., Corby, O., 2015, "Introduction aux ontologies », INRIA : https://www.canal-u.tv/video/inria/introduction_aux_ontologies.20327

GEO (Business Geografic): https://www.business-geografic.com/fr/geo-software/geo-solutions.html

GéoCatalogue national : http://www.geocatalogue.fr/

GeoRezo, forum Geolibre, "Documents pour installation et paramétrage de GeoNetwork v. 3.2.1 » : https://georezo.net/forum/viewtopic.php?id=107404

Goffinet, L., 2010, « Les standards ouverts et la normalisation comme conditions de l'interopérabilité logicielle », *Cahiers Droit, Sciences & Technologies*, p. 165-176.

Greco, A., 2017, « Dates clés de l'Open Data » : https://cdn.knightlab.com/libs/timeline3/latest/embed/index.html?source=1U4hZWJAZGYcX4zRs7yTdRx5Ar8lCiZu6NQ04NGFZkaw&font=Default&lang=fr&initial_zoom=2&height=650

IDGO: www.crige-paca.org/index.php?id=4325

INSPIRE (Directive de 2007) : http://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:32007L0002

INSPIRE Knowledge Base: http://inspire.ec.europa.eu/data-specifications/2892

Isidore: http://www.rechercheisidore.fr/sqe/

ISO: https://www.iso.org/fr/home.html

Kepeklian, G., 2014, "Datalift, une démonstration de la plateforme » : http://www.dailymotion.com/video/x1krni5

Lapalme, G., « Création d'une petite ontologie » (tutoriel), DIRO – Université de Montréal: http://www.iro.umontreal.ca/~lapalme/ift6282/OWL/EtapesCreationOntologie.html

Legifrance.gouv.fr :

Loi CADA : https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000339241

Décret n°2017-638 du 27 avril 2017 (Homologation des licences): https://www.legifrance.gouv.fr/affichTexte.do;jsessionid=D94770F4DA5096AB4EA5436D1A47A5FF.tpdila11v_2?cidTexte=JORFTEXT000034502557&dateTexte=&oldAction=rechJO&categorieLien=id&idJO=JORFCONT000034502533

Article L.323-2 du 7 octobre 2016 (CRPA): https://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000031366350&idArticle=LEGIARTI000033219073

Décret n° 2017-331 du 14 mars 2017 (Données de référence) : https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000034194946&dateTexte=&categorieLien=id

Article L131-3-1 du 3 août 2006 (Code la propriété intellectuelle) : https://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006069414&idArticle=LEGIARTI000006278959

Article L124-4 du 23 octobre 2015 (Code de l'environnement): https://www.legifrance.gouv.fr/affichCodeArticle.do;jsessionid=21FBA36D23A72AA3D0379E05B42EB306.tpdjo12v_3?cidTexte=LEGITEXT000006074220&idArticle=LEGIARTI000006832925&dateTexte=&categorieLien=cid

Loi Lemaire ou pour une République numérique du 7 octobre 2016 : https://www.legifrance.gouv.fr/affichLoiPubliee.do?idDocument=JORFDOLE000031589829&type=general&legislature=14

Metaclic: https://github.com/datakode/metaclic

Neveu, P., Symeonidou, D., Tireau, A., 07-09/02/2017 (formation), "Sémantique des données : interrogation du Web sémantique », INRA Science & Impact, p.43

OpenData France, juin 2016 (version 1), « Les lois régulant la donnée publique », Guide Open Data pour les communes, p.24

OpenData France, juin 2016 (version 1), « Glossaire de la donnée publique », Guide Open Data pour les communes, p.20

OpenDataLab (Occitanie): http://www.opendatalab.fr/

OpenDataSoft: https://www.opendatasoft.fr/

Open Data Support: https://joinup.ec.europa.eu/sites/default/files/e4/49/56/D2.1.2%20Training%20Module%201.4%20Introduction%20to%20metadata%20management%20v1.00_FR.pdf

Open Knowledge Foundation: https://okfn.org/

OpenPACA: http://opendata.regionpaca.fr/opendatalocale.html

Parc naturel régional du Verdon, *Charte 2008-2020*, p.121 : http://parcduverdon.fr/sites/default/files/pnrverdon/pdf/2008_charte.pdf

