



HAL
open science

Mise en place d'une plateforme de catalogage des métadonnées pour les Parcs naturels régionaux d'Île-de-France

Lucien Duflocq

► **To cite this version:**

Lucien Duflocq. Mise en place d'une plateforme de catalogage des métadonnées pour les Parcs naturels régionaux d'Île-de-France. Géographie. 2017. dumas-01859090

HAL Id: dumas-01859090

<https://dumas.ccsd.cnrs.fr/dumas-01859090>

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

« Sciences Géomatiques en environneMent et AménageMent » (SIGMA)

<http://sigma.univ-toulouse.fr>

RAPPORT DE STAGE

Mise en place d'une plateforme de catalogage des métadonnées pour les Parcs naturels régionaux d'Île-de-France

DUFLOCQ Lucien



Maître de stage : GUIHO Jocelyn
Tuteur-enseignant : JÉGOU Laurent

Mars / Septembre 2017

Stage réalisé du 13 mars 2017 au 8 septembre 2017, au sein des locaux du Parc naturel régional du Gâtinais français, sous l'encadrement de :

- **Eric Marot**, chargé de mission SIG et évaluation au Parc naturel régional du Vexin français.
- **Emilie Duc**, responsable SIG, évaluation et TIC au Parc naturel régional de la Haute Vallée de Chevreuse.
- **Pascal Lambert**, chargé de mission SIG et évaluation au Parc naturel régional d'Oise – Pays de France.
- **Jocelyn Guiho**, chargé de mission SIG et évaluation au Parc naturel régional du Gâtinais français.

Résumé

C'est une volonté régionale de favoriser le réseau des Parcs naturels régionaux d'Île-de-France (Gâtinais français, Haute-Vallée de Chevreuse, Oise - Pays de France et Vexin français). Pour la Région, les intérêts de cette mutualisation sont multiples : développer la place et les apports des Parcs ou encore favoriser les transferts d'expériences et de savoir-faire qui y sont développés. Pour les Parcs, c'est l'opportunité de partager des compétences, des moyens financiers et des outils permettant l'ouverture de projets qui étaient jusque-là difficilement envisageables. La mise en place d'un Système d'Information Territorial fut la priorité avec le développement d'une architecture informatique et de deux applications web en 2016 (gestion du patrimoine bâti et gestion des sentiers de randonnée).

Afin de poursuivre le développement du Système d'Information et de faire face aux besoins d'ouverture des données géographiques des Parcs, l'objectif fut de mettre en place un outil de catalogage de métadonnées commun aux quatre Parcs. Porté par le besoin de connaître son patrimoine de données et de le rendre accessible, en respectant les contraintes juridiques qui accompagnent sa diffusion, cet outil devait constituer une plateforme conviviale et facilement administrable. Après avoir analysé et comparé les différentes solutions de catalogage de métadonnées, j'ai choisi celle qui était la plus adaptée aux besoins et à mes compétences. Par la suite, mon travail a consisté à installer, à configurer et à personnaliser l'outil de manière à le rendre opérationnel lors de mon départ. Une documentation détaillée de la mise en place du catalogue est disponible en licence libre sur la plateforme GitHub (<https://github.com/GeoNetworkTeam>).

Mots-clés : métadonnées, catalogue, Parc naturel régional, INSPIRE, open source, SIG.

Abstract

The development of the Regional Natural Park network (Gâtinais français, Haute-Vallée de Chevreuse, Oise - Pays de France and Vexin français) is a regional willingness. This mutualization is interesting for the region for many reasons such as enhancing the place and the contributions of the parks and encouraging the experience and expertise transfers. For Regional Natural Parks, the mutualization is an opportunity to share competencies and financial resources. It also helps them to realize projects, which were hardly thinkable without the mutualization. The implementation of a territorial information system was the priority in 2016 with the development of a computing architecture and the installation of two web applications (one to manage the built heritage and the other one to manage hiking paths).

With the aim of pursuing the information system development and opening the parks geographical database, it has been decided to implement a metadata-cataloguing tool common to the four parks. To reconcile the need to know the heritage of data and to make this data accessible to the public, while respecting the legal requirements for their distribution, this tool should constitute a user-friendly platform and be easy to manage. After having analysed and compared the various solutions of metadata cataloguing, I chose the one who was the most suited to the needs and to my competencies. Afterward, my work has been to install, configure, and personalize the tool, to make it operational for the day of my departure. A detailed documentation of the catalogue implementation is available in free license on the Github platform (<https://github.com/GeoNetworkTeam>).

Mots-clés : metadata, catalogue, regional national park, INSPIRE, open source, GIS.

Remerciements

Je tiens à mentionner toutes les personnes qui ont participé à la réussite de cette expérience et qui m'ont accompagné durant l'élaboration de ce rapport.

Tout d'abord, merci à Mathieu Deperrois pour m'avoir accueilli et encadré durant la première partie de mon stage.

Je remercie ensuite profondément Jocelyn Guiho, chargé de mission SIG au sein du Parc naturel régional du Gâtinais français. Sa prise de poste quelques mois après le début de mon stage ne l'a pas empêché de s'impliquer pleinement dans le projet, d'être à l'écoute et de m'aider à la rédaction de ce rapport.

Je remercie également Eric Marot, Emilie Duc et Pascal Lambert, les responsables SIG du réseau des Parcs naturels régionaux d'Île-de-France pour leur encadrement et leur partage d'expériences.

Je tiens également à remercier Laurent Jégou, pour m'avoir accompagné en tant que tuteur-enseignant.

Enfin, je remercie ma mère, pour son regard extérieur mais pas moins attentif et impliqué. Elle a su me conseiller et m'aider dans de nombreux projets universitaires, de la Licence jusqu'à mon Master 2.

Sommaire

Résumé	3
Abstract	4
Remerciements	5
Sommaire	6
Introduction	8
Partie 1 : Cadre général et description de la mission	9
A) Présentation de la structure d'accueil : le Parc naturel régional	9
1. Qu'est-ce qu'un Parc naturel régional ?.....	9
a. Définition.....	9
b. Fonctionnement et spécificités	10
2. Les Parcs naturels régionaux d'Île-de-France.....	12
a. Quatre Parcs, quatre territoires, quatre services SIG.....	12
b. Mise en place d'un réseau des Parcs en SIG : enjeux, objectifs et état des lieux ...	16
B) Comment animer le réseau des Parcs en poursuivant le développement du SIG ? ..	21
1. Pourquoi et comment traiter les métadonnées au sein du réseau des Parcs franciliens ?	21
a. Qu'est-ce qu'une métadonnée ?.....	21
b. Pourquoi le réseau des Parcs aborde-t-il le sujet des métadonnées ?	22
c. Comment répondre à la demande du réseau des Parcs : objectif du stage	23
2. Solution : le catalogage comme outil de gestion des métadonnées.....	25
a. Qu'est-ce qu'un catalogue de métadonnées ?.....	25
b. Présentation des différentes solutions de catalogage et analyse.....	25
c. Choix de la solution : de geOrchestra à Geonetwork	30
Partie 2 : Mise en place de l'outil	32
A) Installation	32
1. GeOrchestra : retour sur les difficultés rencontrées lors de l'installation	32

2.	Installation de Geonetwork.....	34
a.	Procédé d'installation en monoposte sur un serveur local.....	34
b.	Couplage avec la base de données Postgres.....	34
c.	Paramétrages spécifiques lors de l'installation sur le serveur interParc.....	35
B)	Comprendre le fonctionnement de Geonetwork	36
1.	La place de Geonetwork au sein du serveur interParc.....	36
2.	Obtenir et analyser le Modèle Conceptuel de Données	37
Partie 3 :	Paramétrages, alimentation et perspectives de l'outil	39
A)	Configuration de l'outil	39
1.	Lier des ressources à la métadonnée	39
2.	Diffuser des données via Geoserver	40
3.	Administrer les droits d'accès	42
B)	Alimentation du catalogue	43
1.	Quelles données seront à traiter ?.....	43
2.	Procédé d'intégration des métadonnées	44
C)	Perspectives de l'outil	46
1.	Lien entre la métadonnée et la donnée	46
2.	Optimisation de l'architecture du serveur interParc	48
Bilan personnel	49
Conclusion	51
Liste des illustrations	52
Liste des tableaux	52
Bibliographie	53
Annexes	55

Introduction

« Le défi des gestionnaires de données dans cet environnement de libre circulation consiste à rendre la donnée visible et accessible, à la valoriser. Les deux piliers de cette nouvelle forme de gestion sont le catalogage et les métadonnées. »

Mathieu Becker et Hervé Halbout, Découvrir et comprendre la gouvernance de l'information géographique, Livre Blanc, 2016.

Jusqu'à maintenant, les Parcs naturels régionaux d'Île-de-France (Gâtinais français, Haute-Vallée de Chevreuse, Oise - Pays de France et Vexin français) disposaient de leur propre système de fonctionnement en terme de gestion de métadonnées. Cette configuration, justifiée par un développement du SIG interne au Parc durant de nombreuses années, devient aujourd'hui problématique pour les Parcs naturels régionaux franciliens qui tendent à s'ouvrir vers une base de données commune et des outils partagés. Après avoir mis en place une infrastructure informatique (serveur partagé, système de gestion de base de données commun et serveur cartographique), les PNR réfléchissent à une stratégie de gestion des métadonnées. En effet, la mise en place d'une base de données commune va inévitablement accroître le nombre de données et pousser les PNR à se coordonner et à s'organiser pour garder le contrôle de leur patrimoine de données. D'autant que la mise en place du réseau permet de financer un tel projet.

L'objectif du stage est donc de venir répondre aux attentes du réseau des PNR franciliens en mettant en place un outil de gestion des métadonnées. Cet outil doit être une plate-forme Web de catalogage et de diffusion des données géographiques. Pour un gestionnaire de l'information géographique, le renseignement de métadonnées est en général mal vécu. Il évoque le plus souvent un travail chronophage et peu agréable à cause du manque de convivialité des outils proposés. Les enjeux sont donc importants, il faut proposer une plate-forme qui soit autant performante qu'ergonomique.

Cinq grandes phases sont venues structurer l'organisation du stage. La première phase a consisté à se concerter avec l'ensemble des chargés de mission pour définir les besoins et les enjeux de l'outil de catalogage. La deuxième phase a permis de décider ensemble de la solution la plus adaptée. La troisième phase a été une période test, dans laquelle la solution fut installée et testée au sein d'une machine virtuelle. Cette phase a permis de rédiger une documentation fort utile pour amorcer la quatrième phase, qui s'est concrétisée par l'installation de la solution sur le serveur interParc. Enfin, la cinquième phase a rendu possible le paramétrage de la solution.

Partie 1 : Cadre général et description de la mission

A) Présentation de la structure d'accueil : le Parc naturel régional

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

Parvenir à dresser une définition concise d'un Parc naturel Régional (PNR) n'est pas chose aisée tant cette institution fait écho à de multiples acteurs et domaines d'activités. C'est tout l'enjeu de cette première partie qui va tâcher d'expliquer au mieux ce qu'est un PNR.

a. Définition

La mise en place des PNR s'opère durant les années 1960. Notons qu'à cette période, les rapports entre l'Homme et la nature prennent une nouvelle dimension. Sans parler de prise de conscience, on assiste petit à petit à la disparition d'une vision binaire, qui veut que l'environnement et les sociétés humaines soient assimilés à deux mondes distincts. On évoque alors avec plus d'insistance des relations de cause à effet entre l'Homme et la nature. L'impact de l'un sur l'autre se renforce dans un contexte d'après-guerre, où les projets d'urbanisation et d'industrialisation foisonnent sur le territoire. Une série de facteurs tels que la démocratisation de l'automobile et l'arrivée des congés payés viennent accentuer les déplacements humains dans des zones naturelles jusque-là peu ou non fréquentées. Parallèlement, ce rapport sensible Homme-nature est dénoncé, par le biais de recherches scientifiques et par des associations qui émergent dès les années 1950 dans le paysage français¹.

C'est dans ce contexte que la DATAR², créée en 1963, se met à réfléchir au concept de Parc naturel régional. À l'origine de cette réflexion, deux constats sont établis :

- Les Parcs nationaux sont fréquemment critiqués par les élus ou les acteurs locaux qui estiment que l'institution est trop intrusive et non adaptée aux projets locaux. En effet, les lourdeurs administratives et les potentielles absences de prise en compte des spécificités locales contraignent le développement économique et social de certaines régions.
- Les régions victimes du déclin industriel et minier semblent avoir besoin d'une institution qui les accompagne, avec l'aide des acteurs locaux, dans une croissance de l'urbanisation qui se développe le plus souvent dans des zones naturelles.

Estimant qu'aucune solution préexistante ne pouvait répondre aux problèmes, la Datar a misé sur une nouvelle institution ayant les capacités de répondre simultanément aux problèmes des métropoles, des zones naturelles sensibles et des régions rurales excentrées.

¹ Source : BARON Nacima et LAJARGE Romain, Les Parcs naturels régionaux, des territoires en expériences, 2015.

² Délégation à l'Aménagement du Territoire et à l'Action Régionale : Ancienne administration française dissoute en 2014, elle préparait les orientations et la mise en œuvre de la politique nationale d'aménagement et de développement du territoire.

C'est le 1er mars 1967, après avoir étudié le concept des Parcs à l'étranger et suite à de nombreuses instances de concertation, qu'un décret, signé par le Général de Gaulle, permet l'officialisation des PNR. Ainsi, selon l'article premier du décret n°67-158 du 1er mars 1967, « Le territoire de tout ou partie d'une ou de plusieurs communes peut être classé en Parc naturel régional lorsqu'il présente un intérêt particulier, par la qualité de son patrimoine naturel et culturel, pour la détente, le repos des hommes et le tourisme, et qu'il importe de le protéger et de l'organiser ». Étant un projet local, le classement d'un Parc est validé par décret du 1er ministre, mais c'est également une marque nationale, propriété de l'état.

Parc naturel régional : définition

« Les Parcs naturels régionaux sont créés pour protéger et mettre en valeur de grands espaces ruraux habités. Peut être classé "Parc naturel régional" un territoire à dominante rurale dont les paysages, les milieux naturels et le patrimoine culturel sont de grande qualité, mais dont l'équilibre est fragile. Un Parc naturel régional s'organise autour d'un projet concerté de développement durable, fondé sur la protection et la valorisation de son patrimoine naturel et culturel. »

www.parcs-naturels-regionaux.fr

L'enjeu des PNR est donc de préserver le patrimoine tout en accompagnant un développement raisonné. Au sein d'un Parc, le patrimoine comprend l'ensemble des spécificités et caractéristiques qui constituent le territoire : la biodiversité, la culture, les paysages, l'urbanisme, le développement économique, l'éducation ou encore l'énergie.

Les grandes missions qui caractérisent l'action d'un Parc sont le plus souvent résumées en 5 points :

- Protéger et gérer le patrimoine naturel, culturel et paysager
- Participer à l'aménagement des territoires ruraux, périurbains et extra-métropolitains
- Contribuer au développement social, économique, durable et écoresponsable
- Assurer l'accueil, l'éducation et l'information des publics à l'environnement et aux territoires
- Permettre l'expérimentation et l'innovation

b. Fonctionnement et spécificités

Le fonctionnement d'un PNR peut paraître complexe car il fait intervenir une série d'acteurs et de mécanismes de décisions. Tout d'abord, un PNR est un territoire, c'est un espace ouvert sans réelles frontières physiques. Il ne dispose pas de pouvoir réglementaire, la concertation devient alors une des principaux leviers pour les actions du Parc. Un Parc doit se doter d'un projet de territoire en matière de développement, d'aménagement et de préservation du patrimoine, c'est la charte du Parc. Approuvée par chacune des communes, collectivités territoriales et enfin par l'État, sa durée de vie est de douze années, à son issue, le Parc remet en question son label et réfléchit de nouveau à ses grands objectifs. Un Parc réunit en moyenne 86 communes.

Les collectivités qui composent un Parc (EPCI, communes, départements et régions) forment ce que l'on appelle un syndicat mixte, dont les délégués siègent au comité syndical. C'est au sein de ce comité que se prend la majeure partie des décisions. Le comité syndical élit un président et un bureau, puis met en place des instances de concertation (commission de travail, conseil de développement, conseil scientifique). Les élus décident alors des actions et en confient la réalisation au personnel du Parc et à leurs partenaires (cf. Figure 1). Soulignons que le Parc n'agit pas seul, il intervient davantage en tant qu'animateur, soit en aidant les projets à se concrétiser ou en trouvant des solutions et des financements.

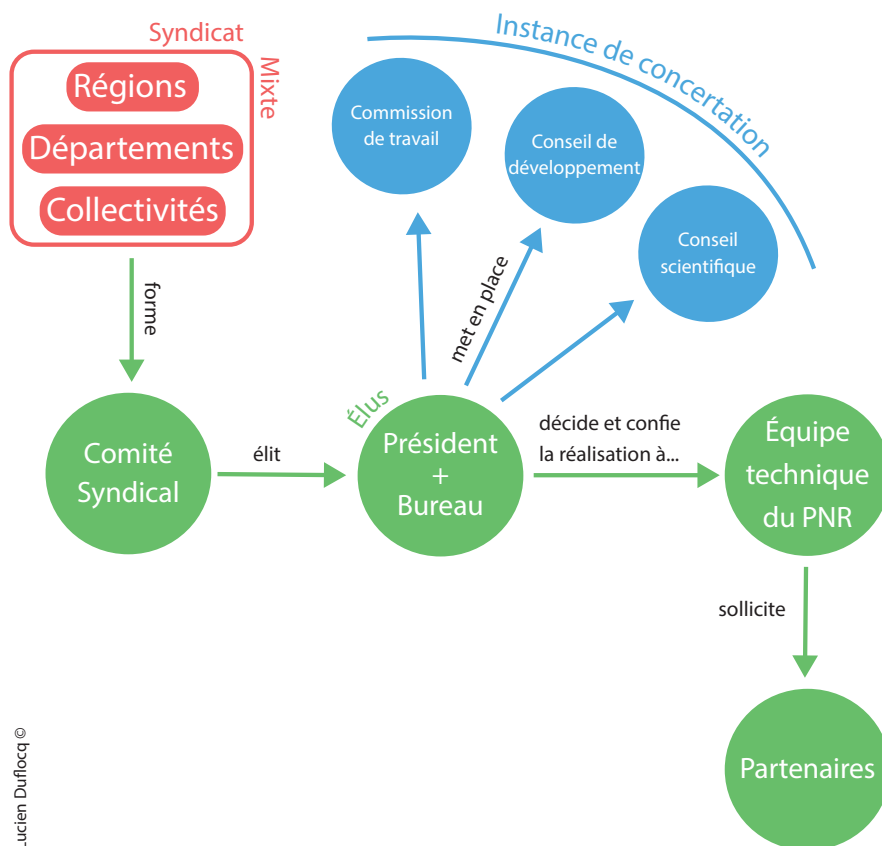


Figure 1 : Organisation d'un Parc naturel régional

La question du financement des projets est un point phare dans la vie d'un Parc. On observe deux sources de financement :

- il peut venir de la part de ses membres qui s'engagent sur le financement d'un certain nombre d'actions portées par le Parc : Europe, État, Département et majoritairement de la part du Conseil régional. Les communes adhérentes au Parc fournissent une cotisation qui repose sur le nombre d'habitants.
- de manière plus ponctuelle, les financements peuvent venir des subventions obtenues lors des créations de projet, auprès d'organismes tels que l'Agence de l'eau, l'ADEME³ ou encore l'Europe (programme LEADER⁴).

³ Agence de l'Environnement et de la Maîtrise de l'Énergie.

⁴ Liaison entre Actions de Développement de l'Économie Rurale : Programme de développement rural destiné à financer des projets pilotes à destination des zones rurales.

Les spécificités d'un Parc semblent reposer sur trois piliers : leurs valeurs, leurs partenariats et leurs méthodes de travail.

Premièrement, en dépit de leurs différences de taille ou de positionnement, chaque Parc partage des valeurs communes, mais également des valeurs propres qui soulignent leur identité. Parmi ces valeurs, nous retrouvons l'attention portée aux biens communs : patrimoine naturel, culturel et paysager par exemple. L'objectif est de créer de l'intérêt général envers ce patrimoine. On retrouve également la volonté d'anticiper, de posséder un temps d'avance sur les actions portées, d'être précurseur mais aussi de savoir proposer des solutions intermédiaires. Ces valeurs, qui évoluent face aux époques et mutations du territoire français, restent un moyen de poursuivre de manière raisonnée leurs actions en restant au cœur des préoccupations et enjeux contemporains.

Deuxièmement, du fait que les missions relient le territoire aux Hommes, les Parcs travaillent de manière transversale et sont en capacité de créer du lien avec des partenaires. Ces partenaires, ce sont les collectivités, les chambres consulaires, les syndicats, les professionnels, mais aussi les associations ou les services techniques. L'obtention de financement via les subventions engendre naturellement des liens étroits avec les acteurs économiques. La mutualisation est aussi au cœur des actions, nous l'évoquerons à la suite de ce rapport, car c'est ce qui a permis d'ouvrir le projet du stage.

Troisièmement, les Parcs se différencient des autres institutions par leurs méthodes de travail. Les équipes sont en moyenne de 35 personnes en 2017 et disposent d'un haut niveau de compétence et de technicité avec 1/3 des agents de niveau bac+5. On retrouve cette difficulté à bien définir la nature d'un Parc de par le foisonnement et l'étendue de leurs interventions. En effet, les interventions dépendent de plusieurs facteurs : orientation de la charte, compétences techniques internes et externes, choix politiques et financiers.

Nous venons de voir que l'ensemble des 51 Parcs présent sur le territoire français est porté par des spécificités communes mais aussi par des particularités qui fondent l'identité d'un Parc. Qu'en est-il des quatre PNR d'Île-de-France ?

2. Les Parcs naturels régionaux d'Île-de-France

La région Île-de-France regroupe quatre PNR : le Parc du Gâtinais français, le Parc de la Haute Vallée de Chevreuse, le Parc du Vexin français et le Parc d'Oise – Pays de France. L'objectif va être de proposer une carte d'identité de chaque Parc. Le Parc du Gâtinais étant ma structure d'accueil principal, nous nous permettrons d'y présenter les différents services qui structurent l'équipe technique.

a. Quatre Parcs, quatre territoires, quatre services SIG

Avec une superficie de 12 000 km² et une population estimée à plus de 12 millions d'habitants, la région Île-de-France possède la plus forte densité de population de France (1001 hab/km²). Les quatre Parcs occupent 18% du territoire francilien et comptent plus de

380 000 habitants⁵. Notons que deux autres territoires, Brie et Deux Morin et Bocage gâtinais, sont en attente de validation pour obtenir le statut de Parc naturel régional. La carte ci-dessous permet de visualiser la répartition des PNR en Île-de-France.

