



**SINP**

Système d'Information  
sur la Nature et le Paysage



## GT Validation des données d'occurrences du SINP

### *Volet 1 : occurrences de taxons*

# Synthèse des réponses au questionnaire pour l'identification des besoins des différents acteurs du SINP en matière de validation

V1

---

**Octobre 2015**



**MUSÉUM**  
NATIONAL D'HISTOIRE NATURELLE

### Historique des versions du document :

<i>Versions</i>	<i>Date</i>	<i>Commentaires</i>
0	18/08/2015	Document préliminaire
1	23/10/2015	Document modifié suite à relecture par les membres du GT Validation SINP

### Rédacteur :

MNHN-SPN	Solène Robert
----------	---------------

### Relecteurs :

MNHN-SPN	Laurent Poncet
MNHN-SPN	Julien Touroult
MNHN-SPN	Jeanne de Mazières
MNHN-SPN	Judith Panijel
FCBN	Johan Gourvil

### Organismes contributeurs :

CBN Alpin, CBN Bailleul, CBN Bassin parisien, DREAL Centre Val de Loire, MNHN-SPN, OAFS, OBHN/CRHN & ONCFS

### Citation recommandée :

Robert S., Poncet L., de Mazières J., Gourvil J., Touroult J. 2015. Synthèse des réponses au questionnaire pour l'identification des besoins des différents acteurs du SINP en matière de validation. Rapport pour le SINP, rapport MNHN-SPN 2015-62, 19 p.

## Contexte

L'enjeu de la validation pour le SINP consiste à mettre à disposition des différents utilisateurs un niveau de confiance associé aux données véhiculées de manière à en permettre un usage adapté.

Dans cet objectif, il est nécessaire que les différents opérateurs en charge de ces procédures de validation dans le cadre du SINP s'accordent sur une organisation opérationnelle, sur une démarche générique, sur une typologie et sur un ensemble d'éléments à ajouter à la DEE occurrence de taxon et à la métadonnée de ces DEE et garantissent toute la traçabilité des opérations de validation réalisées. Pour mener à bien ces travaux, le GT Validation a besoin d'identifier les besoins des différents acteurs du SINP en matière de validation.

## Rappel des objectifs initiaux du questionnaire

Le questionnaire avait pour objectifs d'identifier :

- ❖ Les grands types de besoins fonctionnels pour les **grandes catégories d'acteurs du SINP**, à savoir :
  - producteur de données
  - gestionnaire de plateforme SINP
  - animateur politique SINP
  - administrateur de données (*=intermédiaire, autre gestionnaire, centralisateur de données, fournisseur de données, contributeur*)
  - expert
  - utilisateur de données
  
- ❖ Les priorités en matière de validation.
  
- ❖ Les difficultés fréquemment rencontrées dans la réalisation de la chaîne de contrôle (conformité, cohérence et validation).

## Méthodologie du questionnaire

Le questionnaire a été mis en ligne via un « google form » et s'organisait en trois parties distinctes : l'identification des participants et leur implication dans le SINP, l'identification des besoins et des pratiques existantes (en termes d'informations véhiculées dans les données et métadonnées, en termes d'outils et en termes d'accès à des informations tierces) et l'identification des difficultés rencontrées (pour la standardisation, la conformité, la cohérence et la validation automatique ou manuelle).

Les réponses étaient ouvertes et textuelles de manière à permettre de remonter le maximum d'information et à autoriser une largesse de réponse.

**Date de mise en ligne du questionnaire** : 07 mai 2015

**Date de clôture des réponses au questionnaire**: 20 juillet 2015

## Participation

Le questionnaire a été proposé à l'ensemble des membres du GT Validation qui étaient invités à le proposer aux membres de leur réseau.

