

## Guide de bonnes pratiques pour la collecte et la saisie de données naturalistes

-  
**Novembre 2019**

Ce document liste les opérations à réaliser lors de la collecte de la donnée, qu'il s'agisse de sessions d'observation dites de « terrain », de collecte de données bibliographiques (publications ou carnets de terrain), ou encore de collecte de données de collections et de l'informatisation de ces données (opération de saisie) afin de limiter l'apparition d'éventuelles erreurs. La production de données vise généralement à documenter la biodiversité et conserver dans le temps l'information pour le producteur lui-même, pour le partage avec ses pairs et pour les générations futures. L'objectif est alors d'assurer la production de données naturalistes de la meilleure qualité possible. La liste des préconisations proposées n'est pas exhaustive et pourra naturellement être complétée par d'autres actions jugées pertinentes par le collecteur des données ou son commanditaire.

Les efforts et la rigueur fournis par les observateurs de terrain permettent d'obtenir des données naturalistes de bonne qualité. De fait, il est important que les règles de collecte soient définies et documentées dès le lancement des inventaires de terrain et que soit mis en place un ensemble de mesures pour permettre à l'acteur de terrain de réaliser ses prospections dans les meilleures conditions.

L'utilisation de systèmes de saisie embarqués (sur smartphone, sur ordinateur de poche, etc.) ou d'applications mobiles, compatibles avec le SINP, plutôt que l'usage de documents papiers pourrait limiter les erreurs liées à l'informatisation des données a posteriori. Cela représente également un gain de temps substantiel pour l'ensemble de la chaîne de traitement des données.

Les mesures d'accompagnement passent, entre autres, par le renseignement précis d'un cahier des charges incluant les méthodes et protocoles de terrain à utiliser et les informations à récolter. Une information nécessaire qui n'aurait pas été notée par l'observateur risquerait en effet de nuire aux usages futurs des données ainsi qu'à leur validation scientifique.

Il faut également penser à la manière de livrer ces informations au commanditaire ou tout simplement de les partager avec des tiers. On préconisera ainsi l'utilisation des standards d'échange du SINP (standard national, standards déclinés par les plateformes régionales) ou standard DEPOBIO selon le cadre du projet.

Dans la mesure du possible, et sous réserve de ne pas mettre en péril la population du taxon concerné tout en respectant la réglementation en vigueur, il peut être conseillé aux observateurs de rapporter une preuve (photo, échantillon, enregistrement sonore, etc.). La collecte de preuve doit intervenir dès lors que l'observation en question est singulière (nouvelle observation du taxon pour le territoire, etc.) ainsi que pour les groupes d'espèces difficilement identifiables sur le terrain.

Dans le domaine de l'identification/détermination, il est primordial d'adopter une attitude précautionneuse et de faire preuve d'une rigueur maximale lors de l'identification d'un taxon : il ne faut pas vouloir absolument mettre une identification précise, et au moindre doute, il vaut mieux identifier au genre voire à la famille.

L'usage d'ouvrages de référence (flores, faunes, monographies, guides, etc.), de clés de détermination, de collections de référence ou de site web d'aide à l'identification (i-Naturalist, DORIS, INPN-Espèces, etc.) est également recommandé afin de limiter les erreurs de détermination taxonomique. Il faut prêter attention à utiliser des documents et outils qui traitent toutes les espèces de la zone d'étude, ou qui précisent les risques de confusion.

Dans les groupes d'identification délicate, une détermination effectuée par un tiers « expert » est nécessaire pour renforcer la confiance dans la donnée.

Il est en outre préconisé que l'ensemble des informations récoltées soient bancarisées dans les bases sources et transmises aux plateformes du SINP, y compris l'ensemble des informations complémentaires qui dépassent le classique « où, quand, quoi, qui ». La transmission partielle des informations est limitante pour la mise en œuvre d'une validation scientifique optimale des données.

Il est par ailleurs indispensable de respecter les règles définies pour chaque champ (information attendue) : il ne faut **jamais** détourner les champs, les concepts et les principes prévus.

Par exemple, si un NOM Prénom est demandé, il ne faut pas inscrire de code ou encore de pseudonyme. Si un code nom « CD\_NOM » (identifiant numérique selon TaxRef) est attendu, il ne faut pas inscrire dans ce même champ un identifiant interne. Le respect de ces règles est seul garant d'une utilisation optimale des requêtes et valorisations futures affectant le jeu de données.