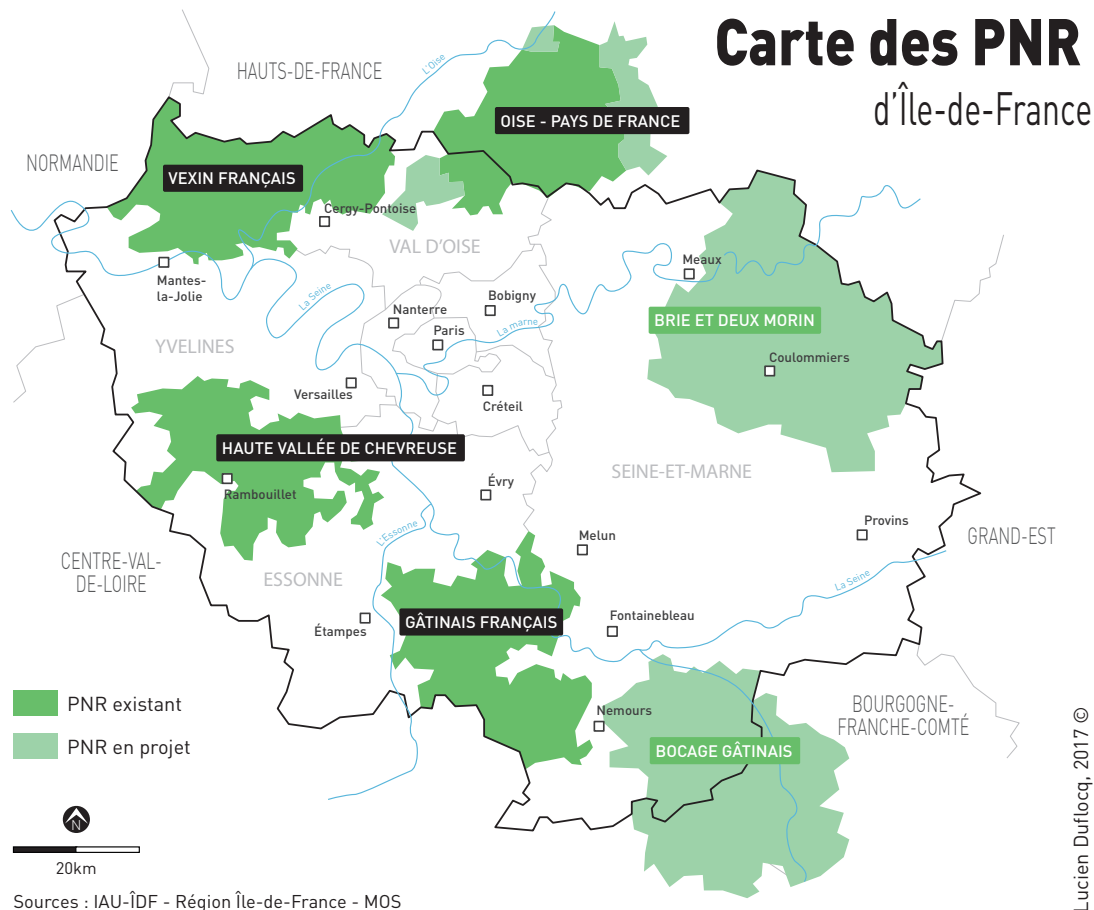


Figure 2 : Carte de localisation des PNR d'Île-de-France

Gâtinais français :

Classé « Parc naturel régional » depuis le 4 mai 1999, le Gâtinais français, aussi appelé « pays des mille clairières et du grès », se situe au sud de la région Île-de-France. Limité par les communes de Nemours (sud-est), Étampes (ouest), Melun (nord-est) et Fontainebleau (est), c'est le Parc qui est doté du plus vaste territoire avec 756 km². Le Parc du Gâtinais s'étend sur deux départements (Essonne et Seine-et-Marne), huit Communautés de Communes et 69 communes (36 en Essonne et 33 en Seine-et-Marne) abritant 82 153 habitants. En terme d'occupation du sol, on compte 58% de terres agricole, 31% de bois et forêts, 7,5% d'urbanisation, 2% de milieux naturels ouverts, 1% de Parcs et jardins et 0,5% de milieux humides. La spécificité de ce Parc réside particulièrement dans la pluralité de ses paysages et dans sa richesse biologique, d'ailleurs reconnu par l'UNESCO, mais également par son activité agricole.

⁵ Source : INSEE, Dossier complet sur le Région Île-de-France, paru le 04/08/2017

À mon départ, l'équipe du Parc était composée de 31 personnes, répartis autour de 5 Pôles, sous l'autorité de la directrice du Parc (cf. organigramme en annexe 1) :

- Le Pôle Développement local (économie durable, éco-tourisme, forêt et projet LEADER)
- Le Pôle Aménagement (urbanisme, architecture, paysage et énergie)
- Le Pôle Environnement (milieux naturels, eau et déchets, bassin versant)
- Le Pôle Administration (gestion, accueil, secrétariat)
- Le Pôle Educ'Info – Culture (patrimoine et animation culturelle, éducation au territoire et SIG – Evaluation)

Haute Vallée de Chevreuse :

Classée « Parc naturel régional » depuis le 11 décembre 1985, la Haute Vallée de Chevreuse est le plus ancien Parc de la région. Situé à l'ouest de l'espace francilien, le long de la vallée de Chevreuse, de la forêt de Rambouillet et du plateau du Hurepoix, sa proximité avec Paris (environ 20 km) en fait une région particulièrement convoitée en Île-de-France, avec une forte pression foncière. À la limite entre le département des Yvelines et de l'Essonne, le Parc s'étend sur 633 km² et compte 51 communes (43 en Yvelines et 8 en Essonne) pour 110 026 habitants. En terme d'occupation du sol, on dénombre 47% de bois et forêts, 37% de terres agricoles, 13% d'urbanisation, 1% de milieux humides et 3% d'autres milieux naturels. Les spécificités du Parc résident dans son important massif forestier, dans sa richesse patrimoniale et dans son activité agricole.

Vexin français :

Classé « Parc naturel régional » depuis le 9 mai 1995, le Vexin français se situe au nord-ouest de la région Île-de-France, limité au nord par la région des Hauts-de-France, à l'ouest par la Normandie, au sud-ouest par la commune de Mantes-la-Jolie et au sud-est par Cergy-Pontoise, il s'étend sur 710 km². Partagé entre le département du Val-d'Oise et des Yvelines, le Parc compte 99 communes (79 en Val-d'Oise et 20 en Yvelines) pour 78 000 habitants. Les spécificités du Vexin reposent principalement sur son vaste territoire agricole où dominant de grandes cultures céréalières (blé, orge, colza) mais aussi sur l'hétérogénéité de ses paysages : plateau calcaire, bois, marais, patrimoine historique.

Oise – Pays de France :

Classé « Parc naturel régional » depuis le 15 janvier 2004, c'est le Parc le plus jeune de la région Île-de-France. Situé au nord de la région, le Parc a la particularité d'être partagé entre deux régions : Île-de-France et Hauts-de-France. Semblable au PNR de Chevreuse, sa proximité avec l'agglomération parisienne (connexion avec le RER A) en fait un lieu stratégique, soumis à une pression foncière importante. Avec une superficie de 600 km², le Parc compte 59 communes (44 dans l'Oise et 15 en Val d'Oise) et abrite 110 000 habitants. Les spécificités du Parc sont tout d'abord son activité agricole, majoritairement à l'est et au sud, puis ses zones forestières à l'ouest, avec les massifs de Chantilly, Halatte et Ermenonville. Enfin, le Parc dispose d'espaces à fortes potentialités économiques, en lien avec sa proximité parisienne et la plate-forme aéroportuaire de Roissy - Charles de Gaulle.

Concernant la gestion et l'animation des Systèmes d'Information Géographique, chaque Parc dispose d'un chargé de mission SIG et Evaluation. La partie évaluation correspond au suivi des actions du Parc. Elle intervient de manière quotidienne, pluriannuelle et au moment de renouveler le classement du PNR. C'est sur cette évaluation et plus particulièrement via des indicateurs que l'on juge le niveau d'avancement d'un Parc. Au niveau national, on estime que la moitié des géomaticiens associent l'évaluation et l'informatique à leur poste. En Île-de-France, les Parcs du Gâtinais et du Vexin s'occupent de la gestion informatique générale du Parc, avec pour le Gâtinais la présence d'un informaticien toutes les deux semaines. Au Parc de Chevreuse, un informaticien a récemment été intégré à l'équipe. Pour la géomatique, bien que les activités du chargé de mission varient d'un Parc à l'autre, nous retrouvons assez naturellement des missions communes telles que :

- Assurer et maintenir le développement du SIG et des bases de données associés au PNR
- Être responsable de l'acquisition, de la production, de la gestion et de l'exploitation des données géographiques
- Participer au traitement des données à travers des analyses et des conceptions cartographiques pour l'ensemble des thématiques abordées au sein du PNR
- Animer l'usage de l'information géographique et apporter un appui technique interne et externe

Concernant les pratiques et outils SIG utilisés, on peut trouver des différences entre les Parcs, selon le profil des chargés de missions, mais aussi selon l'équipement du Parc. Essayons d'en résumer les principales caractéristiques.

Gâtinais français :

En terme de SIG, le Parc est doté d'une licence ArcGIS 10.1 couplée à l'extension Publisher afin que les chargés de mission puissent consulter des cartes. Progressivement, le Parc tend à utiliser la solution libre QGIS. L'application CARMEN, développée par le Ministère en charge de l'écologie et du développement durable, est parfois mobilisée pour le partage de données à travers le web. Concernant la gestion des métadonnées, le Parc utilisait jusqu'en 2010 un tableau Excel. Puis, en 2011, suite au travail d'une stagiaire, un outil de gestion des métadonnées, conforme INSPIRE et structuré via une base Access est mis en place (MetaParc). Cet outil interne au Parc permet de saisir, modifier et rechercher les métadonnées stockées en locale, sur le serveur du Parc.

Haute Vallée de Chevreuse :

En terme de SIG, le Parc utilise majoritairement l'outil ArcGIS. La chargée de mission a mis en place, dans le cadre de l'inventaire du patrimoine bâti, un outil web avec les langages PHP et JavaScript, prenant la forme d'un formulaire de saisie permettant de renseigner un patrimoine bâti dans une base de données MySQL. Pour la gestion des métadonnées, le Parc utilise actuellement Access et stocke les métadonnées en local sur le serveur du Parc.

Vexin français :

En terme de SIG, le Parc a progressivement quitté les outils ESRI pour QGIS, administré via le SGBD PostGIS. Pour la diffusion des données, en WMS/WFS⁶, le Parc utilise ponctuellement

⁶ WMS : Web Map Service (diffusion de cartes raster) ; WFS : Web Feature Service (diffusion d'entités vectorielles)

CARMEN et plus fréquemment le serveur web Geoserver pour des flux internes. Pour la gestion des métadonnées, le chargé de mission a mis en place via Access un formulaire de saisie respectant la norme INSPIRE. Les métadonnées sont stockées en local, sur le serveur du Parc.

Oise – Pays de France :

En terme de SIG, le Parc est semblable au PNR du Gâtinais et de Chevreuse, il est toujours équipé des solutions ESRI (licence ArcMap 10.3) mais il tend à n'utiliser que le couple Postgis/QGIS. En 2007, le Parc achète DynMAP (société Business Geographic) en tant que WebSIG, ce qui lui permet de rendre accessible des données à des utilisateurs extérieurs (c'est le cas du cadastre notamment). DynMAP est aujourd'hui au cœur du système d'information du Parc et rend possible la publication de plus de 300 couches d'informations. Nous allons le voir à la suite du rapport, DynMAP est également un outil utilisé pour le réseau des Parcs franciliens. Le Parc cherche peu à peu à basculer de Dynmap à GEO, nouvelle plateforme géomatique développée par Business Geographic, qui se positionne comme l'outil de convergence en WebSIG moderne.

L'ensemble de ces informations a été obtenu par le biais d'enquêtes réalisées auprès des PNR et grâce au travail réalisé par Fabien Dubois, en 2016, dans le cadre d'un stage au PNR du Gâtinais.

	Gâtinais français	Haute Vallée de Chevreuse	Vexin français	Oise – Pays de France
Outil de gestion des métadonnées	MetaParc (outil développé en interne via Access)	Access	Gestion : Base de données PostGIS/PostgreSQL Interface: Access	ArcCatalogue et Excel
Outil SIG	ArcGIS	ArcGIS	QGIS	ArcGIS

Tableau 1 : Outils utilisés au sein des PNR franciliens

b. Mise en place d'un réseau des Parcs en SIG : enjeux, objectifs et état des lieux

En France, jusque dans les années 2000, la majorité des Parcs développe leur SIG de manière autonome. On observe que la géomatique prend de plus en plus d'ampleur au sein des PNR, mais la branche reste nouvelle et l'apparition de réseaux interParc n'apparaissent que quelques années plus tard. En 2010, se met en place le réseau d'échanges « GéoPNR », dans lequel une liste de discussion commune est instaurée ainsi que des journées de rencontres annuelle. Le réseau s'enrichit avec l'arrivée des Parcs nationaux en 2014. Mais on peut remonter quelques années en arrière pour observer les premières apparitions d'un réseau des Parcs, en 2006, le Système d'Information territorial des PNR de la région Provence-Alpes Côte d'Azur se met en place. À l'origine, ce sont le PNR du Luberon et du Verdon qui s'unissent pour créer une base de données Faune et Flore. Les PNR du Queyras et de Camargue les rejoignent quelques années plus tard. Ce projet, financé dans sa totalité par la Région PACA et l'Europe, se poursuit grâce à la création d'applications métier et lance ainsi la voie aux premières mutualisations entre les PNR.

Application métier

Une application métier est une application informatique qui vient compléter le fonctionnement d'un Système d'Informations Géographiques. Ce sont des applications qui se veulent pleinement intégrées dans les processus métiers des professionnels avec pour objectif de simplifier et d'automatiser leurs actions (saisie de données, analyse, stockage, réalisation cartographique...). Les développeurs qui conçoivent ces applications portent une attention particulière au design de l'interface, qui se doit d'être simple, familier, afin qu'un public non-initié à la géomatique puisse le prendre en main facilement.

En Île-de-France, c'est en 2016 que les PNR et la région entreprennent un projet de réseau des Parcs. Le Parc du Gâtinais, avec l'appui des 3 autres PNR francilien, engage un stage de 6 mois sur le sujet. Une enquête, ouverte en avril 2016, permet d'avoir un écho des projets SIG interParc au sein des PNR. L'enquête compte 15 témoignages de projets SIG en réseau, 14 témoignages de PNR et le témoignage d'un Parc national. Malgré les difficultés évoquées concernant la mise en réseau des Parcs (difficultés à travailler à plusieurs, nécessité d'une animation et d'une bonne communication sur les projets, obligation de parvenir à dégager du temps et à se rendre disponible), la majorité des témoignages souligne les bénéfices de cette mutualisation et la recommande même ouvertement (partages d'expériences, de savoir-faire, de financements et d'outils).

Revenir sur le travail qui a été fait dans le cadre de l'interParc permet de présenter la situation du SIG au début de mon stage mais ce fut également un moyen pour moi de me saisir du travail accompli. Nous avons évoqué lors de la description des 4 services SIG que les Parcs stockaient chacun leurs données en local, sur un serveur interne. Cette organisation est parfaitement adaptée pour une utilisation interne au Parc, mais pour enclencher des dynamiques de mutualisation et de partage d'applications métiers par exemple, il faut s'équiper d'un serveur partagé. Un serveur OVH 2x2 To de disque dur sur Linux a donc été installé, avec l'accompagnement de la société Tranquil-IT-Systems. Ce serveur a l'avantage de pouvoir être subdivisé en autant de machines virtuelles que l'on souhaite.

Il y a actuellement cinq machines virtuelles servant à héberger les deux entités Geotrek, un Wiki, des sauvegardes et un serveur principal avec une base de données, un portail Web et un Geoserver (cf. Figure 3).

Système de virtualisation

Le serveur OVH dispose d'un processeur Intel Xeon DD-1520 4 coeurs. Le système de virtualisation utilisé est XenServer et il permet de créer une multitude de machines virtuelles sur une seule machine physique. XenServer dispose d'une console d'administration permettant de créer et gérer les machines virtuelles. La société Tranqui-IT-System a été en charge de la création de ces machines virtuelles ainsi que de la maintenance de celles-ci. Aujourd'hui, nous n'avons plus de contrat avec cette société.

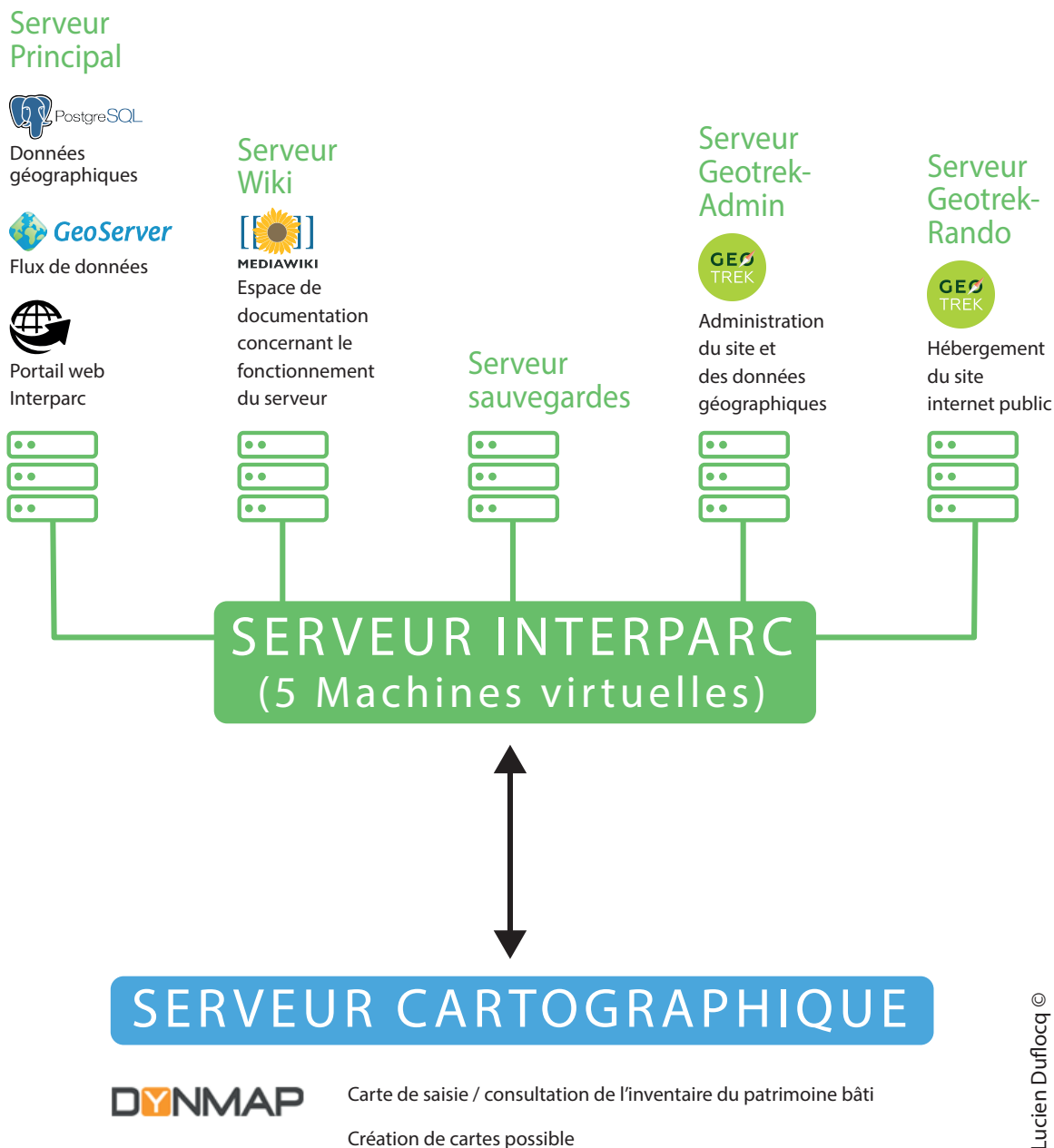


Figure 3 : Organisation du serveur interParc

Premièrement, Geotrek est une application métier. Développé en 2012 par le Parc national des Écrins, du Mercantour, d'Alpi Marittime en Italie et par la société Makina Corpus, c'est un outil « open source » avec des fonctionnalités SIG. Son but premier est de faciliter l'édition et la gestion des itinéraires de randonnées. L'application est organisée en deux parties, la première, destinée aux administrateurs, permet la gestion et la saisie de données et la deuxième, destinée aux utilisateurs, permet d'accéder aux diverses informations thématiques (randonnées, patrimoine, etc.).

Deuxièmement, de façon à disposer d'un espace de documentation en ligne sur le fonctionnement du serveur et plus largement, sur l'organisation du réseau des Parcs, un Wiki a été mis en place (cf. Figure 4). Le Wiki héberge notamment la description de certaines installations, mais aussi des mots de passe permettant l'accès aux outils. L'intérêt de ce Wiki est qu'il est facilement éditable, ce qui permet de l'alimenter aisément au fur et à mesure de l'avancée d'un projet, ce qui fut le cas au cours de mon stage.

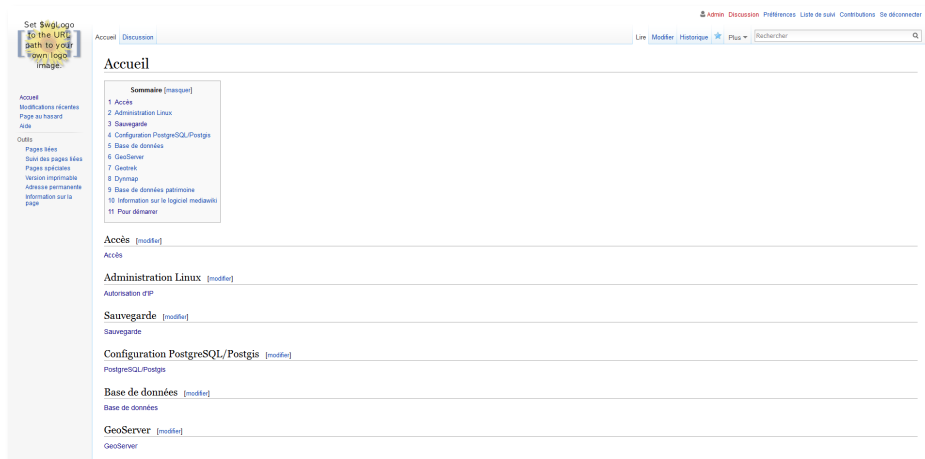


Figure 4 : Page d'accueil du Wiki InterParc

Troisièmement, a été installé sur le serveur principal le Système de gestion de Base de données Postgres afin d'administrer les bases de données. La structuration des bases est la suivante : une base par Parc (intitulés « pnrvf », « pnprop », « pnrhvc », « pnrgrf ») et une base interParc pour les données communes (intitulé « pnridf »). La documentation disponible permet d'appréhender les dénominations des schémas et tables.