**Liste des organismes ayant répondu à l'enquête :** CBN Alpin, CBN Bailleul, CBN Bassin parisien, DREAL Centre Val de Loire, MNHN-SPN, OBHN/CRHN, ONCFS

**Nombre de questionnaires retournés:** 14

Les participants ont généralement retransmis les besoins de leur réseau en renseignant, via un unique questionnaire, les besoins de plusieurs catégories d'acteurs. Les réponses transmises couvrent alors de manière globale l'ensemble des types d'acteurs ayant été identifiés par le groupe.

### **Précaution de lecture**

Ce document se présente comme une synthèse brute (sans analyse spécifique de pertinence ou de priorité) des éléments recueillis, des possibilités d'organisation et des besoins retranscrits par les différents acteurs consultés. **Il ne présume pas des éléments qui seront repris ou non dans le cadre du SINP.**

Les besoins remontés via le questionnaire sont ici retranscrits, indépendamment de l'existence de solutions existantes ou d'ores et déjà en construction.

Le format très libre du questionnaire a donné lieu à une grande variété de réponses dont certaines dépassent le périmètre du GT Validation.

Les éléments recueillis concernant directement le GT Validation sont présentés **en gras** et un **encart** sous chaque thématique permet de renseigner le lecteur sur la solution, l'instance ou le GT concerné par le besoin exprimé.

## Plan du document

1. ASPECTS ORGANISATIONNELS ET PROCESSUS GLOBAL.....	5
2. COLLECTE DE DONNEES.....	5
3. FORMATS DE DONNEES ET DE METADONNEES .....	6
4. INFORMATIONS SUPPLEMENTAIRES ET TRACABILITE.....	7
5. OUTIL DE SAISIE .....	8
6. OUTIL DE CREATION DES DEE ET METHODES DE TRANSFORMATION .....	8
7. OUTIL DE VALIDATION ET SERVICES DEDIES .....	9
8. REFERENTIELS ET BASES DE CONNAISSANCE .....	9
9. MISES A JOUR .....	10
10. DIFFUSION .....	11
11. ANNEXES.....	12
Annexe 1.....	12
Liste brute des éléments transmis via le questionnaire en matière de besoin d’information .....	12
Annexe 2.....	16
Questionnaire mis en ligne pour l’identification des besoins en matière de validation .....	16

## 1. ASPECTS ORGANISATIONNELS ET PROCESSUS GLOBAL

(1) Il ressort des réponses au questionnaire la nécessité de partager une démarche de validation et, tout particulièrement, de s'accorder sur les valeurs des niveaux de validité sur la base d'une typologie commune de niveaux de validité.

(2) Le nombre d'experts et leur disponibilité étant limités au regard du nombre conséquent de données à valider dans le système, il serait utile d'organiser la validation en deux phases : une première phase de validation automatique permettant de pré-valider les données selon des critères automatisables, puis, une validation manuelle portée par le réseau d'experts. Cela permettrait aux experts d'intervenir prioritairement sur les données ayant été jugées douteuses ou à risque par la validation automatique. Les experts auraient cependant toute liberté pour intervenir sur les autres données quel que soit le résultat de la validation automatique.

(3) La démarche de validation automatique peut être réalisée selon une approche itérative (selon la congruence avec d'autres données du système), c'est-à-dire que plus les données du système seraient validées, plus l'opération de validation sera fiable et aisée.

(4) Les données élémentaires d'échange (DEE) devront véhiculer un niveau de validation régionale et un niveau de validation nationale et permettront de véhiculer un niveau de validité producteur. Les différentes échelles de validation (producteur, régionale et nationale) coexistent et ne devraient pas se suppléer entre elles.

(5) Il apparaît nécessaire de limiter au maximum les doubles sollicitations des experts. Pour ce faire, il faudrait véhiculer la traçabilité de la procédure de validation mise en œuvre et le cas échéant indiquer le nom de l'expert recontacté pour la validation par la plateforme qui l'a mis en œuvre.

