



## Gt « Qualification des données biodiversité du SINP »

### Seconde réunion du groupe

Le 15/01/2015, Salle Jean Dorst/MNHN

**Rédaction :** Solène Robert

**Relecteurs :** Yannick Lebeau, Dorothee Meyer, Philippe Landry, Laurent Poncet, Johan Gourvil, Jeanne de Mazières, Alexis Desse

**Pièce-jointe :**

- Support de discussion présenté lors de la réunion (GT\_QUALIF\_SINP\_20150115.pdf)

**Invités :**

ORGANISME/STRUCTURE	Prénom NOM	Présents
MEDDE	Yannick LEBEAU	X
MNHN	Solène ROBERT	X
	Laurent PONCET	X
	Julien TOUROULT	X
	Jeanne de MAZIERES	X
	Bruno DUTREVE	
FCBN & CBN	Alexis DESSE	X
	Johan GOURVIL	X
	Anaïs JUST	X
	Grégory CAZE	
	Jérôme MILLET	
Plate-forme régionale [Réunion]	Frédérique ZELMIRE	<i>Excusé</i>
Plate-forme régionale [Haute Normandie]	Arnaud SALTRE	X
Plate-forme régionale [Aquitaine]	Andy PAPACOTSIA	<i>Excusé</i>
	Marie BARNEIX	X
Plate-forme régionale [Languedoc Roussillon]	Mathieu BOSSAERT	
Plate-forme régionale [Provence Alpes Côte d'Azur]	Dorothee MEYER	X
ONCFS	Philippe LANDRY	X
	Pierre MIGOT	
SFPEM	Christian ARTHUR	
	Stéphane AULAGNIER	
SHF	Jacques CASTANET	X
LPO	Dominique ARIBERT	<i>Excusé</i>
	Gwenaëli QUAINTENNE	X
RESOMAR	Nicolas LAVESQUE	
	Serge HEUSSNER	
	Antoine GREMARE	
	Nathalie SIMON	<i>Excusé</i>
IFREMER	Antoine HUGUET	
ANTAREA	Rumsaïs BLATRIX	
Observatoire Pelagis	Vincent RIDOUX	
GBIF France	Anne-Sophie ARCHAMBEAU	<i>Excusé</i>
	Sophie PAMERLON	X
	Marie-Elise LECOQ	X

## Ordre du jour

1. Rappel du mandat
2. Finalisation de la phase 1
3. Terminologie et périmètre
4. Nom du groupe de travail
5. Définition des besoins
6. Démarche générale
7. Critères/paramètres

### 1. Rappel du mandat

Objectifs des travaux du GT : Porter à connaissance le niveau de confiance associé aux données de manière à permettre une bonne utilisation au regard des différents usages

#### Phase 1 :

Objectif	Produit
Recenser les travaux existants en la matière au niveau national et international	Synthèse avec proposition de définitions des grandes notions

#### Phase 2 :

Objectif	Produit
Partage des définitions conceptuelles	Alimentation du glossaire SINP
Définition du périmètre	Liste d'exigences fonctionnelles
Identification des besoins	
Partage d'une démarche globale	

#### Phase 3 :

Objectif	Produit
Description fine de la démarche	Guide méthodologique de référence (soumission à la CST)
Analyse de la cohérence et de la complémentarité entre les échelles de validation retenues	
Définition des critères et paramètres à prendre en compte	
Propositions à soumettre aux autres GT	

### 2. Finalisation de la phase 1

Le recensement de l'existant est le produit de la phase 1 défini dans le mandat.

Il a donné lieu à un appel à contribution à l'ensemble des contacts du cercle 1 identifiés dans le mandat du GT qui ont également fait appel à leur réseau.

Ce recensement a été soumis aux contributeurs pour validation des informations issues des documentations transmises et pour compléter le cas échéant. 17 systèmes ont été mis à contribution.

Un délai est mis en place pour les relectures des contributeurs (jusqu'au 26 janvier). Un mail (Solène Robert) sera envoyé en ce sens.

