

Guide pratique pour le développement et le choix d'un outil de saisie de données naturalistes

-

Décembre 2019

Les efforts fournis par les observateurs de terrain permettent d'obtenir des données naturalistes de bonne qualité (précises géographiquement, correctement déterminées, etc.). L'utilisation de systèmes de saisie embarqués directement sur le terrain (sur smartphone, sur ordinateur de poche, etc.) plutôt que l'usage de documents papiers peut limiter les erreurs possibles liées à l'informatisation des données a posteriori et faire gagner du temps dans l'ensemble de la chaîne de traitement.

Il n'est pas forcément nécessaire de redévelopper de nouveaux outils car de nombreux outils existent déjà et peuvent convenir aux besoins des utilisateurs. Certains permettent des paramétrages par l'utilisateur ou pour un protocole donné partagé entre plusieurs utilisateurs.

Une liste (non exhaustive) d'outils de saisie répondant aux préconisations prédéfinies 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. Dans le cas du choix d'un outil de saisie existant, il convient de bien s'assurer de la politique de diffusion des données qui est associée à l'outil (politique au choix de l'utilisateur, politique Opendata par défaut, politique restreinte).

Pour qu'un outil perdure dans le temps, il est important que celui-ci évolue en fonction du contexte et des besoins. Il est donc important de prévoir une maintenance et de prendre en considération les retours des utilisateurs. Animer le réseau des utilisateurs d'un outil est un point important pour la pérennité du projet.

Plus les outils seront ergonomiques et faciles d'accès, plus ils seront utilisés.

Il est parfois préférable de prévoir des outils (ou modules) distincts en fonction du besoin (notamment selon le protocole de collecte ou autre) plutôt que de créer un outil « couteau suisse » qui pourrait devenir complexe d'utilisation et difficilement maintenable.

Par ailleurs, plus l'outil développé pour la saisie intègrera de contrôles en amont, plus les données en sortie de l'outil seront de bonne qualité.

A noter cependant que chaque cas est différent et qu'il est important de mesurer les choix pris au regard de l'impact sur les utilisateurs (ne pas trop alourdir/ralentir la saisie par exemple) et de la charge nécessaire à la gestion des mises à jour et aux changements de version.

LISTE DES PRECONISATIONS (non exhaustive)

Gérer les rattachements aux référentiels existants dès la saisie permettra par exemple d'éviter des réconciliations a posteriori et donc l'apparition d'éventuelles erreurs.

Il est possible d'interfacer ses outils avec les référentiels diffusés par API : voir notamment <https://taxref.mnhn.fr/taxref-web/api/doc>

Utiliser les standards du SINP ou des nomenclatures interopérables avec celles proposées par le SINP permettra de limiter l'effort de normalisation lors du partage de ces données. Il est possible d'utiliser des webservices pour récupérer les nomenclatures du SINP ou de les consulter directement à partir du site sur les standards du SINP: <http://standards-sinp.mnhn.fr/>

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.

Gérer régulièrement les montées de version des référentiels permettra de tenir à jour les rattachements et évitera la perte des données rattachées à un référentiel obsolète (fusion de commune, splittage, etc.).

Inclure des contrôles sur les données au niveau applicatif permet à l'utilisateur de vérifier/corriger directement ses données et limitera les allers-retours a posteriori en cas de doute sur les données. Il est également possible de documenter les contrôles effectués et de partager ces règles avec les utilisateurs. Il est possible de consulter la liste des contrôles effectués nationalement: <https://inpn.mnhn.fr/programme/donnees-observations-especes/references/validation>

Proposer des visualisations cartographiques des données (cartes avec localisation des points d'observation) permettra d'éviter la saisie de coordonnées erronées et d'aider l'utilisateur à repérer rapidement les erreurs en lien avec la géolocalisation des données.

Proposer des fonctions de requêtage et de gestion des données à l'utilisateur via l'application permettra à celui-ci de corriger, manipuler, vérifier ses données le cas échéant. Par exemple la possibilité de filtrer les données par date d'observation et de les visualiser sur une carte permettra à l'utilisateur de vérifier facilement qu'il n'y a pas eu d'erreur de saisie des coordonnées GPS.

Communiquer avec les utilisateurs sur le devenir des données (notamment leur circulation au sein du SINP) permettra notamment d'éviter les doubles saisies réalisées par les producteurs (c'est-à-dire de saisir une même donnée sur deux outils qui, tous deux, alimentent *in fine* 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>

Cette communication permettra d'être totalement transparent sur l'usage fait des données vis-à-vis des contributeurs.

Il est ainsi important de bien gérer les questions de droits de diffusion le plus tôt possible, donc au mieux dès la saisie voire à l'inscription des utilisateurs de l'outil.

Pour les noms de personne, imposer un formalisme (NOM Prénom) dans les champs pour renseigner le nom du/des acteurs (observateur, déterminateur, etc.) permettra d'harmoniser l'affichage des sources et permettra par la suite des recherches par acteur (et donc une meilleure valorisation des contributeurs). 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.

Faire le lien avec les métadonnées dès la saisie évitera des opérations de découpage des données à posteriori et permettra au producteur de gérer lui-même l'organisation qu'il souhaite pour ses données.

Une vigilance particulière liée à l'encodage des données (notamment lorsque l'outil propose un module d'import) permettra d'éviter la perte et la dégradation des informations. Des dégradations de données sont souvent constatées dans les attributs textuels de type « commentaire » mais également au niveau du nom des observateurs ou du nom cité des espèces.

Si l'outil demande par exemple un fichier en UTF-8 :

- ENREGISTRER LE FICHIER EXCEL AU FORMAT CSV (SEPARATEUR, POINT VIRGULE), AVEC « ENREGISTRER SOUS »
- UNE FOIS LE FICHIER ENREGISTRE, LE ROUVRIRE AVEC NOTEPAD++ (SI VOUS NE L'AVEZ PAS, INSTALLEZ-LE) EN FAISANT UN CLIC DROIT > OUVRIR AVEC > NOTEPAD++
- DANS LE MENU DE NOTEPAD++, TROUVER « ENCODAGE » > CONVERTIR EN UTF-8
- SAUVEGARDER LE DOCUMENT
- DORENAVANT, LE FICHIER PEUT S'OUVRIRE CORRECTEMENT EN UTF-8 DANS EXCEL (POUR VERIFICATION, LES E/A/U NE SONT PAS TRANSFORMES EN Ã©, Ã, OU Ã¹ PAR EXEMPLE).

Ne pas hésiter à surveiller régulièrement la qualité des données de son système d'information en mettant en place des contrôles/tests automatiques sur la qualité des données.

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. L'outil permet de saisir des données sur la faune, la flore et la fonge. Il est optimisé pour la France métropolitaine et l'Outre-mer. Il permet la gestion des métadonnées et propose des champs paramétrables. Les données alimentent l'INPN et les utilisateurs renseignent les droits de diffusion des données par jeu de données.

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

GeoNature, outil de saisie et de gestion open source (web et mobile) développé par les parcs nationaux de France. La détermination est directement proposée par l'observateur. Il permet la gestion des métadonnées et est interfacé avec les référentiels nationaux en vigueur. Il permet la saisie de données certains protocoles.

[<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. Toutes les données saisies via cette application sont publiques et diffusées au plus précis dans le SINP (en respectant les règles de sensibilité des données à la diffusion).

[<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'outils et d'interfaces de saisie dédiés. 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, demandez conseil aux porteurs des plateformes du SINP (régionales ou nationale).