(6) La circulation des données étant actuellement difficile à maîtriser, il semble inévitable que des doublons apparaissent au sein du système. Bien que ce point ne soit pas forcément dans le périmètre du présent GT, le besoin est remonté de créer un outil permettant de détecter les doublons. Une fois dans le système, il devient très difficile d'identifier ces doublons, il semble qu'il faille donc, en tout premier lieu, axer l'effort sur l'organisation des flux ; à savoir qu'une donnée ne doit entrer qu'une seule fois dans le système et par une unique plateforme.

### Solutions et/ou instance ou GT concerné

1, 2, 3, 4, 5 : GT Validation  
4, 5 : GT Standard de données  
6 : Animation « plateforme »

## 2. COLLECTE DE DONNEES

La conformité, la cohérence, la qualité d'une donnée, l'aptitude à la valider scientifiquement ainsi que la facilité d'usage par les utilisateurs finaux sont dépendants des moyens engagés tout au long du cycle de vie de la donnée, de la phase de construction du programme d'inventaire de terrain jusqu'à la diffusion des données au public.

S'il est en général impossible de revenir sur les données déjà acquises (notamment s'il n'y a pas de preuves associées) il est en revanche possible d'intervenir sur le lancement de nouvelles phases de prospections.

Dans les réponses au questionnaire, nous avons ainsi pu identifier des propositions d'amélioration des procédures dès la phase de collecte.

(1) En tout premier lieu, il a été remonté que les règles de collecte nécessiteraient d'être définies dès le lancement de l'inventaire 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. Cela passerait par le renseignement précis d'un cahier des charges incluant les méthodes et protocoles de terrain à utiliser, les informations à récolter ou encore la manière de livrer les informations au commanditaire (référence aux standards d'échange). Une information nécessaire à la validation qui n'aurait pas été notée par l'observateur risquerait en effet de nuire à la validation future de la donnée.

(2) Ce lancement d'inventaire de terrain pourrait judicieusement être accompagné de sessions de formations des observateurs ou de préconisations quant à l'usage d'ouvrages de référence (flore, etc.) afin de limiter les erreurs de détermination.

#### **Solutions et/ou instance ou GT concerné**

*1, 2 : Animation « métier » (naturaliste)*

### **3. FORMATS DE DONNEES ET DE METADONNEES**

Les plateformes régionales et thématiques sont alimentées par des données sources des producteurs (DSP), par des données sources régionales ou thématiques (DSR/P) ou par des pré-données élémentaires d'échanges (Pré-DEE).

(1) La principale difficulté rencontrée par les gestionnaires ou administrateurs de données porte sur l'hétérogénéité des données reçues de la part des producteurs (DSP) et tient au nombre d'erreurs de normalisation (lieu-dit mentionné à la place de la commune, etc.). Certains producteurs ne sont pas familiarisés avec les concepts de standard d'échange de données ce qui nécessite une assistance non négligeable de la part des gestionnaires de données et administrateurs de la structure receveuse.

La culture de la métadonnée est globalement peu présente chez les producteurs et les jeux de données sont souvent transmis sans description associée ni renseignement du protocole utilisé. Les producteurs auraient donc besoin d'être mieux informés sur les métadonnées, leur intérêt, leur usage, leur format et de bénéficier, le cas échéant d'une assistance de la part d'un gestionnaire de données de plateforme R/T pour la standardisation.

(2) Afin de limiter ces hétérogénéités de format entre les producteurs et de permettre d'obtenir l'ensemble des éléments nécessaires à assurer la conformité future de la DEE, sa complétude et sa validation, les différentes plateformes pourraient préconiser aux producteurs l'usage de formats standard régionaux ou thématiques pour les données et pour les métadonnées, formats basés sur le standard d'échange de données élémentaire d'échange et sur le standard de métadonnées du SINP. Cela pourrait s'accompagner de préconisations quant à la manière de renseigner les informations.