Quatrièmement, le WebService GeoServer⁷ a été installé sur le serveur principal pour permettre la diffusion et la modification de données GeoSpatiales sur le web (flux WMS/WFS notamment).

Cinquièmement, un portail web fait office de page d'accueil pour le réseau interParc (cf. Figure 5). Il permet de rediriger l'utilisateur vers les différents outils mis en place. Ce portail, hébergé sur le serveur principal a été développé en langage de programmation Web HTML et utilise la bibliothèque Leaflet pour la cartographie interactive.

⁷ Geoserver : WebService open source permettant de diffuser des données spatiales selon des protocoles normalisés de l'OGC : WMS, WFS



Figure 5 : Page d'accueil du portail web interParc

Enfin, DynMAP (cf. Figure 6) n'a pas été mis en place dans le cadre du réseau des Parcs, mais il constitue un rôle important, car il a été configuré pour marcher avec le serveur interParc (cf. Figure 3). De fait, les données stockées dans la base Postgis ou dans le Geoserver peuvent être intégrées dans le WebSIG DynMAP. DynMAP est l'exemple parfait qui montre que la mise en réseau des Parcs permet la mutualisation d'outils préexistants. Dans le cadre de l'interParc, DynMAP a permis, par exemple, de développer une application pour l'inventaire du patrimoine bâti au sein des PNR.

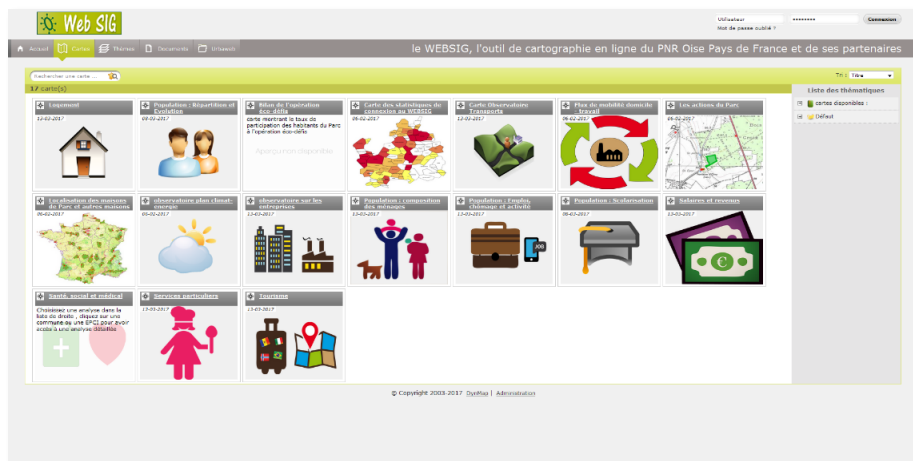


Figure 6 : Page d'accueil du WebSIG DynMAP

En somme, l'étendue du travail qui a été mis en place depuis l'ouverture du réseau des Parcs franciliens n'est pas ici représentée dans sa globalité, de fait, ce n'est pas l'objectif de ce rapport, mais l'intérêt d'y avoir consacré une partie était de montrer que la volonté et la dynamique portées par les 4 PNR est forte. Le but de mon stage est en quelque sorte de prendre le relais en accompagnant le développement du réseau.

Alors de quelle manière va se traduire mon implication et quels vont être la nature de mes objectifs ?

B) Comment animer le réseau des Parcs en poursuivant le développement du SIG ?

1. Pourquoi et comment traiter les métadonnées au sein du réseau des Parcs franciliens ?

Afin de poursuivre le développement du SIG interParc, les 4 chargés de mission ont décidé de traiter mutuellement le cas des métadonnées.

a. Qu'est-ce qu'une métadonnée ?

Une métadonnée est composée de renseignements qui décrivent une donnée, elle est donc indissociable de la donnée elle-même. La métadonnée est présente pour la documentation de l'information géographique, mais aussi pour n'importe quelle autre forme de données (statistique, image, vidéo, etc.). Le contenu des métadonnées rend l'identification et la localisation des données correspondantes possible. De manière courante, on y trouve des informations sur l'authentification, la nature ou la date de la donnée, mais le contenu d'une métadonnée peut être beaucoup plus complet. La mise en place de métadonnées au sein d'un SIG n'est pas anodine, elle traduit généralement une logique de gouvernance⁸ de l'information, le plus souvent définie en amont par les administrateurs. Cette logique de gouvernance permet de pouvoir créer des métadonnées compréhensibles et identifiables par tous et pas uniquement par leur créateur ou par la société qui les édite. L'ère du Big Data rend encore plus important le rôle des métadonnées. En effet, l'accès aux données devenant de plus en plus ouvert, nous avons besoin de clés de lecture pour pouvoir trouver, comprendre et analyser les données.

Big Data

Le Big Data fait référence à l'explosion des données numériques issues des nouvelles technologies. Les 5V du Big Data peuvent permettre de saisir les enjeux du phénomène :

- **Volume** : L'augmentation permanente des volumes de données collectés et analysés.
- **Vitesse** : La vitesse de production des données et leurs traitements qui peut être en temps réel.
- **Variété** : Les données sont d'une grande hétérogénéité (photographies, vidéos, sons, documents, etc.)
- **Véracité** : La fiabilité des données est souvent discutée et il est parfois difficile de reconnaître les fausses informations.
- **Valeur** : L'immensité des données impose de réels efforts pour filtrer et ne sélectionner que les données intéressantes.

La numérisation des métadonnées a permis de rendre plus aisée leur édition et leur diffusion. Néanmoins, il a fallu standardiser le format des métadonnées pour pouvoir les manipuler et les échanger sur différentes plateformes ainsi que sur le web. Un format,

⁸ Source : BECKER Mathieu et HALBOUT Hervé, *Découvrir et comprendre la gouvernance de l'information géographique*, Livre Blanc, 2016.

recommandé par le World Wide Web Consortium, a donc été conçu pour uniformiser et standardiser les métadonnées : c'est le format XMP (eXtensible Metadata Platform), basé sur du langage XML (eXtensible Markup Language). Écrit avec des balises, le XML va respecter une structure qui lui permettra d'être reconnu et interprété par une multitude d'outils. C'est un langage standardisé, simple et configurable qui se doit d'être compris par les auteurs de métadonnées et par les ordinateurs qui vont les faire tourner.

World Wide Web Consortium

« Le World Wide Web Consortium (W3C), est un organisme international qui développe des standards pour le Web afin que les gens puissent communiquer efficacement à travers Internet, autour de formats ouverts garantissant une meilleure inter-opérabilité (c'est-à-dire une meilleure compréhension des systèmes hétérogènes à travers des données et langages standardisés). »

alsacreation.com

Pour résumer, voici les principaux intérêts d'associer des métadonnées à ses données :

- simplifier la recherche de la donnée
- permettre l'ouverture des données au public
- améliorer l'interopérabilité
- renforcer la gestion et l'archivage des données

b. Pourquoi le réseau des Parcs aborde-t-il le sujet des métadonnées ?

On peut citer plusieurs raisons qui ont poussé le réseau des Parcs franciliens à s'intéresser aux métadonnées. Jusque là, nous avons observé assez naturellement que chaque Parc disposait de son propre fonctionnement en terme de gestion de métadonnées. Cette configuration, largement justifiée pour un développement interne au sein d'un PNR, peut devenir problématique dans la perspective d'une mise en commun des données. Dans le cadre de l'interParc francilien, des données vont être collectées et stockées au sein d'une base commune, la gestion de leurs métadonnées se doit donc d'être coordonnée pour que le réseau se développe de manière organisée. Cela rejoint l'idée d'établir une stratégie de gestion des métadonnées afin de rendre pérenne l'évolution du SIG interParc et rendre possible la diffusion de donnée Parc aux partenaires publics. Paradoxalement, la mutualisation interParc impose mais permet également de traiter ce sujet qui de par son coût et sa charge de travail n'avait pas pu se mettre en place précédemment.

Parallèlement, l'arrivée d'une législation telle que la directive européenne INSPIRE⁹ dans le paysage de l'information géographique, pousse petit à petit les institutions qui ont à gérer des données à se doter de métadonnées conformes. On peut également évoquer la conférence de Dublin, en 1995, qui vise à présenter des recommandations pour la description bibliographique des informations présentes sur le web (schéma Dublin Core). Instituée par la Commission européenne, la directive INSPIRE a particulièrement influencé la

⁹ MERRIEN Francis & LEOBET Marc, *La directive Inspire pour les néophytes*, 3^{ème} édition, Mission de l'information géographique (MIG), décembre 2011.

gouvernance de l'information géographique. Obligatoire pour les autorités publiques, recommandée pour les autres organismes, elle impose des spécifications techniques lors de la mise à disposition de données. Pour les métadonnées, la norme impose de renseigner une liste de champs : intitulé, résumé, type, Mots-clés, point de contact, etc. Cela facilite la compréhension pour l'utilisateur mais rend la tâche plus fastidieuse pour le créateur.

c. Comment répondre à la demande du réseau des Parcs : objectif du stage

Après avoir expliqué les différentes motivations qui ont poussé le réseau des Parcs franciliens à mettre en place un outil de gestion des métadonnées, tâchons de présenter les moyens mis en place pour y répondre et donc les objectifs qui ont été fixés pour le stage. On peut se référer au diagramme de Gantt présenté en annexe 2 de ce rapport pour visualiser les différentes phases de mon stage. En premier lieu, j'ai analysé la situation du réseau des Parcs, puis j'ai effectué une étude globale sur les métadonnées. Les connaissances acquises par ces deux analyses sont retranscrites dans les parties précédentes. Elles m'ont permis de comprendre et de saisir la dynamique des PNR afin de pouvoir proposer ensuite les idées les plus cohérentes possible.

En deuxième lieu, une phase de consultation est mise en place avec les chargés de mission SIG des 4 Parcs. L'enjeu de ce projet repose sur une mutualisation des PNR, où l'idée est de faire avancer le projet avec l'accord et les idées de chacun. Cette phase de concertation qui s'est instaurée avait comme priorité de rencontrer les chargés de mission SIG de chaque Parc, de manière à recueillir leurs témoignages concernant l'outil de gestion des métadonnées. J'ai également profité de ces rencontres pour connaître le fonctionnement et les outils SIG mis en place au sein de leur propre Parc (un exemple de fiche préparé dans le cadre de ces rencontres est disponible en annexe 3 de ce rapport). Le recueil de ces témoignages m'a permis de dresser une liste avec l'ensemble des besoins et enjeux souhaités par les PNR. Cette liste est disponible ci-dessous, les attentes ne sont pas classées par ordre d'importance (cf. Figure 7).



Attentes de l'outil de gestion des métadonnées



- La capacité à saisir simplement et rapidement une métadonnée
- La capacité à publier le catalogue en ligne
- La capacité à pouvoir exporter une partie ou la totalité du catalogue en PDF
- La capacité à pouvoir envoyer la métadonnée à un partenaire lors de la mise à disposition d'une donnée
- La capacité à pouvoir récupérer l'ensemble de nos métadonnées en cas de migration
- La capacité à pouvoir moissonner d'autres catalogues
- La capacité à automatiser les tâches (liste des champs, emprise géographique, date...)
- La capacité à paramétrer un moteur de recherche performant (par thématique, mot clé, emprise spatiale)
- La possibilité d'associer l'outil à Postgres (volonté du Parc de se tourner vers un unique SGBD)
- La capacité à visualiser la donnée (aperçu ou dynamique)
- La capacité à télécharger la donnée
- La capacité à respecter les standards (INSPIRE, ISO19110, ISO19115)
- La capacité à pouvoir attribuer un statut à la donnée (données ouvertes à valoriser ou données internes à accès restreint)
- La capacité à relier dynamiquement la donnée et sa métadonnée
- La capacité à pouvoir produire des actes d'engagements lors de la transmission de la donnée
- La capacité à pouvoir construire un historique des modifications sur les données
- La capacité à pouvoir renseigner des métadonnées qui ne respectent pas les standards ou alors pouvoir apporter des annexes (nature des champs par exemple)
- La faculté de posséder une interface web agréable et ergonomique

Figure 7 : Attentes de l'outil de gestion des métadonnées

À la suite de ce travail, l'objectif de mon stage est clair, je dois mettre en place un outil de gestion des métadonnées interParc de manière à terme, à pouvoir mutualiser et rendre accessible l'ensemble des métadonnées produites dans le cadre de l'interParc.

Ce sera également un outil qui pourra servir aux données de chacun des Parcs. Cet outil prendrait les traits d'une plateforme de catalogage et de diffusion des données géographiques, conviviale et adaptée à un large public. L'outil final doit pouvoir retranscrire un maximum des attentes des chargés de mission.

2. Solution : le catalogage comme outil de gestion des métadonnées

Le Géocatalogue participe au bon développement d'un SIG. Bien structuré, il constitue un accès direct à l'étendue des données. Pour l'administrateur des données, le géocatalogue permet de gérer de manière organisée la donnée et sa métadonnée et pour l'utilisateur, il offre un aperçu et un accès simplifié à l'information qu'il souhaite utiliser.

a. Qu'est-ce qu'un catalogue de métadonnées ?

Nous avons déjà présenté en amont du rapport le phénomène de Big Data qui vient accroître la quantité d'informations géographiques. Le phénomène d'Open Data¹⁰ vient lui augmenter les échanges d'informations entre les structures. Si la métadonnée fait office de clé de lecture pour l'information géographique, le catalogue lui va permettre d'harmoniser, classer et structurer ces clés de lecture. Si nous raisonnions par analogie et que nous comparerions le catalogue à un dictionnaire, les définitions seraient les métadonnées et le dictionnaire constituerait le catalogue. Nous retrouvons alors une logique d'indexation au sein du catalogue, qui s'apparente à la recherche par ordre alphabétique pour le dictionnaire. Pour les gestionnaires de données géographiques, la mise en place d'un catalogue est valorisante pour le SIG mais elle constitue un travail conséquent pour l'alimenter et le garder à jour. Soulignons que la présence d'un catalogue de métadonnées est une plus-value. Associer à d'autres outils et services, il devient un complément au bon fonctionnement d'un SIG.

b. Présentation des différentes solutions de catalogage et analyse

Deux hypothèses semblent s'offrir à nous pour le choix du catalogue. La première consisterait à utiliser des outils déjà existants. Ces outils, libres ou payants, permettent pour la plupart de consulter et gérer les métadonnées selon les différents standards (INSPIRE, ISO¹¹). La deuxième hypothèse se traduirait par le développement de notre propre outil de gestion de métadonnées. Cette solution pourrait s'apparenter aux solutions déjà existantes avec toutefois une personnalisation importante selon nos besoins. Pour la première hypothèse, une étude assez large peut être menée, néanmoins, certains critères importants vont venir d'eux-mêmes exclure certaines solutions. Premièrement, nous nous sommes orientés vers des outils répandus, garantissant une certaine fiabilité dans le temps. Pour cela, j'ai effectué une veille technologique sur les différentes solutions en m'appuyant notamment sur le Catalogue des Infrastructures de Données Géographiques (IDG), produit par l'Association Française pour l'Information Géographique (AGIGÉO) et le projet européen eSDI-Net / EUROGI (European Umbrella Organisation for Geographic Information).

¹⁰ Open Data : Effort des institutions pour le partage des données dont elles disposent.

¹¹ ISO : Il définit en détail comment décrire les ressources dans le domaine de l'information géographique tel que les métadonnées. Ce standard précise les descripteurs obligatoires et conditionnels.

Infrastructure de Données Géographiques (IDG)

« Une infrastructure de données géographiques est une structure de mutualisation, d'échange et de diffusion de données géographiques à l'échelle d'un territoire et au bénéfice d'acteurs publics, et indirectement des citoyens. »

Afgeo

On reconnaît une IDG à ses services transversaux : catalogue, Web services, visualiseur, pages Web, etc.

Ce document dresse un panorama des IDG, il en recense 65 (13 nationales, 34 régionales et 18 départementales). Le tableau ci-dessous montre les différents types de catalogue au sein des IDG :

Nom du catalogue	Nombre d'IDG l'utilisant
Géosource	7
Prodige	12
Geonetwork	3
GeOrchestra	3
PIGMA	1
Gamme ESRI	6
Développement interne	3
Géoportail	2
Géocatalogue	1
ADélie	1
Aigle IDS (Business Geographic)	1
Non renseigné	25
TOTAL	65

Tableau 2 : Répartition des solutions de catalogage au sein des IDG françaises

Sur le nombre de catalogue renseignés (40 sur 65), Géosource se place en première position, devant son homologue Geonetwork, puis apparaissent les outils proposés par la gamme ESRI. Précisons que l'outil Prodige utilise la technologie Géosource et que les outils geOrchestra et PIGMA utilisent la technologie Geonetwork. Enfin, trois IDG ont mis en place leur propre solution à travers un développement interne.

Dans l'élan du réseau d'échange « GéoPNR », les 51 Parcs français ont mis en place un tableau commun permettant de renseigner l'outil utilisé pour la gestion de leurs métadonnées. Malgré une demande groupée du 24 mars 2017 pour mettre à jour le document, certains Parcs n'ont pas mis à jour leurs informations depuis plusieurs mois voir plusieurs années. Le tableau ci-dessous résume les types de solution mis en place au sein des PNR français :

Outil de gestion des métadonnées	Nombre de PNR l'utilisant
En réflexion	4
ArcCatalogue	4
Géosource	3
Excel	3
Access	2
DynMAP	2
QGIS (plugin QSphère)	2
PIGMA	1
Outil développé en interne	1
Carmen	1
PRODIGE	1
GéoLimousin	1
GéoNormandie	1
Non renseigné / Aucun	25
TOTAL	51

Tableau 3 : Répartition des outils de gestion des métadonnées au sein des PNR français

On observe qu'il n'y a pas un mode de gestion qui domine, ce sont davantage des outils permettant la création de métadonnées que des outils de catalogage. On retrouve tout de même Géosource et quelques réseaux avec des IDG régionales permettant à ces Parcs de mutualiser les catalogues. Premièrement, nous nous sommes assez naturellement orientés vers une solution libre de manière à garantir son fonctionnement dans le temps, sans avoir à gérer des problématiques de renouvellement de licence. Néanmoins, nous avons conscience que l'installation, l'hébergement et la maintenance sont à prendre en compte dans le calcul de retour sur investissement et que malgré le fait que la solution soit libre, les différents éléments évoqués ci-dessus nécessitent des compétences spécifiques et du temps agent. Ainsi, la gamme ESRI¹², l'outil Aigle IDS¹³ de Business Geographic et ISOGE¹⁴ ont été exclus. Le catalogue Adélie, qui ne peut pas fonctionner avec une base de données de type Postgres est utilisé par très peu d'organismes possédant un SIG. Il a donc été lui aussi écarté. À ce stade, il ne reste que les solutions suivantes : Géosource, Geonetwork, Prodigé et geOrchestra.

Pour orienter à nouveau notre choix, il a été nécessaire dans un premier temps de vérifier si l'outil répondait assez largement aux attentes qui avaient été formulées avec les chargés de mission SIG (cf. Figure 7). Dans un deuxième temps, l'interface et l'ergonomie des solutions qui nous semblaient pertinentes ont été testées via des démos accessibles en ligne et à

¹² <http://desktop.arcgis.com/fr/arcmap/10.3/main/get-started/a-quick-tour-of-arccatalog.htm> (présentation d'ArcCatalog)

¹³ <https://www.business-geografic.com/fr/actualites/communiqués-de-presse/726-busines-geografic-geo-aigle-ids.html> (présentation d'Aigle IDS)

¹⁴ <http://www.isogeo.com/fr/solution/> (présentation d'ISOGE)

travers quelques exemples de plate-forme. Dans un troisième temps, la qualité de la documentation et l'animation de la communauté d'utilisateurs de l'outil ont été prises en compte. L'ensemble des résultats est affiché dans le tableau ci-dessous :

	Geonetwork	Geosource	geOrchestra	Prodige
Open source	✓	✓	✓	✓
Importance de la communauté	+++	++	+++	+
Richesse de la documentation	++	++	++	+
Portée internationale de l'outil	+++	+	+++	+
Couplage avec le SGBD Postgres	✓	✓	✓	✓
Interopérabilité de l'outil	+++	+++	+++	
Métadonnées conformes INSPIRE	✓	✓	✓	✓
Respecte la normalisation ISO	✓	✓	✓	✓
Qualité du Webdesign et de l'interface de navigation	+++	+++	++	+
Capable de saisir des métadonnées	✓	✓	✓	✓
Capable de publier sur Internet	✓	✓	✓	✓
Capable d'exporter le catalogue en PDF	✓	✓	✓	✓
Capable d'envoyer la métadonnée à un partenaire	✓	✓	✓	✓
Capable de récupérer les métadonnées en cas de migration	✓	✓	✓	✓
Propose le Moissonnage	✓	✓	✓	✓
Automatiser l'édition de fiche	✓	✓	✓	✓
Propose un outil de recherche performant	✓	✓	✓	✓
Capable de visualiser la donnée	✓	✓	✓	✓
Capable de télécharger la donnée	✓	✓	✓	✓
Capable d'attribuer un statut à la donnée	✓	✓	✓	✓
Maintenir le lien entre la donnée et la métadonnée	?	?	?	?
Obtenir un historique du catalogue	✓	✓	✓	✓
Capable d'ajouter des documents annexes aux métadonnées	✓	✓	✓	✓

Tableau 4 : Analyse de quatre solutions de catalogage (Geonetwork, Géosource, geOrchestra et Prodige)

Évaluation ordonnée (+++ : très bien, ++ : bien, + : moyen) Évaluation binaire (✓ : la solution remplit l'attente ; X : la solution ne remplit pas l'attente) ? : doute sur la capacité de la solution à répondre à l'attente

Nous nous sommes aperçus que les quatre solutions répondaient à l'ensemble des attentes. Les critères de choix s'affinent alors avec le niveau de documentation, de la communauté d'utilisateurs, mais aussi sur la convivialité et la qualité de l'interface web (style graphique, ergonomie). Nous décidons alors d'exclure Prodiges, qui possède une communauté sensiblement moins active et un Web design moins moderne. Pour affiner notre choix entre les trois dernières solutions une analyse plus fine des critères a été nécessaire.

