



# SINP

Système d'Information  
sur la Nature et le Paysage



## GT « Validation des occurrences »

### Volet 1 : occurrences de taxons

#### Réunion 5

*Le 30/11/2015, Salle Baleine 1 /MNHN*

---

**Rédaction :** Solène Robert (MNHN)

**Relecteurs :** Guillaume Body (ONCFS), Pierre Cellier (Tela Botanica), Dorothee Meyer (DREAL PACA),  
Grégory Caze (CBN SA).

**Pièce-jointe :**

- Support de discussion présenté lors de la réunion (GT\_VALIDATION\_SINP\_20151130.pdf)
-

**Invités :**

| <b>ORGANISME/STRUCTURE</b>                        | <b>Prénom NOM</b>      | <b>Présents</b> |
|---|------------------------|-----------------|
| MEDDE   | Yannick LEBEAU         | X               |
| MNHN  | Solène ROBERT          | X               |
|   | Laurent PONCET         |                 |
|   | Julien TOUROULT        |                 |
|   | Remy JOMIER            | X               |
|   | Jeanne de MAZIERES     |                 |
| FCBN & CBN  | Bruno DUTREVE          |                 |
|   | Alexis DESSE           |                 |
|   | Johan GOURVIL          |                 |
|   | Anaïs JUST             | X               |
|   | Grégory CAZE           | X               |
|   | Jérôme MILLET          |                 |
| Plateforme régionale [Réunion]                    | Frédérique ZELMIRE     |                 |
| Plateforme régionale [Haute Normandie]            | Arnaud SALTRE          |                 |
| Plateforme régionale [Aquitaine]                  | Andy PAPACOTSIA        |                 |
|   | Marie BARNEIX          | X               |
|   | Paul FROMAGE           |                 |
| Plateforme régionale [Languedoc Roussillon]       | Mathieu BOSSAERT       |                 |
| Plateforme régionale [Provence Alpes Côte d'Azur] | Dorothee MEYER         | X               |
| Plateforme régionale [Bourgogne]                  | Damien MARAGE          |                 |
| ONCFS   | Guillaume BODY         | X               |
| SFPEM   | Christian ARTHUR       |                 |
|   | Stéphane AULAGNIER     |                 |
| SHF   | Jacques CASTANET       |                 |
|   | Jean LESCURE           |                 |
| LPO   | Dominique ARIBERT      |                 |
|   | Gwenaël QUANTENNE      |                 |
| RESOMAR   | Nicolas LAVESQUE       |                 |
|   | Serge HEUSSNER         |                 |
|   | Antoine GREMARE        |                 |
|   | Nathalie SIMON         |                 |
| IFREMER   | Antoine HUGUET         |                 |
| ANTAREA   | Rumsaïs BLATRIX        |                 |
| Observatoire Pelagis                              | Vincent RIDOUX         |                 |
| Tela Botanica                                     | Pierre CELLIER         | X               |
| GBIF France                                       | Anne-Sophie ARCHAMBEAU |                 |
|   | Sophie PAMERLON        |                 |
|   | Marie-Elise LECOQ      |                 |

## Ordre du jour

- 1) Point d'information
- 2) Propositions de modification du guide méthodologique
- 3) Listes de valeurs: libellés et définition
- 4) Calendrier

### Point d'information

La synthèse des besoins est en ligne sur le site de Nature France.

Deux nouveaux organismes (Tela Botanica et DREAL Bourgogne) sont intégrés au cercle 1 à leur demande.

Les travaux du GT et leur avancement ont été présentés à la CST SINP/ONB du 9 novembre.

La démarche de validation SINP a été présentée en formation IFORE le 25 novembre.

### Propositions de modification du guide méthodologique

Les participants estiment globalement le guide clair et pertinent.

Les questions/demandes de précisions posées sont les suivants :

- ➔ **Quid de la diffusion ?** Diffusion des données non validées et règles de diffusion partagées par les plateformes.

Les discussions sur les modalités générales de diffusion par les plateformes en fonction des types d'acteurs ne sont pas encore arrêtées (consultation en cours). Ces décisions sont nécessaires pour pouvoir traiter cette question. Dans l'immédiat, le guide V1 ne traitera pas de ce point et le mentionnera clairement en préambule. Les discussions sur cette question mériteront d'être reprises dans un second temps (perspective pour 2016).

