



## GT « Validation des occurrences » Volet 1 : occurrences de taxons

### Réunion 4

*Le 01/10/2015, Salle Jean Dorst/MNHN*

---

**Rédaction** : Solène Robert (MNHN)

**Relecteurs** : Rémi Jomier, Jeanne de Mazières, Yannick Lebeau

**Pièce-jointe** :

- Support de discussion présenté lors de la réunion (GT\_VALIDATION\_SINP\_20151001.pdf)
-

**Invités :**

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

## Ordre du jour

- 1) Synthèse des réponses au questionnaire des besoins
- 2) Guide méthodologique V0
  - Démarche générale
  - Contenu du protocole de validation
  - Mises à jour et corrections
  - Exigences fonctionnelles
  - Perspectives

## Synthèse des réponses au questionnaire des besoins

Ce document est une synthèse des éléments recueillis via le questionnaire mis en ligne (mai – juillet 2015). Il ne présume pas des éléments qui seront repris ou non dans le cadre du SINP.

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

L'ensemble des retours a été regroupé au sein du document en 10 catégories :

- Les aspects organisationnels et le processus global de validation
- La collecte de données
- Les formats de données et de métadonnées
- Les informations supplémentaires à récolter et inclure dans les standards d'échange
- Un outil de saisie
- Un outil de création de DEE + méthodes de transformation
- Un outil de validation et des services dédiés
- Les référentiels et bases de connaissance
- Les mises à jour
- La diffusion

Un encart sous chaque catégorie permet de renseigner sur la solution, l'instance ou le GT concerné.

### **Décisions du groupe :**

Mettre en exergue (**en gras**) sur le document, les éléments qui concernent le GT Validation. Les autres éléments seront communiqués aux instances dédiées ou ont déjà été communiqués (pour les éléments qui concernent les standards données et métadonnées ou GINCO).

**Les relectures par les membres du GT sont à transmettre avant le 20 octobre, pour production d'une version diffusable.**

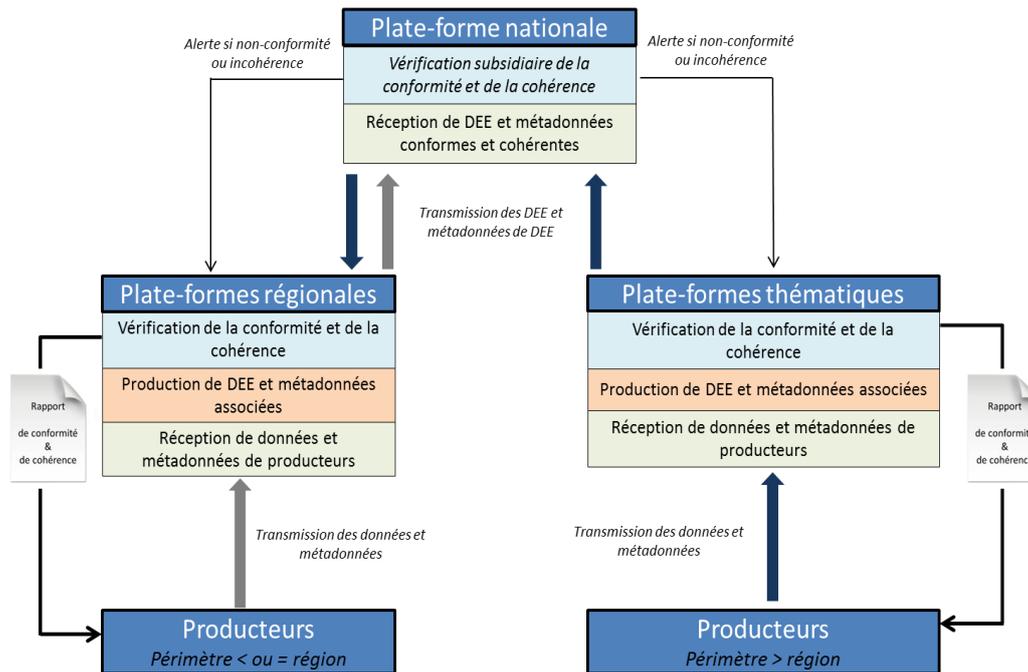
## Guide méthodologique

Une version préliminaire de guide méthodologique pour la validation a été transmise aux membres du groupe.

Ce document se base sur les précédentes réunions du GT, sur le recensement de l'existant, la synthèse des besoins et sur le document de travail sur les critères et paramètres.

## Discussions sur la démarche générale

### Conformité et cohérence



Les DEE circulant au sein du système doivent être conformes et cohérentes.