Développé au cours des années 2000 par la Food Agriculture Organisation (FAO), Geonetwork¹⁵ est devenu une référence en matière de catalogage de métadonnées. La sortie de la version 3.0 en mars 2015 traduit une nouvelle génération pour les catalogues de métadonnées. Davantage dynamique, plus convivial, l'effort apporté sur l'interface utilisateur, avec l'intégration de nouvelles technologies Java et Bootstrap, rend nettement plus séduisante l'approche du catalogue de métadonnées classique jusque-là peu ergonomique. Une situation qui n'incitait pas nécessairement les structures à s'équiper de ce genre d'outil. Ces évolutions garantissent le dynamisme des contributeurs, développeurs et financeurs de Geonetwork et inéluctablement rassure sur la pérennité de l'outil.

Géosource¹⁶, mis en place en 2006 par le Bureau de Recherches Géologiques et Minières (BRGM) et par le Ministère de l'Écologie, du Développement durable et de l'Énergie, l'outil est un fork de Geonetwork, c'est-à-dire qu'il a été créé à partir de son code source. C'est donc une adaptation française de Geonetwork. Hormis le fait que la langue de l'outil et l'importance de la communauté d'utilisateur ne soient pas les mêmes (langue anglaise et communauté plus importante pour Geonetwork), les deux solutions possèdent peu de différences.

GeOrchestra¹⁷ se différencie des deux autres solutions. C'est une infrastructure de données spatiales (IDS) interopérable et modulaire, créée en 2009, pour répondre aux exigences de la directive INSPIRE en Europe. Le principal avantage de geOrchestra est d'être structuré par une suite de modules indépendants et interopérables, parmi lesquels il est possible de choisir pour composer une IDS à la carte. Pour rendre opérables ces liens, les différents modules ont été modifiés afin d'utiliser un unique système d'identification. On y trouve les modules suivants :

- Un catalogue de métadonnées basées sur Geonetwork
- Un serveur cartographique basé sur Geoserver
- Un serveur de tuile basé sur GeoWebCache
- Un visualiseur et éditeur basé sur MapFishapp
- Un extracteur destiné aux téléchargements
- D'autres modules permettant d'administrer la solution

¹⁵ <http://geonetwork-opensource.org> (Site internet officiel de Geonetwork)

¹⁶ <http://www.geosource.fr> (Site internet officiel de Géosource)

¹⁷ <http://www.georchestra.org/fr/> (Site internet officiel de geOrchestra)

L'interopérabilité et la communication entre ces différents modules se font selon les standards de l'Open Geospatial Consortium (OGC) et des API REST (« Application Programming Interfaces », « Representational State Transfer »).

OGC / API REST

OGC : « L'Open Geospatial Consortium, ou OGC, est un consortium international pour développer et promouvoir des standards ouverts, les spécifications OpenGIS, afin de garantir l'interopérabilité des contenus, des services et des échanges dans les domaines de la géomatique et de l'information géographique. »

Wikipédia

API REST : Une API compatible REST est une interface de programmation d'application qui fait appel à des requêtes HTTP pour obtenir (GET), placer (PUT), publier (POST) et supprimer (DELETE) des données. Elle intervient comme un langage permettant de connecter des services Web entre eux.

L'intérêt est donc de relier ces différentes briques et d'avoir des ponts entre elles. Cela va permettre par exemple d'établir un lien bilatéral entre la métadonnée et la donnée, avec un dialogue entre le catalogue et le visualiseur. Pour ce qui nous concerne, geOrchestra propose des possibilités plus larges et plus complètes que les outils de catalogage comme Geonetwork et Geosource. Ajoutons que l'outil est porté par une importante communauté, à l'image de la société CamptoCamp, qui organise chaque année les journées geOcom (<https://www.camptocamp.com/actualite/geocom-2017-programme/>) mais également par plusieurs groupes de discussion très actifs.

c. Choix de la solution : de geOrchestra à Geonetwork

C'est le 5 avril 2017, au cours d'une réunion avec l'ensemble des chargés de mission du réseau des Parcs francilien que nous décidons de la solution à adopter. Tout d'abord, après avoir vu les possibilités offertes par les solutions de catalogage open source, nous décidons de ne pas retenir l'idée de concevoir notre propre outil de manière globale. La crainte de ne pas avoir le temps de rendre opérationnel un outil en six mois vient également conforter cette décision.

Durant cette réunion, nous optons pour l'IDS geOrchestra. En plus de répondre à l'attente initiale qui consiste à mettre en place un catalogue de métadonnées, l'outil offre des possibilités plus larges en matière de développement du SIG. Néanmoins, suite à des difficultés techniques rencontrés lors de l'installation de l'outil que nous évoquerons en page 32 de ce rapport, il a été décidé le 17 mai 2017 de changer de solution et de choisir Geonetwork. Il s'est avéré que l'installation de geOrchestra requérait des compétences techniques importantes en Java et en gestion des serveurs. Le nombre de modules et la gestion des API REST ont laissé supposer un temps de maintenance conséquent, rendant sensible chaque mise à jour de l'outil. Le risque aurait été de léguer une application difficilement gérable pour les chargés de mission SIG. Au vu des délais, une installation de geOrchestra aurait méritée d'être poursuivie à la suite de mon départ, de manière à rendre l'outil opérationnel. Une autre raison plus stratégique est venue conforter ce choix. Il est prévu que le PNR d'Oise – Pays de France s'équipe de l'IDS GEO d'ici peu de temps. Comme

cela a été fait avec DynMAP, le Parc pourrait mutualiser GEO au sein du réseau des Parcs. Il ne paraît donc pas judicieux d'installer un IDS de type geOrchestra, car le risque serait de se retrouver à terme avec deux IDS (geOrchestra et GEO).

La gestion de projet a été une composante importante durant cette période. Il a été utile de planifier dès le début du stage les tâches à réaliser à travers un diagramme de Gantt (cf. annexe 2). Ce document servait de référence pour juger l'état d'avancement du projet et a permis de mettre en évidence le moment où il a fallu changer de solution. Le risque aurait été de rester trop longtemps fixé sur l'installation de geOrchestra et de perdre un temps crucial pour la suite du stage.

Partie 2 : Mise en place de l'outil

A) Installation

1. GeOrchestra : retour sur les difficultés rencontrées lors de l'installation

Suite à la réunion du 5 avril 2017, la prochaine étape fut d'installer geOrchestra. Les objectifs étaient les suivants : tester l'installation, lister les logiciels prérequis et comprendre comment s'imbriquent les différents modules. Nous savons dorénavant que cette installation a été abandonnée, mais il est intéressant d'y revenir afin d'expliquer les raisons techniques qui nous ont amenées à cette décision. Il a été décidé de réaliser cette installation sur une machine virtuelle (VM) Debian.

Machine virtuelle (VM)

« Une machine virtuelle est un fichier informatique, généralement appelé image, qui se comporte comme un ordinateur réel. En d'autres termes, il s'agit d'un ordinateur créé à l'intérieur d'un ordinateur. Elle s'exécute dans une fenêtre, comme tout autre programme, en offrant à l'utilisateur final une expérience identique à celle qu'il aurait sur le système d'exploitation hôte.

La machine virtuelle est isolée du reste du système, les logiciels installés sur la machine virtuelle ne peuvent impacter l'ordinateur hôte. »

azure.microsoft.com

Concernant le matériel, j'ai utilisé mon ordinateur personnel, un MacBook Pro Intel Core i5 64 bits. La machine dont je disposais au sein du Parc avait un système 32 bits, elle ne permettait pas la mise en place de VM. De plus, il me paraissait plus confortable de travailler au sein d'un environnement Mac auquel je suis habitué. Il a ensuite fallu faire en sorte que l'environnement de travail de la VM se rapproche au plus près de l'environnement du serveur interParc. Pour cela, j'ai installé le même système d'exploitation Debian, sans interface graphique. Une première difficulté a été de prendre en main la VM tout en se familiarisant avec les commandes Linux via le Shell¹⁸ de connexion. La pratique régulière de la console Shell et l'aide de plusieurs documentations¹⁹ sur internet m'ont aidé durant cette phase. La documentation la plus riche sur l'installation de geOrchestra a été trouvée sur la plateforme GitHub²⁰. Aux premières lectures, je découvre rapidement un vocabulaire rattaché au domaine du développement informatique qui ne m'est pas familier et je bute naturellement sur des points techniques. Différentes installations sont proposées, selon les objectifs et les attentes que l'on a de geOrchestra. L'installation via Docker²¹ apparaît comme le bon compromis, elle est conseillée pour une prise en main rapide et une utilisation en mode test. Toutefois, n'ayant jamais manipulé Docker, il m'a fallu un certain temps avant de l'appréhender.

¹⁸ Shell : programme servant d'interface entre le système et l'utilisateur.

¹⁹ http://www.di.ens.fr/~pouzet/cours/systeme/unix/commandes_Unix.html (Rappel des commandes Unix)

²⁰ <https://github.com/georchestra/georchestra> (Documentation d'installation pour geOrchestra)

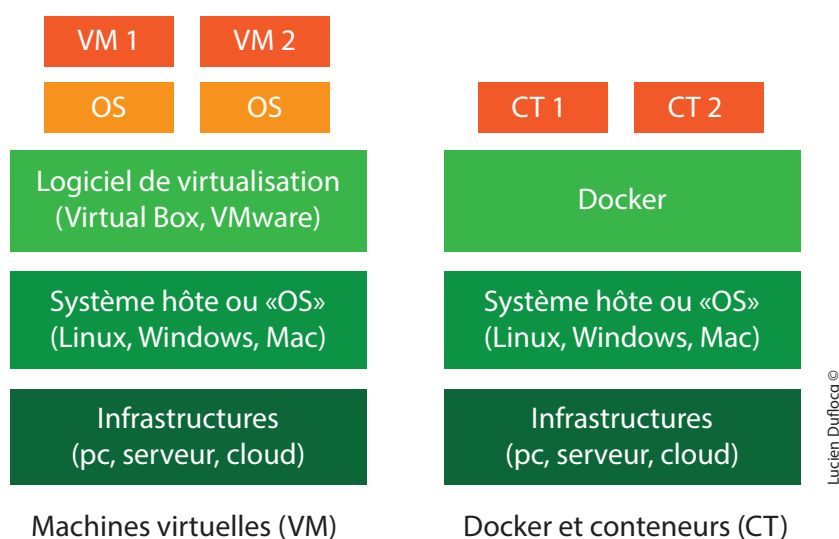
²¹ <http://www.journaldunet.com/solutions/cloud-computing/1146290-docker-tout-ce-qu-il-faut-savoir/> (Définition de Docker)

Docker

« Docker permet d'embarquer une application dans un conteneur virtuel qui pourra s'exécuter sur n'importe quelle machine. C'est une technologie qui a pour but de faciliter les déploiements d'application. Elle est en partie proposée en open source (sous licence Apache 2.0) par une société américaine, également appelée Docker, qui a été lancée par un Français : Solomon Hykes.»

Antoine Crochet-Damais, Journal du Net

La grande différence avec une VM classique, c'est que le conteneur de Docker n'embarque pas de système d'exploitation, il est donc nettement plus léger, mais aussi beaucoup plus facile à migrer (cf. Figure 8).



Lucien Duflocq ©

Figure 8 : Comparaison entre une machine virtuelle et un conteneur Docker

J'ai passé beaucoup de temps à installer Docker. Après avoir testé plusieurs configurations de VM, je me suis aperçu qu'il y avait deux composants Docker à installer : « Docker engine »²² et « Docker compose »²³. Bien que non achevée, la documentation que j'ai produite pour l'installation de geOrchestra via Docker, qui repose en majeure partie sur la documentation disponible en ligne²⁴, est en annexe 4 de ce rapport. Une série d'échanges sur le groupe de discussion Google de geOrchestra m'a permis de résoudre plusieurs soucis de ligne de commandes et d'arriver au bout de l'installation. Malgré tout, geOrchestra ne fonctionnait pas sur ma VM. Suite à cela, le PNR du Vexin français a testé cette installation sous Docker sans succès également. Il semble néanmoins que la source du problème vienne de la configuration d'un fichier permettant d'administrer les connexions aux adresses IP (fichier « hosts »). Ces difficultés techniques, couplées aux différentes raisons évoquées dans la partie précédente (Choix de la solution : de geOrchestra à Geonetwork, page 30), nous ont amené à installer la solution Geonetwork.

²² <https://docs.docker.com/engine/installation/linux/debian/> (documentation d'installation de « Docker engine »)

²³ <http://blog.programster.org/debian-8-install-docker-compose/> (documentation d'installation de « Docker compose »)

²⁴ <https://github.com/georchestra/docker/blob/master/README.md> (documentation d'installation de geOrchestra)

2. Installation de Geonetwork

a. Procédé d'installation en monoposte sur un serveur local

Pour l'installation de Geonetwork, j'ai utilisé la même configuration de VM que pour geOrchestra. Une documentation détaillée a été mise en place en collaboration avec Alexandra Greco, stagiaire au PNR du Verdon mais aussi collègue de promotion du Master SIGMA. Une partie de la documentation est présentée en annexe 5 de ce rapport et sa totalité est disponible sur la plateforme GitHub²⁵. Cette documentation comprend la quasi-totalité des manipulations effectuées au cours de la mise en place de Geonetwork. Elle est bien sûr destinée à laisser une trace de ce qui a été fait au réseau des PNR franciliens, mais aussi à toutes les personnes désireuses de tester Geonetwork sur leur machine. Cette documentation a également constitué la base de l'installation sur le serveur interParc, pour une utilisation en mode production.

L'installation de Geonetwork a pris quatre jours, ce qui est peu comparé au temps passé pour l'installation de geOrchestra. Après avoir trouvé une documentation claire et détaillée sur internet²⁶, je n'ai pas rencontré de difficultés majeures. Le choix du serveur d'application Web a toutefois été important.

Serveur d'application Web

Un serveur Web d'application permet de faire fonctionner des applications Web sur un serveur. Il génère des réponses aux requêtes de l'utilisateur. Les principaux serveurs d'application sont Tomcat, JBoss, JRun et JOnAS.

Par défaut, Geonetwork est installé avec le serveur d'application Jetty, qui a l'avantage d'être très léger et donc de pouvoir être lancé rapidement. Toutefois, le serveur interParc, destiné à accueillir le catalogue, fait tourner les applications Web avec le serveur Tomcat. Plus lourd que son homologue Jetty, il apparaît plus performant lorsqu'il s'agit de gérer de nombreuses applications Web. Il a donc été nécessaire d'installer Tomcat puis de déposer le fichier d'installation Geonetwork (.WAR)²⁷ au sein du serveur Tomcat (documentation disponible sur la plateforme GitHub).

b. Couplage avec la base de données Postgres

Par défaut, Geonetwork est rattaché à une base de donnée H2²⁸. Comme pour le serveur Jetty, H2 a été choisi pour sa légèreté (7mo), sa portabilité (un seul fichier « jar » exploitable sur un système disposant de java) et ses fonctionnalités (schémas multiples, triggers, etc.). L'objectif était d'avoir un fichier .war Geonetwork opérationnel et le plus léger possible. Ainsi, la base de donnée H2 n'est pas conçue pour supporter de gros volumes de données ni un nombre important d'utilisateurs, elle est utile pour réaliser des tests sur des applications notamment. En ce qui nous concerne, la base de données utilisée au sein du serveur interParc est PostgreSQL. Il a donc fallu remplacer H2 par PostgreSQL. Les détails de cette configuration sont présentés dans la documentation en annexe 6.

²⁵ <https://github.com/GeoNetworkTeam> (Documentation open source officielle sur l'installation de Geonetwork)

²⁶ <https://perso-sdt.univ-brest.fr/~cmartin/info/geonetwork/geonetwork-jre8> (Documentation officielle sur Geonetwork)

²⁷ WAR (Web Application Archive) : fichier contenant l'ensemble des composants d'une application Web.

²⁸ <http://www.h2database.com/html/main.html> (Site internet officiel de la base de données H2)

PostgreSQL

PostgreSQL est un Système de Gestion de Base de Données open source, possédant des fonctionnalités puissantes et programmables via divers langages (Python, Java, Perl).

« Une installation de PostgreSQL (nommée cluster) contient une ou plusieurs bases de données, totalement indépendantes. Chaque base de données peut ensuite contenir un ou plusieurs schémas, ou subdivisions logiques, ces derniers contenant les objets courants de base de données (tables, vues, fonctions, requêtes, etc.). »

Laurent Jégou

Quatre services Postgres :

- PostgreSQL : Système de Gestion de Base de Données.
- PostGIS : Module spatial qui confère à PostgreSQL le statut de SGBD spatial.
- PSQL : Interface en ligne de commande permettant la saisie de requêtes SQL.
- pgAdmin : Outil d'administration graphique pour PostgreSQL.

Par défaut, lorsque Geonetwork est configuré avec PostgreSQL et lorsqu'il est lancé pour la première fois via le serveur d'application web Tomcat, il crée l'ensemble de sa base de données dans le schéma « public ». C'est à travers l'outil d'administration graphique pgAdmin que nous avons visualisé et analysé la base de données. Face à ce constat et après discussions avec les chargés de mission des PNR franciliens, il nous est paru plus intéressant pour gérer la base de données, de stocker les tables Geonetwork non pas dans le schéma « public » mais dans un schéma que nous aurions créé spécialement pour lui. En effet, le schéma « public » est un schéma créé par défaut lors de l'installation de PostgreSQL, il n'a pas pour objectif de stocker la base de données d'une application web. L'ensemble des manipulations qui nous ont permis de faire cela est présenté dans la documentation en annexe 7 de ce rapport.

c. Paramétrages spécifiques lors de l'installation sur le serveur interParc

Pour l'installation de Geonetwork sur ma propre VM, je ne me suis pas préoccupé des réglages liés aux adresses IP et au pare-feu. En effet, Geonetwork étant accessible uniquement via l'adresse IP de ma machine, ces configurations n'ont pas été prises en compte directement.

Lors de l'installation sur le serveur interParc, il a fallu par contre gérer ces éléments. Premièrement, une nouvelle VM a été créée sur le serveur pour héberger le catalogue Geonetwork (8Go de RAM lui on été alloué). Concernant le pare-feu, la nouvelle VM bénéficie du même pare-feu que les autres VM, appelé « Fwall », c'est une solution Linux. Le pare-feu est configurable via ce qu'on appelle le « Shorewall », une interface en ligne de commande Shell permettant d'administrer les différents droits d'accès.

Pare-feu

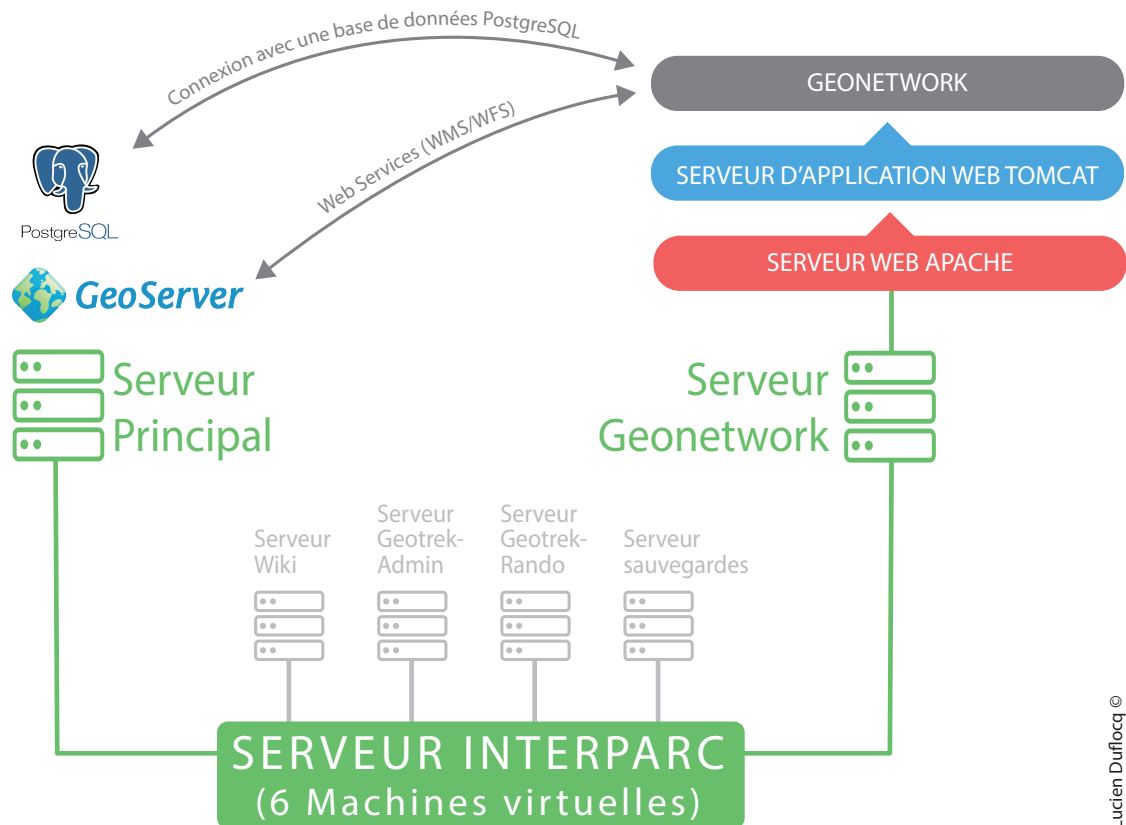
« Un pare-feu est un mécanisme de sécurité réseau, matériel ou logiciel, qui contrôle le trafic entrant et sortant selon un ensemble de règles prédéfinies. »

Margaret Rouse, lemagit.fr

C'est à travers le « Shorewall » que nous avons permis à Geonetwork d'accéder au Web et à la base de données interParc PostgreSQL présente sur une autre VM (cf. annexe 8).

