

DÉBAT : LES DONNÉES NATURALISTES

Romain JULLIARD – CESCO/Muséum national d'Histoire naturelle, Philippe MAINTIGNEUX – Comité de pilotage de Faune-IDF, Laurent PONCET – Service du Patrimoine Naturel (SPN)/Muséum national d'Histoire naturelle, Pierre RIVALLIN – Comité de pilotage de Cettia, Sébastien FILOCHE – Conservatoire botanique national du Bassin parisien (CBNBP)

Animation : Maxime ZUCCA – Natureparif



MUSÉUM
NATIONAL D'HISTOIRE NATURELLE

natureparif

Agence régionale pour
la nature et la biodiversité

iledeFrance

Natureparif, en tant qu'observatoire de la biodiversité francilienne, a notamment pour mission d'animer le Système d'Information sur la Nature et les Paysages (SINP). Il a pour but de centraliser et diffuser les connaissances naturalistes, comprenant les données, pour les rendre accessibles au plus grand nombre. Chaque région organise la récolte et la circulation des données, et les fait remonter au niveau national, à la plate-forme de l'Inventaire National du Patrimoine Naturel (INPN) du Muséum national d'Histoire naturelle. Ce sont essentiellement les naturalistes bénévoles qui saisissent leurs données, mais aussi les professionnels. Les problématiques importantes liées à l'utilisation de ces données ont été soulevées de façon récurrente ces vingt dernières années, et opposent parfois les associations aux bureaux d'études.

Les données d'observations naturalistes permettent d'avoir une meilleure connaissance du territoire et d'orienter ainsi divers projets d'aménagement. Elles permettent également, par exemple, de produire des indicateurs afin d'évaluer l'état de santé de la biodiversité, ou sont encore utiles dans le cadre de programmes de recherche. En Île-de-France, beaucoup de structures naturalistes valorisent les données. Le Conservatoire botanique national du Bassin parisien recense les observations de la flore, dans la base de données Flora, qui appartient à la plate-forme du SINP, et d'autres structures naturalistes assurent l'animation du réseau, la collecte et la centralisation des données. En particulier, depuis 2010, une base de données en ligne, Faune Île-de-France, est animée par le Centre ornithologique Île-de-France (Corif) et la LPO (Ligue pour la Protection des Oiseaux). D'abord axée sur les oiseaux, elle s'ouvre progressivement aux autres taxons. En 2014, Natureparif a déployé, au niveau régional, la base de données lancée préliminairement en Seine-et-Marne par Thierry ROY et Seine-et-Marne-Environnement (SEME), Cettia, avec une philosophie reposant sur l'*open data* pour recueillir les données publiques et autant de données privées que possible, tous taxons confondus.

Quel outil de saisie choisir ?

En Europe, environ 195 outils de saisie en ligne d'observations naturalistes existent. Par exemple, les hollandais qui viennent en France en vacances saisissent leurs données de papillons sur un outil portugais, car c'est celui qu'ils ont l'habitude d'utiliser. Le plus employé dans le monde est eBird, une base de données sur les oiseaux, principalement employée en Amérique du Nord et du Sud, qui est en accès totalement libre et même téléchargeable depuis l'Europe. On peut éventuellement regretter cette profusion d'outils. (RJ)

Cettia et Faune Île-de-France, comme les autres outils en ligne, présentent une ergonomie qui rend obsolète la saisie dans un tableur Excel. L'enregistrement des données, la détection d'erreur ou la possibilité d'émettre des tableaux, sont très efficaces. Si les deux bases de données en ligne ont pour même objectif de fournir un carnet de notes aux naturalistes, elles présentent cependant quelques différences au-delà de l'interface de saisie elle-même. Dans Cettia, le référentiel Taxref du Muséum national d'Histoire naturelle est inclus : ainsi, tous les taxons peuvent être saisis, ce qui n'est pas le cas dans Faune Île-de-France. Faune Île-de-France est issue de deux associations ornithologiques : si elle s'ouvre petit à petit aux autres taxons, sur 1 700 000 données, 98,8 % concernent des oiseaux. Les autres taxons sont assez marginaux donc il y a, à ce niveau, pour l'instant peu de concurrence entre les deux bases. Les deux associations animatrices de Faune Île-de-France s'attachent à maintenir leurs prises de position indépendantes, et leur place en tant qu'associations dans le débat public. Leur ambition est de faire valoir la connaissance associative sur la nature. Le comité de pilotage de Cettia est très pluriel, incluant des associations de protection de la nature, des collectivités et des gestionnaires d'espaces protégés. Faune Île-de-France s'inscrit dans un réseau français et européen (système visionature) ce qui est un atout certain, quand Cettia est pour l'instant uniquement francilienne. Ce dernier point permet cependant d'être plus réactif et de mieux maîtriser le développement de l'outil, en répondant directement à la demande des observateurs du territoire. (PM et PR) Il est à signaler que le CBNBP souhaite également doter Flora d'un outil de saisie en ligne dédié, car les périmètres régionaux actuels des outils existants ne correspondent pas à la totalité du territoire d'action du CBNBP. (SF)