#### Solutions et/ou instance ou GT concerné

1, 2 : Animation « plateforme »

#### 4. INFORMATIONS SUPPLEMENTAIRES ET TRACABILITE

**(1) Différents acteurs jugent qu'il serait utile de demander au fournisseur de données d'autoévaluer les données qu'il transmet sous la forme d'un niveau de validation producteur avec possibilité de commenter ce niveau.**

(2) Il serait alors pertinent de véhiculer, via les métadonnées, l'url du protocole de validation producteur.

**(3) Le producteur pourrait également mentionner qu'il demande une validation spécifique sur des données particulières (notamment lorsqu'il a un doute sur la donnée qu'il transmet).**

(4) Les résultats de l'enquête montrent que les producteurs auraient besoin de récupérer le résultat des validations régionales et nationales et d'être informés sur le fait qu'il manque par exemple des éléments pour valider la donnée.

(5) Les différents acteurs de la validation ont besoin d'être valorisés et que les experts pratiquant la validation soient reconnus. Cela passe notamment par la mise en place d'une chaîne de traçabilité optimale où les rôles de chacun dans la production de la donnée sont clairement mentionnés. La traçabilité, qu'il s'agisse de l'observateur, du déterminateur ou encore du validateur doit donc être assurée. Pour ce faire, il est nécessaire de prévoir le stockage d'un identifiant producteur, du nom de l'observateur, du déterminateur et du validateur.

**(6) Il est souhaité que données et métadonnées véhiculent la trace des validations successives (auteur(s) de la validation, date de validation, niveau de validité de chaque niveau, type de validation appliquée, protocole de validation suivi (automatique, manuel, les deux)) et éventuellement un commentaire sur la validation faite.**

(7) Il a été remonté que des informations nécessaires à la validation seraient absentes des standards d'échange actuels (données et métadonnées).

Les propositions d'ajouts de champs seront remontées aux GT dédiés. Ceux-ci arbitreront sur la pertinence d'ajouter ces champs et sur leur normalisation dans le cadre du SINP.

*La liste des éléments supplémentaires remontée par les enquêtés est disponible en annexe 1.*

(8) Si un certain nombre d'informations sont importantes pour l'expert, elles ne sont pas forcément disponibles, saisies, ou transmises par le producteur, d'où l'intérêt de travailler à la base en incluant ces informations à recueillir directement dans les protocoles d'acquisition.

#### Solutions et/ou instance ou GT concerné

1, 3, 6 : GT Validation

1, 3, 5, 6, 7 : GT Standard de données



*2, 7 : GT Standard de métadonnées*

*4 : GT GINCO/Plateforme*

*5 : Annuaire SINP*

*8 : Animation métier et plateforme*

*6 : GT Architecture*

## **5. OUTIL DE SAISIE**

De retour du terrain avec l'ensemble des éléments nécessaires, les producteurs ont besoin que ces informations puissent être saisies en bases. Cela concerne l'ensemble des informations relatives aux données elles-mêmes mais également les informations relatives aux métadonnées (protocole, vademecum de détermination utilisé, etc.).

Pour le producteur, la difficulté principale qui se dégage dans les réponses au questionnaire tient au peu de moyens alloués à la saisie, à la bancarisation et à la standardisation.

(1) Afin de garantir la création d'une donnée de qualité, les participants indiquent qu'il faudrait veiller à limiter, dans la mesure du possible, les erreurs dès la phase de saisie. Un outil naturaliste pourrait alors judicieusement inclure des contrôles de conformité et de cohérence dès la saisie des informations de manière à informer directement l'opérateur de saisie des manques ou des erreurs manifestes. L'outil pourrait également être doté d'un pré-contrôle de validation scientifique. Pour cela il devrait alors mobiliser un ensemble de couches de référence (selon les contrôles souhaités) et pouvoir interroger d'autres données du système. Le résultat de ce pré-contrôle serait ainsi visible pour l'opérateur de saisie qui pourrait, soit choisir de corriger les informations transmises, soit confirmer son observation.

