

MARS 2019

OPEN- DATA ET BIODIVERSITÉ

POURQUOI ET COMMENT PUBLIER SES
DONNÉES DE BIODIVERSITÉ EN OPENDATA ?

OLIVIER ROVELLOTTI
NATURAL SOLUTIONS

TABLER DES MATIÈRES

- 3** REMERCIEMENTS
- 4** À NOS LECTEURS
- 6** HISTORIQUE DE L'OPEN DATA
- 7** QU'EST-CE QU'UNE DONNÉE PUBLIQUE ?
- 8** POURQUOI PUBLIER SES DONNÉES ?
- 9** POURQUOI DES DONNÉES DE BIODIVERSITÉ ?
- 10** LE CADRE LÉGAL
- 12** QUE DIT LA LOI BIODIVERSITÉ ?
- 15** QUEL FORMAT ?
- 18** QUELLE LICENCE ?
- 19** QUELS OUTILS ?
- 22** ANNEXE 1 : TELEVERSEMENT
- 24** ANNEXE 2 : MODULE GEONATURE EXPORT
- 27** ANNEXE 2 : MODULE ECORELEVE

REMERCIEMENTS

Nous remercions chaleureusement
les participants à ce document qui ont
donné de leurs temps et énergie :

Amandine Sahl (Parc national des Cévennes)

Charles Nepote (Fing)

Camille Monchicourt (Parc national des
Ecrins)

Franck Michel (CNRS)

Simon Chignard (Etalab)

Et les autres ...

À NOS LECTEURS

« OPEN DATA PAR DÉFAUT ! »

L'ouverture et le partage des données publiques ou open data, consistent à mettre à disposition de tous les citoyens, sur Internet, toutes les données publiques brutes qui ont vocation à être librement accessibles et gratuitement réutilisables.

Le droit d'accès à ces données s'impose à l'Etat, aux collectivités territoriales et à toutes les autres personnes de droit public ou de droit privé chargées d'une mission de service public. *L'open data n'est plus une recommandation mais une obligation !*

Depuis samedi 7 octobre 2018, les administrations d'au moins 50 agents sont tenues de mettre en ligne, en open data, toutes les données publiques qu'elles produisent, qu'elles détiennent ou qu'elles ont reçues dans le cadre de leurs missions de services publics et qui ne sont pas des données personnelles ou des données protégées par un secret industriel ou commercial. Il s'agit de dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, instructions, circulaires. Ils devront être publiés sans attendre une quelconque sollicitation. Cette loi a pour objectif de faire de la circulation des données la règle, et non plus l'exception.

Cette obligation qui s'impose en particulier aux collectivités locales de plus de 3500 habitants n'est pas une surprise. Elle était prévue par la loi pour une république numérique, votée le 16 octobre 2016. Elle pourrait toutefois ne pas être suivie d'effets immédiats car les organismes publics ne se sont pas encore mobilisés pour la mettre en œuvre. L'association Opendata France a publié, début octobre 2018, l'Opendata des territoires 2018. Il révèle que sur 4500 territoires concernés seuls 308 ont engagé une démarche de publication de leurs données, une centaine de plus par an. « A ce rythme- là, il nous faudra près de 40 ans pour que cette démarche, pourtant applicable au plus tard le 8 octobre 2018, soit généralisée, regrette Laurence Comparat, présidente d'Opendata France et adjointe au maire de la ville de Grenoble

(La Gazette des Communes, 1^{er} octobre 2018).



AXELLE LEMAIRE

Ancienne secrétaire d'Etat, chargée du Numérique et de l'innovation, associée au sein du cabinet Roland Berger.

Le défi est lourd car l'open data soulève des enjeux d'organisation, de moyens, de ressources, de compréhension, de formation qui sont tous des chantiers à long terme

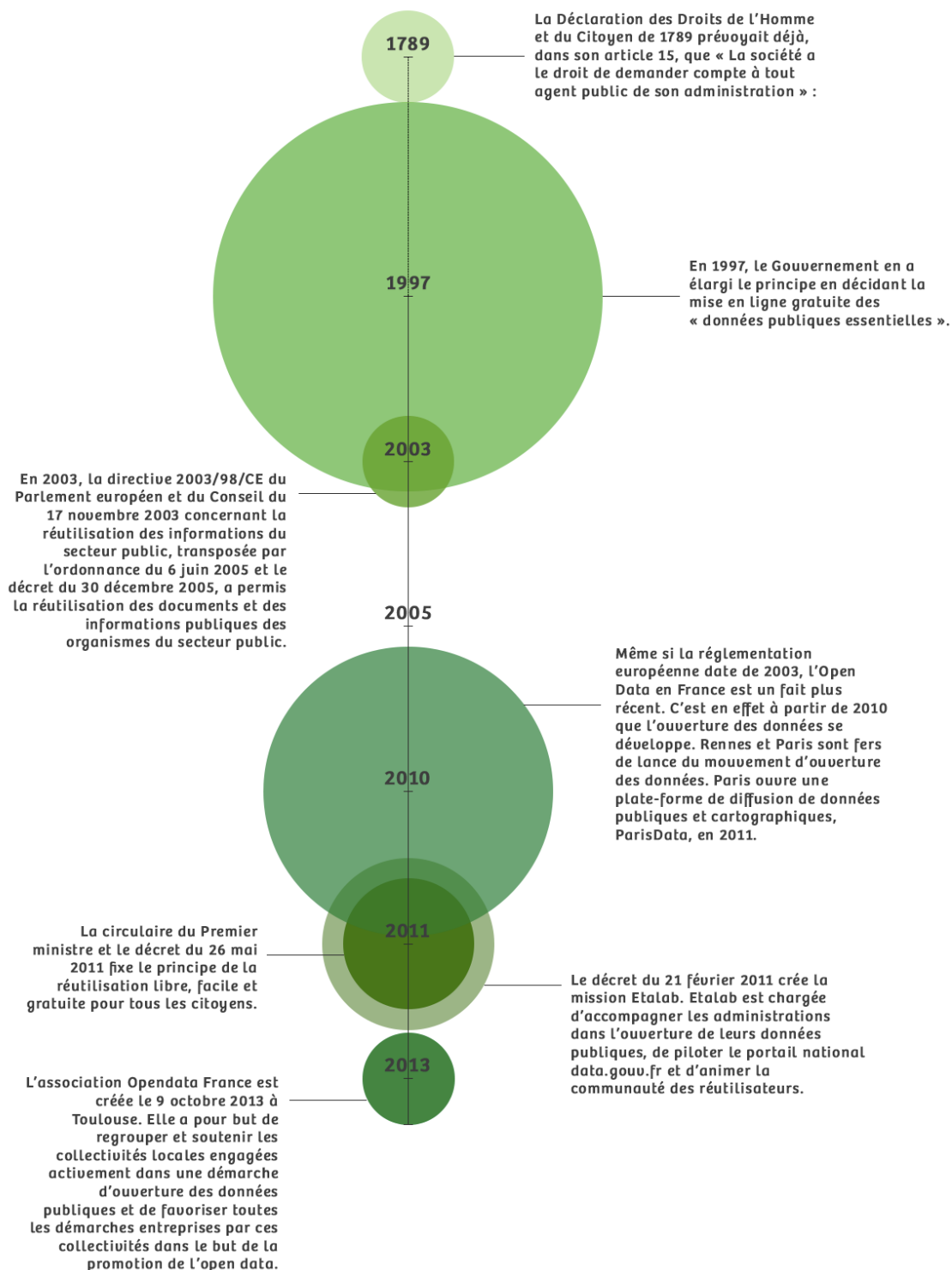
Selon Romain Mazon, rédacteur en chef de cet hebdomadaire, les raisons de cette frilosité sont multiples :

- L'absence de sanction juridique, ou pécuniaire, attachée au non-respect de la loi.
- Cette démarche demande un portage politique encore trop peu répandu.
- Elle bouscule les cultures et les pratiques.

Le défi est lourd car l'open data soulève des enjeux d'organisation, de moyens, de ressources, de compréhension, de formation qui sont tous des chantiers à long terme, estime pour sa part Axelle Lemaire ancienne secrétaire d'Etat, chargée du Numérique et de l'innovation, associée au sein du cabinet Roland Berger.