Suite à la centralisation des retours, une version 0.2 sera transmise à l'ensemble des membres du cercle 1 pour relecture (aux alentours du 5 février). Il sera laissé trois semaines de relecture avant centralisation des retours et diffusion d'une V1 du document qui sera mis en ligne sur le site de Nature France dans la rubrique Action > GT Validation.

### 3. Terminologie et périmètre

Plusieurs propositions de modification sur les définitions rédigées suite à la première réunion du groupe et sur le périmètre en cours de définition :

- ➔ **Remplacement du terme « Statut » de validité par « Niveau » de validité** ; ce terme portant mieux l'aspect graduel (notion de seuils) tel que défini par le groupe.
- ➔ **Faire figurer sur le schéma la rétroaction de la qualification sur la validation.** La validation se base, pour automatiser, sur des couches de référence et des référentiels (expertise amont). La validation permet d'améliorer par exemple la production des couches de référence, la mise à jour du référentiel taxonomique, etc. qui permettent à leur tour de mieux réaliser l'opération de validation.
- ➔ **Mieux distinguer le concept de conformité du concept de cohérence dans les définitions établies.**

Nouvelle proposition de définition :

#### Vérifications de conformité

Cette étape consiste à vérifier la conformité des fichiers échangés au regard des standards d'échange du SINP (données et métadonnées). Ces vérifications concernent les aspects physiques et conceptuels et consistent ainsi à vérifier la présence des champs obligatoires, leur type, les règles et le format des fichiers échangés, mais également la vérification des valeurs faisant appel à un vocabulaire contrôlé sous forme de liste de valeurs ou de référentiels.

#### Vérifications de cohérence

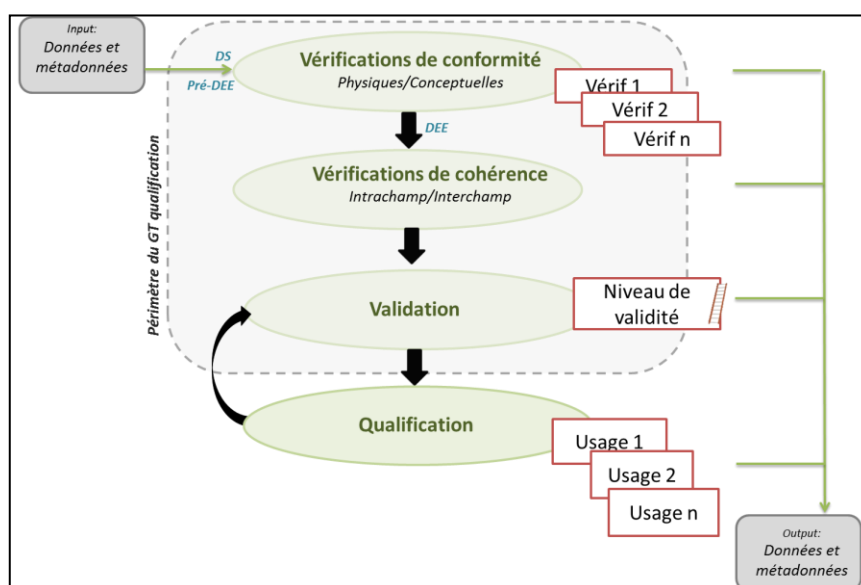
Cette étape consiste à vérifier la cohérence de chaque champ obligatoire (ex : la date est inférieure à la date du jour) ainsi que la cohérence inter-champs (ex : si la localisation a été transmise au point et à la commune alors on vérifie si le point cité (X/Y) est bien situé sur la commune citée (code INSEE).) Les résultats sont multiples, c'est-à-dire que le processus donne lieu à plusieurs résultats binaires (apte / non apte) de vérification par rapport à des règles de gestion.

➔ **Intégrer les vérifications de conformité dans le périmètre du groupe.**

Les vérifications de conformité sont considérées comme un préalable à l'intégration dans le SINP. Les DEE ne pouvant être valides que si elles répondent aux normes définies (cf. standards d'échanges). Les plateformes R/T sont responsables de la standardisation des données sources (DS) ou pré-DEE en données élémentaires d'échange (DEE).