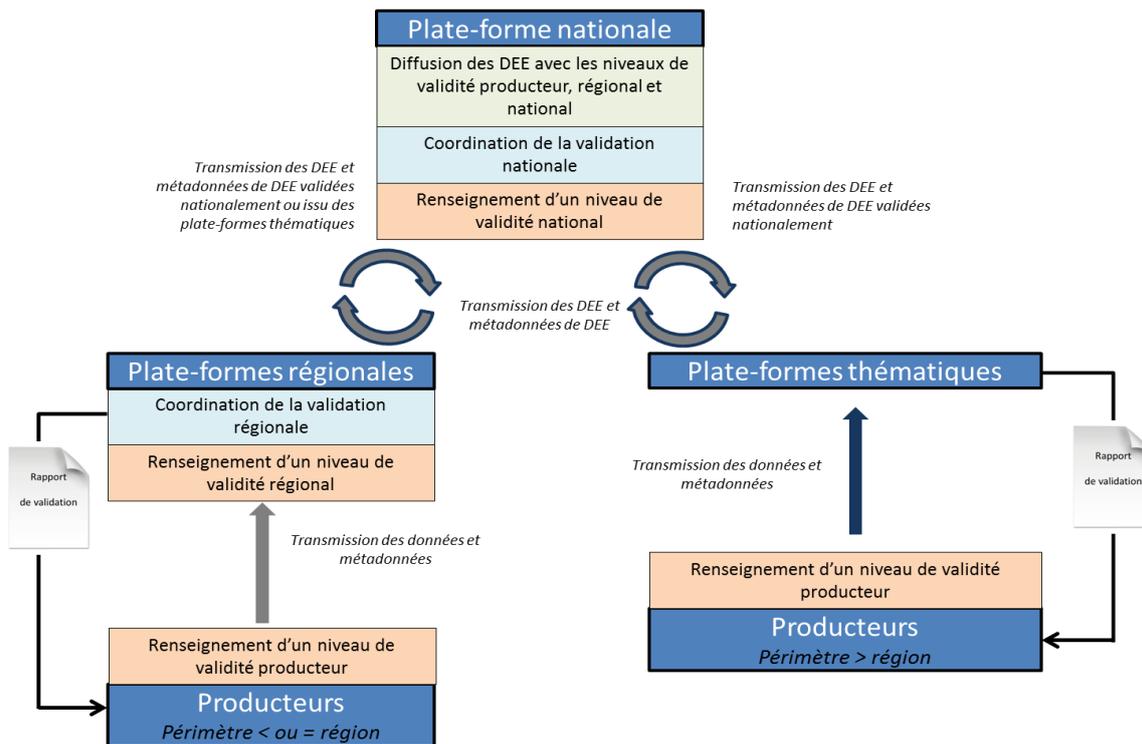
PF R/T portent la responsabilité de la mise en conformité et de cohérence.

PF R/T gèrent les échanges avec les producteurs via le rapport de conformité et de cohérence.

Le groupe maintient la pertinence de réaliser un contrôle de conformité et de cohérence au niveau national ; dans une architecture SINP ouverte, les plates-formes R/T pouvant être de technologie et de maturité variables. La plate-forme nationale éditera alors un rapport de conformité et de cohérence à destination des plates-formes R/T. Le schéma présentant la démarche sera mis à jour pour intégrer plus explicitement la production de ce rapport.

Le groupe s'accorde sur l'utilité de mettre en place et de publier sur le site de l'INPN un protocole de conformité et de cohérence par la PF Nationale. Il sera recommandé aux PF R/T de s'appuyer sur ce protocole pour réaliser leurs contrôles. Dans le cas contraire, il existerait un risque de rejet d'intégration des DEE par la PF Nationale. Il est souhaitable que l'outil GINCO se base sur ce même protocole.

## Validation



Selon l'organisation régionale, il apparaît que, dans certains cas, la validation a lieu avant la transformation des données en DEE. C'est-à-dire que la validation ne porte pas directement sur la DEE mais sur la DSR (donnée source régionale) ou DSP (donnée source producteur). Cela remet en question le fait que l'étape de validation scientifique soit réalisée sur la DEE » et le fait que seules des données conformes et cohérentes soient soumises à validation.

Il est donc acté suite à discussion que le résultat des opérations de validation doit être véhiculé au sein de la DEE. L'opération en elle-même pouvant avoir lieu indifféremment sur une donnée source (DS) ou sur une DEE. Cela implique que toutes les informations de la DS ayant servies à la validation doivent être également présentes dans la DEE de manière à ce que les plates-formes R/T et N opèrent la validation sur la même base d'informations (concrètement, que les PF valident la même chose).

Toutes les données peuvent être soumises à validation si l'organisation en place est telle mais seules les données conformes et cohérentes, quel que soit le niveau de validation, seront échangées entre les plates-formes SINP.

La procédure de validation ne freine pas la circulation des DEE entre PF : les données peuvent être échangées sans avoir subi de validation scientifique.