- ➔ **Quelle est la périodicité prévue pour les échanges et les mises à jour entre les plateformes ?**

Cela ne relève pas du GT Validation mais est du ressort du GT Architecture. La question sera transmise à ce groupe de travail.

- ➔ **Qu'entend-t-on par regroupement ?**

Le concept de regroupement est un nouveau concept (facultatif) du standard de données. Il ne s'agit pas d'une ouverture sur la possibilité de synthèse ou d'agrégation, mais simplement d'une possibilité permettant de rassembler des observations (les lier), tout en gardant à chacune ses caractéristiques individuelles complètes. Cela sera précisé dans le guide pour éviter toute autre interprétation.

Les propositions de modification et d'amélioration proposées et discutées sont les suivantes :

- ➔ **Préciser les terminologies** pour :
- plateformes régionales et thématiques et « pôles régionaux »
  - jeux de données et dispositifs de collecte
  - concept de regroupement

Proposition retenue par le groupe sous réserve que ces termes aient été définis par les GT dédiés. Le GT Validation ne peut produire des définitions sur des concepts qui dépassent son périmètre de travail.

Pour rappel, les définitions sont aussi diffusées au sein du glossaire de Nature France.

- Le guide s'appuie sur divers travaux et rapports existants. Lors ce que cela est possible, **reprendre les éléments de ces rapports au sein du guide** pour en faciliter la lecture et la compréhension.

Proposition retenue par le groupe avec le bémol de la maintenance. Il sera ajouté la référence au document source en précisent que celui-ci peut évoluer.

- **Ajouter en annexe la liste des GT existants**, leur niveau d'avancement et le responsable de chaque groupe.

Cela dépasse le cadre du GT Validation. Un tel tableau évolutif ne peut figurer au sein du guide. La proposition n'est pas retenue.

Cependant, les avancés des différents GT sont consultables sur le site de Nature France (Avancement).

- **Indiquer que les plateformes R/T peuvent mandater les pôles régionaux** pour la réalisation des opérations de validation.

Proposition retenue par le groupe.

- **Ajouter des exemples de protocole de validation scientifique** (exemples concrets) dans la partie dédiée.

La démarche étant nouvelle, il n'existe pas à notre connaissance de protocole de validation respectant complètement les règles établies par le groupe (Cf. analyse de l'existant précédemment réalisée). En l'état il est donc délicat d'y faire référence.

Proposition non retenue par manque de référence adéquate.

- Pour la partie « règles de cohérence »: considérant qu'un protocole de conformité et de cohérence national devra être produit, **il faut considérer les règles énoncées dans le guide comme des exemples** et non comme un ensemble de règles exhaustives. Il est donc proposé de laisser ces informations tout en précisant clairement son statut d'exemple illustratif.

Proposition retenue par le groupe.

- **Supprimer l'obligation de validation indépendante de l'observateur.**

Le groupe s'accorde sur le fait qu'une validation faite par un tiers est la plus pertinente mais la validation « producteur » doit aussi permettre au producteur de mentionner (lui-même) qu'il a un doute sur les données qu'il transmet. Par ailleurs, il peut arriver que des experts soient eux même des producteurs de données et qu'il n'existe pas d'autre expert disponible sur ce taxon/groupe taxonomique.

La proposition est retenue par le groupe. La définition sera modifiée.

- **Proposition d'ajouter un contrôle de doublon à l'entrée des plateformes R/T.**

Le contrôle porterait *a minima* sur l'identifiant unique (associé à la date de mise à jour de la donnée). Un contrôle de redondance plus large et supplémentaire, pour identifier plus finement les doublons (sur un ensemble d'attributs) pourrait être mis en place librement par les plateformes, ne prenant pas en compte cet identifiant unique.

La proposition est retenue par le groupe. Une partie sera ajoutée dans le guide sur ce sujet.

- **Proposition de définir des liens entre validation de périmètre minimal et validation de périmètre élargie**

La validation de périmètre minimal doit porter sur les informations « quoi », « où » et « quand » uniquement. La proposition relative à l'ajout du « qui » dans ce périmètre est rejetée parce que jugée non pertinente pour une validation scientifique (on ne valide pas scientifiquement un observateur), et parce que le contrôle relatif à la présence d'un observateur est déjà porté dans le contrôle de conformité.

