



SINP

Système d'Information
sur la Nature et le Paysage



GT Validation des données d'occurrence du SINP

Volet 1 : occurrences de taxons

Recensement de l'existant pour la validation des données d'occurrence du SINP

Version 1

Avril 2015



MUSÉUM
NATIONAL D'HISTOIRE NATURELLE

Historique des versions du document :

<i>Versions</i>	<i>Date</i>	<i>Commentaires</i>
0	13/10/2014	Document support pour réflexions de la première réunion du GT
0.1	16/12/2014	Document mis à jour suite aux nouvelles contributions
0.2	10/02/2015	Document mis à jour suite à la relecture des contributeurs
1.0	21/04/2015	Document validé suite à la relecture des membres du cercle 1 du GT

Rédacteur :

MNHN-SPN	Solène Robert
----------	---------------

Relecteurs :

MNHN-SPN	V0	Jeanne de Mazières
MNHN-SPN	V0	Laurent Poncet
MNHN-SPN	V0	Julien Touroult
DREAL PACA	V0	Dorothee Meyer
FCBN	V0	Johan Gourvil
IFREMER	V0.1	Emilie Gauthier
CBNBP	V0.1	Sébastien Filoche
DREAL PACA	V0.1	Dorothee Meyer
GBIF France	V0.1	Anne-Sophie Archambeau
SANDRE	V0.1	Florine Leveugle
CEN PACA	V0.1	Julie Delauge
CBN-BI	V0.1	Alexis Desse
ONF	V0.1	Thomas Bouix
OBHN	V0.1	Arnaud Saltré
FCBN	V0.1	Johan Gourvil
MNHN-SPN	V0.1	Julien Touroult
MNHN-SPN	V0.2	Laurent Poncet
CBN-BI	V0.2	Alexis Desse
OBHN	V0.2	Arnaud Saltré
MEDDE	V0.2	Yannick Lebeau
MNHN-SPN	V0.2	Jeanne de Mazières

Contributeurs :

OBHN	Arnaud Saltré
MNHN-SPN	Julie Chataigner
SANDRE	Florine Leveugle
MNHN-GBIF	Sophie Pamerlon
IFREMER	Emilie Gauthier
MNHN-SPN	Pierre Noël
ONF	Thomas Bouix
CEN PACA	Julie Delauge
CBN-BI	Alexis Desse
MNHN-SPN	Arnaud Horellou
MNHN-CBNBP	Sébastien Filoche
MNHN-CBNBP	Anne Plu
MNHN-SPN	Pascal Dupont
NBN	Via Ingrid Marchand (MNHN-SPN)
CBN-FC	Franck Grossiord via Anaïs Just (FCBN)
CBN-PMP	Via Anaïs Just (FCBN)
FCBN	Johan Gourvil
CBN-SA	Grégory Caze
Royal Botanic Garden, Kew	Steven Bachman
CSCF	Yves Gonseth

Citation recommandée :

Robert S., de Mazières J., Poncet L., Desse A., Saltré A., Gourvil J., Meyer D., Touroult J., Leveugle F., Delauge J., Gauthier E., Filoche S., Bouix T., Archambeau A.S. & Lebeau Y. 2015. Recensement de l'existant pour la validation des données d'occurrence du SINP. Rapport pour le SINP, rapport MNHN-SPN 2015-44, 54 p.

Contexte*

Une multitude de jeux de données sur la nature et les paysages est produite chaque année par de nombreux acteurs sur les territoires français. Ces données sont collectées selon des méthodes et protocoles différents et pour des objectifs hétérogènes ; elles concernent également des périmètres taxonomiques et géographiques divers ; les compétences et moyens des observateurs sont tout aussi variables.

Dans le cadre du SINP, les différents utilisateurs des données doivent être en mesure de connaître le niveau de confiance associé à celles-ci (objectifs du SINP listé dans l'Article 3 du protocole de 2013 : « définir et mettre en œuvre des critères de qualité des données »). Cela permet notamment une bonne utilisation des données au regard des usages.

Le SINP doit mettre en place des dispositifs puissants d'échanges, de communication et d'intégration des informations. Cela nécessite notamment le partage de principes scientifiques et règles de contrôle de la qualité et de la validation des données.