B) Comprendre le fonctionnement de Geonetwork

1. La place de Geonetwork au sein du serveur interParc



Lucien Duflocq ©

Figure 9 : Organisation du serveur avec Geonetwork

Le schéma ci-contre (cf. Figure 9) vise à comprendre comment Geonetwork s'articule au sein du serveur interParc. Premièrement, nous retrouvons la structure du serveur interParc déjà évoqué au sein de la Figure 3 page 17. À la différence que le serveur ne possède plus seulement cinq VM mais six. La sixième étant consacrée à Geonetwork, elle est appelée

« Serveur Geonetwork ». Au sein de cette VM, on observe un serveur Web Apache, un serveur d'application Web Tomcat et l'application Web Geonetwork. Le serveur web Apache, ou serveur http, va être en mesure de comprendre et de gérer les requêtes http. Les requêtes http correspondent au langage entre l'utilisateur et le serveur. Que ce passe-t-il par exemple lorsque vous renseignez une adresse internet dans un navigateur web Google et que vous lancez la recherche ? Dans la mesure où l'adresse est bien renseignée et qu'elle existe, la page s'ouvre dans le navigateur. Techniquement, c'est le serveur Web, en l'occurrence Google Web Server (serveur Web qui repose sur une version modifiée d'Apache), qui va recueillir l'adresse et qui va parvenir à l'interpréter de manière à exécuter la commande, ou requête http. Pour Geonetwork, c'est le même processus, il est nécessaire d'avoir un serveur Web pour que l'utilisateur puisse accéder à la page internet de l'application, ici, c'est le serveur web Apache mais il en existe d'autres (Nginx, Lighttpd). Ensuite, le serveur d'application Tomcat, qui a déjà été présenté dans ce rapport, prend le relais et permet l'exécution de GeoNetwork.

Enfin, on observe deux flèches partant de l'application Geonetwork vers deux composants d'une autre VM. Ces deux composants sont le SGBD PostgreSQL et le Webservice GeoServer. En effet, nous l'avons vu précédemment, Geonetwork va établir une connexion avec PostgreSQL, en créant automatiquement une base de données pour y stocker les informations (métadonnées, groupes d'utilisateurs, etc.). Le deuxième composant va permettre de relier des métadonnées présentes dans Geonetwork avec des données du Geoserver (flux WMS/WFS)²⁹.

2. Obtenir et analyser le Modèle Conceptuel de Données

Une autre constatation a été de se rendre compte que nous ne possédions pas le Modèle Conceptuel de Données (MCD) de Geonetwork. Cela peut dépendre de l'implication et du rôle dans la gestion de Geonetwork, mais connaître le MCD va permettre plusieurs choses. Tout d'abord, nous pourrions saisir visuellement la structure de la base de données (champs, clés primaires, clés étrangères, relations entre les tables). Deuxièmement, le MCD est d'autant plus utile, voir indispensable si nous désirons manipuler et personnaliser la base de données, en ajoutant des relations, des triggers ou en effectuant des requêtes SQL par exemple. Et c'est notre cas, nous prévoyons d'intervenir sur la base de données Geonetwork de manière à la relier dynamiquement avec la base de données contenant les données des Parcs³⁰.

Le MCD est introuvable dans les documentations officielles et dans les différents forums qui traitent le sujet. Il a donc fallu trouver une solution pour l'obtenir. De manière générale, lors de la création d'une base de données, il est plus courant de procéder ainsi : élaboration du MCD puis création de la base de données. Dans notre cas, c'est l'inverse, la base de données est déjà créée et nous souhaitons obtenir le MCD. On parle dans ce cas de « reverse engineering », ou rétro conception.

²⁹ Se référer à la page 40 de ce rapport

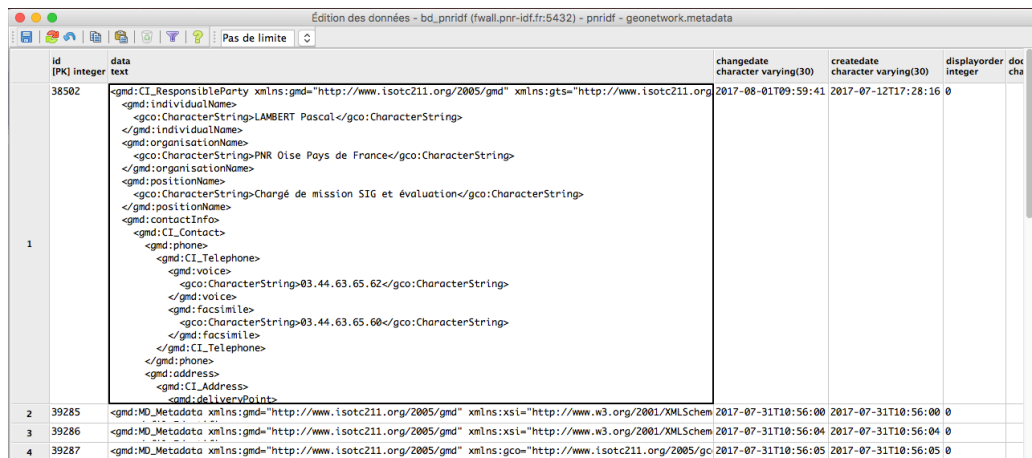
³⁰ Se référer à la page 46 de ce rapport

Reverse engineering ou rétro-conception

L'objectif du «reverse engineering» est d'obtenir un MCD ou un MLD (Modèle Logique de Données) à partir d'un script SQL, contenant l'ensemble des instructions CREATE TABLE utilisées pour définir la structure d'une base de données.

Il existe plusieurs logiciels pour appliquer la rétro-conception (PowerAMC, DB Designer), j'ai retenu pgModeler. C'est un outil libre, multiplateforme, très simple d'utilisation, destiné à l'origine à la conception de base de données sous PostgreSQL. Après l'avoir installé, il suffit d'éditer une connexion à une base de données PostgreSQL (nom de la base, hôte, utilisateur et mot de passe) et pgModeler génère automatiquement le MCD en allant récupérer les informations du script SQL présent dans PostgreSQL. On peut retrouver en annexe 9 de ce rapport le MCD créé (la méthode détaillée pour l'obtenir est disponible sur la plate-forme GitHub³¹). Le MCD de Geonetwork comprend 51 tables, dont 16 tables isolées, c'est-à-dire qui n'ont pas de relations directes avec une autre table, et 35 tables reliées, qui ont au minimum une connexion avec une autre table.

Il a été intéressant de manipuler le catalogue Geonetwork et d'observer comment répondait la base de données. Ces manipulations nous ont permis de voir par exemple dans quelle table se stockaient les informations que nous avons saisies dans Geonetwork. Ainsi, une seconde documentation, dont un extrait est présenté en annexe 10 de ce rapport, a été réalisée dans le cadre de ce stage. Elle est consacrée essentiellement aux MCD de Geonetwork (rôle des tables, clés primaires et attributs). Cette documentation n'est pas complète, mais mérite d'être poursuivie notamment concernant le rôle des tables qui n'est pas encore défini pour toutes. Un des constats les plus importants suite à ces manipulations a été d'observer la table qui stocke la fiche de métadonnées présente au sein du catalogue Geonetwork. Si on observe la Figure 10 ci-dessous, on remarque que cette table vient héberger dans un unique champ l'ensemble du bloc XML de la métadonnée. Cette organisation semble rendre plus complexe de potentiels liens dynamiques entre la métadonnée et sa donnée. Nous reviendrons sur ce point dans la partie 3 du rapport.



id [PK] integer	data text	changedate character varying(30)	createdate character varying(30)	displayorder integer	doc cha
38502	<pre><gmd:CI_Responsibility xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:gts="http://www.isotc211.org/2005/gts"> <gmd:individualName> <gco:CharacterString>LAMBERT Pascal</gco:CharacterString> </gmd:individualName> <gmd:organisationName> <gco:CharacterString>PNR Oise Pays de France</gco:CharacterString> </gmd:organisationName> <gmd:positionName> <gco:CharacterString>Chargé de mission SIG et évaluation</gco:CharacterString> </gmd:positionName> <gmd:contactInfo> <gmd:CI_Contact> <gmd:phone> <gmd:CI_Telephone> <gmd:voice> <gco:CharacterString>03.44.63.65.62</gco:CharacterString> </gmd:voice> <gmd:facsimile> <gco:CharacterString>03.44.63.65.60</gco:CharacterString> </gmd:facsimile> </gmd:CI_Telephone> </gmd:phone> <gmd:address> <gmd:CI_Address> <gmd:deliveryPoint> </pre>	2017-08-01T09:59:41	2017-07-12T17:28:16	0	
39285	<pre><gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:xsi="http://www.w3.org/2001/XMLSchema" 2017-07-31T10:56:00 2017-07-31T10:56:00 0</pre>				
39286	<pre><gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:xsi="http://www.w3.org/2001/XMLSchema" 2017-07-31T10:56:04 2017-07-31T10:56:04 0</pre>				
39287	<pre><gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:gco="http://www.isotc211.org/2005/gco" 2017-07-31T10:56:05 2017-07-31T10:56:05 0</pre>				

Figure 10 : Capture de la table "metadata", issue de la base de données Geonetwork

³¹ <https://github.com/GeoNetworkTeam> (Documentation sur le MCD de Geonetwork)

Partie 3 : Paramétrages, alimentation et perspectives de l'outil

A) Configuration de l'outil

Suite aux nombreuses évolutions de Geonetwork, nous allons voir qu'il est aujourd'hui possible de dépasser la simple compétence de catalogage pour se rapprocher au plus près d'une Infrastructure de Données Spatiales. Seront présentées au sein de ce rapport quelques fonctionnalités, pour avoir un aperçu plus global, vous pouvez vous référer à une troisième documentation³².

1. Lier des ressources à la métadonnée

Dans un premier temps, il est possible de joindre à la métadonnée toutes sortes de documents visibles directement dans la fiche de métadonnées pour une image par exemple ou bien un fichier téléchargeable via un lien. Au moment de recueillir les attentes de l'outil auprès des chargés de mission SIG des 4 PNR francilien, nous avons fait le constat que certaines informations intimement liées à la donnée ne pourraient figurer dans la métadonnée car ils ne sont pas conformes INSPIRE. Pour le PNR de la Haute Vallée de Chevreuse, il s'agit par exemple d'un fichier PDF comprenant une description approfondie des données. Geonetwork rend possible l'ajout de ces informations à la métadonnée.

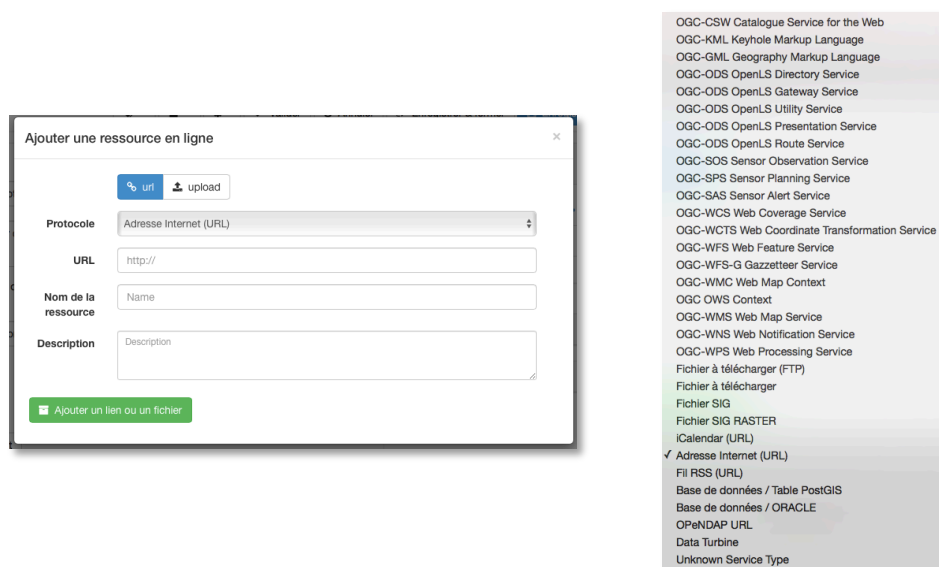
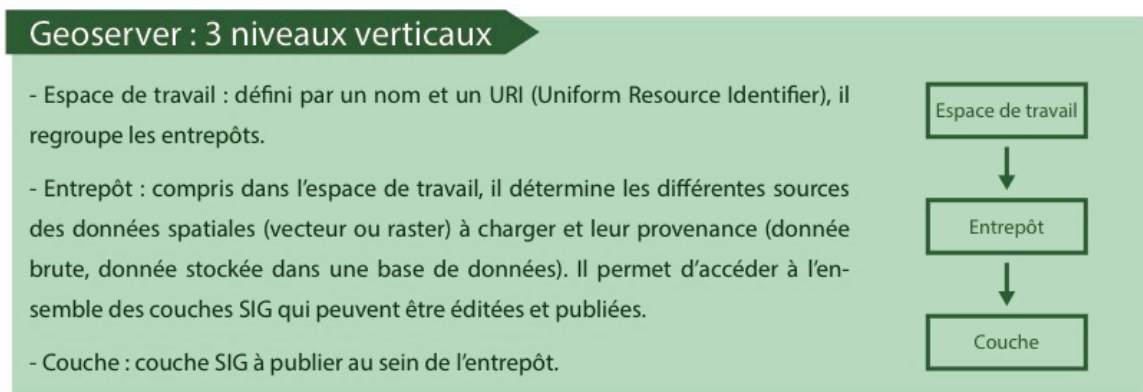


Figure 11 : Interface Geonetwork pour l'ajout d'une ressource dans la fiche de métadonnées

³² <https://github.com/GeoNetworkTeam> (Documentation sur la configuration de Geonetwork)

2. Diffuser des données via Geoserver

Geoserver est un Webservice open source permettant de diffuser des données spatiales selon des protocoles normalisés de l'OGC : WMS, WFS. Geoserver est une application web, au même titre que Geonetwork, elle fonctionne par le biais d'un serveur d'application java, en l'occurrence Tomcat pour les PNR franciliens. Geoserver est déjà installé au sein du serveur interParc (cf. Figure 3 page 17).



Pour bien comprendre, nous savons dans un premier temps que Geoserver va permettre de créer des flux de données à partir de données initialement stockées dans une base Postgres. Un premier lien est donc établi entre PostgreSQL et Geoserver. Dans un deuxième temps, il existe des fiches de métadonnées, visibles et éditables à travers l'interface Web Geonetwork, qui sont stockées dans une base Postgres également. En parvenant à établir une connexion entre Geoserver et Geonetwork, on permet à une donnée de Geoserver d'être reliée à sa métadonnée correspondante dans Geonetwork, voir la Figure 12 ci-dessous :

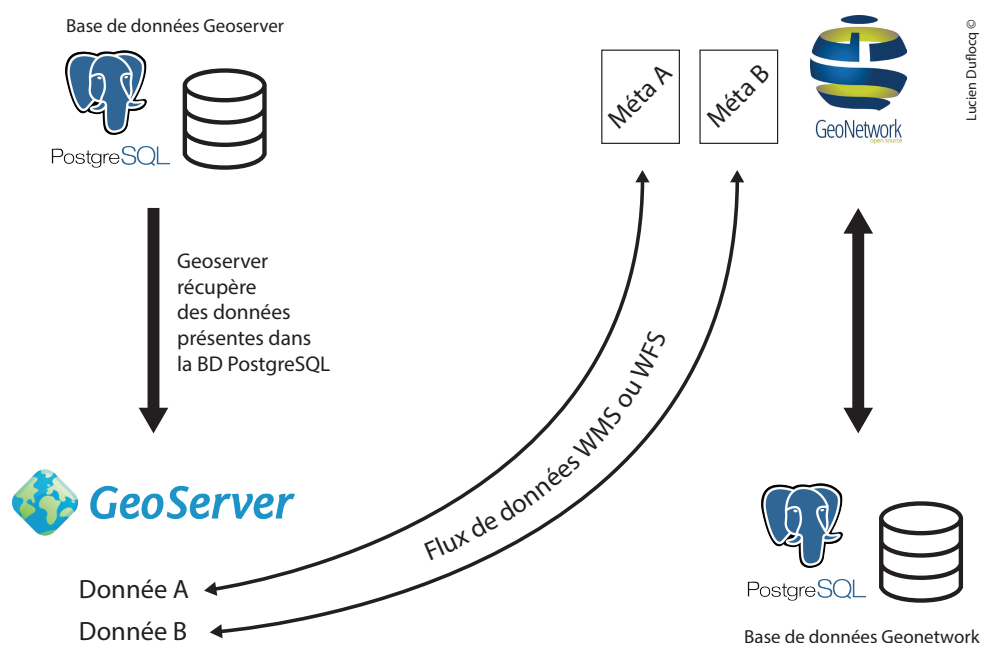


Figure 12 : Connexion entre Geonetwork et Geoserver

Concernant la configuration du Geoserver, je n'ai pas eu à créer d'Espaces de travail car ils avaient déjà été mis en place avant mon arrivée (un Espace de travail par Parc et un espace de travail pour le réseau des Parcs). Au sein de l'Espace de travail du réseau des Parcs, il a par contre fallu ajouter un nouvel Entrepôt, relié à la base de données PostgreSQL du réseau des Parcs. Ensuite, nous avons pu aller récupérer des données dans la base PostgreSQL, et constituer des couches. Puis il a été nécessaire de renseigner d'une part les URL du Geoserver dans le catalogue Geonetwork (cf. Figure 13) et d'autre part les URL de Geonetwork au sein du Geoserver (cf. Figure 14). Une fois ces URL renseignés, les flux WMS et WTS sont opérationnels.

Figure 13 : Renseignement des URL pour connecter un serveur cartographique (Geonetwork)

Type	Format	URL
TC211	text/html	http://geonetwork.pnr-idf.fr/geonetwork/srv/fre/catali Retirer
ISO19115:2003	text/xml	http://geonetwork.pnr-idf.fr/geonetwork/srv/api/recoi Retirer

Ajouter un lien Notez que seules les métadonnées rédigées selon les standards du FGDC et de l'ISO/TC211 peuvent être référencées dans les capacités de WMS 1.1.1

Figure 14 : Renseignement des URL pour connecter un serveur cartographique (Geoserver)

Toutefois, alors que cette connexion fonctionnait au sein de ma VM test, nous avons eu des difficultés pour réitérer le processus sur le serveur interParc, comme si le catalogue n'arrivait pas à accéder aux URL du Geoserver. Après s'y être penché avec le responsable SIG du PNR du Vexin français, nous en avons déduit que le problème pourrait venir du pare-feu du serveur interParc. Nous avons évoqué, au moment de l'installation, la nécessité de configurer le pare-feu de manière à autoriser la VM de Geonetwork à établir des connexions avec la VM de la base de données PostgreSQL. De la même manière, il semblerait que le pare-feu bloque toutes connexions entre Geonetwork et Geoserver. Nous avons vérifié que le problème ne venait pas du Geoserver en testant avec succès une connexion par flux WFS au sein de QGIS. Lors de la rédaction de ce rapport, nous étions alors en contact avec la société qui administrait le serveur interParc (Tranquil-IT-System) pour avoir leur expertise.

3. Administrer les droits d'accès

La gestion des utilisateurs, groupes d'utilisateurs et droits d'accès est une composante importante du catalogue. Nous avons créé dans Geonetwork quatre administrateurs (ou utilisateurs), un pour chaque chargé de mission SIG et cinq groupes d'utilisateurs, un pour chaque PNR et un pour le réseau des Parcs. Chaque chargé de mission est responsable du groupe de son Parc, plus du groupe de l'interParc. Les administrateurs du catalogue auront la possibilité par la suite de créer d'autres utilisateurs. Je pense notamment à la création de compte pour des chargés de mission de leur Parc qui seraient amenés à éditer des métadonnées. Les créations et attributions de rôles sont intéressantes pour plusieurs raisons :

- chaque chargé de mission possède ses propres identifiants de connexions
- chaque publication de métadonnées sera attribuée à l'utilisateur qui a réalisé la manipulation (possibilité de connaître l'auteur et aussi le PNR à l'origine de la publication)
- chaque administrateur aura la possibilité de gérer les droits d'accès par groupe d'utilisateurs. Les différents privilèges autorisés aux groupes sont présentés sur la Figure 15, ci-dessous :

Groupes	Publier	Carte Interactive	Télécharger	Editer	Notifier
pnrgf	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
pnrtvc	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
pnrtcf	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
pnropf	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
pnrvf	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 15 : Gestion des droits d'accès dans Geonetwork

B) Alimentation du catalogue

1. Quelles données seront à traiter ?

À la question « Quels métadonnées seront à faire figurer dans le catalogue ? », les chargés de mission ont évoqué dans un premier temps les données de références présentes dans la base de données interParc (base « pnridf »), c'est à dire des données dont on est sûr de la véracité. On va retrouver par exemple les données cadastrales ou GEOFLA³³. Dans un deuxième temps, l'ensemble des données de la base de données « pnridf » sera à intégrer (cf. Figure 16). Le Parc du Gâtinais français a proposé d'intégrer la base de données des chemins ruraux qui s'est développé cette année en partenariat avec la DRIEE³⁴. Chaque Parc sera évidemment libre d'y ajouter ses propres données.



Figure 16 : Contenu de la base de données "pnridf"

Parallèlement, un travail d'inventaire des données pourrait être utile pour le bon fonctionnement des bases de données, mais aussi pour alimenter proprement le catalogue. Néanmoins, c'est un travail de documentation conséquent qui n'a pas pu être mis en place dans le cadre de mon stage. Il pourrait être envisagé d'y consacrer une personne temporairement pour réaliser ce travail.

³³ Description de l'ensemble des unités administratives nationales (France métropolitaine et DROM)

³⁴ Direction régionale et interdépartementale de l'environnement et de l'Énergie

2. Procédé d'intégration des métadonnées

Au cours d'une réunion le 11 juillet 2017 au PNR du Vexin français, j'ai présenté le maximum de fonctionnalités³⁵ de Geonetwork aux chargés de mission SIG puis nous avons abordé la question de l'intégration des métadonnées au sein du catalogue. Le schéma ci-dessous, Figure 17, montre les trois grands moyens d'alimenter le catalogue Geonetwork :

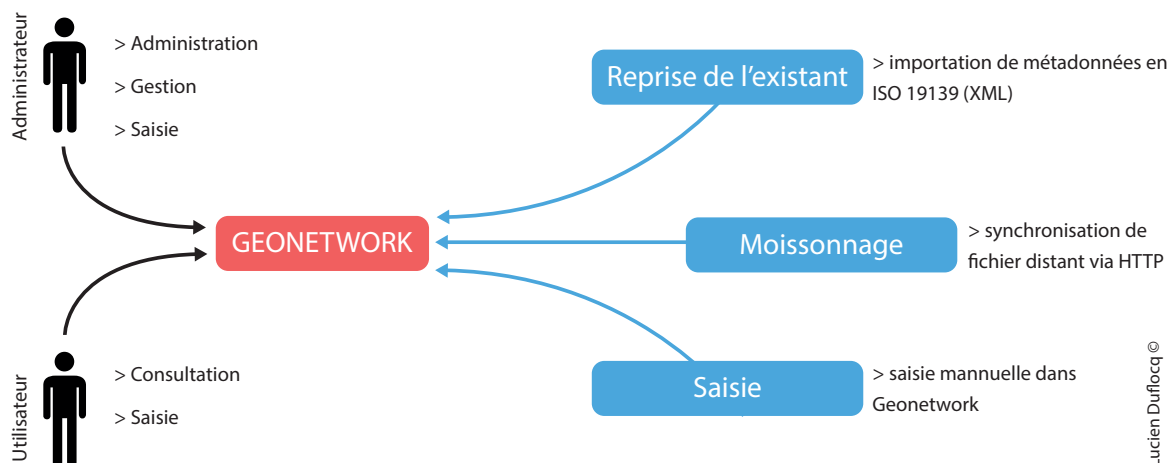


Figure 17 : Trois moyens d'alimenter le catalogue de métadonnées

Premièrement, il est possible d'effectuer une reprise de l'existant, c'est-à-dire de récupérer les métadonnées produites antérieurement et de les importer directement dans le catalogue.

