Gt « Qualification des données biodiversité du SINP »

Réunion de lancement

Le 16/10/2014, Salle Baleine 3/MNHN

Rédaction : Solène Robert (MNHN)

Relecteurs : Julie Chataigner (MNHN), Jeanne de Mazières (MNHN), Yannick Lebeau (MEDDE), Johan Gourvil (FCBN), Philippe Landry (ONCFS)

Pièce-jointe :

- Support de discussion présenté lors de la réunion (GT_QUALIF_SINP_20141016.pdf)

Invités :

ORGANISME/STRUCTURE	Prénom NOM	Présents
MEDDE	Yannick LEBEAU	X
MNHN	Solène ROBERT	X
	Laurent PONCET	*Excusé*
	Julien TOUROULT	*Excusé*
	Julie CHATAIGNER	X
	Jeanne de MAZIERES	X
FCBN & CBN	Bruno DUTREVE	
	Alexis DESSE	X
	Johan GOURVIL	X
	Anaïs JUST	X
	Grégory CASE	X
	Jérome MILLET	
Plateforme régionale [Réunion]	Frédérique ZELMIRE	*Excusé*
Plateforme régionale [Haute Normandie]	Arnaud SALTRE	X
Plateforme régionale [Aquitaine]	Andy PAPACOTSIA	X
	Marie BARNEIX	X
Plateforme régionale [Languedoc Roussillon]	Mathieu BOSSAERT	
Plateforme régionale [Provence Alpes Côte d'Azur)	Dorothée MEYER	X
ONCFS	Philippe LANDRY	X
	Pierre MIGOT	
SFEPM	Christian ARTHUR	
	Stéphane AULAGNIER	
SHF	Jacques CASTANET	
LPO	Dominique ARIBERT	
	Gwenaël QUAINTENNE	X

RESOMAR	Nicolas LAVESQUE	
	Serge HEUSSNER	
	Antoine GREMARE	
	Nathalie SIMON	
IFREMER	Antoine HUGUET	
ANTAREA	Rumsaïs BLATRIX	
Observatoire Pelagis	Vincent RIDOUX	
GBIF France	Anne-Sophie ARCHAMBEAU	

Objectifs

Les objectifs de la première réunion du groupe de travail consistent tout d'abord en la présentation du contexte (protocole et architecture du SINP, etc.) et des éléments définis dans le mandat du groupe de travail (synergies avec d'autres actions du SINP, démarche, périmètre, gouvernance, etc.).
Puis, sur la base du document préparatoire issu du recensement de l'existant (ayant donné lieu à un appel à contribution), il s'agit de discuter des éléments de terminologie et de périmètre dans l'objectif de partager un vocabulaire commun, une base de travail solide et un périmètre cadré.

Contexte

Cf. le support de réunion (GT_QUALIF_SINP_20141016.ppt) et le document préparatoire (RecensementExistant_V0.doc)

Remarques sur ces éléments de contexte :

- ➔ les usages présentés dans le protocole du SINP sont les usages nationaux possibles à partir des données disponibles dans le système. Ils ne reflètent pas tous les usages régionaux, thématiques ou locaux.
- ➔ Le MNHN est coordinateur pour la validation au niveau national et s'appuie pour cela sur un ensemble d'acteurs, d'experts et de partenaires (FCBN, OPIE, SINP Mer, etc.).

Terminologie

Chaîne de traitement : vérification, validation puis qualification.

A l'issu des discussions, les propositions faites via le document RecensementExistant et reprises dans le ppt sont modifiées de la manière suivante :

Vérification

Ce processus comprend deux composantes : une vérification de conformité et une vérification de cohérence.
Les résultats sont multiples, c'est-à-dire que le processus donne lieu à plusieurs résultats de vérification et binaire (apte / non apte).
Ex : vérifications taxonomiques, vérification géographique, etc.

Cette étape préalable permet :
- de 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.

- de 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).)

Validation :

La validation est un processus d'expertise visant à renseigner sur la fiabilité (désigne le degré de confiance que l'on peut accorder à la donnée). Le résultat du processus est unique (une seule valeur possible) et graduel (impliquant la définition de seuils). Ce processus fait intervenir des bases de connaissance et/ou du dire d'expert.