Chacun est libre d'utiliser l'un ou l'autre de ces outils, selon ses besoins. L'important est de lire la charte des deux bases de données et de voir ce qui correspond le mieux à ses propres attentes, notamment concernant la transmission

des données et leur analyse. (PR) Une différence importante entre les deux bases concerne la politique de diffusion des données : Quel accès aux données par les utilisateurs ? Les données sont-elles transmises de manière précise ou floutée aux autres acteurs de la biodiversité ? Est-il possible de travailler librement à partir de l'ensemble des données et essayer de les faire parler, tout en ne perdant pas de vue l'objectif final, à savoir que l'utilisation de ces données doit permettre d'agir en faveur de la biodiversité ? Les données sont-elles utilisées à des fins de conservation ? (LP, RJ)

La diffusion des données

Au niveau national comme régional, il n'y a aucune ambiguïté sur le fait que les données publiques (produites par des structures publiques ou dans le cadre d'études rendues obligatoires par la réglementation) doivent être diffusées pour une utilisation multiple. Concernant les données privées, cela dépend du producteur de la donnée, et un cadre d'échange et de partenariat doit être établi. Dans le cas du SINP, une discussion est actuellement en cours sur les conditions de la diffusion, qui tendent vers de l'*open data* pour toutes les données publiques mais aussi pour les données privées associées à une autorisation d'usage. (LP)

Cettia et Faune Île-de-France permettent toutes deux une consultation précise des données par l'ensemble des utilisateurs inscrits. Flora permet une consultation à la commune ou à la maille. Il est par ailleurs important de différencier l'extraction de la simple consultation de données, qui ne présentent pas forcément les mêmes enjeux. Dans le cas où un organisme demande une extraction et réalise une étude, il paraît légitime qu'il y ait un retour de cette étude en libre accès sur la base ayant fourni les données.

Le comité de pilotage de Faune Île-de-France a une politique de non diffusion de la donnée brute, et assume sa réticence vis-à-vis des bureaux d'études. Le cas d'une étude n'ayant révélé la présence que d'espèces banales sur le site d'étude est cité comme l'exemple d'une démarche peu éthique et d'un manque de prospection permettant de favoriser un projet d'aménagement alors que des espèces à enjeux y sont présentes. L'association ne veut pas se rendre complice de ce genre de démarche. Au lieu d'un envoi de données brutes dont on ne sait pas quelle conclusion en sera tirée, c'est donc une synthèse réalisée à partir des données de l'association qui est proposée. On peut regretter le coût, parfois non négligeable, des données fournies par les associations, y compris pour des acteurs de la conservation de la nature. Pour Faune Île-de-France, effectivement, fournir les synthèses de données a un coût, mais le maintien des bases et donc de l'activité des associations aussi. Si le travail que génère l'exploitation des données n'est pas dénué d'impacts économiques, l'approche du comité de pilotage est aujourd'hui plus militante : il ne s'agit pas d'accepter uniquement un échange financier, puisque les échanges font souvent appel à d'autres conditions. Le comité de pilotage de Faune Île-de-France a pour objectif de suivre un protocole qui se rapproche d'autres institutions scientifiques, en demandant notamment la motivation de la demande et son enjeu, et il se prononce sur la valorisation des données, en proposant de participer à l'étude en question ou de la publier, et, dans tous les cas, d'être associé au projet de façon enrichissante. Récemment, des échanges de données brutes ont cependant eu lieu avec certaines structures associatives. Les demandes doivent toutefois être motivées, correspondre aux objectifs des associations pilotes, et une garantie de sérieux doit être apportée, autant de précautions qui sont indispensables pour lutter contre la perte de contrôle sur l'utilisation des données. C'est un enjeu très important pour les associations comme le Corif, qui cherche à construire sa propre connaissance et un réel savoir associatif. (PM)

