



Journées techniques SINP 20-21 mars 2019

Bilan de la plénière – jour 2

Introduction UMS PatriNat (AFB-CNRS-MNHN) (Laurent Poncet)

Accueil et remerciement des participants pour leur présence. Les journées techniques est un moment d'échange important. Un nouveau format est proposé pour cette deuxième édition des journées techniques du SINP : une table ronde sur les métadonnées, organisée la veille. Rappel de l'existence de la Lettre d'information nationale sur les activités du SINP (LIANS)¹, de la tenue de la réunion des correspondants SINP régionaux organisée en janvier 2019, et des échanges bilatéraux avec l'UMS PatriNat. Au niveau des participants présents aujourd'hui, il y a une grande et bonne diversité de structure. Un des objectifs est également d'avoir des temps d'échange important les temps d'échange. Le dossier participant qui vous a été remis contient un bilan des tables rondes de la veille, un bilan des actions nationales de 2018 et des perspectives 2019.

Contexte national

Politique générale sur la donnée et la connaissance – MTES/DEB (Jean-Jacques Richard)

Rappel sur le SINP : C'est d'abord un système qui apportait des données à l'observatoire national de la biodiversité (ONB). L'ONB fournit des indicateurs pour les décideurs permettant d'orienter les politiques. Pour avoir de bons indicateurs, il faut de bonnes données. Pour avoir de bonnes données, il faut de bonnes métadonnées. La métadonnée, c'est important, notamment dans le contexte de l'augmentation du nombre de données.

Il y a une volonté forte d'ouverture des données ces dernières années, pour la réutiliser et pour l'améliorer de manière intrinsèque. La donnée doit se partager à travers les systèmes en place. Les nouvelles lois vont dans ce sens (Loi Lemaire, RGPD...). Partagez vos données, on en a besoin. On a besoin qu'elle soit de qualité, qu'elle soit cataloguée, on a besoin des métadonnées.

La recherche, l'expertise et les données à l'AFB - AFB/DRED (Xavier Gayte)

Présentation : JTSINP19_01_DRED_Gayte.pdf

Présentation de la Direction pour la recherche, l'expertise et les données (DRED) : la DRED est centrée sur la connaissance au sens large, de la recherche et développement jusqu'à l'expertise et la valorisation de la donnée. Fait le lien entre les données de la recherche et les données des gestionnaires. La DRED est une direction qui intrinsèquement a un partenariat fort avec d'autres structures (UMS PatriNat, pôle R&D, Pôle de coordination des CBN...), élément de convergence avec le SINP. Les entités de la DRED travaillent sur la RDI, les systèmes d'information et l'expertise. Les systèmes d'observation sont au cœur de la DRED.

Les SI fédérateur : l'AFB a la coordination technique des 3 SI fédérateur sur l'eau, la biodiversité et les milieux marins. Ces systèmes se centrent sur le modèle DPSIR (état, pression réponse). Chacun des

¹ Inscription à la LIANS : https://frama.link/inscription_lians



systèmes ont un schéma national de données. Les systèmes d'information métiers alimentent les systèmes d'information métier. Mise en place d'un SI pression – inter-milieu.

OFB : l'Agence française pour la biodiversité et l'Office national de la chasse et de la faune sauvage vont, 1^{er} janvier 2020, former un nouvel établissement, l'Office français pour la biodiversité.

ARB : Les régions et l'AFB peuvent créer des Agences régionales de la biodiversité, pour décliner les politiques publiques biodiversité au sein d'une région. Les ARB peuvent se doter d'un certain nombre de missions, ceux de l'AFB sauf la police. Un principe simple = rien n'est obligatoire et missions à la carte - chaque couple région – ARB décide des missions portées par l'ARB. Plusieurs ARB se positionnent aujourd'hui sur les Observatoires régionaux de la biodiversité (ORB). La question de leur position sur les SINP se pose aujourd'hui.

