



**Coordination scientifique et technique**  
Système d'Information sur la Nature et les Paysages  
Observatoire National de la Biodiversité



**AVIS DE LA CST**  
(SUITE A UNE SAISINE)

**Rappel des étapes :**

- l'avis est pré-rempli par le secrétariat de la CST et envoyé aux rapporteurs avec copie aux contributeurs
- le rapporteur principal, en lien avec le co-rapporteur, rédige une première version de l'avis et recueil des remarques et compléments des contributeurs. Pour assurer un suivi, merci de mettre en copie le secrétariat de la CST lors des échanges de ce document entre rapporteurs et contributeurs.
- une fois consolidé, l'avis est transmis par le rapporteur principal au secrétariat de la CST :

Adresse courriel secrétariat : [secretariatcst@mnhn.fr](mailto:secretariatcst@mnhn.fr)

<b>Numéro de saisine</b>	11 [10/10/2011]
<b>Titre</b>	Schéma global de validation des données du SINP
<b>Rappel des termes principaux de la saisine</b>	<p><b>Question principale :</b> le mandat du groupe de travail associé à l'action est-il adapté pour répondre à la question posée, aux regards des enjeux scientifiques et techniques décrits dans le mandat ?</p> <p><b>Sous-questions :</b></p> <ul style="list-style-type: none"><li>- la composition du groupe de travail est-elle satisfaisante ?</li><li>- les objectifs du groupe de travail sont-ils réalistes ?</li><li>- le calendrier proposé et les livrables permettront-ils de mener à bien le projet ?</li></ul> <p><b>Question optionnelle :</b> La CST peut formuler des recommandations sur la façon de mettre au point un schéma global de validation qui garantisse la fiabilité et la qualité scientifique des données référencées dans le cadre du SINP ?</p> <p><b>Sous-questions :</b></p> <ul style="list-style-type: none"><li>- Quels pourraient être les critères de validation et de qualification des données naturalistes, au niveau local, régional, national ? D'un point de vue scientifique ? D'un point de vue statistique ?</li><li>- Serait-il pertinent de mettre en place une normalisation de type ISO ou une labellisation (interne au SINP) du processus de validation des données ?</li></ul>
<b>Rapporteurs</b>	Frederic <b>GOSELIN</b> , Vincent BOULLET, Amandine ORSINI
<b>Contributeurs</b>	Patrice PRUVOST, Pascal DUPONT, Benoit DAVID, Thomas BOUIX, Pierre MIGOT, Antoine HUGUET, Alain PIBOT

Historique du document

Version	Auteur	Date	Commentaires
V0	Secrétariat CST	13/01/2012	Envoi du document type
V1	Frédéric Gosselin		Rédaction de la première version
V2	Amandine Orsini		Relecture
V3	Thomas Bouix		Ajouts
V4 - consolidée	Frédéric Gosselin		Mise en cohérence
	Pierre Migot		Rappelle ce qu'il a dit en CST
V5-finale	Frédéric Gosselin sur la base des remarques de Vincent Bullet	19/09/2012	Cf. mail du 10 juillet + CR CST

...

### Préconisations à adapter selon la saisine

L'avis doit être synthétique (1 à 3 pages) avec des renvois en annexes si nécessaire (commentaires détaillés...).

Les différents avis des rapporteurs et contributeurs doivent apparaître clairement et être synthétisés sous forme en faisant apparaître les points de consensus et les points de divergences.

On peut conseiller deux approches :

- une analyse par document soumis
- une analyse par question et sous-question de la saisine

Dans tous les cas, une conclusion est demandée afin de répondre aux interrogations de la saisine.

Les conseils et recommandations dépassant le cadre de la saisine doivent bien évidemment être conservés mais seront synthétisés dans un ou des paragraphes spécifiques.

## *Avis*

Nous avons choisi d'organiser notre avis en traitant les deux questions posées successivement : La question principale et ses sous-questions étaient :

**Question principale** : le mandat du groupe de travail associé à l'action est-il adapté pour répondre à la question posée, aux regards des enjeux scientifiques et techniques décrits dans le mandat ?

**Sous-questions** :

- la composition du groupe de travail est-elle satisfaisante ?
- les objectifs du groupe de travail sont-ils réalistes ?
- le calendrier proposé et les livrables permettront-ils de mener à bien le projet ?