Par ailleurs, il est important de ne pas inscrire d'informations dans des champs de commentaire alors qu'il existe un champ dédié. Le champ commentaire ne doit jamais être utilisé pour inscrire des informations obligatoires ou très significatives.

Par ailleurs, il est capital que les acteurs s'organisent afin de ne pas générer de doublons inutiles dans le SINP. Pour cela, en tout premier lieu, il faut éviter les doubles saisies, c'est-à-dire de saisir une même donnée sur deux outils qui, tous deux, alimentent le SINP. La liste des jeux de données disponibles dans l'INPN, plateforme nationale du SINP, est accessible en ligne :

<https://inpn.mnhn.fr/accueil/recherche-de-donnees/metadonnees>

Les modes d'alimentation du SINP sont explicités sur la page suivante :

<https://inpn.mnhn.fr/programme/donnees-observations-especes/contribuer>

Pour finir, il est important de ne pas négliger la rédaction de métadonnées permettant de décrire le contexte de l'acquisition des données. Les métadonnées permettent non seulement d'avoir des informations sur le niveau de cohérence des données entre elles, de connaître les méthodes et programmes d'acquisition mais aussi de mettre en avant la ou les personnes/organismes ayant contribué à l'acquisition des données. Elles permettent ainsi une meilleure réutilisation des données, cohérente avec l'usage recherché. Pour aller plus loin, il existe des publications scientifique appelées « datapaper » qui permettent de décrire un jeu de donnée partagé et ainsi de faciliter sa découverte et sa citation par d'autres scientifiques.

## LISTE DES PRÉCONISATIONS (non exhaustive)

### ❖ Saisie et métadonnées

Utiliser directement un outil de saisie (plutôt qu'un formulaire papier) permet de ne pas avoir à ressaisir les informations a posteriori et donc de limiter les risques d'erreur de saisie lors de la transposition numérique. Certaines informations peuvent être renseignées automatiquement selon les configurations prévues par l'outil ce qui peut simplifier le travail des producteurs. Les outils de saisie peuvent également intégrer des pré-contrôles sur les données et prévenir directement l'opérateur de la saisie d'un problème sur les données. Certains outils intègrent également des visualisations cartographiques des données (cartes avec localisation des points d'observation) ce qui permet d'éviter la saisie de coordonnées erronées en permettant à l'utilisateur de repérer rapidement les erreurs de localisation.

Une liste (non exhaustive) d'outils de saisie est disponible en annexe 1.

Pour connaître les outils régionaux ou locaux, les différents acteurs peuvent se tourner vers les DREAL/DEAL/DRIEE ou contacter les différents pôles régionaux du SINP.

Connaitre les informations minimales pour le SINP (voir standard) à noter avant de se rendre sur le terrain permet d'éviter d'oublier de noter des informations importantes, voire obligatoires dans le cadre du SINP et permet d'utiliser, lorsque c'est possible, directement les nomenclatures (=liste de valeurs) du SINP pour éviter les transpositions (mapping) a posteriori.

En cas de prospection sur la base d'un protocole d'observation ou de suivi, respecter les informations à relever indispensables dans le cadre de ce protocole permettra de produire des données protocolées qualitatives, réutilisables et pertinentes. Utiliser autant que possible un outil de saisie dédié à ce protocole, quand il existe (ou un bordereau de saisie adapté à ce protocole).

Bien organiser les données et renseigner les métadonnées (découpage des données en jeu de données) selon le protocole de collecte utilisé, faire référence à la base nationale des protocoles et méthodes (Campanule : <http://campanule.mnhn.fr/>) si possible – sinon, détailler le protocole dans les métadonnées permettra une meilleure description et contextualisation des données transmises.

*Il est à noter que des guides (profils) pour la transposition des données protocolées dans le standard SINP sont disponibles (données de pêches électriques, IFN et SPIPOLL), et que d'autres sont en cours de production. <http://standards-sinp.mnhn.fr/occurrences-de-taxon-v2-0/>*

En cas de doute sur la façon de transmettre des données protocolées dans les standards du SINP, demander conseil aux porteurs des plateformes du SINP (régionales ou nationale).

Renseigner les métadonnées de la manière la plus complète possible (description, mots clef, financeurs, etc.), c'est-à-dire des descripteurs des jeux de données produits améliorera la visibilité des acteurs et des données, et en facilitera l'utilisation future. Il est à noter que l'usage de certains outils de saisie facilitent grandement cette étape.