Le SINP et le système d'information de la biodiversité (SIB) – AFB/DSOD et UMS PatriNat (Laurent Poncet)

Présentation : [JTSINP19_02_SIB_SINP_Poncet.pdf](#)

Réflexion collective sur les 3 systèmes d'information fédérateur (SIF) pour assurer la cohérence de ces systèmes. Les SIE et SIMM ont également une composante biodiversité, d'où l'importance d'interfacer les 3 SIF. Au sein de l'AFB, co portage PatriNat/DSOD pour le SIB

Points clé du SIB : logique fédérative des SI existants, périmètre très large (DPSIR – le SINP est surtout sur la brique « état » de la biodiversité, cohérence et découplage des données. Organisation revisitée et responsabilisante autour de :

- SI métiers pilotés par un ministère autour d'une politique publique avec objectif de produire, gérer, transformer, utiliser, partager la donnée, avec une organisation propre.
- SI fédérateurs avec une approche englobante, assure la cohérence entre SI métiers, notion de service, mutualisation de plusieurs briques, décrite dans un schéma national

Les documents structurants : le schéma national des données biodiversité (SNDB) publiée dans un arrêté. Consultation large à venir. Il fixe le périmètre, la gouvernance... Schéma annexe des SI métier, publié au bulletin officiel, description plus fine du SI. L'ensemble de la démarche est cadré par une charte d'adhésion décrivant les engagements des parties autour de la contribution à ces SI.

Il y a actuellement 25 SI métiers identifiés dans le SIB, une grande partie de ces systèmes ne sont pas opérationnels. Le SINP est le plus aboutis et le plus complexe. Certains SI métiers sont identifiés dans d'autres SI fédérateurs – ex : données des politiques marines dans le SIMM. S'il y a des données biodiversité, elles doivent également être identifiées dans le SIB.

Continuité d'action autour du SINP.

Objectif d'améliorer la valorisation dans le cadre de l'ONB et de l'INPN

Restructuration de Nature France en réflexion.

Échanges

Question : Peut-on avoir plus de précision du le SI métier « Pression » inter-milieu ?

Réponse : réflexion en cours pour avoir un SI transversal, qui fait écho aux logiques du SINP (en tant que SI métier qui va alimenter les 3 SI fédérateurs). Cette organisation permettra d'avoir plus de « poids »,



avec pour objectif d'avoir par exemple des valorisation croisées état/pression, une carte des pressions attendues par le plan biodiversité....

Question : Le document annexe du SINP remet-il en cause le protocole de 2017 ? Les chartes régionales s'appuient sur le protocole national. Modifier le protocole va donc impacter les chartes régionales.

Réponse : On ne remet pas en cause le fait qu'il y ait des chartes régionales. Le protocole 2017 va évoluer pour prendre en compte les schémas (SNDB et Schéma SINP) et la charte.

Question : Les moyens humains et techniques pour faire tourner ces systèmes sont lourds, l'empreinte carbone est énorme. Aujourd'hui, des choses innovantes existent (Block-Chain, Peer-to-peer...) dans un contexte d'ouverture des données. Ces systèmes sont-ils ouverts à une modification du fond technique pour prendre en compte ces enjeux ?

Réponse : La consommation électrique des serveurs dans le monde est aujourd'hui effectivement énorme (2ème derrière l'industrie automobile). C'est un problème mondial. Le ministère appuie l'innovation dans toutes ses formes pour répondre à cela.

L'organisation proposée ici est plus favorable. Elle se fait dans une logique tournée vers la R&D (cf. DRED). D'autre part, si la gouvernance du SIB souhaite mettre en avant le travail sur la limitation de l'impact carbone du système, elle fera des recommandations pour tous les SI métiers.

La question de l'exemplarité est nécessaire ici. Le « driver » des réflexions actuelles n'est pas celle-là par contre. La DRED qui a un pôle autour de l'interopérabilité ne s'est pas encore posé ces questions, mais il serait intéressant de creuser cette question

Interfacer ses outils avec le SINP – outils et webservices autour des référentiels.

Métadonnées : évolutions et liens avec les référentiels (Frédéric Vest)

Présentation : [JTSINP19_03_métadonnées_Vest.pdf](#)

Question : Existe-t-il des ponts entre les travaux sur le référentiel organismes du SINP et les autres référentiels organismes ?

Réponse : On se base sur le référentiel SIRET de l'INSEE, mais pas de lien avec le référentiel des intervenants du SANDRE pour le moment.

Question : Dans le PNDB, on se focalise sur l'EML. Il y a pas mal de standards qui peuvent être convertis en EML. Au niveau national, est-ce que certaines personnes connaissent totalement l'EML ? Pour comprendre le standard, il y a un coût d'entrée fort. Très peu de personnes peuvent donc maîtriser cela. Comment peut-on faciliter l'appropriation des standards dans le SINP ?