### **Solutions et/ou instance ou GT concerné**

*1 : GT GINCO/Outil naturaliste*

## **6. OUTIL DE CREATION DES DEE ET METHODES DE TRANSFORMATION**

Les plateformes régionales et thématiques ont pour mission de bancariser les données transmises et de les transformer en DEE pour alimenter la plateforme nationale. Ce travail nécessite moyens et outils (ETL ou outil dédié).

(1) Les gestionnaires et administrateurs indiquent avoir besoin de partager, au sein du SINP, les règles de transformation des DSR/DST, notamment pour les choix de rattachement de données acquises sur une localité non administrative stricte à une localité admise dans le SINP. Plus largement, les plateformes ont besoin que les règles de croisements géographiques (croisement commune, mailles, etc.) soient partagées pour la mise en œuvre du floutage des données lors de la transformation en DEE.

### **Solutions et/ou instance ou GT concerné**

*1 : GT Architecture ou GT standard de données*

## 7. OUTIL DE VALIDATION ET SERVICES DEDIES

(1) Les DEE issues des plateformes régionales et thématiques devant être conformes et cohérentes, les gestionnaires et administrateurs de plateformes ont remonté le besoin de mettre en place des services de conformité et de cohérence, utilisables quel que soit l'outil utilisé.

(2) Pour mettre en œuvre la validation des données, les administrateurs de plateformes et gestionnaires de données (pour la partie automatique) ainsi que les experts (pour la validation manuelle) auraient besoin d'un outil simple et convivial facilitant le requêtage sur les données et le renseignement in fine du niveau de validité.

Ils indiquent leur besoin d'accès, via cet outil, à la consultation de l'ensemble des informations portées par les données et les métadonnées. Il leur serait nécessaire d'ajouter des couches d'information tierces (couches de référence, référentiels et bases de connaissance) et de pouvoir, de fait, identifier facilement les données en dehors de la distribution connue.

Ils auraient besoin de disposer d'un ensemble de filtres divers sur les données et métadonnées et sur les couches tierces ajoutées.

Cet outil devrait également permettre d'interroger les autres données du système de manière à permettre d'évaluer la congruence des données entre elles.

### Solutions et/ou instance ou GT concerné

- 1 : COPIL SINP & fonctionnalités de la plateforme nationale
- 2 : GT GINCO/Plateforme & fonctionnalités de la PF nationale

## 8. REFERENTIELS ET BASES DE CONNAISSANCE

(1) Les acteurs ont besoin d'avoir accès à l'ensemble des référentiels nécessaires (taxonomique, géographiques, etc.) pour assurer une standardisation et une exploitation.

(2) Les experts jugeraient utile de produire un référentiel des protocoles d'acquisition incluant une description détaillée du cadre méthodologique ainsi qu'un référentiel de détermination des taxons normalisé par groupe taxonomique.

(3) Les gestionnaires et administrateurs indiquent avoir besoin d'un référentiel observateurs et organismes.

(4) Pour faciliter le travail des experts et prioriser leur action, le besoin est identifié de créer des référentiels/bases de connaissance de cas pouvant permettre de filtrer les données afin d'identifier celles sur lesquelles il convient de porter attention (*exemple : taxon + lieu + date + statut biologique*).

(5) Les différents acteurs indiquent avoir besoin de couches de référence numérisées sur la répartition des espèces et leur probabilité de présence et ce, selon différents niveaux de résolution (commune, maille, département, en fonction de la résolution de la donnée à valider), régulièrement mises à jour

et versionnées. Cela pourrait notamment être réalisé sur la base de modèles probabilistes construits à partir des données disponibles déjà validées.

(6) Les opérateurs de validation indiquent également avoir besoin d'accéder à des référentiels/bases de connaissance sur la difficulté de détermination, sur la rareté, à une base de connaissance sur les confusions fréquentes, à des modèles de distribution, à des bases de connaissance des bornes extrêmes connues et à des bases de connaissance sur la phénologie.