Un guide national de préconisations sur le découpage des métadonnées et leur description est disponible : <https://inpn.mnhn.fr/docs-web/docs/download/244790>

En cas de doute sur la meilleure manière d'organiser les données, demander conseil aux porteurs des plateformes du SINP.

Il faut totalement proscrire l'usage de fichier texte. Lorsque cela est possible (existence d'un outil de saisie répondant aux besoins), il est conseillé d'éviter les saisies directes sur Excel, Libre office Calc. et autres tableurs qui ajoutent une étape supplémentaire dans la chaîne de traitement, susceptible de générer des erreurs. Néanmoins, si des tableurs sont tout de même utilisés, il peut être judicieux d'utiliser un gabarit, avec des listes et contrôles permettant de maîtriser l'information saisie. L'INPN ainsi que de nombreuses plateformes régionales et têtes de réseaux proposent des gabarits à leurs réseaux de partenaires.

Dans la mesure du possible, transmettre l'ensemble des informations recueillies sur le terrain ce qui permettra à la donnée de ne pas être appauvrie par rapport à celle réellement observée (notamment les techniques d'observation, les variables biotiques ou abiotiques éventuellement mesurées telles que le micro-habitat, l'hygrométrie, la température, les effectifs, la taille des ailes, le poids, etc., qui pourront être partagées au sein du SINP via des champs dédiés ou via les attributs additionnels).

Générer dès la base de production l'identifiant unique du SINP (UUID) permettra d'assurer la traçabilité des données dans toute la chaîne de circulation de la donnée et facilitera les échanges et les mises à jour, ainsi que la détection d'éventuels doublons dans le SINP.

Exemple de générateur d'UUID en ligne: <https://www.uuidgenerator.net/>

Générer les UUID directement dans une base de données: <http://standards-sinp.mnhn.fr/faq-identifiant-permanent/>

## ❖ Observations et déterminations

Utiliser des outils d'aide à la détermination permet de fiabiliser la détermination taxonomique et d'éviter les confusions d'espèces sur le terrain.

Il est également important de noter le nom complet de l'espèce le plus fidèlement possible, et éventuellement d'écrire le nom de l'autorité de la description lorsque connu. Il est possible d'ajouter une précision supplémentaire comme le groupe taxonomique si nécessaire dans un champ dédié. Il sera utile d'éviter les abréviations (notamment de nom de genre) pour que les notes de terrain soient compréhensibles par une tierce personne et de ne pas utiliser de caractères spéciaux ( ? ou sp.) pour indiquer l'incertitude (dans ce cas, en rester au rang taxonomique supérieur dans le champ nom cité et indiquer une note dans le champ commentaire pour préciser éventuellement qu'il y a doute par exemple entre plusieurs espèces du genre).

Cela permettra de limiter les erreurs de réconciliation avec les référentiels nationaux (Taxref pour les taxons, Habref pour les habitats). Cela permettra également de revenir sur les données en cas de modification du référentiel taxonomique liée à de nouvelles connaissances scientifiques (recombinaison, splittage, etc.).

Prendre une/des photos du spécimen lorsque le cas s'y prête, cela permet d'attester de la présence de l'espèce en question (notamment en cas d'une nouvelle découverte) et permet de valider la détermination a posteriori. Pour cela, il faudra veiller à bien cibler les photos selon les critères de détermination du groupe taxonomique.

Inscrire le nom du/des observateurs permettra une meilleure traçabilité des données et ainsi, de les rendre visibles afin de valoriser et responsabiliser les contributeurs. Respecter un formalisme (NOM1 Prénom1, NOM2 Prénom2) pour inscrire le nom du/des acteurs engagés dans la collecte/production de la donnée (observateur, déterminateur, etc.) permettra d'harmoniser l'affichage des sources et

permettra des recherches par acteur. Dans les cas où cela est possible, utiliser un référentiel des observateurs peut être judicieux.

Il est pour autant indispensable de respecter le choix des observateurs si ceux-ci ne veulent pas apparaître. Si aucun observateur ne souhaite apparaître : inscrire Anonyme dans la catégorie « observateur ».

Détermination réalisée par une personne tierce. Dans la plupart des groupes biologiques, l'observateur peut être différent du déterminateur. Aussi, il est important de conserver cette information dans un champ distinct et de bien la séparer. La date de la détermination peut venir compléter le nom de l'observateur dans le cas où elle s'applique à des spécimens identifiés dans les collections, qu'elles soient anciennes ou récentes.