Enfin, l'hétérogénéité des sources, la multiplicité des formats, le volume de données, les métadonnées et l'expression de la sémantique des données, pour ne citer que ces points, exigent l'élaboration de méthodologies nouvelles pour collecter, nettoyer, apparier, fouiller, enrichir et diffuser ces données.

HISTORIQUE DE L'OPEN-DATA



QU'EST-CE QU'UNE DONNÉE PUBLIQUE

Le secrétariat général pour la modernisation de l'action publique la définit ainsi : « Il s'agit de données produites ou reçues par une autorité administrative dans le cadre de sa mission de service public, publiées par une autorité administrative ou communicables à toute personne en faisant la demande. Ces informations doivent être présentées sous un format permettant leur traitement automatisé et leur réutilisation. Le droit d'accès et de réutilisation des données publiques concerne donc les textes, mémorandums, documents, tableaux ou statistiques produits par l'administration dans le cadre d'une mission de service public. Il ne concerne pas les documents préparatoires et non définitifs de l'administration en vue de ses délibérations ».

Selon la loi de 2016, quatre types de documents administratifs doivent obligatoirement être publiés en ligne sous réserve qu'ils soient disponibles sous forme électronique :

- Les documents qui ont été communiqués conformément à la loi Cada, à savoir toutes les pièces administratives ayant fait l'objet d'une demande de la part d'un administré.
- Les documents mis à jour, figurant dans le répertoire des principaux documents administratifs.
- les bases de données, mises à jour, produites ou reçues par les administrations.
- Les données, mises à jour, dont la publication présente un intérêt économique, social, sanitaire ou environnemental.

Quelques exceptions toutefois :

- Les documents administratifs dont la consultation porterait atteinte au secret de la défense nationale, à la mise en œuvre de la politique extérieure de la France, à la sûreté de l'Etat, à la sécurité publique, à la sécurité des personnes ou à la sécurité des systèmes d'information des administrations ou encore au secrets protégés par la loi comme le secret professionnel, le secret industriel, etc... Enfin les données personnelles sauf celles qui ont fait l'objet d'une anonymisation, et celles sur lesquelles les tiers détiennent des droits de propriétés intellectuelles ne peuvent pas être ouvertes.

POURQUOI PUBLIER SES DONNÉES ?

Pour fluidifier le partage d'informations entre les services et entre les institutions de différentes tailles qui travaillent au sein d'un même territoire.

Pour faciliter l'accès et le formatage de l'information dans un langage commun qui permettent de décloisonner les services et les collectivités. Les jeux de données standardisés, c'est-à-dire ceux qui respectent des codes connus de tous, ouvrent la voie aux échanges automatisés entre collectivités.

Rendre accessibles et intelligibles les données est un véritable enjeu démocratique car elles permettent de donner tous les outils aux citoyens et aux associations qui veulent s'emparer de problématiques relevant de la collectivité. En ouvrant ses données, on crée des relations avec des publics qu'on n'a pas forcément l'habitude de côtoyer. Beaucoup de collectivités locales s'en sont aperçues : des associations, des créateurs d'entreprises, des acteurs de l'innovation locale vont s'en emparer.

Les données ouvertes constituent une matière première pour une nouvelle industrie des biens et des services en dehors même du domaine où elles ont été créées.

Elles constituent un matériau précieux pour les chercheurs dans toutes les disciplines des sciences et de la culture. Dans beaucoup d'organisations la donnée a été plutôt utilisée pour nourrir des tableaux de bord et faciliter la décision. On voit arriver des usages de données couplées à des algorithmes qui ne se contentent plus d'informer un décideur mais de lui faire des recommandations. On va vers des outils susceptibles de prédire les phénomènes.

POURQUOI DES DONNÉES DE BIODIVERSITÉ ?

Pour plus de transparence

La protection de l'Environnement doit être l'affaire de tous, pour cela il est impératif de communiquer clairement sur l'enjeu lié à cette protection. Le public ainsi informé sera d'autant plus réceptif à la contrainte liée à ces enjeux.

Pour toucher de nouveau public

La Biodiversité est encore trop souvent une affaire d'expert. Afin d'optimiser notre mission de sauvegarde, il est nécessaire de toucher un public le plus large possible. Les plateformes open data sont certes à l'heure actuelle réservées à un usage très « geek » mais les initiatives telles que les insola ont pour objectif d'ouvrir ces enjeux à un public plus large.

Pour permettre des réutilisations inédites

Votre institution n'a probablement pas le moyen de financer l'application mobile grand public, l'outil d'interprétation et la plateforme touristique nécessaires à une bonne communication avec des nouveaux publics. L'open data peut agir comme un point de contact et susciter des usages qui n'ont pas été envisagés par les producteurs des données.

LE CADRE LÉGAL

L'ouverture des données publiques ne s'est pas faite en un jour. Elle a commencé en 1978, avec la loi CADA, loi qui donnait à chaque citoyen le droit de solliciter une administration pour lui demander un document en sa possession. Ce texte a été modifié plusieurs fois pour être adapté aux évolutions de la société et des technologies.

En 2016, la loi pour une République numérique est venue inverser les rôles définis par la loi CADA, puisqu'elle oblige désormais les administrations à ouvrir spontanément leurs données, sans attendre qu'un citoyen en fasse la demande.

« Seuls les documents numériques sont concernés, précise Etalab. Le principe d'ouverture spontanée s'applique à tous les documents administratifs qui existent sous une forme numérique, c'est-à-dire ceux qui n'ont pas besoin d'être passés au scanner ou sortis des archives pour être mis en ligne. La publication des documents peut se faire sur data.gouv.fr ou sur le site de l'administration concernée ».

Pour faciliter l'action des administrations, l'ensemble de leurs obligations est rassemblé et codifié dans le Code des Relations entre Public et Administration) : le CRPA (livre III, article 300-1 et suivants).

La CNIL et la CADA ont annoncé la publication, d'ici la fin de l'année 2018, d'une synthèse du cadre juridique de l'open data : rappels des différentes règles, tant en matière de mise en ligne des documents administratifs que de protection des données personnelles. Il permettra de clarifier l'interprétation de l'ensemble des dispositions présidant à la publication et à la réutilisation des données contenues dans les documents administratifs.

UN PRÉCURSEUR DE LA «DATAVISUALISATION» DE L'INFORMATION.



Il s'appelle **John Snow**. En 1854, il superposa sur une carte les adresses des victimes du choléra et constata qu'elles coïncidaient avec les accès aux points d'eau. Il découvrit l'origine des épidémies de choléra.



Quelle réutilisation ?

La réutilisation des données publiques peut susciter le développement de nouveaux services comme les applications mobiles, des sites Internet, des visualisations de données ou « datavisualisation » notamment par la presse, etc. Les données publiques peuvent être aussi réutilisées par les chercheurs, les enseignants, les étudiants, les responsables associatifs, les citoyens, pour construire de nouveaux points de vue sur la société ou sur l'action publique. Bref : cette réutilisation, qui doit être libre et gratuite, peut aussi avoir lieu à d'autres fins que celles pour lesquelles les données ont été produites ou reçues. « La loi pour une République numérique oblige les administrations à ouvrir leurs données, sans préjuger de leur utilisation potentielle, rappelle Brice Boulesteix, sur le blog d'Etalab. Ce qu'il est possible de faire avec les données publiées dépend de la licence qui en régit l'utilisation, mais dans le cas d'une licence ouverte, les données peuvent alors être :

- Réutilisées : par exemple copier les données, les mettre en forme dans un tableau, ou sur un fond de carte ;
- Redistribuées : par exemple publier les données sur un blog, dans un livre, ou encore sur Twitter ;
- Adaptées : par exemple changer le titre d'une colonne dans un fichier CSV pour en faciliter la compréhension
- Exploitées commercialement : par exemple faire une carte à partir de données ouvertes, l'imprimer, et la vendre.

Sous réserve de :

- Mentionner la source des données réutilisées ;
- Inclure la date de dernière mise à jour des données.

QUE DIT LA LOI BIODIVERSITÉ ?