Réponse : On a fait le même constat avec le standard occurrences, et le DarwinCore. L'EML est assez proche du standard de métadonnées du SINP. On a des spécificités sur le concept de « rôle ». Il y a également les normes ISO, mais qui sont très génériques. On intervient au TDWG. Un champ de progression est identifié sur le lien entre ces standards.



Taxref et Base de connaissance (Olivier Gargominy)

Présentation : [JTSINP19_04_TAXREF_BDC_Gargominy.pdf](#)

Question : les données de répartitions sont-elles dans la réponse des webservice ?

Réponse : il n'y a pas d'occurrence dans ces webservice (Le projet « requeteur » répondra à cela). Il y a par contre des statuts biologiques dans les territoires (dans TAXREF). Il y a également l'Atlas de la biodiversité départementale et des secteurs marins (ABDSM) dans la BDC statut. Sur ce programme, on travaille par groupe opérationnel. La couche de référence quand 100% du groupe est fait. L'année dernière, il y avait 2 ou 3 groupes, ce qui n'était pas suffisant pour l'ouvrir sur l'API. Pour la prochaine version de TAXREF, on devrait les publier.

Question : Y a-t-il une clé d'API ? Quelles sont les limites d'usage et modalité d'accès à cette API ?

Réponse : tout est accessible, il n'y a pas de clé d'API.

Question : les interactions entre espèces sont-elles téléchargeables en csv (pour du hors-ligne) ?

Réponse : non, tout est accessible à travers l'API. Il existe des convertisseurs JSON en CSV si besoin. Les seuls documents produits en CSV sont Taxref et la BDC Statut en csv.

Question : Les standards SINP en Linked open data, c'est pour quand ?

Réponse : Pas pour le moment

Question : Les groupes fonctionnels de Taxref (groupe INPN1 et INPN2) ont-ils vocation à évoluer et aller plus loin ?

Réponse : Ces groupes INPN vont jusqu'au groupe 3. On assure la pérennité de ces groupes (exemple, le groupe des reptiles). Il existe également les groupes opérationnels, qui sont des groupes de taxon créés selon le besoin (ex = poissons d'eau douce de Nouvelle-Calédonie) – travaux en cours là-dessus, en lien avec Taxref et la BDC.

Question : À quel pas de temps elles sont mis à jour les statuts de menace et de protection des taxons? Quelles sont les infos disponibles ? Informations sur la rareté ?

Réponse : On couvre toutes les listes rouges, de mondiale à régionale, en fonction de leur mise à jour propre.

Référentiel des organismes du SINP (Rémy Jomier)

Présentation : [JTSINP19_05_OrgRef_Jomier.pdf](#)

Question : est-il prévu de faire un lien avec la base des espaces protégés (ex : quel organisme gère quel espace)?

Réponse : oui, ce référentiel a vocation à être le référentiel des organismes pour le SIB. Ce n'est pas le cas actuellement, mais c'est la perspective.

Question : comment sont gérées les évolutions des zones géographiques des organismes ?

Réponse : la zone géographique est modifiée. Par contre, le référentiel ne garde pas l'historique sur ce point.



Question : comment vous allez gérer les fusions d'organismes, l'historique sera géré ?

Réponse : création de la fiche du nouvel organisme et gel des anciens. Par contre, pas de lien prévu entre les anciens organismes fusionnés et le nouvel organisme. C'est un choix délibéré face à la complexité de gestion de ces problématiques. NB : Si vous êtes un organisme avec un code SIRET, le lien est fait par l'INSEE. Une API est proposée par l'INSEE pour faire le lien historique entre tous ces codes SIRET.

Question : avez-vous fait le lien avec le référentiel des organisations administratives de l'État produit par ETALAB en RDF ?

Réponse : oui, il est prévu de discuter avec le SANDRE et Etalab pour faire des correspondances. Le Sandre gère un référentiel des interlocuteurs avec plus de 160 000 interlocuteurs.

Question : est-ce que vous prévoyez sur un référentiel des personnes pour gérer proprement les données?

Réponse : ce n'est pas à l'ordre du jour au regard de la complexité du montage et de la gestion d'un tel référentiel.

Question : l'accès facilité au référentiel IGN par les associations n'est pas mis en œuvre aujourd'hui alors que c'est une contrepartie importante de la participation au SINP. On n'arrive pas avoir de réponse opérationnelle.