Si la validation de périmètre minimal porte sur ces trois informations et peut être traitée comme telle de manière automatique, il est en revanche d'usage que les experts intervenant sur la validation manuelle considèrent l'ensemble des informations transmises pour statuer sur le niveau de validité minimal. Les experts, au regard de la cohérence des autres attributs, vont disposer d'éléments pour orienter leur validation de périmètre minimal. Concrètement, même si la validation ne porte que sur ces trois informations, elle peut tout de même être réalisée en considérant les autres champs. Cela est de nature à générer des disparités entre les validations expertes des différents niveaux (régional, national), puisque subjectives et basées sur l'interprétation de l'expert à partir des informations disponibles mais c'est là le propre d'une validation humaine.

La proposition de limiter l'accès des experts sur ces trois informations pour la validation minimale n'est pas retenue.

Pour la question de la connexion (rétro-application) entre les validations minimales et étendues (dans les deux sens), la proposition de fixer les liens dans le guide n'est pas retenue. Le choix de ces rétroactions est laissé à la liberté des plateformes en fonction des protocoles qu'elles appliquent pour ces deux périmètres. Les connexions devront cependant être décrites au sein des protocoles de validation.

La validation de périmètre élargi permet notamment au SINP de s'appuyer sur des démarches de validation existant dans le cadre de programmes spécifiques (inventaires, atlas, etc.). L'utilisation de ces protocoles devra être décrite dans le protocole de validation « SINP » et préciser son implémentation pour la validation SINP.

- ➔ **Proposition de retirer la partie sur les dépendances entre données, jeux de données et dispositifs de collecte** considérant que cette question n'est pas encore complètement figée et que le GT Métadonnées est encore ouvert.

Proposition non retenue. Ce point sera laissé dans le guide mais en faisant néanmoins référence au GT Métadonnées et en mentionnant une réserve quant au choix final dont la responsabilité est portée par ce groupe dédié.

- ➔ **Proposition de ne pas trop rentrer dans les détails de la démarche de mise à jour considérant que cela est du ressort du GT Architecture.**

Proposition retenue. La partie est maintenue mais ne décrira pas plus finement la démarche.

## Listes de valeurs: libellés et définitions

L'implémentation (codages et nomenclatures) est du ressort du GT standard de données.

La définition plus fine des niveaux est du ressort de chaque plateforme qui décrira alors les règles d'attribution de ces niveaux au sein de son protocole de validation. Ces règles ne doivent cependant pas dénaturer la définition originale (générique) définie par le groupe de manière à ne pas engendrer de disparités manifestes entre les plateformes.

Le niveau de validité correspond à un état à un moment donné et doit donc toujours être associé à une date de validation.

Discussions sur la base des propositions préliminaires de définition présentées en séance. (Cf support de présentation) :

- La proposition de sectorisation des valeurs selon quatre niveaux est jugée pertinente.
- Les deux niveaux supplémentaires « Non réalisable » et « Non réalisé » sont acceptés par le groupe.
- Le groupe considère qu'il est très difficile de formuler une définition générique qui convienne à la fois à un processus de validation manuel/combinatoire et à un processus de validation automatique. Il est donc proposé de distinguer la définition (voire le libellé) en fonction du processus appliqué. Les discussions relatives aux libellés ne trouvent pas consensus en séance. Le MNHN enverra prochainement aux membres du groupe une synthèse des scénarii discutés en séance. Les membres du groupe pourront réagir directement sur cette synthèse. Le guide méthodologique devra ensuite intégrer le résultat de ce travail.

## Calendrier

Les modifications proposées et discutées en séance, ainsi que les retours ayant été transmis par mail seront pris en compte pour produire une nouvelle version du guide. Le MNHN proposera (mi-décembre) cette nouvelle version au groupe pour un ultime tour de relecture.

Le groupe aura jusqu'à début janvier pour transmettre ses retours.

Une V1 sera alors produite et soumise à relecture par la CST (saisine).

## Point complémentaire

Un problème a été détecté pour la diffusion des informations par le MEDDE quant à la sortie de la nouvelle version du standard de données (DEE).

Le nouveau standard de données est actuellement diffusé sur le site de Nature France et le MEDDE a diffusé l'information dans la lettre d'information SINP/ONB. Le MEDDE indique qu'il rencontre des problèmes liés à la taille de la liste de diffusion qu'il utilise et qu'il tâchera de les résoudre.

Le MNHN propose, dans l'immédiat et pour pallier temporairement à ce constat, de transmettre la nouvelle version du standard aux membres du groupe validation.