(7) Ils indiquent qu'il leur serait nécessaire de disposer de couches d'occupation du sol, couches climatiques, couches pédologiques, aux réseaux hydrographiques, aux couches de relief, aux couches d'habitat, à la couche IGN des communes.

#### **Solutions et/ou instance ou GT concerné**

*1 : Fonctionnalités de la plateforme nationale. Les référentiels cités y sont d'ores et déjà diffusés.*

*2 : Travaux « métiers ». Travail sur un catalogue des méthodes actuellement en cours au MNHN*

*3 : GT GINCO. Travail prévu dans le cadre de l'annuaire SINP*

*4 : Travaux « métiers »*

*5 : Travaux « métiers ». Coordination MNHN pour l'élaboration des couches de référence nationale*

*6 : Travaux « métiers »*

*7 : Certaines couches sont accessibles, d'autres sont payantes*

## **9. MISES A JOUR**

(1) Pour mettre en place les formats régionaux et thématiques, les différents acteurs ont besoin de disposer de versions stabilisées du standard de données DEE et du standard de métadonnées et de pouvoir en prévoir les évolutions (fréquence de mise à jour et calendrier, lieu de publication).

**(2) Tous les acteurs ont besoins que le processus de mise à jour des validations et leur fréquence soit acté et décrit, qu'il s'agisse d'une mise à jour de routine car les connaissances évoluent ou d'une mise à jour liée à d'éventuelles modifications ou corrections de la DEE ou encore une mise à jour liée à la sortie d'une nouvelle version d'un référentiel (notamment taxonomique).**

(3) Ils ont besoin d'être alertés sur les modifications mises en œuvre entre les versions des différents référentiels (notamment taxonomique) et sur les éventuelles spécificités de gestion qu'elles induisent.

#### **Solutions et/ou instance ou GT concerné**

*1 : GT Standard de données et de métadonnées. Fonctionnalités de la plateforme nationale*

*2 : GT Validation/GT Architecture*

*3 : Taxref & Habref/ Fonctionnalités de la plateforme nationale*

## 10. DIFFUSION

(1) Les utilisateurs indiquent avoir besoin de visualiser l'ensemble des DEE selon les règles du protocole SINP. Ils ont besoin de pouvoir sélectionner les données qu'ils visualisent ou téléchargent selon différents attributs et notamment selon le niveau de validité (régional et national).

**(2) Ils ont également besoin de connaître l'échelle géographique (résolution) à laquelle la validation de la donnée a été faite.**

(3) Les utilisateurs souhaitent pouvoir alerter la plateforme qui diffuse les données sur une erreur ou faire une demande de validation spécifique sur des données particulières.

### **Solutions et/ou instance ou GT concerné**

*1 : COPIL SINP*

*2 : GT Validation, GT Standard de données*

*3 : Fonctionnalités de la plateforme nationale*

## 11. ANNEXES

### Annexe 1

#### Liste brute des éléments transmis via le questionnaire en matière de besoin d'information

**Rubrique** : besoin en termes d'informations portées par les données

Observateur
Identifiant de la base de données
Nom cité du taxon
Dates
Lieu-dit
Localisation GPS
Commune (code INSEE et libellé)
Précision sur la localisation (précise/imprécise)
Protocole d'acquisition de la donnée
Matériel utilisé (jumelle, longue vue, etc...)
Méthode de détermination
Bibliographie de référence /ouvrage pour l'identification
Bibliographie de référence pour la nomenclature
Identificateur
Organisme de l'observateur
Organisme de référence
Adresse mail de l'observateur ou du détermineur
Preuve/présence d'une preuve/liens vers la preuve
Commentaire permettant à l'observateur de mentionner que la donnée est « hors norme » dans les connaissances actuelles et qu'il en a conscience
Dépôt en collection
Numéro de collection
Collection dépositaire
Url forum
Date de publication