Cette étape de vérification de la conformité est identifiée dans la chaîne de traitement. Elle sera donc décrite dans la note finale (livrable du GT) (en faisant référence aux documents standards), et pourra faire état de signalements aux GT standards le cas échéant mais ne donnera pas lieu à une méthodologie spécifique dans le cadre du groupe de travail.

*Nouvelle proposition de schéma général du processus de contrôle*



Il est proposé que ces modifications soient validées via la validation de ce compte-rendu et que le glossaire du SINP soit mis à jour après intégration et validation de ces modifications.

#### **4. Nom du groupe de travail**

Le périmètre du GT, tel qu'il est à présent défini, ne concerne pas l'utilisation des données en fonction d'un usage (défini sous le terme de qualification). De fait, l'intitulé du groupe porte à confusion et son renommage est proposé de la manière suivante :

**GT Validation des occurrences / Volet 1 : Occurrences de taxons**

Ce changement d'intitulé nécessite la révision du mandat qui intégrera, de fait, les terminologies présentement définies.

## 5. Définition des besoins

La phase 2 du mandat intègre l'identification des besoins fonctionnels : Issus de la réflexion collective du groupe de travail, devront être identifiés et formalisés les besoins des différents acteurs du SINP en matière de validation.

Liste des acteurs du SINP identifiés (concerne toute la chaîne du flux de données):

- Producteur de données
- Administrateur de données / intermédiaire (autre gestionnaire) /Centralisateur de données/Fournisseur de données / Contributeur
- Commanditaire (= « propriétaire », « maitre d'ouvrage » ou « financeur »)
- Gestionnaire de plate-forme
- Expert en charge de la validation
- Utilisateur des données

Suite à discussion, il est défini que l'identification des besoins passera par le biais d'un questionnaire qui circulera en interne (les membres du cercle 1 représentant d'ores et déjà un large panel d'acteurs). La possibilité est cependant laissée aux différents membres de contacter d'autres partenaires (éventuellement par le biais d'entretiens téléphoniques) de manière à couvrir de manière uniforme le panel d'acteurs identifiés.

L'objectif est d'identifier les grands types de besoins fonctionnels par grandes catégories d'acteurs. Exemple : un expert aura besoin de telles informations pour assurer la validation manuelle, un usager aura besoin du niveau de validité et devra avoir les outils pour requêter facilement les données selon le / les niveaux qu'il souhaite conserver, etc.

Une ébauche de questionnaire sera prochainement transmise (Solène Robert) au cercle 1 pour contribution à son élaboration, validation puis renseignement.

Un recensement des besoins sera rédigé suite à centralisation des questionnaires.

## 6. Démarche générale

### **Vérifications de conformité :**

Sont réalisées par :

- Plate-formes régionales
- Plate-forme thématique occurrence

La plate-forme nationale pourra réaliser un contrôle complémentaire de conformité mais ce sont bien les plate-formes R/T qui portent la responsabilité de la conformité des données élémentaires d'échange.

### **Vérifications de cohérence :**

Sont réalisées par :

- Plate-formes régionales
- Plate-forme thématique occurrence

La plate-forme nationale pourra réaliser un contrôle complémentaire de cohérence mais ce sont bien les plate-formes R/T qui portent la responsabilité de la cohérence.

**Validation :**

Est réalisée par :

- ➔ Plate-formes régionales
- ➔ Plate-forme nationale

Les plate-formes régionales organisent la validation et diffusent le résultat sous la forme d'un statut de validité régional. La procédure est réalisée plus proche des systèmes producteurs et des experts régionaux.

La plate-forme nationale organisent la validation et diffusent le résultat sous la forme d'un statut de validité national.

Il s'agit d'une validation globale s'appuyant sur les réseaux d'experts nationaux et les usagers (via le site de l'INPN, via les travaux de qualification pour les programmes, etc.).