La loi pour la biodiversité, désormais votée par le parlement, prévoit -entre autres- un libre accès aux données naturalistes contenues dans les bases de données de l'Inventaire national du patrimoine naturel (INPN). Si cette décision est une bonne nouvelle du point de vue de la transparence et de la démocratie d'une part ; de la préservation du patrimoine naturel grâce à sa meilleure connaissance, d'autre part, elle suscite toutefois de nombreuses interrogations sur sa mise en œuvre.

Avec ses 3 millions de données de faune, de flores sauvages et d'habitats naturels, l'Inventaire national du patrimoine naturel (INPN) est la plus grosse base de données de ce type au niveau de notre pays. D'autant qu'elle est enrichie chaque jour par plus de 5000 données versées par plusieurs dizaines de contributeurs en France, auxquels s'ajouteront prochainement celles des maîtres d'ouvrage et des collectivités, au titre de la Loi pour la reconquête de la biodiversité, de la nature et des paysages. Concrètement, chaque citoyen pourra ainsi avoir accès à l'intégralité de ces informations, librement consultables en ligne, ce qui « préfigure une forme « d'Open data pour la biodiversité » comme le soulignait un article du Monde.

Des études d'impacts plus transparentes

Faut-il s'en réjouir ? Incontestablement, puisque cela aboutira à une démocratisation des données concernant la biodiversité des territoires. Une partie d'entre elles, notamment celles « récoltées » par les maîtres d'ouvrages privés ou publiques, à l'occasion d'études d'impact avant la construction d'infrastructures n'étaient jusqu'à présent pas accessibles, ces derniers n'étant pas tenus de la rendre public. La loi permettra concrètement de valider ou d'invalidier la pertinence des études menées par des structures spécialisées et diminuer le risque qu'elles dissimulent ou minimisent la découverte d'espèces protégées, pour ne pas remettre en cause les travaux.

Protéger les données sensibles

Mais cela offrira aussi aux bureaux d'études spécialisés une base de travail existante, ce qui est loin d'être négligeable, en termes de temps et de moyens à mettre en œuvre pour réaliser les études d'impacts. Les collectivités locales bénéficieront enfin d'une base existante pour réaliser les Atlas de la biodiversité des communes (ABC). La loi pourrait, de ce point de vue, donner un coup d'accélérateur à ce projet du ministère de l'Environnement, dont peu de communes et d'intercommunalités se sont emparées, faute de moyens financiers et humains suffisants. La mise à disposition publique des données permettra, enfin, de concevoir ou d'améliorer les outils de saisie ; de cartographie et de relevés biologiques existants, à l'image des produits développés par Natural Solutions.

Protéger les données sensibles

Mais cela offrira aussi aux bureaux d'études spécialisés une base de travail existante, ce qui est loin d'être négligeable, en termes de temps et de moyens à mettre en œuvre pour réaliser les études d'impacts. Les collectivités locales bénéficieront enfin d'une base existante pour réaliser les Atlas de la biodiversité des communes (ABC). La loi pourrait, de ce point de vue, donner un coup d'accélérateur à ce projet du ministère de l'Environnement, dont peu de communes et d'intercommunalités se sont emparées, faute de moyens financiers et humains suffisants. La mise à disposition publique des données permettra, enfin, de concevoir ou d'améliorer les outils de saisie ; de cartographie et de relevés biologiques existants, à l'image des produits développés par Natural Solutions.

Sauvegarder les espèces protégées

La mise en œuvre de la Loi pour la reconquête de la biodiversité, et en particulier son volet concernant la mise à disposition publique des données, suscite en revanche de nombreuses interrogations. Qui par exemple des données sensibles, concernant par exemple des espèces protégées ou bénéficiant d'un statut de conservation forte ?

Trop de transparence aurait incontestablement un effet contreproductif, dans ce cas précis : des garde-fous restent encore à définir.

Harmoniser les modes de saisies naturalistes

La question des licences, liées à l'exploitation des futures données récoltées et consultables sur le site de l'INPN, reste également à déterminer. Idem pour la pérennité des entreprises dont le modèle économique est basé sur la récolte des données naturalistes :

il devra être éclairci rapidement. Enfin, un dernier point -et pas des moindres- reste à régler : celui de l'interopérabilité des taxinomies et des protocoles de saisie, autrement dit la manière de récolter les datas sur le terrain. Leur harmonisation, entre bureaux d'études et associations naturalistes, devra nécessairement être étudiée pour garantir la fiabilité des futures informations contenues dans les bases de données de l'Inventaire national du patrimoine naturel.

ENTRETIEN AVEC OLVIER ROVELLOTTI, DIRECTEUR DE NATURAL SOLUTIONS

“ Quelles plateformes existent pour collecter les données naturalistes ? ”

Il existe différentes échelles de plateformes où publier ses données. Les plateformes locales, à l'instar de Géonature que nous sommes en train de développer, permettent de publier ses données à l'échelle de son parc, de sa réserve, de son association ou de sa collectivité. Il existe aussi des plateformes régionales comme Silene, le portail public des données naturalistes en Paca, gérée par la Dreal Paca. Ensuite ces plateformes locales ou régionales se chargent de les communiquer à la plateforme nationale du SINP, Ginco - Gestion d'Information Naturaliste Collaborative et Ouverte-. L'inter-moissonnage entre les différentes plateformes qui collectent les données de biodiversité reste un problème. Il faudrait qu'elles puissent communiquer entre-elles pour agréger ces différentes données.

“ Ces données sont-elles ouvertes à monsieur et madame tout le monde ? ”

Non. Car il existe la problématique des données sensibles. C'est la station d'orchidée, c'est le lieu de nidification des vautours, c'est la peur du chasseur. Le SINP a beaucoup réfléchi sur cette notion de données sensibles. Il s'agit de données dont la consultation ou la communication peut porter atteinte à la biodiversité. Le protocole national du SINP prévoit que les données estampillées « sensibles » soient uniquement accessibles aux autorités publiques. Il existe un référentiel national des données sensibles qui a ensuite été décliné dans chaque région. Résultat : sur Silene, par exemple, monsieur et madame tout le monde peuvent avoir accès à une partie des données mais les autres comme la localisation sont floutées. La donnée est dégradée.

“ Quelles relations existe-t-il entre les communautés biodiversité et opendata ? ”

Ces deux communautés n'ont pas encore vraiment fait le lien et communiquent peu. Et c'est dommage. La communauté biodiversité a beaucoup avancé sur la question de l'interopérabilité avec le standard SINP et Darwin core et ce n'est pas forcément le cas dans tous les domaines. Il faut le souligner. Les échanges de données existent depuis plus de 20 ans entre experts. Le mouvement Opendata est arrivé beaucoup plus tard, il y a dix ans environ.

“ Quel est l'enjeu du futur ? ”

Le web sémantique connaît un désintérêt de la communauté scientifique actuellement mais il reviendra probablement sous une autre forme. Parce que l'enjeu de l'interopérabilité et de l'inter moissonnage est là. Les chercheurs se sont beaucoup concentrés sur le big data, sur le deep learning et beaucoup moins sur l'intelligence artificielle symbolique. On s'est éloigné de tout ce qui était ontologie ... Cette vision a pris du plomb dans l'aile.

On en est arrivé à un tel niveau de complexité que mettre en place les plomberies du web sémantique est devenu vraiment difficile ! Il faut en revenir à plus de pragmatisme sur l'annotation des datas sets pour mettre en place des solutions de moissonnage qui soient simples et efficaces.

QUEL FORMAT ?

Pour toutes les données mises à disposition, les administrations doivent recourir à « un standard ouvert, aisément réutilisable et exploitable par un système de traitement automatisé ». En clair, pas question de mettre en ligne des fichiers PDF absolument inexploitable sur le plan technique. « La démarche devrait être de proposer en priorité un format ouvert comme CSV ou XML », poursuit Simon Chignard. Ou à défaut des formats dits standards de fait comme XLS. D'une manière générale, il ne faut pas que les formats retenus limitent les possibilités de réutilisations. Ainsi, on fait le choix de multiplier les formats de mise à disposition plutôt que de se restreindre à un seul format. « Il faut comprendre que l'open data évolue » souligne Charles Nepote, responsable du programme OpenData Impact, à l'occasion d'un entretien accordé à la Gazette des Communes le 8 novembre 2018.