Lorsque l'observation est faite sur plusieurs jours (pose de piège par exemple), saisir deux dates pour les observations (une date min de début d'observation et une date max de fin d'observation, dans deux colonnes séparées) permet de gérer facilement les observations réalisées sur une période donnée. De même, il est recommandé de mesurer le temps d'une prospection en un lieu donné ; cela permet de quantifier l'effort de recherche dans une station ou un habitat en particulier.

Même dans le cas d'observations « simples » (sans protocole particulier, ni objectif d'étude précis), il est conseillé de noter l'ensemble des espèces présentes d'un groupe taxonomique et pas uniquement les espèces remarquables. En effet, cette information sur la communauté d'espèces observée offre des perspectives d'exploitations statistiques beaucoup plus étendues et permet de mieux quantifier la rareté relative des espèces.

### ❖ Précision géographique

Utiliser un GPS pour connaître les coordonnées évite les imprécisions géographiques/erreurs de pointage lorsque les coordonnées sont récupérées a posteriori. Il est également important de connaître et renseigner la précision du GPS utilisé (voir notice de l'appareil). Transmettre les coordonnées dans le système de projection utilisées directement pour le géoréférencement sur le terrain permet d'éviter les éventuelles imprécisions liées à des transpositions dans d'autres systèmes de projection. Dans le cas où les coordonnées n'auraient pas été relevées sur le terrain, il est possible d'utiliser des outils a posteriori. La plupart des outils de saisie naturalistes proposent un système de localisation. Il est également possible d'utiliser le site de l'IGN pour obtenir les coordonnées d'un lieu prospecté : <https://www.geoportail.gouv.fr/>

Renseigner la précision de l'objet géographique transmis permet de limiter les erreurs de transposition selon d'autres référentiels géographiques (niveau communal, niveau maille 10x10km, etc.). Il est à noter que le standard SINP propose un champ (attribut) dédié pour la transmission de cette imprécision.

Dans la mesure du possible, transmettre l'information géographique la plus précise possible limitera les dégradations des données relatives aux croisements géographiques et améliorera les possibilités d'exploitation futures des données.

## EXEMPLES D'ERREURS ou MAUVAIS USAGES FRÉQUENTS

Ne pas noter les coordonnées sur le terrain et se référer ensuite à un centroïde (par exemple d'une commune).

Conséquence : l'information collectée est dégradée – peu précise. Elle permettra de valoriser la donnée à l'échelle de la commune mais entraînera des erreurs potentielles dans le cas de valorisation selon d'autres résolutions (présence dans une aires protégée, cartographie dans un atlas...). Cette imprécision risque de plus de ne pas être connue s'il n'y a pas de rayon d'imprécision associée à la donnée (fausse précision).

Bonne pratique : je note les coordonnées de l'observation ou, si je ne peux indiquer que la commune, je transmets le nom de la commune et son code INSEE.

Numériser (redessiner) ou transmettre des polygones qui relèvent d'un référentiel.

Exemple : ma donnée est située sur une commune.

Bonne pratique : je transmets le code INSEE de la commune puisqu'un référentiel existe déjà.

Mauvaise pratique : je dessine ou transmet un polygone de la commune qui est redondant avec la couche géographique des communes produite par l'IGN (BD Topo).

Je dessine ou transmet un polygone de maille (national ou européenne) redondant avec la couche géographique existante.

Cette pratique surcharge les bases de données et ralentie les processus d'exploitation.

Mélanger site d'observation et site de stockage de spécimen

Les coordonnées de l'observation saisie doivent correspondre à l'endroit où l'espèce a été observée. Si le lieu de stockage des spécimens en collection est important, noter l'information dans un champ dédié et ne pas mélanger plusieurs types d'information dans une même colonne.

Mélanger date d'observation et date de saisie

La bonne pratique consiste à ne jamais mélanger date d'observation et date de saisie dans un même champ.

Mélanger site d'observation et lieu où l'individu se trouvait réellement

Les coordonnées de l'observation saisie doivent correspondre à l'endroit où l'espèce a été observée et non à l'endroit depuis lequel l'espèce a été observée. Si peut donc s'avérer nécessaire de corriger, avant ou pendant la saisie, les coordonnées acquises automatiquement par les appareils munis d'un GPS (appareil photo, téléphone mobile, etc.).

Ceci est important par exemple pour les organismes observés en mer à partir de la côte (dune, falaise) ou inversement organisme à terre alors qu'on est sur un bateau en mer. Dans le premier cas par exemple, le lieu d'observation GPS est terrestre alors que l'organisme est marin. Exemples : oiseaux (Fou de Bassan, Laridés...) ou mammifères (phoques, cétacés) observés au loin.