Réponse : Il y a une restructuration des engagements de l'IGN en cours de discussion portée par la DEB. Une question centrale est le nombre de structures concernées par l'accès à ces données. L'annuaire est là pour cela, notamment à travers le remplissage du champ « adhérent ». On a aujourd'hui quasiment aucun retour de la part de chacune des régions.

Élargissement

Question : Un référentiel des individus est compliqué juridiquement et techniquement à mettre en place, mais il est fondamental pour la mise en qualité des données. Ça fait partie de la valorisation des utilisateurs, notamment des experts individuels. Je suis étonné que ça ne fasse pas partie des champs de travail imminent.

Réponse : La traçabilité reste un point majeur du SINP, ce n'est pas parce que ce n'est pas sous la forme d'un référentiel que ce n'est pas important. C'est un travail lourd qui n'est pas une priorité aujourd'hui, car ça n'apportera pas plus en visibilité. Cependant, si cela vous semble prioritaire, ce débat peut être ouvert. D'autre part, on propose des règles de bonnes pratiques et ce sujet peut être discuté dans le GT Standard.

Question : Il semble problématique qu'il n'y ait pas d'historisation sur les changements d'organismes et de territoire.

Réponse : il y a historisation, mais pas historisation des liens. C'est un choix délibéré qui a été pris avec le groupe de travail sur le sujet au vu des moyens à mettre en œuvre pour gérer cela.

Question : sur le RGPD, y a-t-il un accompagnement prévu là-dessus ?

Réponse : relance du GT juridique prévu dans ce cadre, piloté par la DEB. La CADA a produit un livret sur la mise en place du RGPD dans les structures. Il y a une première réflexion au niveau du ministère, côté DGALN également. Il faut avoir à l'esprit qu'il y aura des liens entre ces réflexions.



Question : y aura-t-il des évolutions de Taxvern, qui liste les noms vernaculaires, pour avoir un nom vernaculaire de référence ?

Réponse : C'est un projet de longue date sur lequel on n'a jamais réussi à prioriser par manque de temps. En fonction des différents réseaux d'acteur, ce n'est pas le même nom de référence. Il faudrait un cadre méthodologique pour trouver un consensus, ce qui n'est pas forcément simple.

Il faut cependant noter que le premier nom dans la liste des noms vernaculaires est très généralement celui qui fait référence. Il correspond au nom scientifique français (rapport 1-1 avec l'entité biologique).

Question : existe-t-il des exemples de BDC dans d'autres pays ?

Réponse : sur les interactions entre espèces, on est plutôt originaux. Pour les traits de vie, l'enjeu est d'arriver à un modèle de données standardisé avec les autres bases qui existent et d'identifier les services à proposer au niveau national.

Autour des données et de leur partage

Dépopbio : premier bilan sur le dépôt légal des données brutes de biodiversité (Judith Panijel – Mathias Laville)

Présentations : [JTSINP19_06a_Dépopbio_Panijel.pdf](#) et [JTSINP19_06b_Dépopbio_Laville.pdf](#)

Question : avez-vous vérifié si les points correspondants à la localisation des observations n'étaient pas les centroïdes des communes ?

Réponse : on a 316 000 données associées à un point, correspondants à 70 000 points uniques (même localisation). Il y a des chances que certains correspondent à des centroïdes de commune. Un test est prévu.

Question : retour d'expérience « administrateur SINP » - de nombreux bureaux d'étude nous contactent, car ils n'arrivent pas à déposer les fichiers sur DEPOBIO. Ce qui est dommage, c'est pour résoudre leur problème, ils suppriment tous les attributs facultatifs, et ils donnent juste la commune (car ils n'arrivent pas à véhiculer le WKT). Il y a une grosse perte d'information. N'est-il pas possible de bloquer l'import en obligeant la fourniture d'une géométrie sur le logiciel ? Est-ce qu'un peu plus de flexibilité ne serait pas possible pour permettre de partager ces informations ? Une solution serait de passer en amont par les plateformes techniques du SINP pour formater les données au bon format, ce qui implique un intermédiaire avant le dépôt sur DEPOBIO.

Réponse : On a volontairement été permissif sur le format de données d'import pour donner de la souplesse aux bureaux d'étude quant à l'obligation de dépôt dans des délais très contraints des études d'impacts. Nous souhaitons progressivement restreindre cette permissivité, être plus contraignant en termes de qualité des données.