Concernant Cettia, si une requête d'extraction est faite par une structure sur son territoire d'étude, comme une association régionale qui veut faire une synthèse, un animateur de site Natura 2000 ou un bureau d'études dans la mesure où ils justifient l'étude qu'ils doivent réaliser, une demande est faite au comité de pilotage, dans lequel chaque structure membre a la possibilité de voter dans un délai de 10 jours. Si la majorité a répondu positivement, la demande est satisfaite, ce qui a, jusqu'ici, toujours été le cas. En contrepartie, il est demandé que les données produites par la structure en question soient saisies ou versées dans Cettia. Il est considéré que rendre accessible les données brutes, en particulier aux bureaux d'études, permet justement de s'assurer que les enjeux d'un site seront mieux pris en compte avec le maximum de connaissances, et que le bureau d'études, en présence d'information bibliographique riche, sera obligé d'apporter une connaissance supplémentaire à l'existant, par exemple en recherchant d'autres taxons, une preuve de reproduction, ou un statut particulier. (PR)

Le comité de pilotage de Faune Île-de-France s'est organisé pour publier, tous les six mois, la liste des utilisations qui ont été faites des données des observateurs, et un outil spécial sera publié pour que chaque utilisateur puisse avoir accès à cette information. C'est également le cas sous Cettia, l'ensemble des extractions effectuées durant l'année 2015 ayant été publié en fin d'année sous forme de « nouvelle ». L'idéal à long terme serait d'éditer une synthèse pour chaque contributeur qui l'informerait de la part de ses observations qui a été utilisée et dans quel objectif, comme l'apport de connaissances sur un site Natura 2000 ou la mise en protection d'un site. (PM et PR)

Des inquiétudes vis-à-vis de l'*open data* subsistent car on s'interroge sur sa capacité à réduire l'influence des associations de terrain dans les différents débats. Cependant, le fait d'ouvrir totalement l'accès aux données permettrait

aux associations d'être consultées pour interpréter ces informations et d'être beaucoup plus critiques en termes d'aménagement : c'est donc leur donner un rôle plus important. Par rapport aux prises de décision, l'*open data* a l'objectif très clair, intrinsèque, de rendre accessible à tout acteur tous les éléments, afin de pouvoir contester une décision publique ou privée beaucoup plus simplement. *(LP)*

Le cas particulier des espèces sensibles

Il peut paraître logique et normal de ne pas diffuser les données liées à une espèce sensible, mais actuellement, on détruit plus par méconnaissance que par diffusion de l'information. Le problème du dérangement d'espèce reste un sujet important, qui nécessite d'être traité à part entière, presque au cas par cas, avec des enjeux spécifiques à décrire, d'où l'importance d'intégrer au mieux les connaissances sur la biodiversité dans les discussions pour mener aux bonnes décisions. Se posent légitimement des questions éthiques sur le traitement des espèces rares et menacées, et sur l'accessibilité des données les concernant : il peut arriver d'assister à des affluences de photographes ou d'observateurs peu soucieux du dérangement occasionné sur des sites sensibles, car l'observation d'une espèce rare y a été précisément localisée. Certaines espèces présentent un enjeu uniquement lors de certaines étapes de leur cycle de vie, ou sur certaines localités. En raison de la complexité du sujet, un guide méthodologique national de production de liste de sensibilité régionale est en cours de réalisation dans le cadre du SINP. Ce dernier fournira un cadre général dans l'établissement de cette liste, basé sur des critères internationaux. Il reviendra ensuite aux régions de s'approprier ce guide et de produire ces listes, au regard de la sensibilité des espèces. *(LP)*

Dans la base de données Cettia, l'observateur est laissé comme juge de la sensibilité de sa donnée : il peut choisir de cacher ou non une observation s'il considère que sa communication risque de porter davantage atteinte à l'individu ou à la population concernée, que sa mise sous silence. Il n'existe pas de liste d'espèces automatiquement masquées à la saisie, mais les observateurs sont par exemple invités par le comité de pilotage à masquer leurs données de gîtes de chiroptères si cela n'a pas été fait lors de la saisie. Un observateur peut également considérer un site privé comme « sensible » et masquer les données s'y rapportant, souvent à la demande du propriétaire. *(PR)*

Dans la base Faune Île-de-France, cinq espèces – un mammifère et quatre oiseaux – sont masqués systématiquement, à la période de l'année qui peut s'avérer critique. Les données les concernant sont utilisées dans des cartes de synthèse à large échelle. De plus, chaque utilisateur peut choisir de cacher certaines de ses données et le comité de pilotage également. *(PM)*

Quelle interopérabilité entre les bases franciliennes ?