Dans le SINP, il existe plusieurs niveaux de qualification :

- le producteur, s'il a produit lui-même les DEE, certifie une donnée exacte et fiable, contrôlée selon les normes et vérifications techniques et scientifiques mises en place par lui-même. La qualification initiale des données-source est de la responsabilité des producteurs de données ;

- la plate-forme régionale ou thématique, qualifie techniquement et scientifiquement les DEE qu'elle génère ou consolide. Les plateformes régionales ou thématiques assurent également l'identification et la qualification des métadonnées. Les métadonnées doivent notamment comporter les informations relatives à la qualité et au mode de qualification des séries de données qu'elles décrivent.

- le Muséum national d'Histoire naturelle (MNHN), coordonne au niveau national, la qualification des DEE pour l'établissement des données de référence nationale publiées par la plate-forme nationale. Il s'appuie pour cela sur un ensemble de partenaire ;

- les données recueillies par le GBIF et issues de bases de données étrangères mais concernant le territoire français sont qualifiées « données internationales » et seront mises à disposition dans le SINP.

Les conclusions du GT Validation serviront à mettre en place les principes du système d'information du SINP. La validation des processus et activités de validations au sein du SINP devront être conformes aux conclusions du groupe.

Les DEE conformes au SINP seront alors tagguées notamment par un ou plusieurs attributs de qualification. Une évolution du standard d'échange de données sera alors réalisée.

Les métadonnées décrivant les services et les jeux de données (DS, DEE et Données de synthèse) doivent préciser leur mode de qualification.

Le format d'échange de données comprend des champs optionnels et des champs obligatoires (notamment la qualification de la donnée (validation)).

**Éléments extraits du document d'architecture SINP, du protocole SINP et du mandat du GT Validation.*

Objet du présent document

Ce document est produit dans le cadre du lancement du Groupe de Travail (GT) sur la validation des données d'occurrence du SINP. Le recensement de l'existant est le produit de la phase 1 définie dans le mandat.

Suite à l'étape préalable de recensement non exhaustif des travaux existants en matière de contrôle, de validation ou de qualification de données, au niveau régional, national et international, ce document présente une synthèse des éléments analysés de manière à alimenter les discussions et les décisions du GT Validation.

Préambule

Le recensement de l'existant a donné lieu à un appel à contribution à l'ensemble des contacts du cercle 1 identifiés dans le mandat du GT Validation. Nous remercions l'ensemble des contributeurs pour la transmission de documentations qui ont permis la réalisation de cette synthèse.

Ce recensement non exhaustif est soumis aux contributeurs pour validation des informations issues des documentations transmises et pour complément le cas échéant puis à la validation collégiale du présent groupe de travail.

Éléments de terminologie

Les éléments de terminologie étant à définir par le groupe de travail, le terme générique « contrôle » sera utilisé dans ce document.

Le terme « données » est utilisé dans le document au sens observation (ensemble d'informations composant l'observation (où, quand, qui, quoi, d'où, comment, etc.)).

Le terme « métadonnées » est utilisé dans le document dans le sens d'informations permettant de décrire un ensemble de données.

I.	SYSTEMES ANALYSES.....	7
1.	Liste des systèmes analysés	7
2.	Présentation des systèmes analysés	8
II.	SYNTHESE ET ANALYSE	11
1.	Terminologies utilisées par les différents systèmes.....	11
2.	Types de contrôle.....	16
3.	Niveau de mise en œuvre des procédures de contrôle	18
4.	Sources utilisées dans les processus de contrôle	20
5.	Objet portant le résultat de l'analyse.....	22
6.	Valeurs attribuées dans les processus de contrôle.....	23
7.	Sectorisation des contrôles	36
8.	Critères/Paramètres pris en compte dans le processus de contrôle	37
9.	Gestion des données inexploitable ou douteuses pour la diffusion	41
10.	Retour aux utilisateurs ou aux producteurs du résultat de contrôle.....	43
11.	Gestion de la traçabilité des corrections.....	45
III.	ELEMENTS COMPLEMENTAIRES ISSUS DE LA BIBLIOGRAPHIE.....	47
1.	Identification des besoins.....	47
2.	Éléments de terminologie	47
3.	Traçabilité de la procédure.....	48
4.	Délais	48
5.	Types d'erreurs.....	48
6.	Types de contrôles	49
7.	Critères et paramètres	49
8.	Codages	50
9.	Feedback.....	50
10.	Augmentation de la qualité des données	50
11.	Mises à jour	51
IV.	BIBLIOGRAPHIE.....	52