Les 3 niveaux de validation (producteur, régional et national) coexistent et sont indépendants. En l'absence de validation, le niveau de validité prend la valeur 0 c'est-à-dire « validation non réalisée ».

Ce sont les PF R/T qui communiquent aux fournisseurs les résultats des validations via le rapport de validation.

La PF nationale ne communique pas directement avec les producteurs. Elle transmet aux PF R/T le résultat de la validation nationale. A partir des éléments transmis, les PF R/T établissent le rapport de validation à destination du producteur. L'organisation de la transmission du rapport (un rapport de validation régionale puis un rapport de validation nationale ou un rapport conjoint) est laissée au libre choix des PF R/T.

Le schéma concernant la démarche de validation sera mis à jour afin de mieux distinguer les flux de données et les flux de validation.

**Seule la PF R (ou organisme mandaté par elle) dispose d'un droit d'écriture dans les champs de validation régionale.**

**Seule la PF N (ou organisme mandaté par elle) dispose d'un droit d'écriture dans les champs de validation nationale.**

Que valide-t-on exactement ? Est-ce l'ensemble des informations de la donnée (jusqu'aux effectifs par exemple) ?

Le besoin en matière de « périmètre » de validation sera différent en fonction de l'usage visé. Les usages des données dans le cadre SINP sont multiples.

Suite à discussion, il est alors proposé que le guide de validation indique que l'on valide, à minima, l'information de présence, à savoir l'assemblage, taxon/lieu/date. Une validation complète des informations peut également être réalisée. Il est alors nécessaire de pouvoir distinguer les niveaux de validité attribués en fonction de la « précision » de validation appliquée (à minima ou complète). Un champ commentaire libre permettra également de préciser la démarche au besoin. Cela sera pris en compte et viendra compléter les exigences fonctionnelles.

Les protocoles de validation sont de la responsabilité des PF régionales et nationales. Il peut donc y avoir des variabilités entre régions. La validation nationale, couche supplémentaire de validation indépendante des validations régionales, permettra de proposer une validation harmonisée d'envergure nationale. Une analyse des disparités pourra être faite a posteriori de manière à identifier les hétérogénéités et rechercher une homogénéisation.

L'essentiel dans l'immédiat consiste à s'accorder sur l'attribution d'un niveau de validité, sur une liste de valeurs communes et sur une définition générique des niveaux. Il faudra notamment penser à prévoir un niveau pour indiquer que la plate-forme n'a pas pu statuer. Cette information est à distinguer de l'absence de validation.

La liste des valeurs de l'attribut « niveau de validité » du standard DEE fera l'objet d'un point spécifique à l'ordre du jour de la prochaine réunion du GT.

Il faudra également veiller à ce qu'une interopérabilité soit possible avec les normes INSPIRE (en particulier la spécification technique « species distribution »).

## Protocole de validation

Chaque PF SINP réalisant la validation doit décrire et publier le protocole de validation qu'elle applique sur les données. Ce document devra décrire :

- Les organismes ayant été mandatés par la PF le cas échéant pour réaliser ces opérations et leur périmètre d'action respectif (ex : flore pour un CBN, etc.)
- La démarche de validation qu'il s'agisse de la partie automatique ou de la partie manuelle
- Les critères et les paramètres pris en compte
- Les outils techniques utilisés
- Les référentiels, bases de connaissance ou couches de référence utilisés (incluant leurs résolutions (commune, maille, etc.))
- La méthode d'attribution des niveaux de validité à partir de ces éléments.
- La définition fine des valeurs de ces niveaux pour la PF

## Mises à jour

Différents cas ont été identifiés comme susceptibles de nécessiter une mise à jour des champs de validation :

- Corrections et ajouts d'information
- Edition d'une nouvelle version d'un standard d'échange
- Evolution des connaissances et de l'écologie des espèces (MàJ des référentiels et couches de référence).

Attention cependant : une nouvelle version d'une couche de référence (pour une autre emprise que la précédente) ne nécessitera pas de mise à jour des données de l'emprise précédente. Dans le cas des couches de référence, seules les corrections de ces couches pour une même emprise nécessiteront une correction éventuelle du niveau de validité. Exemple : sortie d'un atlas faisant référence pour 2010-2015 : ne nécessite pas de revalider les données de 1950 à 2009.

Dans le cas des mises à jour de référentiel, il serait nécessaire de pouvoir identifier les cas nécessitant une mise à jour. Exemple dans le cas de TAXREF : analyse des cas à la sortie d'une nouvelle version du référentiel. Idem pour les évolutions INSEE (fusion ou scission (splitage) de communes, etc.). Ce travail est jugé important et impacte la validation des données. La mise en pratique de ces travaux dépasse cependant le cadre du présent GT et doit donner lieu à des discussions spécifiques supplémentaires en d'autres lieux.