Devant la multiplicité des bases de données, certains observateurs souhaitant bien faire les choses sont tentés de saisir leurs observations dans plusieurs outils. L'existence d'échange entre les bases est évidemment souhaitable. Plusieurs types de passerelles existent.

La possibilité d'importer des données saisies dans d'autres systèmes fonctionne dans Cettia, mais implique une transformation manuelle. De nombreuses données provenant d'autres acteurs franciliens ou nationaux détenteurs de bases de données sont en cours d'import ou le seront prochainement. Inversement, les données de Cettia peuvent tout à fait être importées dans Faune Île-de-France à terme s'il existe une réciprocité. *(PR)* Les concepteurs de Biolovision travaillent également sur la mise au format pour injecter directement dans Faune Île-de-France de gros jeux de données. Cependant, les administrateurs de Faune Île-de-France ne souhaitent pas que leur outil devienne un tuyau à données. *(PM)* Pour la flore, des passerelles vont prochainement exister entre Cettia et Flora, cette dernière ne disposant pas, à l'heure actuelle, de dispositif de saisie en ligne. Des échanges réguliers seront effectués d'une base à l'autre, mais pas en temps réel. *(SF)*

Une des solutions est d'avoir un outil de restitution au niveau national qui puisse être complet et ouvert, régulièrement alimenté par le lot de données provenant de tout le pays : c'est le rôle de l'Inventaire National du Patrimoine Naturel (INPN) hébergé par le Muséum national d'Histoire naturelle. *(LP)* L'outil utilisé pour la saisie importe finalement peu, la vraie problématique résidant dans l'outil de consultation des données : pourra-t-on voir toutes les données dans une même base, telle que celle de l'INPN, et en temps réel ou presque ? *(SF)* L'important est bien sûr que les données soient intégrées à l'échelle nationale, mais aussi qu'elles reviennent aux régions, car les actions doivent être mises en place localement. La mise en place d'un identifiant unique (et pérenne !) de la donnée facilitera le passage d'une base à l'autre en minimisant les doublons. *(LP)*

Idéalement, il serait très pratique de concevoir deux types de bases pour la faune et la flore, avec une saisie nationale type Faune France et Flore France, contenant des passerelles entre les régions, et à terme, pourquoi pas, entre les pays. Cependant, la création d'un outil commun national semble malheureusement idéaliste, car actuellement, chaque région

présente des associations qui ont leur propre système, leur propre vision de l'utilisation qui est faite des données – sujet encore très clivant – et qui ont un besoin d'existence lié à la donnée. *(PR)* Aujourd'hui, les standards d'échanges de données entre les niveaux national et régional sont très opérationnels : des échanges entre régions pourraient donc se baser sur l'existant, mais il appartient aux régions de régler ces questions. *(LP)* On peut toutefois noter qu'il existe déjà un portail de saisie national orienté sur la migration, appelé Migraction, et qu'une coordination de tous les portails Faune régionaux est également à l'étude. *(PM)* L'échelon régional, en outre, est le mieux adapté à la validation des données.

La validation des données

Une autre problématique essentielle qui touche les données naturalistes est celle de leur validation. Cettia et Faune Île-de-France y répondent par deux approches différentes. Sous Cettia, le principe de la validation collaborative est privilégié, c'est-à-dire que tout contributeur peut interroger une donnée et y apporter son avis. *(PR)* Dans Faune Île-de-France, la validation est faite par des spécialistes, membres ou non, des deux associations animatrices, choisis par le comité de pilotage et qui se coordonnent. Cette approche fonctionne très bien pour les oiseaux, mais moins pour d'autres taxons, comme les insectes, pour lesquels les contributeurs sont moins nombreux. Il existe un partenariat avec la Société Française pour l'Étude et la Protection des Mammifères (SFPEM) pour les mammifères qui fonctionne très bien, et tout expert sur les autres taxons est le bienvenu pour rejoindre les comités de validation. *(PM)*

La validation est une démarche qui doit être bien structurée, mais sans constituer un frein à la diffusion et à l'échange. L'essentiel, lors d'échanges de données d'une base à une autre, est que leur statut de validation soit bien codifié. Par exemple, il existe des groupes taxonomiques pour lesquels très peu de spécialistes existent, et les données ne peuvent pas être validées systématiquement. Il ne faut pas malgré tout considérer que de telles données, non validées, ne peuvent pas pour autant alimenter les bases de données nationales et internationales. *(LP)* Il faut cependant rester vigilant à ce qu'une validation effectuée dans une base de données, soit répercutée dans les autres bases contenant cette donnée. C'est notamment le rôle de l'identifiant unique SINP, mais cette problématique s'avère très complexe. *(SF)*

Quelle qualité de la donnée ?