Il est possible également d'utiliser l'outil GéoNature de DEPOBIO, qui permet aux bureaux d'étude et au MOA de faire la saisie des données. Le fichier de sortie est standardisé, avec toute la richesse de la donnée.

Question : plusieurs inquiétudes, notamment sur la génération de doublon. A priori, les bureaux d'étude récupèrent des données sur les plateformes régionales et les reversent dans DEPOBIO avec des échelles qui peuvent varier (doublons pas tout à fait équivalents, donc plus difficilement identifiables).



D'autre part, sur le niveau de restitution, si les données proviennent des plateformes régionales et sont déposées sur DEPOBIO à la précision maximale, c'est en incohérence avec un certain nombre de chartes régionales qui donnent accès aux données précises uniquement aux utilisateurs authentifiés.

Réponse : L'analyse est notamment là pour répondre à ces inquiétudes, notamment pour comprendre ce qui se passe exactement autour de ce projet, qui n'est pas un projet SINP. L'enjeu ici est de raccrocher ce projet à une logique SINP : s'il y a de nouvelles données, elles doivent intégrer le SINP. Il y a un intérêt de faire également une évaluation collective des données qui sont utilisées dans le cadre des décisions publiques.

D'autre part, les identifiants SINP ne sont effectivement pas utilisés, il y a donc une problématique de pratique. L'analyse doit faire ressortir cela : identifier les doublons, pourquoi n'y a-t-il pas utilisation des identifiants SINP ?

Il y a besoin également d'accompagnement, car l'identifiant SINP n'est pas obligatoire (pour les nouvelles observations). De plus, nombreux de ces acteurs ne sont pas des partenaires du quotidien – il faut les amener sur des problématiques SINP, ce qui n'est pas acquis. Le contrôle des fichiers doit se faire en lien avec les services instructeurs.

Sur la précision des données : toutes les données brutes utilisées dans une étude doivent être déposées en l'état sur la plateforme DEPOBIO.

Question : Quel est niveau de la précision géographique de la donnée diffusée par DEPOBIO ?

Réponse : les données versées sont publiques et diffusées dans toute leur précision, hors sensibilité.

Question : Concernant le délai des 15 ans suite au dépôt des données, que se passe-t-il après ? Les données sont-elles supprimées ou conservées ?

Réponse : Ces données sont publiques et seront dans le SINP.

Question : un adhérent SINP qui met à disposition ses données en pensant qu'elles sont floutées pour une diffusion grand public à une maille de 10km va se retrouver à la voir diffusée précise sur DÉPOBIO, car elle est utilisée dans le cadre d'une étude d'impact. Il pourrait avoir un effet contre-productif – un adhérent ne souhaitera plus partager ses données au SINP de peur qu'elle soit diffusée précisément.

Réponse : c'est ce que dit la loi effectivement. Pour les données privées, il faut les informer.

Question : Ça nous met en défaut par rapport aux engagements que l'on prend sur la mise à disposition des données aux bureaux d'étude. Et il est hors de question d'arrêter de la mettre à disposition aux bureaux d'études, car ils en ont besoin pour faire leur étude d'impact.

Réponse : c'est une question collective intéressante. Lorsqu'il y a une loi biodiversité comme celle de 2016 qui impose de nouveaux dispositifs, ce n'est pas à la loi de s'adapter au SINP, mais bien l'inverse. Comment adapter intelligemment le SINP au nouveau dispositif réglementaire ? L'analyse des données est vraiment le support à cette réflexion que l'on doit porter ensemble.

Question : DEPOBIO a été ouvert en 2018. Cependant, dans l'analyse, si on regarde la répartition des occurrences par date d'observation, on voit que la donnée nouvellement produite est clairement

minoritaire. La majorité des données déposées sont donc certainement des données qui proviennent des autres plateformes.

Réponse : une étude est un processus long sur plusieurs mois ou plusieurs années. Ce n'est donc pas totalement en décalage. Mais l'analyse doit bien apporter son éclairage sur cet aspect.

Question : concernant l'accompagnement des acteurs, c'est souvent les bureaux d'études qui déposent les données. Les bureaux d'études sont très professionnels et savent manipuler de la donnée naturaliste, et donc la déposer, n'est-ce pas ?

Réponse : Il y a aujourd'hui une grande variété des bureaux d'études, des très gros qui sont capables et des très petits qui ont aujourd'hui du mal à mettre les données au format.