- Evolution du protocole de validation de la PF

La démarche de mise à jour proposée pour les contrôles de conformité et de cohérence est la suivante :

La plate-forme, pour les données dont elle est en charge pour ces aspects (régionale ou thématique), est responsable de la mise à jour des contrôles de conformité et de cohérence. Les dates de réalisation des contrôles sont mises à jour en conséquence.

La démarche de mise à jour proposée pour la validation est la suivante :

**Cas où la plate-forme juge que la modification n'est pas de nature à modifier le niveau de validité; la modification est dite mineure :**

- Le niveau de validité reste inchangé.
- La date de validité est mise à jour.

**Cas où la plate-forme juge que la modification est de nature à modifier le niveau de validité; la modification est dite majeure :**

- Le niveau de validité est mis à zéro.
- La date de validité est mise à jour.
- Les données repassent alors par le circuit de validation.

Les données dont le niveau de validité a été mis à jour doivent être re-livrées par la plate-forme source à la plate-forme cible qui peut alors choisir de mettre à jour la validation des données selon la démarche décrite ci-avant.

La plate-forme « cible » (qui reçoit) doit pouvoir identifier les données sur lesquelles il y a eu mise à jour du niveau de validité par la plate-forme « source » (qui livre).

Il est acté suite à discussion qu'il n'est pas nécessaire que les DEE portent l'information de l'ancien niveau de validité et la date de cette ancienne validation. Cependant, les plates-formes ont toute liberté pour archiver ces informations si elles le souhaitent, sans toutefois livrer ces informations à la plate-forme « cible ».

## **Exigences fonctionnelles**

### Pour l'étape de conformité et de cohérence

Les informations proposées dans le guide conviennent et sont jugées suffisantes par le groupe.

A savoir :

Au niveau de la donnée :

- la date de réalisation du contrôle sur la donnée
- les attributs non conformes ou incohérents et leurs valeurs (déportation des champs par l'administrateur de la PF)

Au niveau de la métadonnée :

- la date de réalisation du contrôle de la métadonnée
- les attributs de métadonnée non conformes ou incohérents et leurs valeurs (déportation des champs par l'administrateur de la PF)
- le nom de la plate-forme ou de l'organisme ayant réalisé le contrôle

### Pour la validation « producteur »

Les informations proposées dans le guide conviennent et sont jugées suffisantes par le groupe.

A savoir :

A l'échelle de la donnée :

- la date de réalisation de la validation de la donnée
- le niveau de validité producteur (les valeurs sont laissées au libre choix du producteur)
- le nom, le prénom et l'organisme du validateur

A l'échelle de la métadonnée :

- le protocole de validation du producteur (url, référence bibliographique ou texte libre)

### Pour la validation régionale et nationale

Il est discuté de l'utilité de tracer et de véhiculer, au sein de la DEE, les résultats des validations intermédiaires (automatiques et manuelles) en complément du résultat final. A l'instar des décisions concernant la mention de l'ancienne valeur de validité (lors d'une mise à jour), il est proposé que la DEE ne véhicule que le résultat final en renseignant sur la procédure ayant été réalisée (est-ce le résultat d'une validation automatique, manuelle, ou le résultat d'une combinaison entre les deux résultats ?). Les PF ont cependant toute liberté pour stocker dans leurs bases de données ces informations complémentaires si elles le souhaitent.

La validation (régionale et nationale) nécessitera les informations suivantes au sein de la DEE :  
Au niveau de la donnée :

- la date de réalisation de la validation
- le niveau de validité
- le type de validation retenue (automatique, manuelle ou combinatoire)
- l'échelle de validation de la donnée (=l'échelle de la DEE « précise »)
- le périmètre de la validation appliquée (minimale ou complète)
- le lien vers le protocole de validation et sa version
- un commentaire libre

**Les modifications discutées en séance et les commentaires déjà transmis seront centralisés pour produire une nouvelle version du guide qui sera, ensuite, soumis à relecture à l'ensemble des membres du GT.**

### **Perspectives**

Décliner le guide aux autres thématiques du SINP ? Une partie du guide pourrait être générique?

Ce point sera mis à l'ordre du jour de la prochaine réunion

Décliner plus précisément les règles de conformité dans les standards (ou via FAQ dédiées). Cela est prévu par le GT standard de données.

Réaliser des modèles :

- de rapport de conformité & de cohérence
- de rapport de validation
- de protocole de validation

Mettre en place des services Web de conformité et de cohérence ?

A discuter dans le cadre des fonctionnalités de la PF nationale.

Mettre en place un « outil de validation » qui aurait pour objectif de faciliter le requêtage sur les données, le croisement avec des couches de référence et référentiels et l'attribution du niveau de validité et des champs associés ?

**Prochaine réunion à prévoir pour fin novembre : un doodle sera prochainement transmis.**