L'émergence de la saisie en direct sur le terrain avec les nouveaux outils technologiques apporte une précision extrême aux observations. Si, pour les oiseaux ou les papillons, cette précision présente moins de valeur, elle en a beaucoup pour la flore et les insectes, qui témoignent par exemple de l'état ou de la nature d'un milieu en un lieu et un instant très précis. *(RJ)*

D'autre part, la donnée reste une notion complexe. On catégorise une observation dans la nature en un lieu, une espèce, une date... mais on oublie parfois que les individus ont des caractéristiques propres, et que les espèces ne vivent pas isolément les unes des autres : elles existent dans une communauté, c'est-à-dire un assemblage d'espèces, dont la composition est une information que l'on a tendance à dénaturer. En effet, cette dernière est détruite lorsqu'on la ramène à des données individuelles dans des bases. Il convient de souligner l'importance de prendre en compte le comportement des individus : qu'étaient-ils en train de faire lors de l'observation ? La recherche de nourriture ou la reproduction sont autant d'informations capitales. Quand un insecte est observé sur une fleur, cette relation trophique est beaucoup plus intéressante que les deux informations prises séparément. Il convient donc de saisir et surtout d'observer utilement par rapport à un objectif d'action, car toutes ces informations présentent de forts enjeux de connaissance et d'action. *(RJ)*

Cettia a l'avantage par rapport à d'autres outils de permettre la saisie de données autres que les observations opportunistes, comme les données obtenues à partir d'un protocole standardisé. C'est notamment intéressant pour les gestionnaires d'espaces naturels qui peuvent mettre en place des suivis sur leurs sites et les retranscrire dans la base. Cette entrée par protocole se fait via un formulaire adapté et ce système va être amené à se développer de plus en plus pour répondre aux besoins spécifiques des professionnels. Il existe, à l'heure actuelle, une trentaine de protocoles dans Cettia. La saisie par polygones est par exemple en train d'être adaptée aux besoins des formulaires de terrain du CBNBP, à la demande des botanistes. Cettia offre, en outre, la possibilité de relier deux observations par une interaction écologique, telle que la pollinisation, par exemple. *(PR)*

Faune Île-de-France a également une entrée par protocole avec le Suivi Temporel des Oiseaux Communs (STOC). Et elle vient de développer un système de saisie par formulaire journalier : une liste d'espèces est fournie à l'observateur pour s'assurer qu'il renseigne bien toutes les espèces, y compris les plus communes, ainsi que la durée de sa session d'observation. Faune Île-de-France reste basée sur les données opportunistes mais tend autant que possible à orienter les observateurs vers quelque chose de plus cadré. *(PM)*

Les bases de données comme système d'alerte

Multiplicité et interopérabilité des outils, échelle de saisie, validation, sensibilité des données sont autant de problématiques pour lesquelles les réponses actuellement apportées ne sont que partiellement satisfaisantes. La possibilité de visualiser les données brutes et synthétisées en direct, sur des interfaces telles que Faune Île-de-France, Cettia et ses atlas dynamiques de la biodiversité, ainsi que Flora, constitue une révolution récente dans le monde naturaliste. La création d'une plate-forme agrégeant en temps réel les données des différents outils n'est pour l'instant pas encore d'actualité, les obstacles étant probablement plus d'ordre politique que technique.

Malgré les bonnes volontés existant çà et là, en l'absence d'un outil central de restitution en temps réel, les délais impliqués par les échanges de structure à structure vont à l'encontre du besoin d'agir rapidement. On peut en effet imaginer également ces outils comme des systèmes d'alerte : par exemple, si moins d'alouettes sont observées un printemps, c'est peut-être qu'une pratique agricole se met en place et a un impact, ce qui nécessite une surveillance efficace des pratiques actuellement menées. Des intuitions vont émerger lorsque l'on assiste à un phénomène sur le moment, intuitions que l'on aura oubliées si l'on s'interroge dessus 5 ou 10 ans plus tard, période à laquelle, qui plus est, le combat sera bien souvent déjà perdu. La mise en place d'un système collectif centralisant les observations, en les analysant simultanément pour assurer une veille, permettrait à de nombreux niveaux d'agir efficacement en faveur de la biodiversité. (RJ)



Salle comble pour les 7^{es} Rencontres naturalistes d'Île-de-France.
© Natureparif