Il y a dix ans, je disais « **ouvrez vos données et Dieu reconnaîtra les siens** ». On ne peut plus continuer comme ça. Il faut savoir ce que l'on nous demande en priorité et s'y adapter. Cela passera sans doute par plus d'énergie pour répondre aux demandes de certains secteurs. Les données urbaines, par exemple, sont particulièrement appréciées et l'immobilier entre dans une véritable réutilisation des données ouvertes. Le transport est également allé très vite grâce à des standards existants comme GTFS, Neptune et Trident. À l'inverse, la culture et l'éducation restent les parents pauvres. Le tourisme est en demi-teinte : le journal Ouest-France peut faire un agenda culturel commun sur son territoire, mais dans de nombreux endroits, les acteurs sont encore très frileux et préfèrent penser à des conventions plutôt qu'à l'ouverture. La disparité est grande mais il est important de noter que, dans certaines conditions, ce n'est pas tant le gratuit qui est important que le normalisé.

ENTRETIEN AVEC SIMON CHIGNARD, CONSEILLER STRAGÉGIQUE POUR ETALAB ET AUTEUR D'OPEN DATA, COMPRENDRE L'OUVERTURE DES DONNÉES PUBLIQUES (EDITIONS FYP).

“ Pourquoi la standardisation des données est-elle importante ? ”

Ce travail de standardisation est essentiel. Car si l'on veut qu'un utilisateur puisse comparer deux territoires, une thématique au niveau national ou international, il faut qu'il puisse s'appuyer sur des standards et que les données soient décrites de manière identique. Cette standardisation avance à des vitesses différentes selon les sujets. Dans le domaine des transports et plus particulièrement de l'information voyageurs, des modèles de standards ont émergé et en particulier un modèle de données appelé Gpfs. Le Gpfs est un standard d'échange de données « usagers » dans un cadre d'interopérabilité des réseaux de transport. Cette norme a été mise en place afin de simplifier la transmission des informations « usagers » entre les différents acteurs du transport. Un autre exemple : la description des bornes de recharge des véhicules électriques.

“ Où en est la standardisation des données ? ”

Open data France a travaillé sur le socle commun des données locales. Les standards d'une dizaine de données dont ont été définies : cela va aussi bien de comment s'appelle ce jeu de données, au contenu et à sa sémantique. Mais les standards sont encore loin de recouvrir l'ensemble des champs sur lesquels l'open data permet d'ouvrir les données. Il n'y a pas des standards sur tout. Ce qui a été normalisé au niveau international, c'est la manière dont on décrit un jeu de données, encore appelé métadonnée. Cette description doit comprendre à minima un titre, une date de première publication et de mise à jour, une description.

“ Et la biodiversité ? ”

La circulation de l'information ne va pas forcément de soi. Une grande partie de l'open data en France, est basée sur le principe d'accès aux documents administratifs, acté par la loi dès 1978, qui a créé la commission d'accès aux documents administratifs, la CADA. Cette institution a 40 ans, comme la Commission informatique et liberté (Cnil). Cette levée du secret administratif, cette transparence, n'est pas encore partout et tout le temps, complètement acquise. Les responsables politiques peuvent s'inquiéter d'une perte de souveraineté, d'être jugés. Certains redoutent aussi que les données complexes soient mal interprétées. Que se passerait-il si les utilisateurs les employaient à mauvais escient volontairement ou involontairement ? L'ouverture et le partage des données publiques doit donc être en permanence encouragée. Certes, la loi pour une république numérique permet de le faire, mais beaucoup d'autres outils également. Le plus efficace encore est de réussir à faire prendre conscience à celui qui produit une donnée et qui va la faire circuler, qu'il bénéficie lui aussi des données que les autres ont ouvertes.

“ Quels sont les freins au développement de l’open data ? ”

La circulation de l’information ne va pas forcément de soi. Une grande partie de l’open data en France, est basée sur le principe d’accès aux documents administratifs, acté par la loi dès 1978, qui a créé la commission d’accès aux documents administratifs, la CADA. Cette institution a 40 ans, comme la Commission informatique et liberté (Cnil). Cette levée du secret administratif, cette transparence, n’est pas encore partout et tout le temps, complètement acquise. Les responsables politiques peuvent s’inquiéter d’une perte de souveraineté, d’être jugés. Certains redoutent aussi que les données complexes soient mal interprétées. Que se passerait-il si les utilisateurs les employaient à mauvais escient volontairement ou involontairement ? L’ouverture et le partage des données publiques doit donc être en permanence encouragée. Certes, la loi pour une république numérique permet de le faire, mais beaucoup d’autres outils également. Le plus efficace encore est de réussir à faire prendre conscience à celui qui produit une donnée et qui va la faire circuler, qu’il bénéficie lui aussi des données que les autres ont ouvertes.

“ Quels pourraient être les nouveaux champs de l’open data ? ”

La première extension du concept original de l’open data, c’est la donnée produite, coproduite ou améliorée par les utilisateurs, ce que l’on appelle le crowdsourcing. Ce qui est très marquant est la création de bases de données collaboratives à l’instar de OpenStreetMap ou Open food facts (1) qui ont popularisé l’idée que nous sommes tous producteurs de données.

C’est clairement dans le champ des sciences participatives que l’on a vu, les plus riches initiatives de production de données citoyennes. L’astronomie a beaucoup utilisé les observations citoyennes, par exemple. L’inventaire de la biodiversité est un autre domaine pour lequel la participation citoyenne s’est beaucoup développée.

(1) Cette base de données des produits alimentaires, répertorie les ingrédients, les allergènes, la composition nutritionnelle et toutes les informations présentes sur les étiquettes des aliments. Elle est alimentée par 9000 contributeurs de 200 pays qui utilisent leur smartphone pour scanner les codes-barres et envoyer des photos des produits et de leurs étiquettes. 600 000 produits y sont répertoriés.

Conclusion

« La communauté biodiversité a beaucoup avancé sur la question de l’interopérabilité »

QUELLE LICENCE ?

Pour définir ce que le public peut faire des données publiées, l'administration à l'origine des données peut prévoir une licence. Celle-ci fixe le cadre juridique qui s'applique aux données publiées. Ce qu'il est possible d'en faire avec dépend de la licence qui en régit l'utilisation.

Une licence libre est une licence s'appliquant à une œuvre de l'esprit par laquelle l'auteur concède tout ou partie des droits que lui confère le droit d'auteur, en laissant au minimum les possibilités de modification, de rediffusion et de réutilisation de l'œuvre dans des œuvres dérivées.

Dans le cadre de la politique du Gouvernement en faveur de l'ouverture des données publiques, Etalab a conçu la « Licence Ouverte / Open Licence ». Cette licence, élaborée en concertation avec l'ensemble des acteurs concernés, facilite et encourage la réutilisation des données publiques mises à disposition gratuitement.

S'il s'agit d'une licence ouverte, les données peuvent alors être :

- Réutilisées : par exemple copier les données, les mettre en forme dans un tableau, ou sur un fond de carte ;
- Redistribuées : par exemple publier les données sur un blog, dans un livre, ou encore sur Twitter ;
- Adaptées: par exemple changer le titre d'une colonne dans un fichier CSV pour en faciliter la compréhension ;
- Exploitées commercialement : par exemple faire une carte à partir de données ouvertes, l'imprimer, et la vendre.

Sous réserve de :

- Mentionner la source des données réutilisées ;
- Inclure la date de dernière mise à jour des données ».

Les licences les plus utilisées en France pour les portails OpenData sont :

- La Licence ouverte conçue par Etalab, dans le cadre de la politique du Gouvernement en faveur de l'ouverture des données publiques. Elle a été élaborée en concertation avec l'ensemble des acteurs concernés.

- La licence ODbL est ouverte également mais inclut une clause de réciprocité, si les données publiées répondent à un objectif d'intérêt général. En clair, les utilisateurs des bases de données sous licence ODbL doivent offrir à leur tour ce qu'ils en ont produit.

Il s'agit d'une obligation de partage selon les mêmes conditions. « L'avantage de la licence ODbL, est d'inciter à la création d'un cercle vertueux de la réutilisation et de la participation, souligne Simon Chignard, conseiller stratégique pour Etalab et auteur d'Open data, comprendre l'ouverture des données publiques (Editions fyp).

C'est par ailleurs un standard international reconnu.