Le premier point sur lequel nous souhaiterions insister est **que la question posée au groupe de travail – et les objectifs lui sont assignés – ne sont pas suffisamment explicités**. Il semble que parmi ces objectifs figurent :

- "l'utilisateur final a besoin de connaître le niveau de confiance associé à chaque jeu de données disponible.

- "la standardisation des définitions, des critères de qualification, des niveaux de confiance taxonomiques, géographiques, méthodologiques... est aujourd'hui nécessaire dans le cadre des avancées du SINP"

- une démarche partagée aboutissant à un consensus.

Si ces éléments semblent globalement pertinents, **nous nous demandons si la question n'est pas incomplète** et s'il ne faut pas ajouter au critère de confiance – associé à la notion de qualité des données – la question de savoir ce que signifient les données – associée notamment au temps de prospection, au plan d'échantillonnage et – pour les espèces – l'existence de données d'absence. Il y aurait ainsi besoin

dans ce mandat d'une prise en compte de la donnée dans son itinéraire global, incluant la méthodologie associée à la prise de données.

Notre première recommandation est donc de préciser la question – ce qui pourrait être une première étape de la démarche.

De manière plus générique pour le fonctionnement de la CST on peut se demander si pour ce type de demandes il ne faudrait pas inclure une première question à la CST, pour préciser les formulations qui paraissent scientifiquement intéressantes (cf. discussions CST du 8/03).

Cette saisine se situe entre les deux, peut être à l'origine de la difficulté d'émettre un avis.

Le Second point est qu'on ne comprend pas très bien quels sont **les moyens alloués** à ce groupe de travail (en termes de compétence, un peu, mais en termes d'ETP, ce n'est pas du tout clair). Toujours sur la question des moyens – non pas du groupe de travail mais alloués à la validation des données du SINP – il nous paraît nécessaire que le groupe de travail propose un (ou des) scénarios se positionnant clairement, soit dans des solutions les plus satisfaisantes scientifiquement, indépendamment du coût nécessaire à la mise en œuvre, soit dans des solutions dans un cadre de moyens et de temps définis préalablement (par les pouvoirs publics ?) qui conditionnent des choix méthodologiques et scientifiques pour la validation

De plus, la rédaction des phases de travail laisse penser que la phase 2 mélange deux étapes :

\* l'étape mentionnée dans le titre de la phase 2 "recensement des besoins et des critères", phase qui nous semble tout à fait utile et a priori bien placée après la phase 1 (relative au recensement de l'existant) ; lors de cette étape, il serait utile de préciser les attentes et la terminologie utilisées ;

\* l'étape mentionnée sous le titre de la phase 2 - étape qui consiste à jauger les outils/procédures identifiés dans la phase 1 aux besoins et critères définis dans la phase 2. Nous distinguerons donc cette troisième phase de la phase 2.

Au sujet des sous-questions sur le **calendrier, les livrables et les objectifs**, nous avons donc du mal à répondre, les moyens n'étant pas clairement connus et la question n'étant pas très claire non plus. Ce qui nous semble évident c'est qu'on ne peut pas espérer avoir produit à la fin de l'année un guide qui fasse le tour de la question. Les rédacteurs en sont d'ailleurs conscients puisqu'ils décrivent un ordre de priorité des données à aborder (haut de la page 2).

Les phases 1 et 2 (nouvelle version) nous semblent très importantes – il faut bien connaître ce qui se fait en France et à l'étranger, bien définir les différentes notions associées à la validation des données du SINP, puis bien recenser les besoins et les critères ; peut-être faudrait-il leur donner plus de temps.

Peut-être faut-il enfin concevoir ce projet comme devant être évolutif et donc clairement afficher qu'il devra être continué à la fin de l'année, non seulement en accompagnement comme c'est précisé, mais en réflexion et développement.

Au sujet de **la sous-question de la composition groupe de travail**, si l'objectif est d'avoir une démarche partagée qui aboutisse à un consensus, il nous semble qu'il faudrait ouvrir le 1er cercle à certains des organismes du second cercle - si ceux-ci l'acceptent – nous pensons notamment à la LPO et à des représentants de plateformes régionales. Il faudrait probablement aussi ouvrir la liste à d'autres établissements scientifiques (CNRS notamment) et à d'autres organismes de suivi de la nature (IGN-IFN?), voire à des partenaires étrangers, peut-être francophones.