Qualification :

La qualité étant d'une façon générale, l'aptitude d'un produit ou d'un service à répondre aux usages qui sont attendus de lui, la qualification est faite en fonction de l'usage que l'utilisateur veut faire de la donnée. Le résultat consiste à indiquer si un ensemble de données peut répondre à un ou plusieurs objectifs d'exploitation définis par l'utilisateur. Le résultat est binaire (apte à l'usage/non apte à l'usage) et peut être multiple (une donnée pouvant être qualifiée différemment pour plusieurs exploitations).

Périmètre

Le présent groupe de travail n'a pas vocation à travailler sur de la donnée dite « de synthèse » (atlas, carte de répartition, etc.) qui sont des usages ou exploitations spécifiques, mais sur les données d'occurrence (élémentaire) permettant la réalisation notamment de ces synthèses. Le contrôle, entre autres, des couches de référence, est ainsi considéré comme hors périmètre du groupe de travail.

Le périmètre du groupe de travail porte sur les contrôles des DEE que ce soit :
- au niveau des plate-formes R/T lorsque la donnée est fournie par le producteur en pré-DEE : DEE standardisée mais sans identifiant permanent donné par les plate-formes SINP ni obligatoirement au format GML
- au niveau de la plate-forme nationale lorsque la DEE est fournie par les plate-formes régionales ou thématiques (R/T).

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). En cas de difficultés ou de demandes de précision, il faut se tourner vers les GT standard de données et standard de métadonnées.
Les plate-formes R/T sont en effet responsables de la standardisation des données sources (DS) 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 mention des « bonnes pratiques » si jugé nécessaire par le groupe, mais ne donnera pas lieu à une méthodologie spécifique dans le cadre du groupe de travail.

Les vérifications de cohérence (intra et interchamps) seront traitées par le groupe.

L'étape de validation sera intégralement traitée par le groupe.

Au sujet de la qualification : la qualification, telle que définie par le GT, est directement liée aux usages que l'utilisateur veut faire de la donnée. Ces usages sont multiples et indéfinis ; ils sont propres à chaque thématique (occurrence d'espèces, espaces protégés, habitats, géologie,...). L'utilisateur, expert ou non, est ainsi responsable de l'utilisation qu'il fait des données et doit lui-même définir si ces données conviennent à l'usage qu'il en fait.
Le groupe s'accorde donc sur l'exclusion de la qualification dans le périmètre de travail du groupe. Il ne s'agira donc pas de qualifier les données/métadonnées mais de fournir à l'utilisateur les moyens de sélectionner efficacement les données en fonction de son objectif d'exploitation.
Pour ce faire, le groupe devra ainsi identifier les besoins des utilisateurs pour leur permettre d'apprécier pleinement la pertinence des données en fonction de son usage. Il s'agira de fournir à l'utilisateur les informations nécessaires à la sélection des données.

A l'issue du travail du groupe, les informations nécessaires à faire porter par le standard de données et métadonnées seront listées. Il conviendra de voir avec les différents groupes de travail impactés comment concrètement échanger ces informations au niveau national (GT standard de données, GT métadonnées) et régional et thématique.

Échelles de travail :
Données et métadonnées seront concernées par la chaîne de contrôle (vérification de la cohérence et validation) et porteront les résultats des contrôles définis par le groupe. Cela aura probablement un impact sur les standards de données et de métadonnées du SINP.
Pour la prochaine réunion, afin d'identifier les contrôles possibles, seront listés les champs des standards de données et de métadonnées mentionnés comme obligatoires pour l'échange.