- Les licences CC-by (Creative Commons):

La loi pour une République numérique a prévu la création d'une liste de licences qui peuvent être utilisées par les administrations pour la réutilisation à titre gratuit de leurs informations publiques (<https://www.data.gouv.fr/fr/licences>).

Il faut choisir entre 3 licences, seulement 3, c'est très peu finalement. Donc si, il y a consensus.

La licence Etalab assure une conformité avec les institutions françaises, la licence ODbL est un standard international reconnu, et les licences CC ont l'avantage d'être les plus connues de tous.

QUELS OUTILS ?



Quels outils ?

Si vous êtes un organisme public de type Parc National ou régional, la plateforme GeoNature est particulièrement adaptée à vos besoins.

GeoNature est un ensemble d'applications web et mobiles pour saisir, gérer et diffuser des données de biodiversité (Faune/Flore). Plus précisément, GeoNature permet de déployer un système d'informations complet pour la gestion des données Faune/Flore d'une structure allant de :

- La gestion des référentiels (taxonomiques et utilisateurs),
- À la saisie web et mobile dans différents protocoles,
- À la gestion de leurs métadonnées,
- À l'intégration de données de partenaires,
- À l'export des données selon les formats attendus par chaque partenaire,
- À la synthèse des données des différents protocoles sous forme de DEE (Donnéesélémentaires d'échange)
- À la diffusion des données sur un portail lweb grand public.

Il permet aussi de structurer, agglomérer et diffuser des données de biodiversité provenant de différentes sources, différents protocoles.



ecoReleve

Professionnels, Natural Solutions met à votre disposition EcoRelevé, progiciel disponible en ligne pour saisir des données sur le terrain lors de vos études, pour les classer, les trier et les exporter ensuite. Un outil parfaitement adapté pour transmettre vos datas au SINP (système d'information sur le nature et les Paysages), comme le veut la loi sur la biodiversité.

ENTRETIEN AVEC FRANCK MICHEL, INGÉNIEUR DE RECHERCHE AU LABORATOIRE D'INFORMATIQUE, SIGNAUX, SYSTÈME DE SOPHIA-ANTIPOLIS (UNIVERSITÉ CÔTE D'AZUR, CNRS, INRIA).

“ Les défis de l'Opendata ? ”

Que l'on utilise des vocabulaires et des formats standards pour publier les données ouvertes. C'est un énorme challenge. On a besoin de ces outils pour pouvoir croiser des données issues de communautés différentes, et répondre à des questions comme par exemple: quel est l'impact du changement climatique sur telle espèce, dans telle zone géographique.

“ Comment y arriver ? ”

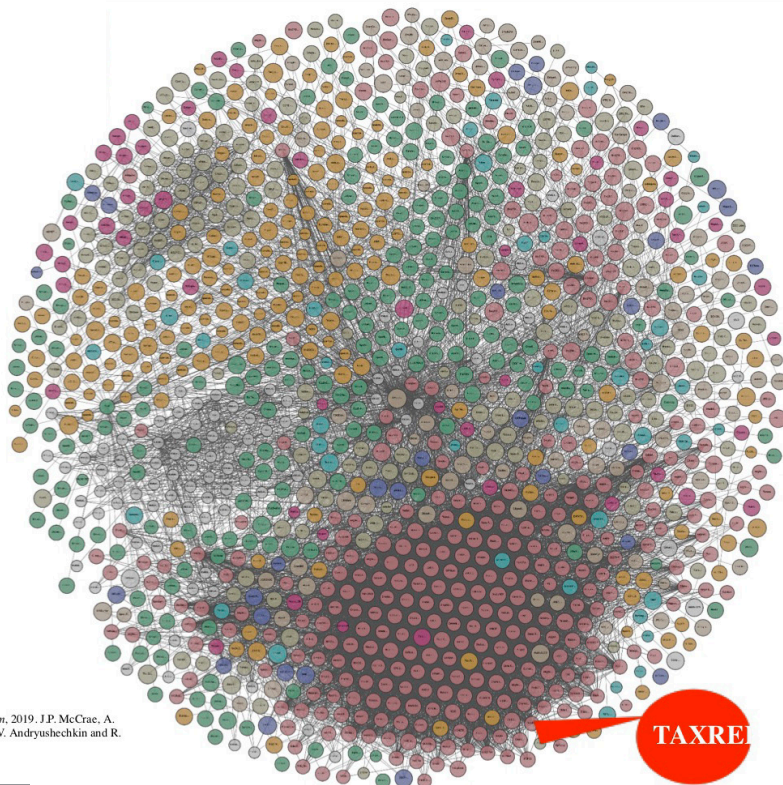
Il s'agit de représenter les données pour que les machines puissent les traiter et les agréger pour nous. Car cela devient impossible à l'échelle humaine tant les données sont nombreuses, volumineuses et de natures différentes. Et pour qu'elles puissent le faire, il est nécessaire de représenter de façon explicite la signification des données et des relations entre elles. Autrement dit, au lieu de décrire les choses dans des PDF ou des pages web, nous devons les décrire selon un format standard, manipulable par une machine, en utilisant des vocabulaires eux aussi standards et dont la signification soit comprise et partagée par tous les acteurs. De cette façon, on étend le web que nous connaissons, un web de documents destinés aux humains, avec un web de données destiné aux machines. Les principes de ce web de données ont été énoncés par l'inventeur du web, Tim Berners-Lee. Il propose d'attribuer des identifiants uniques à toutes les informations, les idées, les concepts que nous manipulons : une plante, une personnalité publique, ... Il a aussi proposé que l'on utilise le web comme infrastructure pour échanger les descriptions des ressources. Dans ce contexte, les identifiants uniques sont des URI http. Il suffit alors à un humain d'entrer un URI dans un navigateur pour recevoir une page web décrivant la ressource, pendant qu'une machine peut utiliser le même URI pour obtenir une représentation de la ressource dans un format standard qu'elle sait manipuler (RDF).

“ Vous travaillez avec le Muséum National d’Histoire Naturelle ...? ”

Cette institution gère et maintient à jour TAXREF, le référentiel taxonomique pour la France métropolitaine et l’outre-mer. Il recense environ 270 000 espèces et compte plus de 550 000 noms scientifiques. C’est un inventaire et une classification, par genre, par famille, par ordre etc., auquel s’ajoutent des informations sur les habitats, les observations, les programmes de conservation de la biodiversité. Nous travaillons à transformer TAXREF dans les standards du web de données, le RDF, en donnant notamment des URI à chaque espèce. Notre objectif est aussi de lier chaque URI d’espèce aux URI correspondants dans d’autres référentiels internationaux, afin de favoriser l’interopérabilité entre tous les jeux de données qui s’appuient sur ces référentiels.

“ Et à plus long terme? ”

L’enjeu, dans l’avenir, est aussi de pouvoir répondre à des questions qui exigent de collecter et requêter des données issues de champs scientifiques très différents, et qui traversent donc des champs multiples de la connaissance. Est-ce que tel gène de telle espèce a un impact sur la construction des écosystèmes où elle vit ? Aujourd’hui, on ne sait pas encore le faire. Car il s’agit de réconcilier des données représentant la complexité du vivant à des niveaux multiples (molécule, protéine, gène, cellule, phénotype, espèce, écosystème...). Les recherches actuelles visent à créer les outils nécessaires pour explorer cette transversalité.



Linking Open Data cloud diagram, 2019. J.P. McCrae, A. Abele, P. Buitelaar, A. Jentsch, V. Andrushechkin and R. Cyganiak. <http://lod-cloud.net/>

Legende

- Domaine croisé
- Geographie
- Gouvernement
- Sciences de la vie
- Linguistique
- Média
- Publications
- Réseaux sociaux
- Généré par l'utilisateur

TÉLÉVERSEMENT

Prenons l'exemple de maîtres d'ouvrages ou bureaux d'études qui doivent réaliser une étude d'impact dans le cadre d'un projet de construction sur un espace naturel. Cette étude demande de réaliser des observations (généralement de terrain) sur la biodiversité, c'est à dire de lister la présence (ou absence !) de l'ensemble des espèces de l'espace étudié, à une date précise.

Concrètement ?

Si le principe est simple, concrètement c'est un peu plus compliqué ! En effet, la plate-forme SINP n'accepte pas le versement de données brutes de terrain telles quelles.