Deuxièmement, il est possible de moissonner d'autres catalogues pour récupérer des métadonnées distantes.

Moissonnage

« Procédé utilisé pour permettre à des catalogues de données de s'échanger leurs contenus. Ce n'est alors pas ce contenu lui-même qui est moissonné mais sa description dans le catalogue : on parle alors de moissonnage de métadonnées. Ainsi un catalogue B moissonné par un catalogue A permet d'afficher son contenu dans ce catalogue A. »

GéOInformations, Espace Interministériel de l'information géographique, Glossaire.

Troisièmement, il est possible de saisir manuellement les fiches de métadonnées grâce à l'interface Geonetwork.

Concernant le premier moyen, il faudrait d'une part que les métadonnées soient dans le format d'interopérabilité des métadonnées qui est le XML et d'autre part, il faut être sûr d'avoir des métadonnées à jour et conformes INSPIRE. Précisons alors que peu de métadonnées déjà existantes remplissent ces critères. Une personne pourrait donc être en

³⁵ Se référer à la documentation : <https://github.com/GeoNetworkTeam>

charge de la future intégration des métadonnées au sein du catalogue. Quelques pistes pour faciliter ce travail sont possibles. Tout d'abord, l'outil payant ISO GEO³⁶ offre la possibilité d'effectuer un inventaire des données, à travers un scan, de manière à sélectionner plus judicieusement les données à intégrer au catalogue. Mais il existe aussi un inventaire automatique gratuit avec Talend Spatial³⁷, développé par les Parcs nationaux et la société Titellus. Cet outil permet la génération de métadonnées respectant les normes, grâce à un scan agissant directement dans la base de données. Cette technologie permettrait d'obtenir des métadonnées homogènes pour chacun des PNR franciliens. Pour le deuxième moyen, il n'a pas semblé pertinent d'être moissonné, néanmoins, il pourra être utile de moissonner quelques grands catalogues comme le Géoportail³⁸. Le troisième moyen est celui qui sera le plus employé. Nous nous sommes donc concertés avec les chargés de mission SIG sur la manière dont vont être saisies les métadonnées dans le catalogue. Trois objectifs ont été fixés :

- Saisir des fiches conformes aux ISO 19115 – 19139 afin de standardiser et d'assurer l'interopérabilité des métadonnées
- Saisir des fiches conformes à INSPIRE
- Homogénéiser les fiches produites au sein du réseau des PNR franciliens

Pour répondre aux objectifs, je me suis intéressé à l'IDG GéoBretagne³⁹, qui a mis en place un Wiki⁴⁰ dans lequel est décrit comment gérer l'intégration des métadonnées au sein de leur catalogue. Suite à cela, j'ai proposé au réseau des PNR franciliens deux actions :

- Préparer une fiche de métadonnées modèle pour le réseau des Parcs
- Préparer un guide de recommandations pour la saisie des fiches de métadonnées

La préparation du modèle de métadonnée a été la priorité. Pour la création de ce modèle type, nous sommes partis d'un modèle conforme à l'ISO 19139 proposé par défaut au sein de Geonetwork. Puis, de manière concertée, nous avons repris un à un les différentes informations à remplir dans le modèle et nous avons noté celles que nous voulions modifier, supprimer ou ajouter. Après avoir défini ces ajustements, le modèle a été exporté en XML depuis Geonetwork puis modifier dans un logiciel d'éditeur de texte⁴¹. Enfin, le modèle a été enregistré et importer dans Geonetwork pour être disponible au moment de la saisie de fiche de métadonnées. Afin d'assurer la plus grande homogénéité entre les fiches, il est recommandé aux éditeurs d'utiliser prioritairement ce modèle. Chaque PNR reste libre de produire de nouveaux modèles.

³⁶ HALBOUT Hervé, *Catalogage des données géographiques : Isogeo ou Géosource ?* Géomatique Expert, N°86, Mai-Juin 2012

³⁷ <https://docs.google.com/document/d/10GyQLL9XORawurllI6AUCpP7PDfcQ3D3nkBO1Le3UI8/edit> (Présentation de l'outil Talend Spatial)

³⁸ Portail national de la connaissance du territoire mis en œuvre par l'IGN, il a pour vocation de faciliter l'accès à l'information géographique de référence.

³⁹ <https://cms.geobretagne.fr> (Site internet de l'IDG GéoBretagne)

⁴⁰ <http://kartenn.region-bretagne.fr/wiki/doku.php?id=accueil> (Wiki du GéoBretagne)

⁴¹ Se référer à l'annexe 11

Concernant le guide de recommandations, c'est un travail qui n'a pas pu être mis en place au cours de mon stage par manque de temps. Néanmoins, cela reste une perspective intéressante pour Geonetwork, car cela permettra de garantir une aide et une référence pour la saisie des fiches, utiles pour les prochains chargés de mission SIG, notamment ceux qui hériteront du catalogue. De la même manière que pour GéoBretagne, ce guide pourrait être mis en place à travers un Wiki, dans lequel des conseils seront proposés afin d'aider l'éditeur de métadonnées (explications des champs à remplir, montrer des exemples, montrer ce qu'il ne faut pas faire, etc.).

C) Perspectives de l'outil

1. Lien entre la métadonnée et la donnée

Lors de la mise en place des attentes de l'outil de catalogage (cf. Figure 7 page 23), la capacité à pouvoir relier dynamiquement la donnée à la métadonnée avait été un point important. Les Parcs souhaitaient que suite à l'ajout ou à la mise à jour d'une donnée au sein de la base PostgreSQL, la métadonnée liée à cette donnée soit créée ou mise à jour également, et ce de manière automatique. Ce processus relève alors d'une connexion directe entre la base de données PostgreSQL des données interParc et la base de données PostgreSQL créée par Geonetwork. Nous savons que ce processus est rendu possible via l'objet DBLink, qui permet d'exécuter des requêtes sur une autre base de données, qu'elle soit physiquement présente sur la machine où qu'elle soit distante.

Suite à l'analyse des différentes solutions de catalogage, nous nous étions aperçus qu'aucune solution ne proposait ce lien dynamique. Nous avons évoqué précédemment que certaines solutions offraient des configurations rapprochant au plus près les données à la métadonnée. On peut citer ISOGEO et son inventaire automatisé de la base de données ou geOrchestra et son lien dynamique entre la métadonnée et la visualisation de la donnée. Mais aucune solution clé en main ne proposait de lien d'automatisation entre la base de données et la base Geonetwork. Nous avons donc décidé de créer ce lien nous-même. Une première hypothèse consistait à établir des triggers entre les deux bases (base de données des données interParc et base de données Geonetwork) de manière à ce que lorsque l'on modifie une information sur une base, cela déclenche une modification automatique au sein de la seconde base.

Trigger

« Un trigger, ou déclencheur, est une spécification précisant que la base de données doit exécuter une fonction particulière quand un certain type d'opération est traité. Les fonctions déclencheurs peuvent être définies pour s'exécuter avant ou après une commande INSERT, UPDATE ou DELETE, soit une fois par ligne modifiée soit une fois par expression SQL. Si un élément déclencheur se produit, le gestionnaire de déclencheurs est appelé au bon moment pour gérer l'événement. »

Documentation PostgreSQL 8.1.23

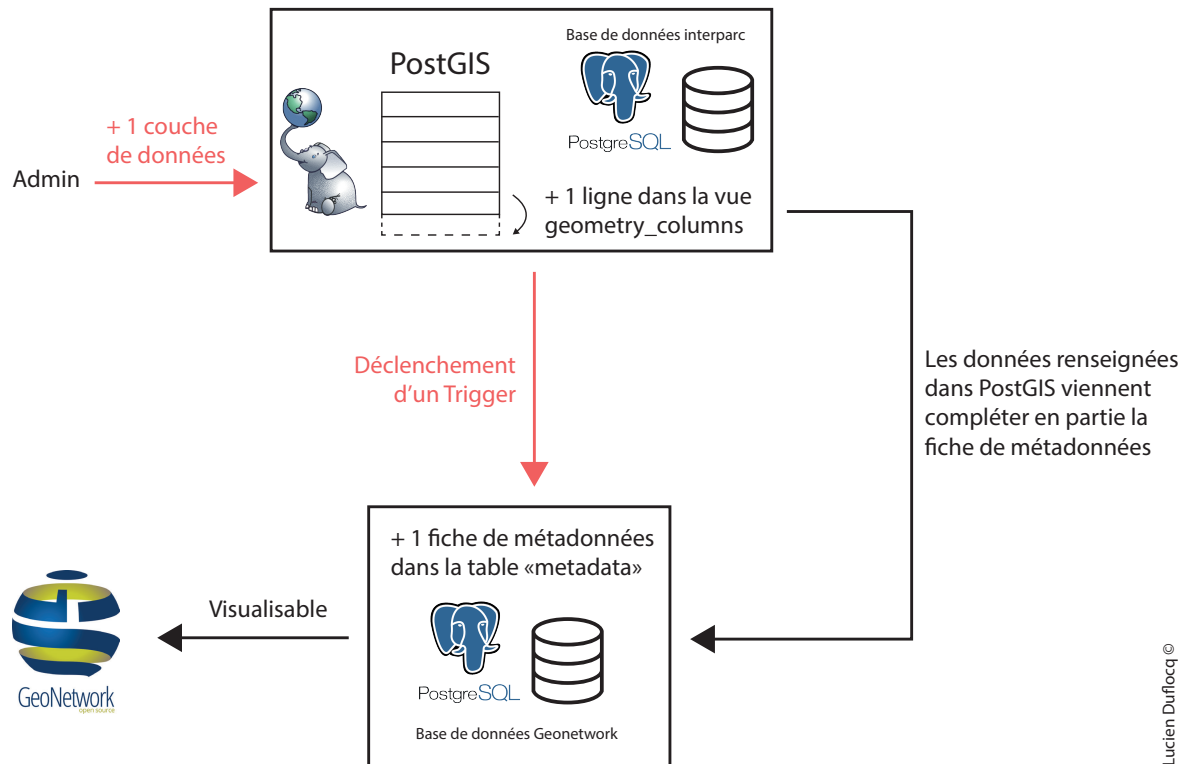


Figure 18 : Création d'une métadonnée via un Trigger

Lors de la rédaction de ce rapport, Eric Marot, chargé de mission SIG au PNR du Vexin français, est parvenu à mettre en place un processus permettant d'ajouter une fiche de métadonnées dans la base de données Geonetwork à partir de l'ajout d'une donnée dans la base de données interParc via PostGIS (cf. Figure 18).

Concernant les mises à jour de métadonnées en cas de modification de la donnée dans la base, cela s'avère plus complexe. En effet, nous avons vu avec la Figure 10 page 37, que la base de données Geonetwork stocke la métadonnée dans une table en un seul bloc XML. Il est donc plus compliqué de cibler et de modifier une information particulière dans ce grand bloc XML. Ce travail reste une des principales perspectives d'évolution pour le catalogue.

2. Optimisation de l'architecture du serveur interParc

« Vers quelle architecture pour le serveur interParc ? », c'est la question qui a été posée lors de la réunion du 11 juillet 2017. Plusieurs éléments sont à prendre en compte pour répondre à la question. Premièrement, Geonetwork est relié à deux instances qui sont elles-mêmes dans un serveur distant : Geoserver et la base de données PostgreSQL (cf. Figure 9 page 35). Deuxièmement, l'organisation actuelle avec la base de données PostgreSQL, le portail web et Geoserver dans le même serveur convenait jusque-là car Geoserver était finalement peu mobilisé. Avec la mise en place du catalogue Geonetwork et des flux de données WMS/WFS, Geoserver va être beaucoup plus utilisé. En prenant en compte le fait que Geoserver est une application sensible et consommatrice en ressource, il a été envisagé de créer un nouveau serveur (le septième en tout) destiné à ne faire tourner que Geoserver (cf. Figure 19). Une autre modification consisterait à créer deux instances Tomcat, une pour Geonetwork et une pour le Geoserver. Cette configuration aurait l'avantage d'augmenter la stabilité de l'application, en favorisant notamment le débogage en cas de problème lié au serveur, puisque l'on isole les applications et répartit la charge. En effet, si par exemple Geoserver perturbe la machine virtuelle, il ne gênera pas le fonctionnement des autres applications.

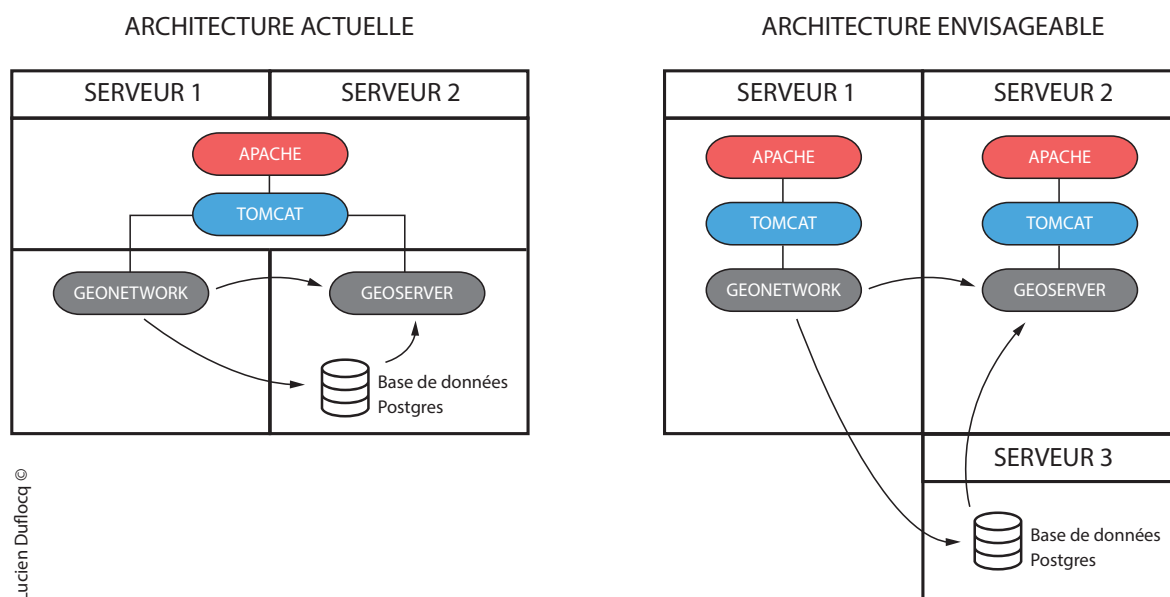


Figure 19 : Perspectives d'amélioration du serveur interParc

Bilan personnel

Bien que mon stage ne soit pas terminé au moment de la rédaction de ce rapport, je vais tâcher de présenter un bilan personnel de cette expérience. Je vais commencer par les compétences acquises durant le stage. Il est certain que les thématiques des métadonnées et du catalogage ont été pleinement abordées. Approchées durant la formation du Master SIGMA, ce sont deux sujets qu'il m'a paru intéressant d'approfondir et de mettre en pratique au cours d'un stage. Cela a été un bon moyen de me pencher sur les aspects théoriques de la législation, des normes et des standards en vigueur. Ce sont des aspects souvent peu attrayants et difficilement abordables, mais le cadre du stage permet de prendre du temps pour s'y pencher. L'approche du catalogage et sa mise en place au sein d'un service SIG est aujourd'hui quelque chose que j'aborde sereinement, tant du point de vue méthodologique que conceptuel.

Techniquement, je maîtrise l'installation et l'utilisation du catalogue Geonetwork. Malgré les difficultés rencontrées lors de l'installation de geOrchestra, je connais les particularités de l'outil ainsi que ses avantages et ses inconvénients. Les installations de Geonetwork et de geOrchestra m'ont permis de développer mes compétences en commande Linux, que ce soit pour l'installation ou pour la gestion des serveurs. Je dois dire que l'installation de geOrchestra m'a posé de réelles difficultés. Sans maintenance informatique, il a été compliqué d'assimiler les prérequis nécessaire en langage Java et en administration serveur. Néanmoins, cette épineuse phase d'installation de geOrchestra s'est révélée salvatrice au moment de l'installation de Geonetwork qui fut abordée de manière plus méthodique et efficace. Le changement de solution de catalogage en cours de stage a été délicat à gérer, car il a fallu faire le choix de stopper ce qui avait été mis en place afin de ne pas perdre trop de temps sur l'installation. Toutefois, cela a été pour moi un moyen de saisir l'importance de la gestion de projet. Car au-delà du fait de ne pas avoir pu installer la solution choisie à l'origine, le principal fut de savoir réagir rapidement.

J'ai pu approfondir mes connaissances avec le serveur cartographique Geoserver, plus particulièrement sur la mise en place de Web Services WMS/WFS au sein du catalogue Geonetwork. J'ai également manipulé le format XML dans le cadre de la création d'un modèle type de métadonnées en ISO 19139.

Enfin, l'installation de Geonetwork a permis d'élargir mes compétences en gestion des serveurs, que ce soit les serveurs Web (Apache) ou les serveurs d'applications (Tomcat).

Je tenais à souligner l'importance des échanges et de la mutualisation pour ce type de projet. En effet, la mise en réseau ne révèle pas uniquement son potentiel à travers les outils et logiciels, mais aussi grâce à la mise en réseau des acteurs. Les circonstances ont fait qu'Alexandra Greco, stagiaire au sein du PNR du Verdon et collègue de Master, travaillait comme moi sur l'installation de Geonetwork. La mise en relation de nos travaux a permis l'élaboration de trois documentations actuellement disponibles sur la plateforme GitHub⁴² (« Installing Geonetwork on Debian Machine », « Conceptual Data Model of Geonetwork »),

⁴² <https://github.com/GeoNetworkTeam>

« Configuring the catalog Geonetwork »). Des échanges réguliers ont permis de partager nos difficultés, de confronter nos résultats et de trouver les solutions les plus adéquates. J'ai également apprécié de pouvoir laisser une trace écrite de mon travail, accessible en open source, elle est à destination du réseau des PNR franciliens mais aussi des futurs utilisateurs de Geonetwork, ou même de ceux qui voudraient simplement tester la solution sur une machine virtuelle. Aussi, la quasi-totalité des documents produits dans le cadre du stage ont été stockés sur un Google Drive de l'interParc (notes, présentation des réunions, MCD, etc.).

Conclusion

À la fin de mon stage, le réseau des Parcs naturels régionaux d'Île-de-France dispose d'un outil opérationnel pour le catalogage de ses métadonnées. Basée sur Geonetwork, la plateforme permet d'accéder à un outil de catalogage des données, un serveur cartographique et un menu d'administration. Un modèle de métadonnées, répondant à la directive INSPIRE et conforme à la norme ISO 19139, a été élaboré afin d'uniformiser les métadonnées et d'en faciliter la saisie par les utilisateurs. Enfin, trois documentations ont été rédigées à propos de l'installation de Geonetwork, l'analyse de son MCD et les paramétrages de l'outil. Une partie des documentations est présentée en annexe de ce rapport et la totalité est disponible sur la plateforme GitHub⁴³.

La mise en place du catalogue offre de belles perspectives d'évolution. Premièrement, les PNR franciliens vont pouvoir amorcer l'alimentation du catalogue, c'est un travail conséquent qui risque de prendre du temps car il va s'accompagner d'un inventaire des données. Il s'agira alors d'analyser les données disponibles afin de distinguer les données à cataloguer des données qui ne le sont pas, ce sera également l'occasion de supprimer les éventuels doublons ou les données qui n'auraient plus lieu d'être dans la base de données des PNR. L'organisation d'un futur stage pour la réalisation de cette tâche est envisageable. Deuxièmement, les PNR peuvent poursuivre leurs recherches concernant une potentielle modification automatique des métadonnées en cas de modifications des données dans la base. Un processus qui permettrait au chargé de mission SIG de ne pas avoir à éditer une fiche de métadonnées dans le catalogue Geonetwork suite à une information modifiée dans le SIG (couches PostGIS par exemple). Troisièmement, la réflexion autour de l'optimisation de l'architecture du serveur interParc dans lequel se trouve l'outil de catalogage Geonetwork mérite d'être poursuivie. On évoque par exemple la possibilité de créer une machine virtuelle spécialement destinée à accueillir le serveur cartographique Geoserver, jusque-là dans la même machine virtuelle que la base de données PostgreSQL. L'objectif étant d'améliorer les performances et la stabilité du serveur interParc.

Toutes les figures et illustrations présentes dans ce rapport, et dont la source n'est pas mentionnée, ont été réalisées par Lucien Duflocq.

Lucien DUFLOCQ, 2017.

⁴³ <https://github.com/GeoNetworkTeam>

Liste des illustrations

Figure 1 : Organisation d'un Parc naturel régional.....	11
Figure 2 : Carte de localisation des PNR d'Île-de-France.....	13
Figure 3 : Organisation du serveur interParc	18
Figure 4 : Page d'accueil du Wiki InterParc	19
Figure 5 : Page d'accueil du portail web interParc	20
Figure 6 : Page d'accueil du WebSIG DynMAP	20
Figure 7 : Attentes de l'outil de gestion des métadonnées.....	24
Figure 8 : Comparaison entre une machine virtuelle et un conteneur Docker.....	33
Figure 9 : Organisation du serveur avec Geonetwork.....	36
Figure 10 : Capture de la table "metadata", issue de la base de données Geonetwork.....	38
Figure 11 : Interface Geonetwork pour l'ajout d'une ressource dans la fiche de métadonnées	39
Figure 12 : Connexion entre Geonetwork et Geoserver	40
Figure 13 : Renseignement des URL pour connecter un serveur cartographique (Geonetwork)	41
Figure 14 : Renseignement des URL pour connecter un serveur cartographique (Geoserver)	41
Figure 15 : Gestion des droits d'accès dans Geonetwork	42
Figure 16 : Contenu de la base de données "pnridf"	43
Figure 17 : Trois moyens d'alimenter le catalogue de métadonnées	44
Figure 18 : Création d'une métadonnée via un Trigger	47
Figure 19 : Perspectives d'amélioration du serveur interParc	48

Liste des tableaux

Tableau 1 : Outils utilisés au sein des PNR franciliens	16
Tableau 2 : Répartition des solutions de catalogage au sein des IDG françaises.....	26
Tableau 3 : Répartition des outils de gestion des métadonnées au sein des PNR français	27
Tableau 4 : Analyse de quatre solutions de catalogage (Geonetwork, Géosource, geOrchestra et Prodigie)	28

Bibliographie

Ouvrages

BARON Nacima & LAJARGE Romain, *Les Parcs naturels régionaux, des territoires en expériences*, Versailles, Éditions Quac, 2015, 247p.