Standard de données : [Enquête, guides sur les données protocolées, standard occurrence habitat \(Rémy Jomier\)](#)

Présentation : [JTSINP19_07_Standards_Jomier.pdf](#)

Question : Quels sont les exemples d'attributs additionnels qui reviennent régulièrement ?

Réponse : Quelques exemples : le nom de l'étude, personne responsable de la saisie, organisme financeur, sexe de l'individu, habitat. Souvent, dans les attributs additionnels, nous avons des attributs qui correspondent à des concepts qui font partie du cœur du standard, mais qui suivent des nomenclatures différentes.

Question : beaucoup des attributs additionnels que l'on produit sont des informations liées à des protocoles (température, longueur de l'individu...). Est-il prévu d'avoir des extensions de format pour des groupes taxonomiques particuliers, par exemple les chiroptères ?

Réponse : cela rejoint la notion de profil (ex : profil pêche électrique, profil STOC...). L'enjeu n'est pas forcément d'ajouter des attributs, mais d'identifier comment utiliser le standard dans le cadre d'un protocole particulier. Pour les chiroptères, il y a une étude de préfiguration qui va être menée par le CESCO qui vise à voir comment utiliser les standards du SINP pour véhiculer de la donnée, notamment sur des protocoles de suivi de mortalité.

Remarque : on ne peut pas avoir un unique standard, avec un unique référentiel et une unique nomenclature qui vient répondre à tous les besoins de tout le monde sur tous les taxons. C'est peut-être plus intéressant d'avoir des standards adaptés à chacun des protocoles. Sinon, pour une occurrence simple, on se retrouve noyé dans 90 champs avec 800 possibilités à chaque fois. Vouloir ajouter des champs à chaque fois n'est pas forcément pertinent. Il vaut mieux parfois commencer un nouveau standard à côté.

Réponse : L'objectif de l'analyse est bien d'éviter de complexifier le standard de manière permanente. On va essayer de le maintenir à des niveaux compréhensibles. On a intégré dernièrement la notion de standard « minimaliste » pour faire de l'occurrence simple – avec un document qui fait 30% du document de base. Pour ce qui est des protocoles, l'objectif est d'avoir des profils / cas d'usage pour faciliter l'utilisation du standard dans ces cas particuliers. On ne va pas effectivement ajouter des champs pour des protocoles très particuliers représentant 1000 données.



Pour rappel, les standards du SINP sont là pour faire des échanges entre des outils informatiques. La complexité relative aux standards peut se comprendre.

Question : Effectivement, le standard est avant tout fait pour échanger des données entre plateformes. Ceci-dit, nous avons des producteurs vertueux qui nous demandent à ce que ces standards soient stables, car ils les intègrent dans leurs outils, nous également. Quelle est la stabilité du standard et qu'est-ce que l'on vise en termes d'évolution ?

Réponse : Pour l'occurrence de taxon, on est sur des évolutions sur environ 3 ans, avec une version mineure intercalaire éventuellement entre les 2 sachant que la version mineure est compatible avec la version précédente.

Question : Les standards sont bien là pour l'homogénéisation des données pour des outils informatiques. Ceci dit, si le standard peut être utilisé dès la production de données, ça évite de faire la transformation des données et le plus souvent une altération des données (perte de précision). Plus les standards sont utilisés proches du producteur, plus on aura des données fiables à analyser.

Réponse : Le standard correspond plus à un dictionnaire de données avec plusieurs formats d'échanges, qui peuvent être utilisées à différents niveaux. Plusieurs régions ont leur propre standard, qui est un format d'échange qui respecte le dictionnaire de données.

Data paper : [une publication scientifique pour décrire un jeu de données \(Sophie Pamerlon\)](#)

Présentation : [JTSINP19_08_DataPaper_Pamerlon.pdf](#)

Question : c'est quoi le top de l'impact factor d'un data paper ?

Réponse : c'est le data paper sur l'herbier du MNHN qui a été publié dans Nature. Sur gbif.org, il y a la liste des revues. Au début, les premières revues n'avaient pas un grand impact factor (pensoft, biodiversity data journal). Maintenant, beaucoup plus de revues avec de bon impact factor acceptent les data paper.

Question : le logiciel que l'on peut décrire dans un data paper doit-il être open source ?

Réponse : ce n'est pas obligatoire, mais c'est vrai que la mouvance va dans ce sens. À voir avec les comités de lecture.