Il faut que les données brutes soient transformées en un format précis et rigide pour que la plateforme SINP les accepte. C'est ce qu'on appelle le « standard occurrences des taxons » du SINP.

Sans entrer dans les détails, un fichier respectant le standard SINP contiendra une vingtaine de colonnes obligatoires (et beaucoup d'autres optionnelles) pour chaque observation. Ces informations obligatoires portent principalement sur l'observation elle-même (identifiant, dates, ...), l'identification du spécimen (quelle espèce ?) et sa localisation géographique.

Et en plus de ce fichier respectant le standard SINP, il faudra fournir d'autres informations sur le projet et le jeu de données en se connectant à d'autres plateformes web.

Pour résumer, il y a 5 étapes à suivre obligatoirement :

Pour résumer, il y a 5 étapes à suivre obligatoirement :

1

Déclarer le dossier sur demarches-simplifiees.fr : le site demande de créer un login et de remplir une dizaine de champs (coordonnées, objet du dossier, ...). Normalement c'est l'étape la plus simple !

2

Se connecter à l'application Métadonnées de l'INPN (en créant un autre identifiant) et fournir des informations sur chaque jeu de données (cadre d'acquisition et jeu de données lui-même). Avec le remplissage de ces informations, le maître d'œuvre obtient l'identifiant du cadre d'acquisition, l'identifiant SINP et l'identifiant du jeu de données.

Ces informations seront indispensables car elles vont être demandées lors de l'étape suivante.

3

Vous allez devoir organiser vos données au standard SINPN en les mettant dans un fichier .CSV (voir paragraphe suivant pour les explications).

4

C'est l'étape du téléversement en lui-même, c'est donc là que vous allez fournir vos données formatées magnifiquement au standard SINP ! Pour cela, il faut se rendre sur la plate-forme web GINCO (Gestion d'Information Naturaliste Collaborative et Ouverte), utiliser les identifiants obtenus jusque maintenant et sélectionner le jeu de données renseigné à l'étape 2. Vous pourrez ensuite fournir votre fichier de données. Si aucune erreur n'est détectée (= fichier standard SINP parfait !), un certificat de conformité est délivré et un lien URL (vers vos données) est fourni.

5

Se reconnecter à demarche.simplifiee.fr et fournir le lien URL obtenu. Au niveau national, la procédure de téléversement va maintenant permettre d'avoir un point d'entrée unique des données brutes de biodiversité : le SINP reçoit les données brutes et les filtre sur la base de leur qualité (vérification de la conformité, cohérence et validité scientifique des observations). Le téléversement intervient sur les étapes de conformité et de cohérence. Si la qualité des données brutes est validée par le SINP, celles-ci sont transmises à l'INPN pour diffusion au grand public (Inventaire National du Patrimoine Naturel).