I. SYSTEMES ANALYSES

1. Liste des systèmes analysés

SYSTEME	Organisme gestionnaire/ coordinateur	Contact ayant transmis la documentation	Emprise spatiale
Atlas floristique de la FCBN	FCBN	Johan GOURVIL (FCBN)	Nationale (FR)
BDN	ONF	Thomas BOUIX	Nationale (FR)
BRAHMS dans le cadre de l'usage par Kew	Royal Botanic Gardens, Kew	Steven BACHMAN (Kew)	Nationale (UK)
CBN PMP -> Liste Rouge régionale des plantes vasculaires	CBN Pyrénées et Midi-Pyrénées	Via Anaïs JUST (FCBN)	(supra-)Régionale (FR)
CSCF	CSCF	Site internet	Nationale (Suisse)
Digitale2 (24)*	CBNBI	Alexis Desse (25)* via Arnaud SALTRE	Régionale (Nord - Pas de Calais, Picardie et Haute-Normandie) (FR)
FLORA	MNHN-CBNBP	Sébastien FILOCHE / Anne PLU	(supra-)Régionale (FR)
GBIF France	GBIF : secrétariat basé à Copenhague et point nodal français au MNHN-GBIF France	Sophie PAMERLON / Anne-Sophie Archambeau	Internationale/Nationale (FR) (point nodal)
INPN	MNHN-SPN	Groupe de travail interne	Nationale (FR)
NBN Record Cleaner	National Biodiversity Network (NBN)	Helen ROY, Steve WILKINSON, Lynn HEELEY, Paula LIGHTFOOT (via Ingrid MARCHAND)	Nationale (UK)
Observatoire de la Flore Sud-Atlantique	CBN Sud-Atlantique	Grégory CAZE (CBN-SA)	(supra-)Régionale (FR)
ODIN	OBHN	Arnaud SALTRE	Régionale (FR)
SIE	ONEMA	Florine LEVEUGLE et Julie CHATAIGNER	Nationale (FR)
Sigogne	PPNMEFC	Franck GROSSIORD via Anaïs JUST	Régionale (FR)
SILENE	DREAL PACA/ CR PACA/ CEN PACA / CBN-Med	Julie DELAUGE	Régionale (FR)
SI Quadrige ²	IFREMER	Emilie GAUTHIER	Locale et Nationale (FR)
VerNet	Equipe incluant les universités de Californie, du Colorado, du Kansas et de Tulane + réseaux de partenaires (EOL, MOL, ADW, NEOMAP, etc.) / Financier : National Science Foundation (NSF)	Javier OTEGUI & site internet	Internationale

*Cf. BIBLIOGRAPHIE (IV)

La variabilité des objectifs et périmètres des systèmes est importante. Cependant, ils traitent tous de consolidation de données d'occurrence espèces, à différents stade du processus de partage de la donnée, depuis la production jusqu'à la diffusion nationale ou internationale, et avec différentes emprises territoriales.

Tous ces systèmes mettent en œuvre des procédures de contrôle ; cependant, le présent recensement de l'existant a permis d'identifier que les pratiques méthodologiques sont souvent peu documentées ou de manière incomplète.

2. Présentation des systèmes analysés

Atlas floristique de la FCBN :

Projet de notice sur la méthodologie de qualification de la complétude et de la cohérence des données de présence.

<http://www.fcbn.fr/>

BDN:

La Base de Données Naturalistes (BDN) a pour objectif de mettre à disposition des gestionnaires forestiers les données faune, flore et habitat. Cet outil de gestion intègre un système de vérification des données avant diffusion.

<http://www.onf.fr/>

<http://outils-naturalistes.fr/applications/fiche/bdn>

BRAHMS :

Le Botanical Research And Herbarium Management System (BRAHMS) est un système de gestion de base de données pour les chercheurs et les herbiers botaniques. Développé à Plant Sciences, Université d'Oxford, il développe des fonctionnalités pour recueillir, éditer, analyser et publier des données botaniques.

Le Royal Botanic Gardens, Kew utilise ces fonctionnalités qu'il adapte en fonction des différents projets dont il a la charge.