Question : Pourrait-on inventer un « GINCO » du chercheur qui produit un data paper à chaque fois ?

Réponse : c'est certainement faisable d'un point de vue informatique. À L'INRA, il existe une plateforme qui permet de déposer des jeux de données et de créer un eml automatiquement donc c'est faisable. Cela dit, il faut des métadonnées bien renseignées pour publier un data paper. Un data paper demande un gros travail de rédaction, ce n'est pas du tout les métadonnées d'aujourd'hui qui pourrait uniquement alimenter ces data paper. Quand on fait un data paper, on rédige spécialement des métadonnées, notamment en anglais, il y a une démarche scientifique. Ce travail est intéressant sur des jeux de données de suivi, protocolées, de long terme...

Valorisation des données : [indicateurs ONB, atlas, rapportage, travaux de recherche... \(Julien Touroult\)](#)

Présentation : [JTSINP19_09_Vvalorisation_Touroult.pdf](#)

Pas de questions particulières

Dernier temps de discussion

Question : Qu'en est-il du domaine marin ?

Réponse : La diversité des acteurs dans le milieu marin et dans le milieu terrestre n'est pas la même sur ce champ. L'organisation des données et de la connaissance n'est pas forcément dans le cadre des politiques publiques « Inventaires », mais plutôt dans des politiques publiques « marines ». Les acteurs sont plus nationaux plutôt que régionaux. Les enjeux sont assez différents : on n'est pas à la commune, les standards ne sont pas les mêmes. C'est un sujet d'interface entre les 3 SI fédérateur, notamment le SIMM et le SIB. Malgré cela, le SINP est aussi marin, car il répond à l'inventaire du patrimoine naturel qui est terrestre et marin comme le dit la loi.

Question : Malgré les nombreuses vérifications, il existe toujours des erreurs sur l'INPN, le GBIF ou ailleurs. En tant que scientifique, je regrette qu'on ne puisse pas corriger, signaler ou masquer une erreur.

Réponse : Le travail de validation est un travail continu. Dans le SINP, il y a aujourd'hui une méthode de validation sur 3 niveaux : producteur, régional et national. Elle se fait de manière collective et documentée, pour avoir un consensus scientifique. Cela évite que cela ne soit pas le premier ou dernier qui parle qui ait raison. Elle ne se fait pas au compte-goutte.

Aujourd'hui, on a peu d'outils pour faire cette validation. On y travaille. On commence à spécifier un outil de validation national qui se greffera sur le projet requeteur.

Remarque : c'est important d'avoir un processus de validation qui soit continu. À l'échelle régionale et locale, certaines données peuvent paraître étonnantes. On risque de perdre la confiance de nos partenaires si quelqu'un vient invalider arbitrairement la donnée. Le cadre de la validation continu est nécessaire.

Remarque : 2 exemples d'erreur qui semblent pourtant évident : une donnée d'albatros nicheur dans le Parc national des Cévennes. Des espèces déterminantes ZNIEFF en Corse qui n'existaient qu'en Corse, avec pour ces noms d'espèces, des sources uniquement en Corse. Cela semble des erreurs de codage.

Réponse : Sur les données ZNIEFF (cadre du programme dédié) il y a un circuit régional prévu à cet effet. Il faut informer le responsable du programme en région qui administre les données directement dans un outil partagé.

Question : Est-ce que sur le site de l'INPN il y a un contact pour permettre de remonter des remarques sur certaines données ?

Réponse : il y a une rubrique contact sur l'INPN. Pour un partenariat sur un groupe entier, il faut voir directement avec l'équipe données qui mettra en place le processus adéquate.

Question : Sur le choix des contenus des présentations, c'était très intéressant, mais très national. Les journées techniques du SINP, c'est également l'occasion pour les acteurs régionaux de faire passer des messages, de faire des témoignages, d'échanger sur ce qui se fait un peu plus au niveau local. C'est une suggestion pour la suite.

Réponse : Oui, on a bien conscience de cela et on souhaite évoluer dans un format plus co-construit. Malheureusement, pour cette séance, on s'y est pris trop tard pour intégrer des interventions régionales. On intégrera cela pour les prochaines journées. Il est également envisageable que ces journées soient



organisées par une région en particulier. Si une région souhaite organiser la prochaine séance, n'hésitez pas à nous contacter.

Merci à tous pour ces échanges