MODULE GEONATURE EXPORT

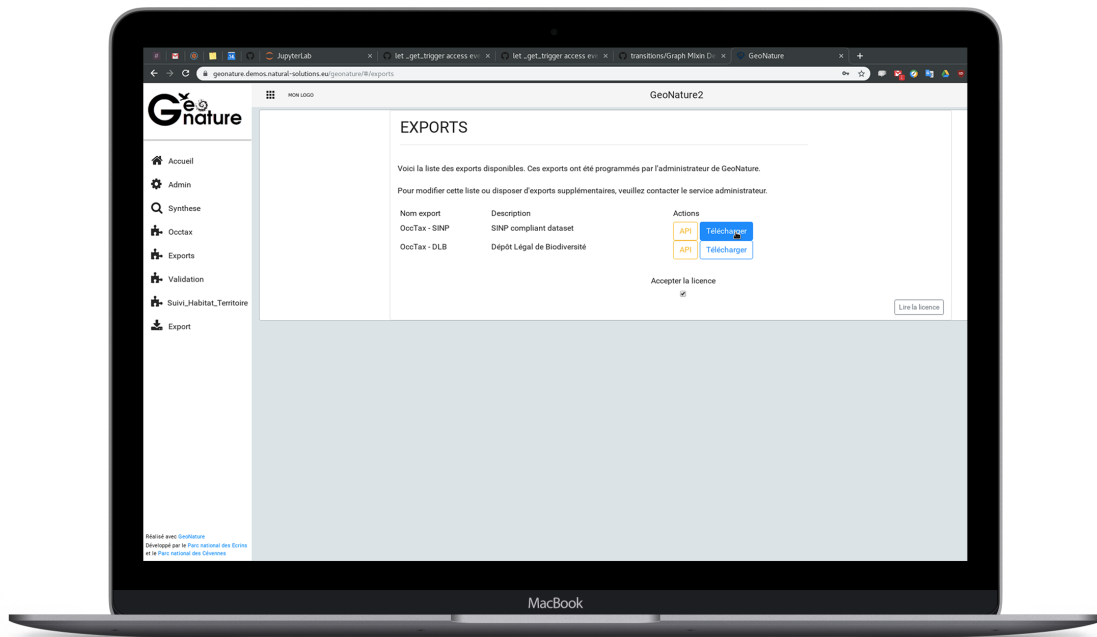


FIGURE 3 : CONSULTATION DE LA LICENCE

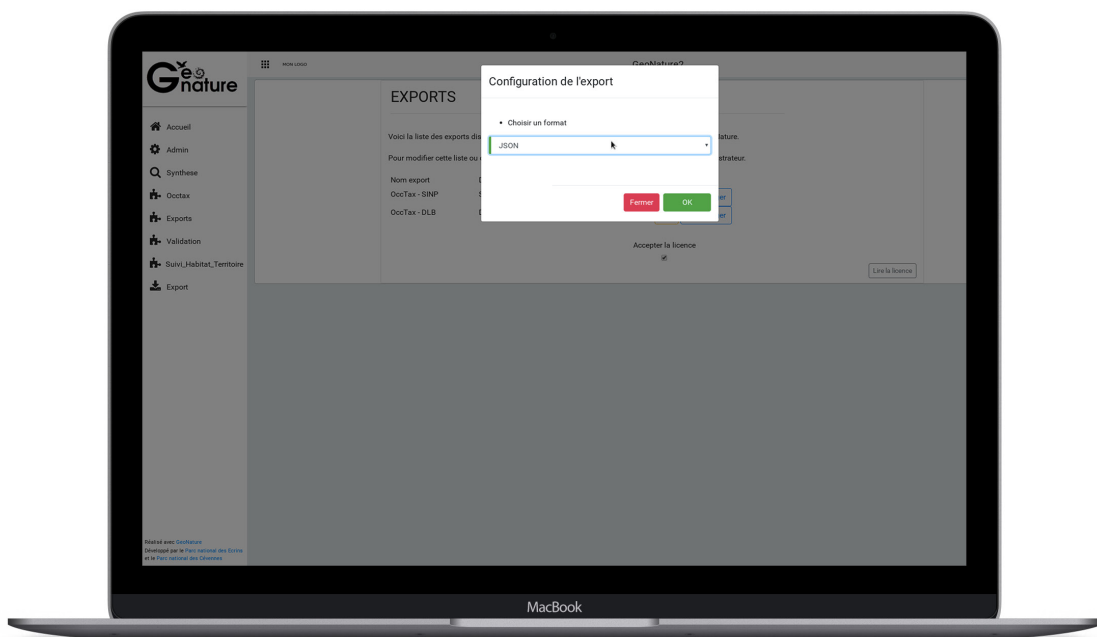


FIGURE 4 : CHOIX DU FORMAT D'EXPORT

MODULE GEONATURE EXPORT

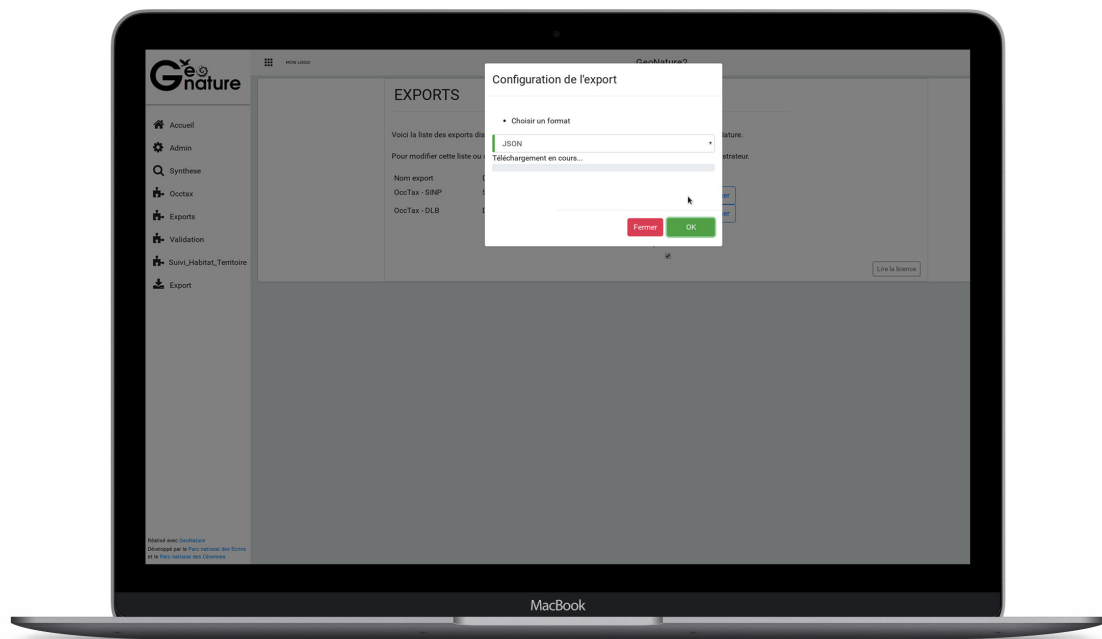


FIGURE 5 : TÉLÉCHARGEMENT DU FICHER

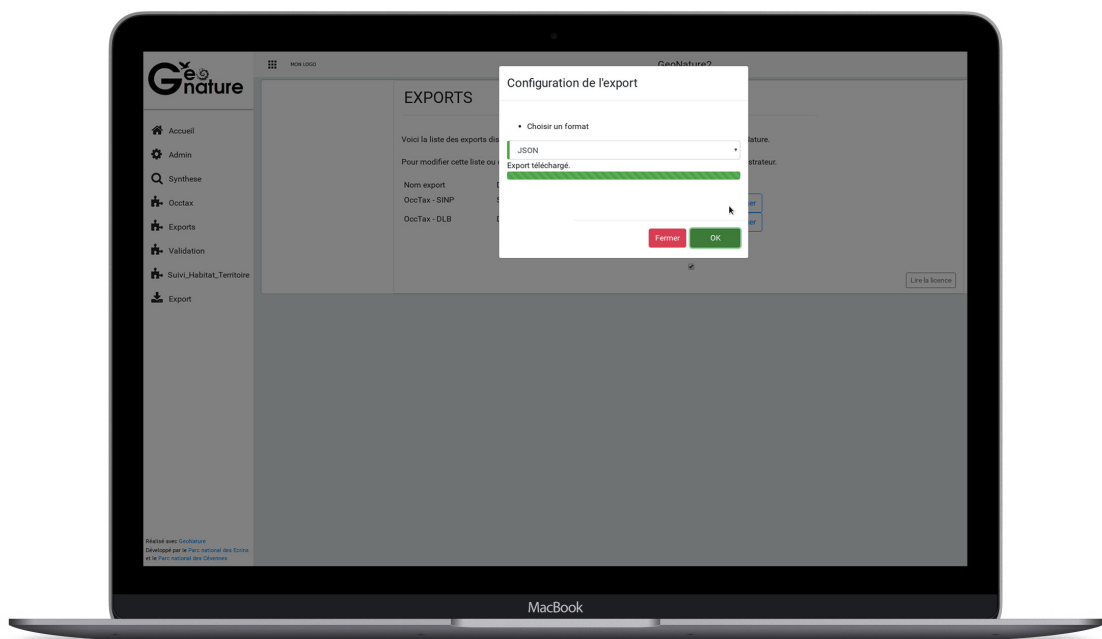


FIGURE 6 : TÉLÉCHARGEMENT EFFECTUÉ

MODULE GEONATURE EXPORT

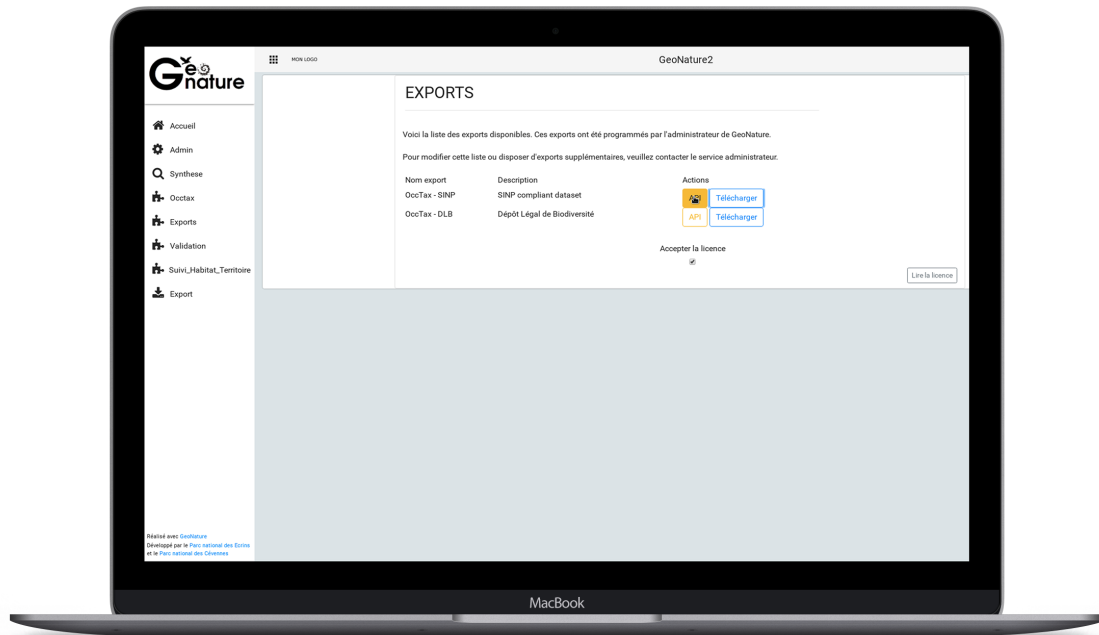