Conséquence : décalage de la distribution, une espèce marine est affichée en terre et inversement. Les cartes de distribution/répartition s'en retrouve faussées. Les données sont susceptibles d'être invalidées.

Écrire la mention « NON OBSERVÉ » en commentaire ou oublier de préciser qu'il s'agit d'une « non observation » (= absence de contact avec l'espèce recherchée). Un champ dédié à cette information existe dans le standard du SINP. Ne pas l'utiliser correctement implique qu'une donnée de non observation pourrait être considérée comme une donnée d'observation, avec toutes les conséquences

que cela pourrait avoir. Dans le cas d'une non observation, il est précieux de détailler le protocole qui a été utilisé pour attester de la non observation au sein des métadonnées.

Ecrire effectif =0 pour signifier une absence (ou non observation)

Un champ dédié à cette information existe dans le standard du SINP. Ne pas l'utiliser correctement implique qu'une donnée de non observation pourrait être considérée comme une donnée d'observation, avec toutes les conséquences que cela pourrait avoir.

Inscrire les coordonnées du site de l'observation dans un champ non dédié

Par exemple : inscrire les coordonnées dans les champs commentaires : ces informations ne seront pas exploitables pour localiser la donnée et, si un floutage de la localisation est opéré sur la donnée par la suite (par exemple en cas de sensibilité de la donnée pour la diffusion), ces champs resteront lisibles.

Transmettre des données uniquement floutées car l'espèce est sensible : la sensibilité dans le SINP est définie par une méthodologie et par un ensemble de règles qui permettent de ne pas diffuser largement les informations lorsque la donnée est sensible. Des listes de sensibilité sont établies dans ce cadre et validées par les CSRPN.

Ne pas transmettre la localisation précise des observations des espèces sensibles aux porteurs de plateformes du SINP impliquera que ces données ne pourront pas être considérés dans les études et analyses, avec le risque que ces espèces, par méconnaissance, ne soient pas prises en compte dans les projets, voir soient impactées négativement par des projets d'aménagement.

Plus d'informations sur la sensibilité au lien suivant: <https://inpn.mnhn.fr/programme/donnees-observations-especes/references/sensibilite>

## Annexe 1

### Liste (non exhaustive) d'outils de saisie gratuits proposés au niveau français pour la saisie et le partage des données naturalistes sur les espèces dans le cadre du SINP

CarNat – outil de saisie sur mobile associé à CardObs, pour naturaliste « confirmé ». La détermination est directement proposée par l'observateur.

[<https://cardobs.mnhn.fr>]

GeoNature, outil de saisie et de gestion développé par les parcs nationaux de France. La détermination est directement proposée par l'observateur.

[<https://geonature.fr/>]

INPN-Espèces – outil fondé sur la photographie, recommandé pour les naturalistes débutants et pour les taxons qu'un expert maîtrise moins (exemple pour qu'un lépidoptériste puisse transmettre ses observations de coléoptères). La détermination est validée par un ou plusieurs experts de l'INPN.

[<https://inpn.mnhn.fr/informations/inpn-especes>]

i-Naturalist – sur le même principe qu'INPN-espèce, mais permettant de saisir des photographies d'espèces observées dans le monde entier, France comprise. La détermination est validée par une majorité des deux-tiers d'utilisateurs d'iNaturalist ayant exprimé leur avis sur la photographie soumise. Les données sont diffusées sur le GBIF et ensuite récupérées dans le SINP.

[<https://www.inaturalist.org>]

De nombreux protocoles disposent d'outil et d'interface de saisie dédiée. On peut citer notamment les sites des programmes vigieNature [<http://www.vigienature.fr>] :

- Suivi Temporel des Libellules (STELI) : <http://steli.mnhn.fr>
- Suivi Photographique des Insectes Pollinisateurs (SPIPOLL) : <https://www.spipoll.org>
- Suivi Temporel des Rhopalocères de France (STERF) : <http://sterf.mnhn.fr/>
- VIGIE-CHIRO : <http://www.vigienature.fr/fr/page/participer-vigie-chiro>
- VIGIE-FLORE : <http://www.vigienature.fr/fr/participer-3473>
- ...

En cas de doute sur l'outil à utiliser, demander conseil aux porteurs des plateformes du SINP (régionales ou nationale).