Au total, nous manquons donc d'éléments pour répondre à la question principale. Les principaux éléments manquants sont les moyens alloués à ce groupe et une question de départ plus précise. Nous avons néanmoins fait des propositions en termes de calendrier, de définition des phases du travail et de composition du groupe de travail, qui doit probablement être revu.

Au sujet de **la sous-question de la normalisation ou la labellisation du processus de validation**, il faudrait y inclure la qualification : c'est même sur ce point que devrait porter la normalisation ou la labellisation, la validation en étant dépendante.

Nous en venons à la question optionnelle :

**Question optionnelle :**

La CST peut formuler des recommandations sur la façon de mettre au point un schéma global de validation qui garantisse la fiabilité et la qualité scientifique des données référencées dans le cadre du SINP ?

**Sous-questions :**

- Quels pourraient être les critères de validation et de qualification des données naturalistes, au niveau local, régional, national ? D'un point de vue scientifique ? D'un point de vue statistique ?
- Serait-il pertinent de mettre en place une normalisation de type ISO ou une labellisation (interne au SINP) du processus de validation des données ?

Quelques pistes sur ces points :

- pourquoi ne pas employer des catégories de probabilité d'erreur des différents processus de production de la donnée (observation, détermination, saisie...)
- un critère "complémentaire" pour "valider" ou qualifier les données peut être de mesurer plusieurs fois la même chose avec différents opérateurs ou systèmes. On pourrait ainsi avoir une idée du niveau d'erreur global, voire proposer des systèmes de validation ou qualification intégrant explicitement ce principe, comme c'est fait dans certains systèmes de suivi de biodiversité (par exemple, Lépidoptères en Grande-Bretagne).
- faut-il inclure les "collections de référence" (i.e. ici, stockage des organismes déterminés pour possibilité de vérification future) dans le périmètre du groupe de travail ?
- il devrait être possible d'intégrer une auto-qualification des données par le preneur de données suivant une méthodologie qui définit des questions et des critères qui permettent au preneur de la donnée de qualifier celle-ci.

## ***Hors Avis***

Eléments supplémentaires hors saisine à verser au dossier du futur groupe de travail :

Même si ce n'est pas le rôle de la CST sur une saisine, voici des propositions d'éléments à prendre en compte :

- Faire le lien avec les métadonnées nécessaires pour valider les données (protocole, inventaire, plan d'échantillonnage et quantités associées...)
- Distinguer « contrôle de cohérence » des données : à faire systématiquement et « validation » sensus stricto qui peut analyser la qualité de l'inventaire réalisé (cf. remarque sur système de validation)
- Détermination de la qualité de l'inventaire -à distinguer de la validité de la donnée elle-même - portant sur : la pression d'inventaire, le taux de détection, le protocole utilisé...
- Nécessité de « valider » les protocoles s'ils sont précisés (labellisation ?)
- Identifier les méthodes de validation internes des producteurs quand elles existent (à décrire dans les métadonnées ?)
- Proposer une « validation » sur un jeu de donnée (pas sur la donnée)

De manière plus générale les producteurs de données naturalistes du SINP sont divers et il faut un système de contrôle qui valorise les producteurs et ne les excluent pas : a priori toute donnée est intéressante, quel que soit son niveau de validation.

- Quelques règles de base (contrôle de cohérence sur les données naturalistes à l'ONF)

Les points clés à prendre en considération pour un contrôleur sont, dans l'ordre :

- L'auteur : les compétences reconnues de l'auteur ;
- La facilité de déterminations du taxon (caractéristiques morphologique et risque de confusion, souvent à associer avec la zone géographique) ;

- La chorologie de l'espèce :
  - la présence d'observations déjà signalées dans le secteur géographique (massif forestier) ou
  - la cohérence avec l'aire de répartition française connue ;
- D'autres critères écologiques : saison et type d'habitat.
  - Les données fournies par un observateur confirmé sont à valider, sauf si elles s'avèrent aberrantes d'un point de vue chorologique ou écologique.
  - Les données fournies par un observateur non spécialiste sont à valider si l'espèce est simple à déterminer dans le contexte géographique
  - Les données fournies par un observateur non spécialiste pour une espèce moyennement simple à déterminer sont à valider si elles se situent dans une station avérée de l'espèce

○