<http://herbaria.plants.ox.ac.uk/bol/>

<http://www.kew.org/>

CBN-PMP (cadre de l'élaboration de la liste rouge régionale de la flore vasculaire) :

Le document transmis concerne la méthodologie appliquée pour l'élaboration de la liste rouge régionale de Midi-Pyrénées.

Les catalogues floristiques départementaux ont été établis automatiquement sur la base de la validation des données élémentaires suivant une clé de décision présentée dans ce document. Cette partie relative à la validation est utilisée pour alimenter le recensement de l'existant.

<http://cbnpmp.blogspot.fr/>

CSCF :

Le Centre Suisse de Cartographie de la Faune (CSCF), dans le cadre de ses différentes activités, œuvre notamment à la mise en place d'un système coordonné de diffusion, dans le réseau informatique mondial, des données rassemblées dans les collections des différents musées de Suisse ou dans les centres de coordination nationaux (projet GBIF).

<http://www.cscf.ch/cscf>

Digitale2 :

Digitale2 est le Système d'information sur la flore et la végétation développé au Conservatoire Botanique National de Bailleul (CBNBI).

Digitale2 regroupe toute l'information sur les plantes sauvages et la végétation des régions de Haute-Normandie, Picardie et Nord-Pas-de-Calais. Cet outil, au-delà de la conservation des données, est un outil pour aider les aménageurs et collectivités territoriales dans leurs prises de décisions liées à la gestion du patrimoine naturel.

<http://www.cbnbl.org/ressources-documentaires/Digitale2-une-base-de-donnees>

FLORA :

Système de gestion des données du Conservatoire Botanique National du Bassin Parisien (CBNBP), service scientifique du Muséum National d'Histoire Naturelle.

C'est un dispositif public d'observation de la flore sauvage des régions Bourgogne, Centre, Champagne-Ardenne, Ile-de-France et du département de la Sarthe.

Il vise à collecter, rassembler, gérer, valider et diffuser toutes les informations récentes ou anciennes sur les plantes sauvages, produites sur ces régions.

<http://cbnbp.mnhn.fr/cbnbp/>

GBIF :

Le Global Biodiversity Information Facility est le système mondial d'Information sur la biodiversité qui a pour objectif l'accès libre et gratuit aux données de la biodiversité.

Le GBIF France, en tant que point nodal national, met à disposition les outils et les services créés par le GBIF. Il a également vocation à connecter au portail international les données de spécimens et d'observations hébergées en France, que celles-ci concernent les territoires français ou le reste du monde. Le GBIF utilise différents outils pour améliorer la qualité des données (taxonomique, géoréférencement...) qui sont recensés sur le site et notamment dans le Data Quality Hub:

<http://www.gbif.fr/page/ressources/outils-du-gbif> et <http://www.gbif.es/BDQ.php>.

<http://www.gbif.fr>

INPN :

L'Inventaire National du Patrimoine Naturel (INPN) est un système mis en place afin d'assurer de manière standardisée la restitution de données de synthèses nécessaires à l'expertise, à l'élaboration de stratégies de conservation et à la diffusion d'informations et de rapports nationaux et internationaux sur le patrimoine naturel français (espèces végétales et animales, milieux naturels et patrimoine géologique).

<http://inpn.mnhn.fr/accueil/presentation-inpn>

NBN Record Cleaner :

NBN Record Cleaner est un logiciel développé par le National Biodiversity Network et notamment utilisé par le Joint Nature Conservation Committee (JNCC) et le Biological Records Center (BRC). Ce logiciel est conçu pour vérifier les données de manière cohérente et simple.

<http://www.nbn.org.uk/>

Observatoire de la Flore Sud-Atlantique :

L'Observatoire de la Flore Sud-Atlantique (OFSA) est un dispositif public d'observation de la flore sauvage des régions Aquitaine et Poitou-Charentes.

Il vise à collecter, rassembler, gérer, valider et diffuser toutes les informations récentes ou anciennes sur les plantes sauvages, produites par le réseau des observateurs de l'Observatoire sur ces deux régions.

<http://www.ofsa.fr/>

ODIN :

L'Outil de Diffusion de l'Information Naturaliste de Haute-Normandie (ODIN) est l'application support de la plate-forme de Haute-Normandie. Cet outil constitue le point d'accès et de gestion de

l'ensemble des données de biodiversité (point d'accès uniquement pour les données élémentaires relatives à la flore et aux habitats).

