

## Guide pratique pour la standardisation des données naturalistes

-  
**Décembre 2019**

Les traitements réalisés sur les données, notamment la standardisation, doivent être le moins possible source de dégradation/dénaturation des informations transmises.

Les informations importantes collectées par les producteurs doivent être transmises (la donnée produite doit être aussi proche et complète que possible de la donnée source). Une information qui n'aurait pas été transmise risquerait en effet de nuire aux usages futurs des données ainsi qu'à la validation scientifique.

Les données doivent par ailleurs être associées à des métadonnées (description du contexte d'acquisition des données et déclaration des droits de diffusion).

### LISTE DES PRECONISATIONS (non exhaustive)

Identifier le standard à utiliser en fonction du contexte :

- Le standard SINP pour l'échange « informatique » entre plateformes du SINP
- Le standard INPN pour la transmission de données d'un producteur à l'INPN (cadre échange avec la plateforme thématique SINP)
- Le standard DEPOBIO dans le cadre du dépôt légal (données des études d'impact et autres études réglementaires)
- Les standards régionaux pour la transmission de données d'un producteur à une plateforme ou un pôle régional du SINP

En cas de doute sur le standard à utiliser et pour obtenir des gabarits (templates) d'échange adaptés, demander conseil aux porteurs des plateformes du SINP (régionales ou nationale).

Utiliser si possible la version la plus récente du standard permettra de faciliter les échanges de données et d'assurer la complétude de l'échange.

Veiller à respecter les listes de valeurs (nomenclatures) des standards pour assurer la conformité des 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/>

Veiller à standardiser au maximum les informations importantes et veillez à les transmettre lors des échanges (soit dans les champs dédiés du standard lorsqu'ils existent, soit dans les attributs additionnels).

En cas de lacune du standard pour normaliser certaines informations importantes (notamment au regard d'un protocole particulier), il est possible de faire remonter la demande d'évolution au GT

Standard de données en transmettant la demande par l'intermédiaire des correspondants SINP (contacter votre DREAL/DRIEE/DEAL) ou directement auprès de la plateforme nationale (sinp@mnhn.fr). L'ajout d'attributs dans le standard sera évalué au regard de l'enjeu et de la récurrence du besoin.

En cas de doute et de questions sur la manière de standardiser les données :

- Consulter la FAQ des standards. Adresse FAQ nationale : <http://standards-sinp.mnhn.fr/category/faq/>
- Contacter la plateforme régionale SINP référente.
- Contacter l'assistance (spécifique pour DEPOBIO) : [assistance.depobio@afbiodiversite.fr](mailto:assistance.depobio@afbiodiversite.fr)
- Contacter le référent SINP sur les standards (sinp@mnhn.fr)

Appliquer les mêmes transformations sur l'ensemble des données traitées afin de produire des lots de données uniformes.

Documenter si possible les procédures appliquées et garder les scripts afin de les réutiliser ou de permettre de revenir sur des traitements a posteriori. Garder un historique des modifications réalisées sur les données afin d'assurer la traçabilité de ces changements.

Limiter autant que possible les floutages entre la donnée source et la donnée du SINP (DEE). On entend par floutage l'opération qui consiste à dégrader la précision géographique de la donnée source (transmettre une commune alors que l'information originale était un point GPS par exemple). Plus les données seront partagées précises, plus celles-ci seront intéressantes et utilisables. Par ailleurs, l'opération de floutage peut engendrer des erreurs de rattachement et complexifier le format de la donnée (une observation = n communes = n mailles) et donc son exploitation.

Si la standardisation n'est pas réalisée par le producteur, il est important de le renseigner sur les traitements appliqués sur les données ou lui faire valider les traitements particuliers (entraînant des modifications par exemple).

Veiller à l'encodage lors de la standardisation 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. Le SINP recommande vivement l'usage de l'encodage UTF-8.

Si l'identifiant unique du SINP (UUID) n'est pas présent dans la donnée source, proposer au producteur des données la création de l'identifiant unique et le report dans sa propre base de données. Cela 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/>

Ne pas multiplier les outils de reprojection utilisés et stocker si possible la projection native. La précision avec laquelle la donnée a été numérisée sera conservée ce qui réduira les éventuels décalages dues à l'utilisation de trop nombreux outils ou de reprojections successives.

Lorsque l'observation est faite sur plusieurs jours (pose de piège par exemple) et que le producteur a transmis cette information dans une seule colonne, il sera nécessaire de transformer les dates d'observation dans deux colonnes distinctes : une date min de début d'observation et une date max de fin d'observation. Si la date d'observation est une date précise à la journée, la date min et la date max doivent être égales. Dans le cas contraire, la date min correspond à la date de début d'observation et la date max à la date de fin d'observation.