Etudes, articles scientifiques et thèses

AFIGÉO – EUROGI / eSDI-Net, *Catalogue des IDG 2014, Catalogue des Infrastructures de Données Géographiques françaises*, version actualisée le 27 mai 2014.

BECKER Mathieu et HALBOUT Hervé, *Découvrir et comprendre la gouvernance de l'information géographique*, Livre Blanc, 2016.

DUBOIS Fabien, *Développement d'un Système d'Information InterParc, Mise en place de l'architecture informatique et des deux premières applications métier : L'inventaire du patrimoine bâti et Geotrek*, Rapport de stage, 2016.

HALBOUT Hervé, *Catalogage des données géographiques : Isogeo ou Géosource ?* Géomatique Expert, n°86, Mai-Juin 2012.

LE LOEUFF Emilie, *Structuration des données environnementales du Parc naturel régional du Gâtinais français*, Rapport de stage, 2012.

MERRIEN Francis & LEOBET Marc, *La directive Inspire pour les néophytes*, 3^{ème} édition, Mission de l'information géographique (MIG), décembre 2011.

RADISSON Laurent, « *Gérer un Parc naturel régional, c'est accepter une concertation permanente* », dans Environnement et technique, n°368, avril 2017

Sources en ligne

Site internet officiel de la fédération des Parcs naturels régionaux d'Île-de-France :
<http://www.Parcs-naturels-regionaux.fr>

Site internet de la Région Île-de-France, carte interactive des Parcs naturels régionaux :
<https://www.iledefrance.fr/fil-actus-region/carte-Parcs-naturels-ile-france>

Site internet officiel du catalogue Géosource :
<http://www.geosource.fr>

Site internet officiel du catalogue Geonetwork :
<http://geonetwork-opensource.org>

Site internet de la documentation officiel de Geonetwork :

<http://geonetwork-opensource.org/manuals/trunk/eng/users/index.html>

Site internet officiel de l'Infrastructure de Données Spatiales geOrchestra :

<http://www.georchestra.org/fr/>

Groupe de discussions Google à propos de l'Infrastructure de Données Spatiales geOrchestra :

<https://groups.google.com/forum/#!forum/georchestra>

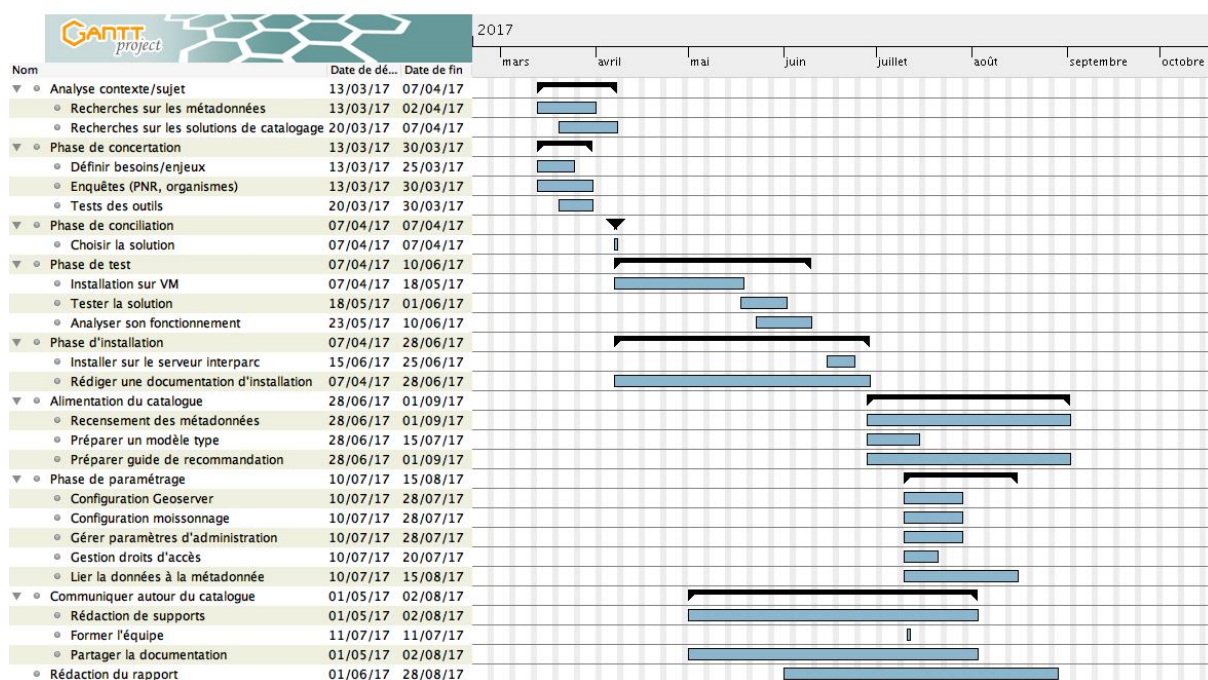
Annexes

Annexe 1 : Organigramme du Parc naturel régional du Gâtinais français	56
Annexe 2 : Diagramme de Gantt et dates clés	56
Annexe 3 : Extrait d'un formulaire ayant servi à recueillir les besoins et les attentes de l'outil de catalogage auprès des chargés de mission SIG	58
Annexe 4 : Extrait de la documentation d'installation créée pour geOrchestra.....	60
Annexe 5 : Extrait de la documentation pour l'installation de Geonetwork.....	63
Annexe 6 : Extrait de la documentation pour connecter Geonetwork à PostgreSQL	66
Annexe 7 : Extrait de la documentation pour changer le schéma de la base de données Geonetwork.....	71
Annexe 8 : Configuration du pare-feu pour le serveur interParc.....	74
Annexe 9 : Modèle Conceptuel de Données de Geonetwork (tables liées et tables isolées) .	75
Annexe 10 : Extrait de la documentation sur le MCD de Geonetwork	76
Annexe 11 : Modèle type pour la fiche de métadonnées (format XML).....	77

Annexe 1 : Organigramme du Parc naturel régional du Gâtinais français



Annexe 2 : Diagramme de Gantt et dates clés



Dates importantes

- 16/03/17 : Réunion interParc (en présence des 4 chargés de mission SIG des PNR).
- 21/03/17 : Déplacement au PNR d'Oise – Pays de France.
- 23/03/17 : Déplacement au PNR du Vexin français.
- 24/03/17 : Demande aux PNR français de mettre à jour leurs informations concernant l'outil de gestion des métadonnées.
- 30/03/17 : Déplacement au PNR de la Haute Vallée de Chevreuse.
- 05/04/17 : Réunion interParc au PNR du Vexin pour déterminer la solution de catalogage.
- 12/05/17 : Documentation d'installation pour GeOrchestra terminée.
- 17/05/17 : Bilan sur l'installation de GeOrchestra et changement de solution pour Geonetwork.
- 18/05/17 : Demande de devis auprès de la société CamptoCamp.
- 22/05/17 : Installation de Geonetwork en local avec une configuration par défaut.
- 02/06/17 : Paramétrage de Geonetwork avec la SGBD Postgres.
- 12/06/17 : Déplacement de la base de données Geonetwork dans un nouveau schéma.
- 14/06/17 : Documentation d'installation pour Geonetwork terminée.
- 26/06/17 : Installation de Geonetwork sur le serveur interParc.
- 11/07/17 : Réunion interParc au PNR du Vexin pour faire une formation de Geonetwork et se concerter sur la gestion et l'intégration des métadonnées.
- 02/08/17 : Mise en ligne de toutes les documentations sur GitHub et partage sur le forum Georezo.

Annexe 3 : Extrait d'un formulaire ayant servi à recueillir les besoins et les attentes de l'outil de catalogage auprès des chargés de mission SIG

Projet Interparc - 2017

PNR du Gâtinais français

PNR d'Oise Pays de France – 21/03/17

Objet de la rencontre : Cette rencontre avec Pascal Lambert, chargé de mission SIG/Evaluation au sein du parc naturel régional Oise – Pays de France, s'inscrit dans le cadre du projet de développement d'un outil de catalogage des métadonnées Interparc. Cet outil a pour enjeu de répondre aux attentes et aux caractéristiques propres à l'ensemble des quatre Parc naturels régionaux. Une des priorités est donc de rencontrer les chargés de missions SIG de chaque parc de manière à recueillir leurs témoignages, en vue d'atteindre un objectif commun. Cette rencontre permettra également de connaître le fonctionnement et les outils SIG mis en place au sein du parc.

Points à aborder :

➤ Les métadonnées :

- Comment sont produites les métadonnées au sein du parc ?

Anciennement, le service SIG du PNR Oise – Pays de France éditait ses fiches de métadonnées via un formulaire Excel. Cette solution a été abandonnée lors du passage d'Excel 97 à 2003 car le formulaire ne fonctionnait plus.

Puis, étant équipé d'une licence ESRI, la solution a été d'utiliser le système ArCatalogue. Cette solution permet assez aisément d'éditer et de relier une fiche de métadonnées à la donnée.

Enfin, le parc utilise également la solution Dynmap pour renseigner des fiches de métadonnées. Toutefois, la saisie de métadonnées ne peut s'opérer essentiellement sur des données stockées sur l'outil Dynmap.

- Comment sont stockées les métadonnées au sein du parc ?

Les métadonnées sont stockées de différentes manières, certaines sont stockées en locales, sur un serveur interne, il s'agit essentiellement des métadonnées conçues avec l'ancien formulaire Excel puis avec ArCatalogue. D'autres métadonnées sont stockées via l'outil Dynmap, et donc présentent physiquement sur le serveur Interparc.

- Comment sont gérées les métadonnées au sein du parc ?

Actuellement, les métadonnées sont donc principalement administrer via ArCatalogue et Dynmap.

- Vers une BD commune/unique PostGis (pgAdmin3) ?

La mise en place d'une base de données commune Interparc via PostGis est souhaitable de manière à faciliter la gestion des données/métadonnées.

- Quels profils de métadonnées sont à faire figurer dans le catalogue ?

Le profil des métadonnées à faire figurer dans le catalogue est un point important mais qui pourra être traité plus tard. L'idée reste néanmoins de pouvoir y intégrer dans un premier temps les métadonnées des données de référence et puis à terme la totalité des données.

- Utiliser le référentiel spatial (ou emprise) correspondant à l'espace sur lequel porte les données en vue de permettre la recherche de métadonnées via son emprise spatial ?

C'est une fonctionnalité proposée quasiment par l'ensemble des solutions de catalogage qui nécessite néanmoins d'avoir renseigné l'emprise de la donnée.

➤ Le catalogue de métadonnées :

- Quelles en seront les utilisateurs ?

Au sein des PNR, en interne, mais également au grand public.

- Les attentes de l'outil ?

-saisie « simple et rapide » d'une métadonnée

-la publication d'un catalogue en ligne (et possibilité d'exporter le catalogue complet en PDF)

-l'envoi d'une métadonnée à un partenaire lors de la mise à disposition d'une donnée (lien vers la fiche de métadonnée en ligne, métadonnée au format XML ou CSV)

-la capacité à ne pas dépendre d'une solution fermée (récupération des métadonnées présente dans l'outil de catalogage)

-la capacité d'être moissonnée par d'autres catalogues (geoportail, datagouv.fr)

-une automatisation des tâches de saisies (liste des champs, emprise géographique, date de dernière modification...)

-la capacité de recherche par thématique, mot clé, emprise spatiale

-la capacité de visualiser la donnée

-la capacité de télécharger la donnée

-le respect des standards (INSPIRE, ISO19110, ISO19115)

-la capacité à distinguer les données « valorisables » à l'extérieur du Parc des données « internes » à ne pas cataloguer

-lien dynamique entre la métadonnée et la donnée : mise à jour de la métadonnée lors d'une modification sur la donnée (catalogue ↔ BD PostGis)

➤ Enquête auprès des PNR Français :

Questionnaire via Google Doc ou demande auprès des PNR de mettre à jour leurs informations concernant le choix de la solution de catalogage.

➤ Dynmap :

Présentation de l'outil Dynmap, développé par la société business Geographic, remplacé par la nouvelle solution Geo. Explication du travail réalisé, dans le cadre d'un stage notamment, pour permettre une bonne gestion des données à travers PostGis et Dynmap.

Annexe 4 : Extrait de la documentation d'installation créée pour geOrchestra

6. Installation de Docker sur la machine virtuelle

Dans un premier temps, il est utile de vérifier la version et le nom de la distribution Debian de manière à suivre les bonnes commandes d'installation. :

```
$ uname -r
```

```
$ lsb_release -cd
```

```
root@debian:/home/lucien# uname -r
3.16.0-4-amd64
root@debian:/home/lucien# lsb_release -cs
jessie
```

Dans notre cas, la version de Debian est la 3.16.0, amd64, Jessie.

Installer Docker engine :

<https://docs.docker.com/engine/installation/linux/debian/>

Installer les paquets suivants :

```
$ sudo apt-get install \  
    apt-transport-https \  
    ca-certificates \  
    curl \  
    gnupg2 \  
    software-properties-common
```

Ajouter la clé officielle GPG de Docker :

```
$ curl -fsSL https://download.docker.com/linux/debian/gpg | sudo apt-key  
add -
```

Vérifier que la clé ID est bien : 9DC8 5822 9FC7 DD38 854A E2D8 8D81 803C CD88

```
$ sudo apt-key fingerprint 0EBFCD88
pub 4096R/0EBFCD88 2017-02-22
    Key fingerprint = 9DC8 5822 9FC7 DD38 854A E2D8 8D81 803C 0EBF
CD88
```

```
uid          Docker Release (CE deb) <docker@docker.com> sub
4096R/F273FCD8 2017-02-22
```

Ajouter un dépôt pour Docker :

```
$ sudo add-apt-repository \  
"deb [arch=amd64] https://download.docker.com/linux/debian \  
jessie \  
stable"
```

Mettre à jour :

```
$ sudo apt-get update
```

Installer Docker :

```
$ sudo apt-get install docker-ce
```

Vérifier que l'installation c'est bien déroulé avec la commande suivante :

```
$ sudo docker run hello-world
```

Installer Docker compose :

<http://blog.programster.org/debian-8-install-docker-compose/>

Installer les paquets suivants :

```
$ sudo apt-get install python-pip -y  
sudo pip install --upgrade pip  
sudo pip install docker-compose
```

Vérifier que l'installation c'est bien déroulé avec la commande suivante :

```
$ docker-compose --version
```

7. Faire fonctionner GeOrchestra avec Docker

<https://github.com/georchestra/georchestra/blob/master/docs/docker.md>

Télécharger docker-compose.yml et docker-compose.override.yml :

```
wget https://raw.githubusercontent.com/georchestra/georchestra/master/docker-  
compose.yml  
wget https://raw.githubusercontent.com/georchestra/georchestra/master/docker-  
compose.override.yml
```

Créer un dossier dans la machine hôte et s'attribuer les droit de superutilisateur pour ce dossier :

```
sudo mkdir /etc/georchestra  
sudo chown unprivileged_user /etc/georchestra
```

Cloner l'image Docker de GeOrchestra dans le dossier préalablement créer :

```
git clone -b docker-master https://github.com/georchestra/datadir.git  
/etc/georchestra
```

Télécharger l'image :

```
sed -i '/build:/d' docker-compose.* && docker-compose pull
```

Créer et lancer les conteneurs:

```
docker-compose up
```

Éditer le fichier « /etc/hosts » :

```
127.0.1.1          georchestra.mydomain.org
```

Dans votre navigateur web, ouvrir le lien suivant :

<https://georchestra.mydomain.org/>

Utiliser les identifiants suivants :

- testuser / testuser
- testadmin / testadmin

2. MISE EN PLACE DE GEONETWORK

2.1. Installation de openjdk-8-jre-headless

Premièrement, installez *openjdk-7* afin d'éviter des problèmes de dépendances liées aux différents paquets :

```
apt install openjdk-7-jre-headless
```

Deuxièmement, récupérez le paquet *ca-certificates-java_20170531_all.deb*, permettant également d'éviter les conflits entre les différents paquets :

```
wget http://ftp.fr.debian.org/debian/pool/main/c/ca-certificates-  
java/ca-certificates-java_20170531_all.deb
```



ATTENTION : le nom du paquet peut changer avec les nouvelles versions et la date peut donc changer. Si cela se produit, vous aurez logiquement un message vous informant que le paquet est introuvable. Il vous faudra alors aller sur le site <https://packages.debian.org> pour vérifier le nom du paquet.

Installez le paquet *ca-certificates-java_20170531_all.deb* qui vient d'être récupéré :

```
dpkg -i ca-certificates-java_20170531_all.deb
```



Il est possible que Debian ne localise pas le paquet venant d'être récupéré. Dans ce cas, *avant de lancer l'installation*, indiquez à la console le dossier dans lequel a été récupéré le paquet *ca-certificates-java_20170531_all.deb*. Généralement, avec la commande *wget*, le paquet est déposé par défaut dans les dossiers personnels, soit :

```
cd /home/ « votre nom d'utilisateur »
```

Installez *openjdk-8-jre-headless*¹ :

```
echo deb http://http.debian.net/debian jessie-backports main >>  
/etc/apt/sources.list  
apt update && apt install openjdk-8-jre-headless
```

Supprimez le paquet *optnjdk-7* qui n'est plus utile :

```
apt purge openjdk-7-jre-headless
```

¹ Sous Debian 9, on peut se passer de l'installation *openjdk-7* et commencer directement par la version *openjdk-8* en faisant : `apt install openjdk-8-jre-headless`.

Configurez Java (généralement, il n'y a aucune configuration) :

```
update-alternatives --config java
```

2.2. Installation de Tomcat8

Il est important de mettre à jour régulièrement la liste des fichiers disponibles dans les dépôts APT présents dans le fichier de configuration `/etc/apt/sources.list`, pour cela, rentrez la commande suivante :

```
apt-get update
```

Installez le serveur Tomcat8 :

```
apt-get -y install \  
    tomcat8 libpostgresql-jdbc-java \  
    libpostgis-java postgresql-client
```

Il est recommandé de fournir de la mémoire à Tomcat ainsi qu'à la machine virtuelle Java pour faire fonctionner GeoNetwork (minimum 2Go). Dans notre cas, nous donnerons 2048 de RAM, soit 2 Go à la JVM. Pour effectuer des tests, cette répartition est suffisante, toutefois, lors d'un passage en production, il faudra attribuer davantage de mémoire.

```
sleep 2  
systemctl stop tomcat8  
sed -i -e 's/^\(JAVA_OPTS=.*)-Xmx\[ ^ \]*\(.*)/\1-Xms2048m -Xmx2048m -  
XX:+UseConcMarkSweepGC\2/' /etc/default/tomcat8
```

2.3. Installation de Geonetwork 3.2.1¹

Attribuez le dossier `/var/lib/tomcat8` à tomcat8 car GeoNetwork va tenter de créer une base de données H2 dans ce dossier :

```
find /var/lib/tomcat8 -ls >/var/lib/tomcat8/files-and-users  
chown -hR tomcat8: /var/lib/tomcat8
```

Téléchargez le `.war` via le lien suivant :

https://sourceforge.net/projects/geonetwork/files/GeoNetwork_opensource/v3.2.1/geonetwork.war/download. Enregistrez le `.war` dans le dossier `/tmp`.

Puis, déposez-le dans `/var/lib/tomcat8/webapps/geonetwork.war` :

```
cp /tmp/geonetwork.war /var/lib/tomcat8/webapps/geonetwork.war  
Démarrez Tomcat8 :
```

¹ Dernière version datant de juin 2017.

```
systemctl start tomcat8
```

Vous avez la possibilité de suivre les traitements grâce à la commande `top`¹.

```
top - 19:02:18 up 41 min,  2 users,  load average: 0.73, 0.67, 0.31
Tasks:  77 total,   1 running,  76 sleeping,   0 stopped,   0 zombie
%Cpu(s): 94.6 us,  5.4 sy,  0.0 ni,  0.0 id,  0.0 wa,  0.0 hi,  0.0 si,
0.0 st KiB Mem:  8195964 total, 1714800 used, 6481164 free,   992
buffers KiB Swap:  385020 total,    0 used,  385020 free. 1260812
cached Mem

PID USER      PR  NI   VIRT   RES   SHR  S  %CPU  %MEM    TIME+  COMMAND
13782 tomcat8   20   0 5186796 312088 17312 S  99.9   3.8   0:13.92 java
1 root      20   0  28696   4748   3072 S   0.0   0.1   0:01.12 systemd
2 root      20   0     0     0     0 S   0.0   0.0   0:00.00 kthreadd
3 root      20   0     0     0     0 S   0.0   0.0   0:00.30
ksoftirqd/0

5 root      0  -20     0     0     0 S   0.0   0.0   0:00.00
kworker/0:0H

6 root      20   0     0     0     0 S   0.0   0.0   0:00.27
kworker/u2:0
```

Figure 35. Suivi des traitements suite au démarrage de Tomcat8

2.4. Ouvrir GeoNetwork dans un navigateur web

Se rendre à l'adresse suivante : <http://votreserveur.votre-univ.fr:8080/geonetwork/>



ATTENTION : Il faut remplacer « `votreserveur.votre-univ.fr` » par le nom de domaine de votre ordinateur. Un moyen de connaître son nom de domaine est de taper la commande « `hostname` » dans la console.

Pour le premier lancement, l'ouverture de la page peut mettre plusieurs minutes. Une fois sur la page d'accueil de GeoNetwork, vous avez la possibilité de vous connecter en tant qu'administrateur avec les identifiants par défaut :

User : admin

Mot de passe : admin



Vous pouvez vous déconnecter et fermer la page de GeoNetwork sur votre navigateur Web pour suivre les étapes suivantes d'installation.

¹ Taper « `q` » pour sortir du suivi des traitements.