Référentiels utilisés
Type de suivi
Code des métadonnées
Stade de vie
Statut biologique
Statut de la donnée (publique/privée)
Statut d'indigénat
Sexe
Comportement de l'individu observé
Nombre d'individus
Milieu
Cortège d'espèces
Auto-évaluation de la certitude d'identification
Mention de la mise en œuvre d'une correction sur la donnée
Niveau de validation régionale
Niveau de validation nationale
Type de validation (automatique/manuelle)
Protocole utilisé pour la validation

**Rubrique** : besoin en termes d'informations portées par les métadonnées

Bibliographie de référence /ouvrage pour l'identification
Référentiels utilisés
Existence d'une validation « producteur »
Protocole de la validation « producteur »
Caractère public ou privé du jeu
Identification des différents acteurs et rôles (producteurs, commanditaires, financeurs, coordinateurs, etc.)
Résolution source et résolution de la DEE (savoir si l'information a été floutée)
Emprise taxonomique du jeu de données
Descriptif du cadre de récolte (programme, etc.)
Fiabilité du programme d'acquisition

Système de projection
Version du référentiel INSEE
Version du référentiel TAXREF
Existence d'un contrôle de conformité
Existence d'une validation
Historique de la validation
Coordonnées de la structure productrice des données
Droits d'utilisation
Emprise temporelle
Emprise géographique
Fiabilité du jeu : rapport données douteuses/données valides
Signification des attributs associés à des niveaux de validation

**Rubrique : besoin en termes de traçabilité**

Dates d'utilisation de la donnée
Trace des validateurs successifs et du déterminateur initial, de l'observateur
Gestionnaire des données sources
Nom de la plateforme par laquelle la donnée entre dans le SINP
Identifiant producteur
Identifiant interne de la plateforme
Identifiant SINP
Identifiant des intermédiaires
Niveau de validité régionale
Date de création de la donnée
Date de modification de la donnée s'il y a lieu
Type de validation appliquée (automatique, manuelle, les deux)
Date de la validation
Auteur de la validation
Commentaire sur la validation
Echelle d'exploitation pour laquelle la donnée a été validée

Producteur de la donnée
Reconnaissance des experts intervenant en fin de chaîne
Bibliographie lorsque la donnée en est issue
Niveau de validité nationale



## Annexe 2

### Questionnaire mis en ligne pour l'identification des besoins en matière de validation



**Identification des besoins en matière de validation**

L'objectif de ce questionnaire, produit dans le cadre du GT Validation des données d'occurrences du SINP, est d'identifier :

- Les grands types de besoins fonctionnels par grandes catégories d'acteurs dans le cadre du SINP
- Les priorités en matière de validation pour chaque type d'acteur
- Les difficultés fréquemment rencontrées dans la réalisation de la chaîne de contrôle (de la conformité à la validation)

Merci de votre participation

---

Définitions préalables:

Contrôles de conformité - conformité des données transmises au regard d'un standard de données. Ces vérifications peuvent consister à vérifier la présence de champs obligatoires, leur type, les règles et le format des fichiers échangés, mais peut également consister à vérifier des valeurs faisant appel à un vocabulaire contrôlé sous forme de liste de valeurs ou de référentiels.

Contrôles de cohérence - cohérence entre les informations transmises (ex : la date est inférieure à la date du jour, etc.)

Validation scientifique automatique - validation faisant appel à des résultats d'expertise amont (des référentiels, des bases de connaissance, etc.). Ce processus est réalisé de manière automatique (informatique)

Validation scientifique manuelle - validation des informations faisant appelle à une expertise aval et manuelle (dire d'expert suite à l'analyse des données transmises).