Il sera nécessaire de bien gérer la traçabilité de la procédure pour éviter les sollicitations croisées des experts (nom de l'expert contacté, date, etc.).

Les corrections devront être réalisées en accord avec le producteur des données et en étroite coordination avec le point d'entrée dans le SINP (PF R ou T).

Il existe des groupes taxonomiques orphelins pour lesquels il n'y a pas d'experts disponibles en région comme au niveau national.

La démarche ne doit pas être bloquante pour la circulation des données dans le système.

La validation par les plate-formes a lieu sur deux niveaux (régional et national). Les données circulent et portent, in fine, le résultat des deux validations. La signification des valeurs (niveau de validité) en sortie pourra ne pas être totalement identique. Ce point sera à discuter dans les futures réunions du groupe et permettra de préciser les définitions proposées pour la validation régionale et nationale ainsi que la mise en œuvre du processus de validation pour ces deux niveaux.

Des discussions relatives aux conditions de diffusion des données élémentaires d'échange par la plate-forme nationale sont en cours. Il est probable que ces choix auront un impact sur le travail du présent groupe (notamment un potentiel décalage entre l'échelle de la validation et l'échelle de diffusion). Cet impact sera évalué et discuté lorsque les documents SINP seront validés (sur la base d'une note sur les modalités de diffusion des DEE sur la plate-forme nationale du SINP).

Concernant l'aspect diffusion, la question est posée de connaître le traitement des données hors du territoire français mais produites par un organisme français. En l'état actuel, ces données pourraient être bancarisées et gérées par le SINP sans subir de validation. Ce point dépasse cependant le périmètre de travail du présent GT.

## 7. Critères/paramètres

### Travail sur la cohérence :

- ➔ Au niveau « métadonnées »
- ➔ Au niveau « données »
- ➔ Entre données et métadonnées

Une limitation des contrôles de cohérence sur les métadonnées et entre données et métadonnées est soulevée. Cette limitation est liée à l'absence d'implémentation informatique des concepts par le GT Métadonnées (quels champs, format des champs, précision sur les cardinalités pour ces champs, etc.). En l'attente de l'éventuelle amélioration de ces spécifications, seuls les champs concrètement spécifiés et obligatoires des métadonnées pourront être utilisés pour ces contrôles.

Cela pose également la question de l'implémentation de ce standard par les plate-formes.

La liste des contrôles de cohérence possible est présentée. Une colonne « remarque » permet de mentionner les limites de chaque contrôle.

Ce tableau sera mis à jour et sera diffusé à l'ensemble des membres cercle 1 pour un travail collectif qui sera ensuite centralisé et discuté en réunion.

### Travail sur la validation :

Les limites à la validation, notamment son automatisation, sont :

- la connaissance (répartition, traits de vie, etc.)
- La variabilité existante entre taxons, en fonction de la saison, au cours du temps, etc.
- La standardisation des paramètres utilisés pour la validation

La liste des critères et paramètres pouvant être pris en compte pour la procédure de validation est transmise (« Identification des critères et paramètres pour la validation sur la base des éléments transmis dans le cadre de l'élaboration du recensement de l'existant »). Cette liste est issue du recensement de l'existant (réalisé pour la phase 1 du mandat). Ce tableau permet également d'identifier les paramètres utilisables selon l'échelle d'application (régionale ou nationale) et d'évaluer les possibilités d'automatisation.

Une colonne « remarque » permet de mentionner les limites de chaque paramètre.

Ce tableau sera mis à jour et sera diffusé à l'ensemble des membres cercle 1 pour un travail collectif qui sera ensuite centralisé et discuté en réunion.

### Les prochaines étapes :

- Identifier les besoins
- Définir les critères et paramètres
- Définir les valeurs en sortie pour chaque type de contrôle

- Définir les champs utiles à chaque contrôle/et à chaque niveau
- Gérer la traçabilité (date, validateur, etc.)

*Prochaine réunion prévue pour début avril 2015.*

---