3. CONNECTER GEONETWORK A UNE BASE DE DONNEES POSTGRES

3.1. Installation de Postgres

```
apt-get install postgresql postgresql-client
```

3.2. Installation de Postgis

```
apt-get install postgis
```

3.3. Création d'un utilisateur Postgres

On se connecte à Postgres¹ :

```
su - postgres
```

On passe en mode SQL :

```
psql
```

On crée un superutilisateur avec un mot de passe :

```
CREATE ROLE lucien LOGIN PASSWORD 'lucien' SUPERUSER INHERIT CREATEDB  
CREATEROLE REPLICATION;
```

3.4. Création d'une base de données Postgres

On crée une base de données administrée par le superutilisateur créé préalablement :

```
CREATE DATABASE dbgeo WITH OWNER=lucien ENCODING='UTF8' CONNECTION  
LIMIT=-1;
```

On quitte notre connexion à Postgres en tant qu'utilisateur « postgres » :

```
\q
```

¹ Sous Debian 9 : su postgres

3.5. Création des extensions Postgis

On se connecte à la base Postgres créée préalablement, en tant qu'utilisateur de cette base :

```
psql -U lucien dbgeo
```

1



ATTENTION : En cas de message d'erreur pour se connecter, modifiez le fichier « pg_hba.conf » afin de se connecter avec d'autres comptes que ceux du système (il est nécessaire de rebooter le serveur après les modifications) :

```
nano /etc/postgresql/9.4/main/pg_hba.conf
```



Il se peut que vous ayez une version de Postgresql différente que la 9.4, dans ce cas, vérifiez dans vos fichiers la version ou alors utilisez la tabulation.

```
lucien@debian: ~
Fichier  Édition  Affichage  Rechercher  Terminal  Aide
GNU nano 2.2.6 Fichier : ../postgresql/9.4/main/pg_hba.conf

# Noninteractive access to all databases is required during automatic
# maintenance (custom daily cronjobs, replication, and similar tasks).
#
# Database administrative login by Unix domain socket
local all postgres peer
#
# TYPE DATABASE USER ADDRESS METHOD
#
# "local" is for Unix domain socket connections only
local all all md5
# IPv4 local connections:
host all all 127.0.0.1/32 md5
# IPv6 local connections:
host all all ::1/128 md5
# Allow replication connections from localhost, by a user with the
# replication privilege.
#local replication postgres md5
#host replication postgres 127.0.0.1/32 md5
#host replication postgres ::1/128 md5

^G Aide ^O Écrire ^R Lire fich ^Y Page préc ^K Couper ^C Pos. cur.
^X Quitter ^J Justifier ^W Chercher ^V Page suiv ^U Coller ^T Orthograp
```

Figure 36. Aperçu du fichier « pg_hba.conf » dans le terminal

Pour pouvoir autoriser la connexion, il est nécessaire de modifier l'accès « peer » en « md5 ».

Pour enregistrer la modification du fichier : CTRL + O

Pour quitter le fichier : CTRL + X

¹Exemple de message d'erreur : Attention ! psql: FATAL: authentification peer échouée pour l'utilisateur « lucien ».

Rebooter maintenant votre machine virtuelle :

```
reboot
```

Reconnectez-vous à Postgres :

```
su - postgres
```

Connectez-vous à la base Postgres créée préalablement, en tant qu'utilisateur de cette base :

```
psql -U lucien dbgeo
```

Créez les extensions :

```
CREATE EXTENSION postgis;  
  
CREATE EXTENSION postgis_topology;
```

Quittez Postgres avant de configurer Tomcat :

```
\q  
\exit
```

3.6. Configurer Tomcat8

Éditer le fichier srv.xml

Arrêtez le serveur tomcat8 :

```
systemctl stop tomcat8
```

Accédez en mode édition au fichier srv.xml :

```
nano /var/lib/tomcat8/webapps/geonetwork/WEB-INF/config-node/srv.xml
```

Décommentez la ligne suivante (dire à Geonetwork de se connecter à une BD Postgres) :

```
<import resource= « ../config-db/postgres.xml »/>
```

Commentez la ligne suivante (dire à Geonetwork de ne plus se connecter à une BD H2) :

```
<import resource= « ../config-db/h2»/>
```

Enregistrez les modifications et enregistrez le fichier srv.xml.

Éditer le fichier postgres.xml

Accédez en mode édition au fichier postgres.xml :

```
nano /var/lib/tomcat8/webapps/geonetwork/WEB-INF/config-  
db/postgres.xml
```

Apportez les modifications suivantes au fichier postgres.xml (en gras) :

```
<?xml version="1.0" encoding="UTF-8"?>  
  
<beans default-lazy-init="true"  
  
xmlns=http://www.springframework.org/schema/beans  
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"  
  
xsi:schemaLocation="  
  
http://www.springframework.org/schema/beanshttp://www.springframework.org/sc  
hema/beans/spring-beans.xsd  
">  
  
<import resource="defaultJdbcDataSource.xml" />  
  
<import resource="postgis-alternate-datasource.xml" />  
  
<bean id="jpaVendorAdapterDatabaseParam" class="java.lang.String">  
  
<constructor-arg value="POSTGRESQL"/>  
  
</bean>  
  
<bean id="jdbcDriverClassName" class="java.lang.String">  
  
<constructor-arg value="org.postgis.DriverWrapper"/>
```

```
</bean>
```

```
<bean id="jdbcURL" class="java.lang.String">
```

```
  <constructor-arg
```

```
    value="jdbc:postgresql_postGIS://${jdbc.host}:5432/${jdbc.database}"/>
```

```
</bean>
```

```
</beans>
```

Éditer le fichier jdbc.properties

Accédez en mode édition au fichier jdbc.properties :

```
nano /var/lib/tomcat8/webapps/geonetwork/WEB-INF/config-  
db/jdbc.properties
```

Modifiez le fichier jdbc.properties en fonction des identifiants de l'utilisateur et de la base de données créés sur Postgres :

```
jdbc.usernam=lucien  
jdbc.password=lucien  
jdbc.database=dbgeo  
jdbc.host=localhost  
jdbc.port=5432
```



C'est par l'intermédiaire de ce fichier et de ces paramètres que vous pourrez changer de base de données si besoin est.

Relancez le serveur Tomcat8¹ :

```
systemctl start tomcat8
```

¹ **Attention**, si vous souhaitez modifier le schéma de la base de données GeoNetwork (Cf. partie 5

5. MODIFIER LE SCHEMA DE LA BASE DE DONNEES GEONETWORK

Par défaut, lorsque GeoNetwork est configuré pour Postgres et lorsqu'il est lancé pour la première fois, il crée l'ensemble de ses tables dans le schéma « public » de Postgres. Pour des raisons de gestion de la base de données, il peut être intéressant de stocker ses tables non pas dans le schéma « public » mais dans un schéma que vous aurez créé préalablement, réservé à GeoNetwork.

Pour réaliser cette configuration, il faut avoir une base de données vierge, c'est-à-dire dans laquelle GeoNetwork n'a pas encore été déployé.



ATTENTION : Pour avoir cette base de données vierge, vous pouvez suivre les parties suivantes de la documentation : « Création d'un utilisateur Postgres », « Création d'une base de données Postgres » et « Création des extensions Postgis ». Vous pouvez également faire ces manipulations directement dans pgAdminIII.

Ensuite, au sein de cette base de données vierge, suivez les étapes qui suivent.

5.1. Renommer le schéma « public »

En suivant la partie « administrer la base de données Postgres avec l'interface pgAdminIII », connecter-vous à votre base de données afin d'accéder à son contenu.

Puis, modifiez le nom du schéma « public » créé par défaut lors de la création de la base de données.

Clic droit sur le schéma « public » > Propriétés > Modifier le nom (vous pouvez par exemple mettre « public_old »).

5.2. Créer un nouveau schéma « public »

Clic droit au niveau des schémas > Ajouter schéma

Affecter à ce nouveau schéma les mêmes droits que l'ancien schéma « public » que vous avez renommé juste avant.

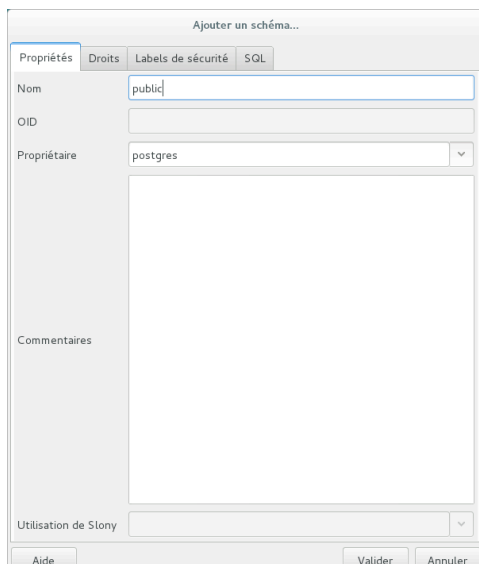


Figure 39. Nom du nouveau schéma

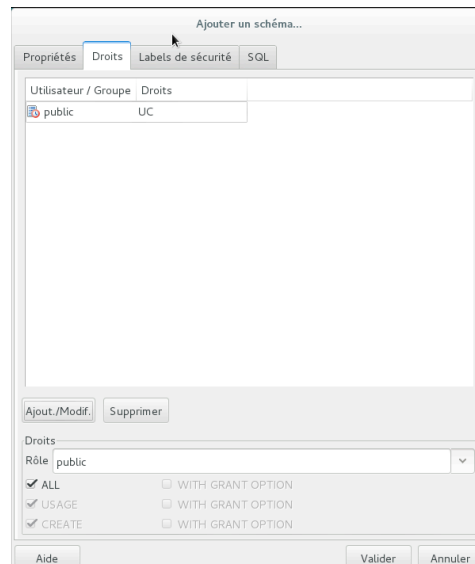


Figure 40. Droits du nouveau schéma

5.3. Relancer Tomcat8

Vous pouvez maintenant relancer Tomcat8 :

```
systemctl restart tomcat8
```

Puis, vous pouvez ouvrir GeoNetwork dans votre navigateur web afin que GeoNetwork déploie ses tables dans le nouveau schéma « public » qui vient d'être créé.

Dans pgAdminIII, vous pouvez vérifier dans le schéma « public » que les tables de GeoNetwork sont bien présentes.

5.4. Renommer le schéma « public »

À présent, renommez le schéma « public » pour qu'il soit à présent identifié à GeoNetwork.

Clic droit sur le schéma « public » > Propriétés > Modifier le nom (vous pouvez par exemple mettre « geonetwork »).

5.5. Paramétrer votre BD pour la connecter au nouveau schéma « geonetwork »

Il faut maintenant attribuer à la BD la possibilité d'aller chercher les tables non pas seulement dans le schéma « public » mais également dans le schéma « geonetwork ». Pour cela, exécutez la requête SQL suivante :

```
ALTER DATABASE « nom de votre BD » SET search_path = public, geonetwork;
```

5.6. Renommer le schéma « public_old »

Vous pouvez maintenant renommer le schéma « public_old » en schéma « public » afin de retrouver le schéma d'origine « public » dans votre base de données.

Clic droit sur le schéma « public_old » > Propriétés > Modifier le nom en « public »

Vous pouvez relancer le serveur Tomcat8 :

```
systemctl restart tomcat8
```

Ouvrez maintenant GeoNetwork dans votre navigateur web et vérifiez que le lien entre GeoNetwork et la BD fonctionne. Pour cela, vous pouvez par exemple créer un nouvel utilisateur dans GeoNetwork, puis vérifiez dans pgAdminIII au sein de la table « users » que votre nouvel utilisateur est bien présent.

Annexe 8 : Configuration du pare-feu pour le serveur interParc

`vi /etc/shorewall/params` Ajouter les lignes suivantes:

```
SRVGEONETWORK="10.95.10.12"
```

Modifier les règles du parefeu:

```
ACCEPT net:$SRVWEB1 fw tcp 22,5432

#redirection pour geonetwork
DNAT net:$SUPPORT_NET loc:$SRVGEONETWORK:22 tcp 45
DNAT net:$SUPPORT_PNR loc:$SRVGEONETWORK:22 tcp 45
DNAT net loc:$SRVGEONETWORK:8080 tcp 8080
```

Vérification des règles et mise à jour du shorewall

```
shorewall check
```

```
shorewall restart
```

Il faut ensuite configurer la machine geonetwork

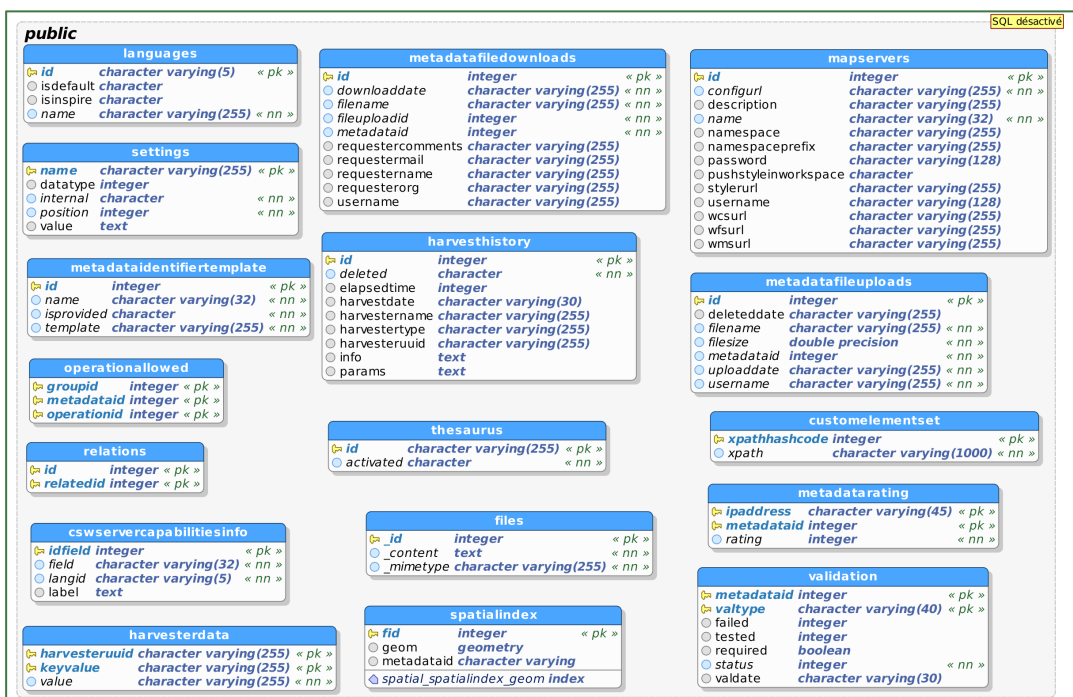
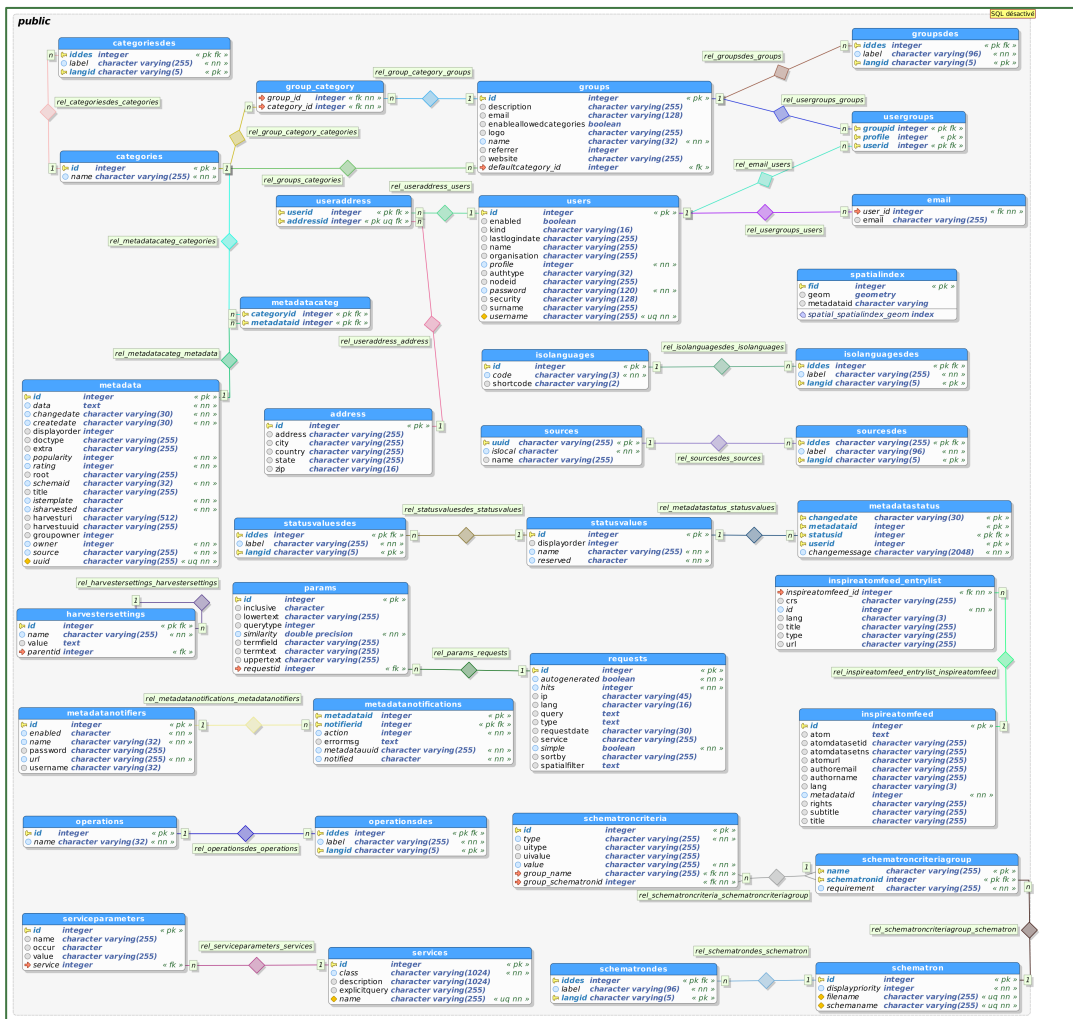
Dans le fichier `/etc/network/interfaces`, ajoutez la ligne

```
# The primary network interface
auto eth0
iface eth0 inet static
    address 10.95.10.12
    netmask 255.255.255.0
    gateway 10.95.10.254
```

Enfin, appliquez les modifications

```
/etc/init.d/networking restart.
```

Annexe 9 : Modèle Conceptuel de Données de Geonetwork (tables liées et tables isolées)



Annexe 10 : Extrait de la documentation sur le MCD de Geonetwork

3.1. address

Table	address
Relations	useraddress (1 ; N) <i>Une adresse peut avoir plusieurs utilisateurs mais un utilisateur ne peut avoir qu'une adresse</i>
Clé primaire	Identifiant unique id (entier)
Attributs	id (entier) ; adress (caractère) ; city (caractère) ; country (caractère) ; state (caractère) ; zip (caractère)
Rôle	Stocke les informations concernant l'adresse des utilisateur

3.2. categories

Table	categories
Relations	metadatacateg (1 ; N) <i>Une catégorie peut avoir plusieurs metadatacategorie mais une metadatacategorie ne peut avoir qu'une catégorie</i> categoriesdes (1 ; N) <i>Une catégorie peut avoir plusieurs categoriesdes mais une categoriesdes ne peut avoir qu'une catégorie</i> group_category (1 ; N) <i>Une catégorie peut avoir plusieurs group_category mais un group_category ne peut avoir qu'une catégorie</i> groups (1 ; N) <i>Une catégorie peut avoir plusieurs groups mais un groups ne peut avoir qu'une catégorie</i>
Clé primaire	Identifiant unique id (entier)
Attributs	id (entier) ; name (caractère)
Rôle	

3.3. categoriesdes

Table	categoriesdes
Relations	categories (N ; 1) <i>Une categoriesdes peut avoir qu'une catégorie mais une catégorie peut avoir plusieurs categoriesdes</i>
Clé primaire	langid (caractère)
Clé étrangère	iddes (clé primaire de categories)
Attributs	id (entier) ; adress (caractère) ; city (caractère) ; country (caractère) ; state (caractère) ; zip (caractère)
Rôle	

3.4. groups

Annexe 11 : Modèle type pour la fiche de métadonnées (format XML)

```
modele_type_metadata.xml
1 <gmd:MD_Metadata xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:gml="http://www.opengis.net/gml" xmlns:gco="http://www.isotc211.org/2005/gco" xmlns:
  xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.isotc211.org/2005/gmd http://www.isotc211.org/2005/gmd.xsd http://
  www.isotc211.org/2005/gmx http://www.isotc211.org/2005/gmx.xsd http://www.isotc211.org/2005/srv http://schemas.opengis.net/iso/19139/20060504/srv/
  srv.xsd">
2 <gmd:fileIdentifier>
3 <gco:CharacterStrings>937bbb6e-c4c2-43e9-aafe-f8b2595f839d</gco:CharacterString>
4 </gmd:fileIdentifier>
5 <gmd:language>
6 <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/" codeListValue="fre"/>
7 </gmd:language>
8 <gmd:characterSet>
9 <gmd:MD_CharacterSetCode codeListValue="utf8" codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/codeList/
  ML_gmxCodeLists.xml#MD_CharacterSetCode"/>
10 </gmd:characterSet>
11 <gmd:contact>
12 <gmd:CI_ResponsibileParty>
13 <gmd:individualName>
14 <gco:CharacterStrings>Nom de la personne</gco:CharacterString>
15 </gmd:individualName>
16 <gmd:organisationName>
17 <gco:CharacterStrings>Nom de l'organisation</gco:CharacterString>
18 </gmd:organisationName>
19 <gmd:positionName>
20 <gco:CharacterStrings>Chargé de mission SIG et évaluation</gco:CharacterString>
21 </gmd:positionName>
22 <gmd:contactInfo>
23 <gmd:CI_Contact>
24 <gmd:phone>
25 <gmd:CI_Telephone>
26 <gco:CharacterStrings>--- Numéro ---</gco:CharacterString>
27 </gmd:phone>
28 </gmd:voice>
29 <gmd:facsimile>
30 <gco:CharacterStrings>Fax</gco:CharacterString>
31 </gmd:facsimile>
32 </gmd:CI_Telephone>
33 <gmd:phones>
34 <gmd:address>
35 <gmd:CI_Address>
36 <gmd:deliveryPoint>
37 <gco:CharacterStrings>Adresse</gco:CharacterString>
38 </gmd:deliveryPoint>
39 <gmd:city>
40 <gco:CharacterStrings>Commune</gco:CharacterString>
41 </gmd:city>
42 <gmd:administrativeArea gco:nilReason="missing"/>
```

XML

Tab Size: 4

Line 1, Column 1