



GT « Validation des occurrences »

Volet 1 : occurrences de taxons

Réunion 3

Le 15/04/2015, Salle Baleine 4/MNHN

Rédaction : Solène Robert (MNHN)

Relecteurs : Jeanne de Mazières (MNHN), Yannick Lebeau (MEDDE), Dorothée Meyer (DREAL PACA), Rémy Jomier (MNHN), Jean Lescure (SHF)

Pièce-jointe :

- Support de discussion présenté lors de la réunion (GT_VALIDATION_SINP_20150415.pdf)

Invités :

ORGANISME/STRUCTURE	Prénom NOM	Présents
MEDDE	Yannick LEBEAU	X
MNHN	Solène ROBERT	X
	Laurent PONCET	X
	Julien TOUROULT	
	Remy JOMIER	X
	Jeanne de MAZIERES	X
FCBN & CBN	Bruno DUTREVE	
	Alexis DESSE	X
	Johan GOURVIL	X
	Anaïs JUST	X
	Grégory CASE	Excusé
	Jérôme MILLET	
Plateforme régionale [Réunion]	Frédérique ZELMIRE	Excusée
Plateforme régionale [Haute Normandie]	Arnaud SALTRE	Excusé
Plateforme régionale [Aquitaine]	Andy PAPACOTSIA	Excusé
	Marie BARNEIX	Excusée
	Paul FROMAGE	X
Plateforme régionale [Languedoc Roussillon]	Mathieu BOSSAERT	
Plateforme régionale [Provence Alpes Côte d'Azur]	Dorothée MEYER	X
ONCFS	Philippe LANDRY	X
	Pierre MIGOT	
	Guillaume BODY	X

SFPEM	Christian ARTHUR	<i>Excusé</i>
	Stéphane AULAGNIER	
SHF	Jacques CASTANET	
	Jean LESCURE	X
LPO	Dominique ARIBERT	
	Gwenaël QUAINTENNE	
RESOMAR	Nicolas LAVESQUE	
	Serge HEUSSNER	
	Antoine GREMARE	
	Nathalie SIMON	
IFREMER	Antoine HUGUET	<i>Excusé</i>
ANTAREA	Rumsaïs BLATRIX	
Observatoire Pelagis	Vincent RIDOUX	
GBIF France	Anne-Sophie ARCHAMBEAU	
	Sophie PAMERLON	X
	Marie-Elise LECOQ	

Objectifs de la réunion

- Faire le point sur les travaux en cours, leur état d'avancement et les échéances.
- Préciser la démarche pour la conformité, la cohérence et la validation.
- Préciser et compléter la liste des critères et paramètres.

Recensement de l'existant

Le document est finalisé, les relectures sont closes. Il sera prochainement diffusé sur le site Nature France et aux membres de cercle 1 qui pourront le diffuser dans leurs réseaux respectifs.

Questionnaire

Objectifs:

- Faire état des besoins par type d'acteurs,
- recueillir des informations sur les pratiques actuelles par type d'acteurs,
- permettre, in fine, de croiser ces besoins avec la faisabilité pour prioriser les paramètres.

Peu de retour sur la trame transmise. Un délai supplémentaire (jusqu'au 27 avril) est accordé.

Remarques remontées en séance :

- ajouter un type d'acteur « animateur politique SINP »
- ajouter un encart sur les informations tierces nécessaires (couches géographiques, etc.)

Une version pour diffusion sera transmise au groupe pour renseignement par eux-mêmes et transmission à leur réseau de manière à couvrir le panel d'acteurs du SINP.

Retour des questionnaires renseignés souhaités avant le 4 juin.

Pour les organismes assurant plusieurs « rôles », il est possible de remplir une seule fois le questionnaire ou plusieurs si souhait de distinguer les réponses par rôle.

Concernant le format physique du questionnaire (word, pdf, Lime survey, autre outil). Ces outils seront évalués rapidement afin de voir quelle serait la solution optimale pour faciliter la récupération des éléments des questionnaires.

Travail sur l'identification des critères et des paramètres

Un document intitulé « Identification et analyse des contrôles de conformité, de cohérence et des critères/paramètres pour la validation automatique et manuelle » a été proposé comme support de travail évolutif pour centralisation des réflexions du groupe. Chaque membre est invité à le compléter avec ses propres réflexions.

Objectifs/perspectives pour ce document :

- étoffer la liste des contrôles pour conformité et cohérence de manière à tendre vers l'exhaustivité.
- pour chaque contrôle (conformité et cohérence), préciser les règles d'application et l'impact de ce contrôle sur la mise en circulation de la donnée, c'est-à-dire l'échange entre plate-formes (circulation étant à distinguer de la diffusion).

Proposition du groupe : baser les règles de circulation sur le caractère obligatoire et facultatif des champs des standards.

Les champs nécessaires à la phase de validation (sensus stricto) doivent également être identifiés de manière à inclure des vérifications sur ces champs lors des contrôles de conformité et de cohérence.

Pour le document de travail : réaliser un tableau de synthèse pour chaque étape avec les règles et le caractère bloquant ou non.