Parc naturel régional du Verdon : http://parcduverdon.fr/

Parcs naturels régionaux de PACA, *Charte Géomatique SIT-PNR-PACA,* p.24

Plateforme d'aide à la décision pour l'ouverture des données publiques des Parcs naturels régionaux : http://193.55.175.126/sigmaetu10/PNRV/

Plateforme Inspire nationale : https://inspire.data.gouv.fr/

Poupeau, G., décembre 2010, ISIDORE et le web de données, Université d'hiver du TGE Adonis, Valprè

Poupeau, G., 2010, « Isidore à l'interface de la rencontre des SHS et du Web de données », p. 37 : https://fr.slideshare.net/AntidotNet/isidore-linterface-de-la-rencontre-des-shs-et-du-web-de-donnes

Pouyllau, S., Minel, J.-L., Capelli, L., décembre 2010, ISIDORE : Présentation générale du projet, Université d'hiver du TGE Adonis, Valprè

Région PACA, juin 2011, « A la découverte des Parcs naturels régionaux », p. 68

Roche, M., décembre 2014, « Guide pratique d'ouverture des données du Parc naturel régional du Vercors », p. 16

SILENE: http://www.silene.eu/index.php?cont=accueil&args=acces

SIT du Parc naturel régional du Verdon : http://www.pnrpaca.org/index.php?cont=login

The Open Definition: http://opendefinition.org/

Touroult, J., « Données sensibles, naturalistes prudents ? », *Espaces naturels*, juillet 2017, n°59. Disponible sur : http://www.espaces-naturels.info/donnees-sensibles-naturalistes-prudents

Trombinoscope du Parc naturel régional du Verdon : http://parcduverdon.fr/sites/default/files/pnrverdon/pdf/2017_trombinoscope_mars_2017.pdf

Vandenbussche, P.-Y., Vatant, B., Charlet, J., 2012, *Linked Open Vocabularies, un écosystème encore fragile*, p. 13 : https://liris.cnrs.fr/QetR2012/site/wp-content/uploads/2012/02/qetr2012_1.pdf

Wikipedia:

DBpedia: https://fr.wikipedia.org/wiki/DBpedia
SparqlEndpoints: https://www.w3.org/wiki/SparqlEndpoints
Open source: https://fr.wikipedia.org/wiki/Open_Source_Definition

LISTE DES FIGURES

Figure 1 Le territoire du Parc naturel régional du Verdon et ses paysages 6
Figure 2 Schéma des instances du Syndicat Mixte de gestion du Parc du Verdon 8
Figure 3 Schéma des commissions et autres instances opérationnelles du Syndicat Mixte de gestion du Parc du Verdon 9
Figure 4 Répartition des tâches selon les instances 9
Figure 5 Tableau récapitulatif des réunions de Pôle au Parc du Verdon - sensibilisation Open Data - 10
Figure 6 Evolution du SIT depuis sa création 11
Figure 7 Evolution de l'infrastructure de données spatiales 12
Figure 8 Architecture du système d'information et des outils utilisés par le Parc du Verdon 13
Figure 9 Organigramme des tâches 16
Figure 10 Diagramme de Gantt 17
Figure 11 Conditions de diffusion des documents administratifs 19
Figure 12 Conditions pour faire de l'Open Data 21
Figure 13 Orientations des axes de la Charte du Parc du Verdon en lien avec l'Open Data 26
Figure 14 Critères de définition d'une espèce sensible 28
Figure 15 De l'observation à la connaissance partagée : le chemin de la donnée naturaliste 33
Figure 16 Engagement des PNR dans la démarche Open Data 34
Figure 17 Etapes clés de la démarche Open Data 35
Figure 18 Chaîne de traitement – De la donnée à sa publication 38
Figure 19 Comparaison entre les principales normes internationales et standards 40
Figure 20 Cas d'utilisation possibles de Docker 46
Figure 21 Instances d'ontologie 49
Figure 22 Anatomie d'une requête en SPARQL 50
Figure 23 Perspectives pour le SI du Parc du Verdon : vers le Web de données 51