---

Plan du questionnaire:

- > Identité et participation au SINP
- > Identification des besoins et des pratiques (à renseigner selon les rôles exercés)
- > Identification des difficultés rencontrées (à renseigner selon les rôles exercés)

---

**NOM**

**Prénom**

**Adresse mail**

**Structure/Organisme**

**Vous ou votre structure est adhérent(e) au protocole SINP:**

**Participation à des groupes de travail SINP (plusieurs réponses possibles):**

- GT Architecture du SINP
- GT Base espaces protégés
- GT Données sensibles
- GT INSPIRE
- Projet OGAM
- GT Validation des données
- GT Standard de données
- GT Standard de métadonnées
- GT Juridique
- GT Gouvernance indicateurs
- GT I-BD<sup>2</sup>
- GT Indicateurs SNB
- GT Portrait
- GT Publics
- GT Réseau des observatoires
- Other:

**Vous êtes (plusieurs réponses possibles):**

- Producteur de données
- Gestionnaire de plate-forme SINP
- animateur politique SINP
- Administrateur de données (=Intermédiaire, autre gestionnaire, centralisateur de données, fournisseur de données, contributeur)
- Commanditaire (=propriétaire, maître d'ouvrage ou financeur)
- Expert
- Utilisateur de données
- Other:

## Identification des besoins et des pratiques

En tant que producteur, renseignez vos besoins/pratiques en matière de validation, de cohérence ou de conformité

En tant que gestionnaire de plate-forme SINP renseignez vos besoins/pratiques en matière de validation, de cohérence ou de conformité

En tant qu'animateur politique SINP renseignez vos besoins/pratiques en matière de validation, de cohérence ou de conformité

En tant qu'administrateur de données, renseignez vos besoins/pratiques en matière de validation, de cohérence ou de conformité

En tant que commanditaire, renseignez vos besoins/pratiques en matière de validation, de cohérence ou de conformité

En tant qu'expert, renseignez vos besoins/pratiques en matière de validation, de cohérence ou de conformité

En tant qu'utilisateur de données, renseignez vos besoins/pratiques en matière de validation, de cohérence ou de conformité

Besoin en termes d'informations portées par les données (P):

Besoin en termes d'informations portées par les métadonnées (P):

Besoin en termes d'outil (P):

Besoin en termes d'informations tierces (ex : couches géographiques, etc.) (P):

Besoin en termes de traçabilité (P):

Autre (P):

## Identification des difficultés rencontrées

**Vous êtes amenés à réaliser (plusieurs réponses possibles):**

- Des opérations de standardisation/normalisation
- Des contrôles de conformité
- Des contrôles de cohérence
- De la validation scientifique automatique
- De la validation scientifique manuelle
- Other:

## Pour les opérations que vous réalisez, indiquez les difficultés rencontrées ...

Pour la standardisation:

Pour les contrôles de conformité:

Pour les contrôles de cohérence:

Pour la validation automatique:

Pour la validation scientifique:

Commentaire libre:



# SINP

Système d'Information  
sur la Nature et le Paysage



Une multitude de jeux de données sur la nature et les paysages est produite chaque année par de nombreux acteurs sur les territoires français. Ces données sont collectées selon des méthodes et protocoles différents et pour des objectifs hétérogènes ; elles concernent également des périmètres taxonomiques et géographiques divers ; les compétences et moyens des observateurs sont tout aussi variables.

Dans le cadre du SINP, les différents utilisateurs des données doivent être en mesure de connaître le niveau de confiance associé à celles-ci. Cela permet notamment une bonne utilisation des données au regard des usages.

Ce document est produit dans le cadre du Groupe de Travail (GT) sur la validation des données d'occurrences du SINP. Pour mener à bien ses travaux, le GT doit être en mesure d'identifier les besoins des différents acteurs du SINP en matière de validation. La synthèse de l'identification des besoins est un produit de la phase 2 défini dans le mandat du groupe de travail.

Suite à la mise en ligne d'un questionnaire dédié à l'identification des grands types de besoins pour les différents acteurs du SINP, à l'identification des priorités en matière de validation et en la mise en exergue des difficultés fréquemment rencontrées par les acteurs pour la réalisation de la chaîne de contrôle (conformité, cohérence et validation scientifique), ce document présente la synthèse des éléments transmis par les enquêtés.



**MUSÉUM**  
NATIONAL D'HISTOIRE NATURELLE