FIGURE 7 : CONSULTATION DE L'API

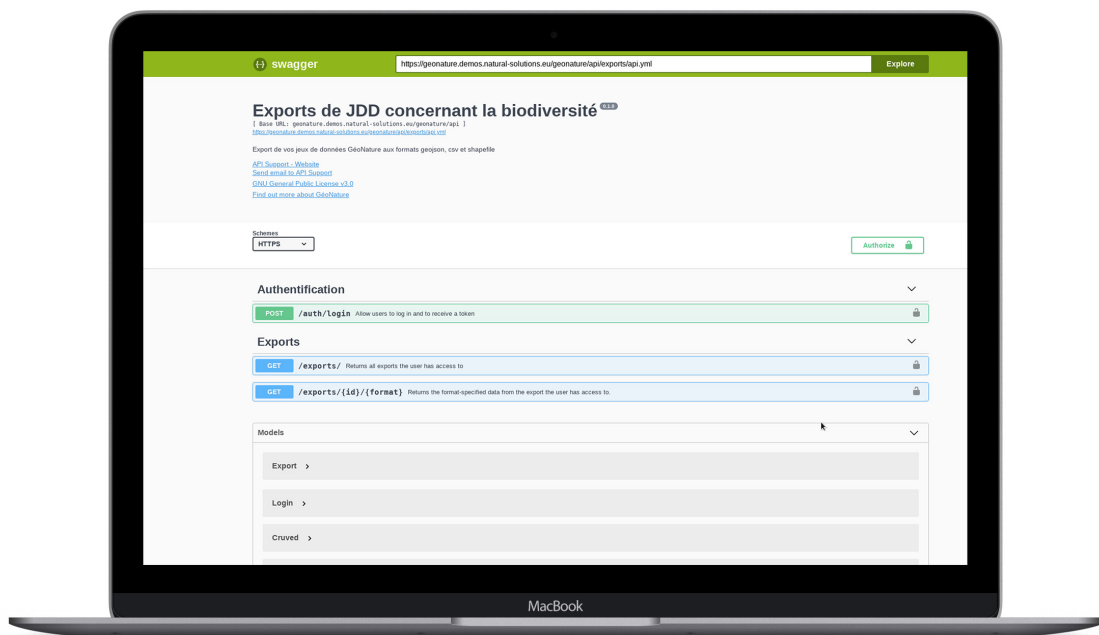


FIGURE 8 : DOCUMENTATION DE L'API

MODULE ECOREVELE

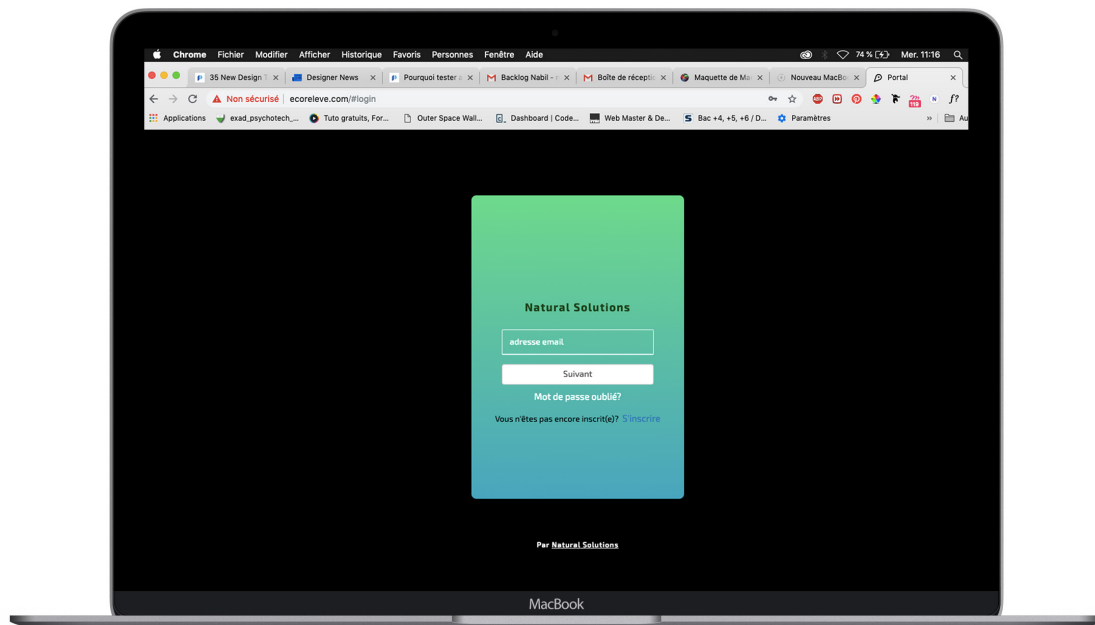
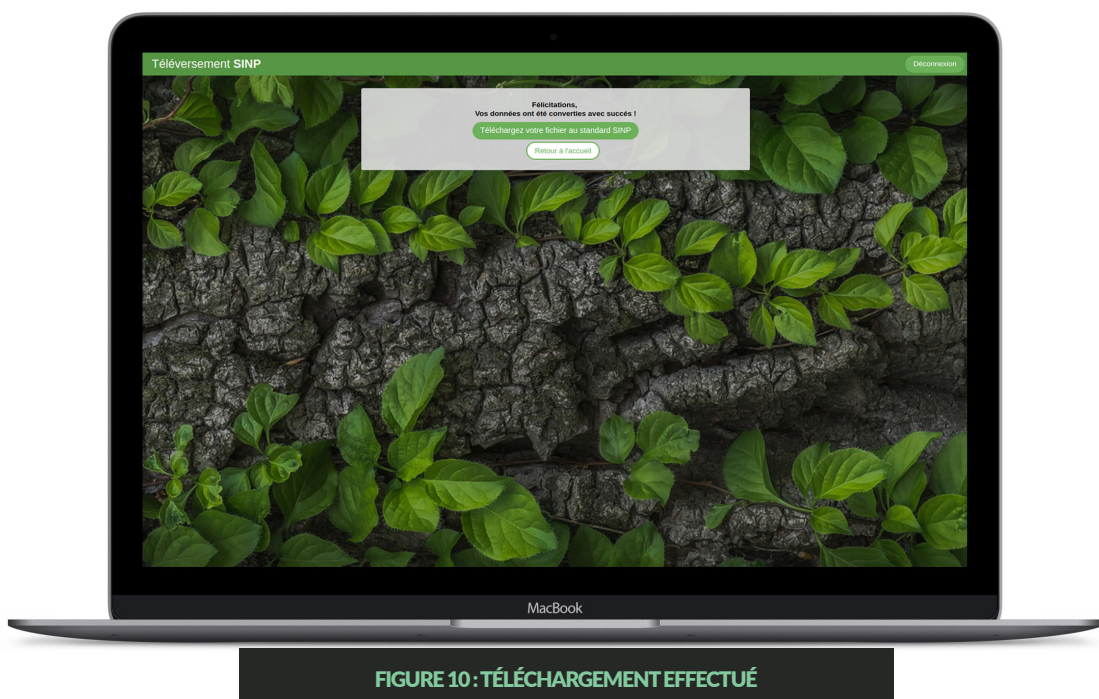
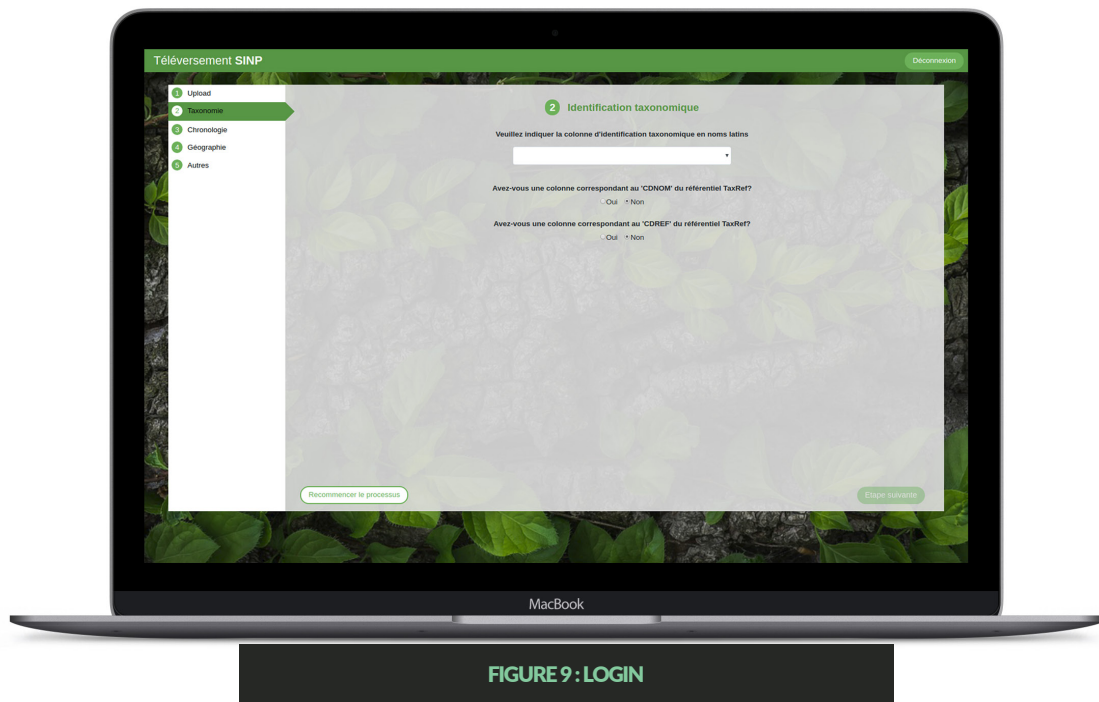


FIGURE 9 : LOGIN



FIGURE 10 : TÉLÉCHARGEMENT EFFECTUÉ

MODULE ECORELEVE



NOS PARTENAIRES



**AGENCE FRANÇAISE
POUR LA BIODIVERSITÉ**
MINISTÈRE DE L'ENVIRONNEMENT