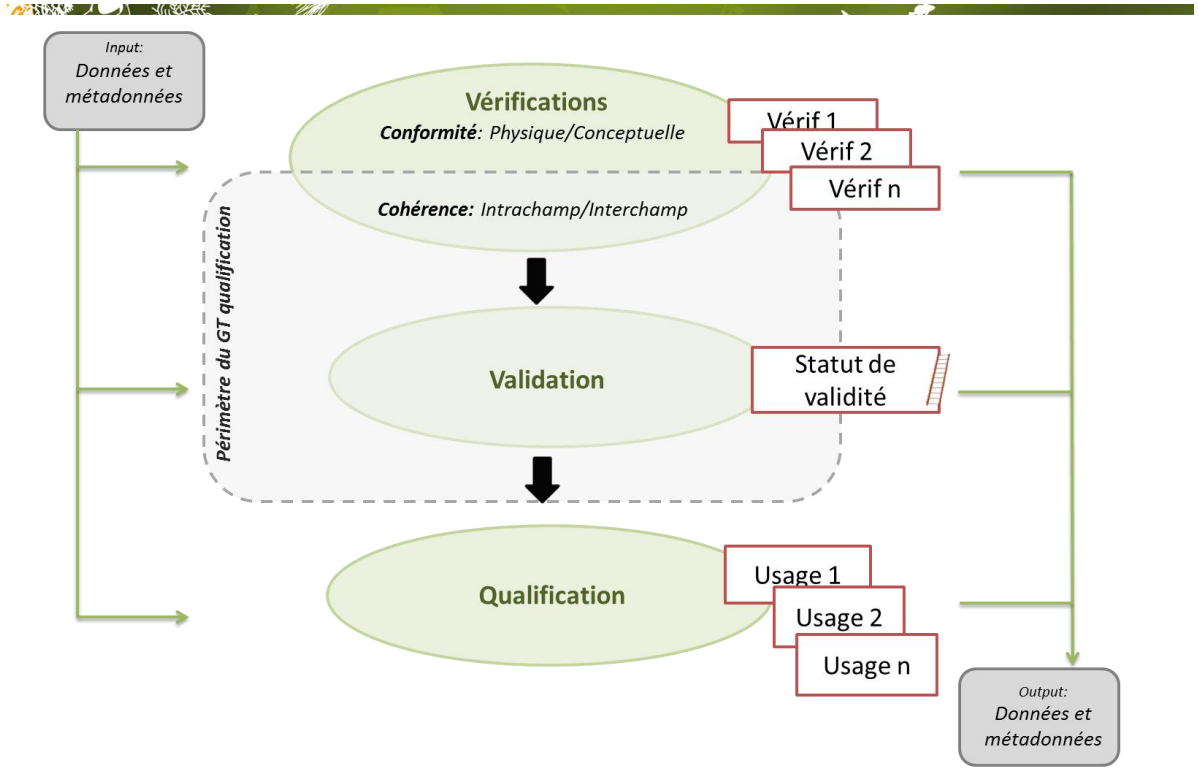

Schéma présentant les grands concepts de contrôle discutés lors de la réunion et leur intégration dans le périmètre de travail du groupe.

Accessibilité des données considérées comme erronées ou douteuses après la procédure de validation :
Les membres du groupe partagent l'intérêt de garder ces données mais les avis divergent quant à la diffusion publique de ces données.
La diffusion des données erronées (statut définitif) pourrait être différente de celle des données douteuses (statut temporaire).
Il a été demandé de rajouter cette thématique dans le document de recensement de l'existant pour voir comment cette problématique est gérée par les systèmes existants.

Décision : Les normes INSPIRE relatives à l'accessibilité des données seront portées à la connaissance du groupe pour alimenter les futures discussions sur ce sujet.

Intégration du niveau producteur :
Lorsque le producteur réalise une validation, il doit pouvoir le mentionner au niveau de la donnée. Les métadonnées pourront porter, en complément, des informations relatives à la méthodologie appliquée pour cette étape.

Intégration du niveau international :
Les données internationales (produites à l'étranger mais concernant le territoire français) intègrent le SINP par la plate-forme Thématique occurrence de taxon. Elles seront soumises aux mêmes procédures que les autres données entrant dans le SINP par cette plate-forme, qui sont les données

des producteurs dont le territoire couvre plus d'une région et qui ne souhaitent pas redécouper leurs jeux de données par région.

La mention « données internationales » pourra être parallèlement portée par les métadonnées du jeu.

	Vérifications		Validation	Qualification
Niveau/opérateur	Conformité	Cohérence		
Producteur	Hors périmètre	Hors périmètre	X	
Régional et thématique	Hors périmètre/conformité aux standards d'échange	X	X	
National	Hors périmètre/conformité aux standards d'échange	X	X	
International (=Idem Thématique)	Hors périmètre/conformité aux standards d'échange	X	X	
Utilisateur				Hors périmètre

Tableau récapitulatif des opérations de contrôle selon les niveaux opérateurs

Planning

Phase 1 :

-Finalisation du document RecensementExistant suite aux modifications de définitions, aux nouvelles contributions et à la relecture des contributeurs.

Les nouvelles contributions devront être transmises à srobert@mnhn.fr avant le 1er novembre.

-Le document sera ensuite soumis à relecture aux membres du cercle 1 puis le document modifié selon les relectures sera diffusé aux membres du cercle 2.

Phase 2 :

-Prochaine réunion pour finaliser les définitions discutées et aborder la définition des besoins utilisateurs ainsi que la démarche globale (dec. 2014/janv. 2015).