Rappel : Autant que faire se peut, la validation ne doit pas bloquer l'échange de données entre plate-formes R/T et nationale. C'est pourquoi les données ayant passé avec succès les contrôles obligatoires bloquants seront mises en circulation avec des informations indiquant leur degré de validation.

Les règles de diffusion ne sont pas du ressort du présent GT.

Données et métadonnées sont concernées par les contrôles de conformité et cohérence. Il est très probable que les résultats des contrôles nécessitent la création de champs dédiés au niveau de la donnée comme au niveau de la métadonnée (ex : statut_conformité : conforme/non-conforme + commentaire conformité : mauvaise standardisation du champ « denombrement »). Cela se matérialisera par une liste d'exigences fonctionnelles transmises aux GT dédiés pour implémentation.

Question du blocage unitaire ou en masse en cas de détection d'erreurs (conformité et cohérence) dans un jeu :

L'analyse se fait au niveau de la donnée, il est possible de ne bloquer que les données erronées.

Le gestionnaire de la PF, en fonction du pourcentage ou des caractéristiques des données erronées dans un même jeu, pourra définir la méthodologie de blocage la plus pertinente (masse, unitaire, par « lot », etc.).

Cet élément est trop complexe et variable (cas de figures multiples) pour faire l'objet d'une règle définie dans le cadre de ce GT mais des recommandations pourront être fournies dans le document final du GT.

Les plateformes devront, de fait, apprécier l'intégrité d'un lot de données notamment dans le cas d'erreurs sur une donnée faisant partie d'un ensemble (dans le cas d'un relevé par exemple). Une donnée erronée dans un relevé pouvant entraîner l'invalidation de l'ensemble des données de ce relevé.

[Discussion sur les erreurs de formalisme concernant des champs non obligatoires :](#)

Il est proposé de ne pas véhiculer les informations facultatives mais non-conformes ou non cohérentes dans les DEE et donc de mettre à blanc les attributs correspondants. En complément, il serait prévu de mentionner ces anomalies dans un champ récapitulatif dédié permettant à l'utilisateur de savoir que d'autres informations, entachées de défauts, étaient présentes dans la donnée source mais qu'elles n'ont pas été transcrites dans la DEE.

Ces attributs facultatifs pourraient être rectifiés à posteriori et les données concernées seraient alors mises à jour. Les informations de validité doivent pouvoir indiquer ce changement d'état (par le statut et la date de validation notamment).

[Discussion sur les mises à jour et corrections :](#)

Les contrôles, tout comme les mises à jour, doivent être datés. Si les mises à jour concernent des champs obligatoires des standards, alors les contrôles doivent être mis à jour également. Si les mises à jour concernent des champs facultatifs, c'est au gestionnaire de la PF ou à l'expert de décider si les mises à jour influencent les résultats des contrôles ou non et modifier, en fonction, la date des contrôles.

[Discussion sur la possibilité d'inclure la conformité et la cohérence sur les données sources \(DS\) au périmètre de travail du groupe :](#)

Les standards régionaux sont hétérogènes, il sera, de fait, difficile de définir des règles globales pour les DS. Le besoin des régions n'est pas encore assez précis pour que cela puisse être pris en compte.

L'inclusion de paramétrages au sein d'OGAM pour la gestion de contrôle sur les DS n'est pas du ressort du présent GT.

[Discussion autour du degré d'expertise de l'observateur :](#)

Ce critère est subjectif mais est bien souvent pris en compte pour la validation experte.

Le degré d'expertise peut également être lié à la formation donnée aux observateurs avant/pendant les phases d'acquisition de données ou à la présence de fiches d'aide à la détermination.

Il faudrait prévoir de pouvoir véhiculer le protocole d'acquisition décrivant les éléments mis en place pour aider les observateurs ainsi que le protocole de validation du producteur. Ces informations relèvent plutôt de la métadonnées décrivant un jeu de données.

Démarche

Il est proposé au GT de travailler sur les règles globales :

- lister les paramètres et leurs déclinaisons,
- faire des préconisations quant à l'application de ces paramètres,
- définir les valeurs pour le niveau de validité (liste des valeurs et définition de chacune d'elles),
- définir un modèle de documentation pour le protocole de validation appliquée.

Il est proposé que chaque PF R/T :

- définisse les déclinaisons des paramètres qu'elle applique sur les données (et les couches de contrôle qu'elle utilise) pour la partie « automatique » tout comme pour la partie « manuelle »,
- décrive le protocole de validation utilisé et le diffuse sur la PF,
- attribue, en fonction du protocole, le niveau de validité selon les valeurs définies par le GT.

Cette proposition permet de ne pas être intrusif dans les protocoles de validation régionaux et de définir un cadre commun de manière à limiter les hétérogénéités régionales.

La recherche d'une cohésion inter-régionale doit être visée de manière à limiter les disparités et l'effet de « frontières ».

Prochains travaux :

Le groupe est invité à compléter le document sur l'identification des contrôles, critères et paramètres. Le document mis à jour en fonction des discussions de la présente réunion sera prochainement transmis au groupe.

Les retours sur la trame de questionnaire sont à transmettre avant le 27 avril.

Prochaine réunion à prévoir pour fin mai/début juin : un doodle sera prochainement transmis.