<http://www.biodiversite.hautenormandie.fr/SINP/Outil-de-Diffusion-de-l-Information-Naturaliste>

Pêches électriques BDMAP (SIE) :

Le Système d'Information sur l'Eau (SIE) est un dispositif créé par l'État pour le partage et la mise à disposition des données sur l'eau du secteur public.

Ces données concernent la ressource en eau, les milieux aquatiques, leurs usages et en particulier les services publics d'eau et d'assainissement pour la France métropolitaine et les départements d'outre-mer.

Sera plus particulièrement présenté le cas des données de pêches électriques

<http://www.eaufrance.fr/3/a-propos-d-eaufrance/le-systeme-d-information-sur-l-eau>

Sigogne :

Sigogne est le système d'information du SINP régional en Franche-Comté.

L'animation de la plate-forme informatique est confiée à la maison du patrimoine Naturel de Franche-Comté.

<http://www.sigogne.org/>

SILENE :

SILENE est le portail régional des données naturalistes en Provence-Alpes-Côte d'Azur faisant fonction de plateforme régionale du SINP. Plusieurs modules permettent la consultation des données de faune, de flore (et prochainement d'habitats). Flore et Habitats sont également renseignés pour le Languedoc-Roussillon (aire de compétence du CBNMed).

C'est un outil public et collectif au service de la prise en compte de la biodiversité.

<http://silene.eu/>

SI Quadrige² (SIE):

Le Système d'Information Quagrige² (Q²) contient des données issues de la surveillance sanitaire et observations environnementales des eaux littorales, dans le cadre de réseaux d'observation et de surveillance locaux et nationaux. Q² contient notamment toutes les données acquises pour l'évaluation DCE de la qualité des masses d'eau littorales.

<http://envlit.ifremer.fr/resultats/quadrige>

VerNet :

Vertnet est un projet collaboratif financé par la National Science Foundation (NSF). VerNet a pour objectif de rendre accessible gratuitement, sur le web, les données de biodiversité (vertébrés) de manière rapide, durable et évolutive et propose des applications notamment pour l'amélioration de la qualité des données et leur visualisation.

<http://portal.vertnet.org/>

II. SYNTHÈSE ET ANALYSE

1. Terminologies utilisées par les différents systèmes

«-» = Pas d'information

SYSTEME	Typologie utilisée*	Définition associée
Atlas floristique de la FCBN	Qualité des données	La qualité des données est un terme générique décrivant les caractéristiques des données : complètes, fiables, pertinentes, à jour, cohérentes, mais aussi des processus qui permettent de garantir ses caractéristiques. Les données sont dites de qualité si elles satisfont aux exigences de leurs utilisations. En d'autres termes, la qualité des données dépend autant de leur utilisation que de leur état. Les principales dimensions de la qualité des données : utilité, interprétabilité, accessibilité et crédibilité.
BDN	Validation	Organisation, anté-diffusion, d'un contrôle de qualité des données. La donnée est valable et peut être validée, un doute subsiste ou elle est aberrante et donc refusée.
BRAHMS	Contrôle de qualité	-
	Vérification de données	-
	Normalisation	-
CBN-PMP	Validation des données élémentaires	-
	Validation départementale	Validations départementales calculées à partir des validations des données élémentaires réalisées par le CBN-PMP. <i>Il s'agit ici d'une validation pour l'élaboration d'une synthèse.</i>
CSCF	Contrôle de cohérence	Ce contrôle automatisé, vise essentiellement à détecter les fautes de report des coordonnées ou d'encodage des nombreux champs identifiant l'espèce ou la source de l'information.
	Contrôle de validité scientifique	Contrôle des informations par spécialistes interposés.
Digitale2	Vérification	Conformité des informations saisies par rapport à la source et aux concepts de Digitale 2.
	Pré-validation	Traitement de pré-validation (réconciliation avec les référentiels ; calculs d'état, affectation d'un niveau de fiabilité) incluant la notion de précision.
	Validation	Validation par expertises (faisant appel à des référents) Associe la notion de niveau de validité et la notion de fiabilité. La validation se fait sur les différents niveaux de précision : à la région, à la commune et à la

		localisation citée.
	Niveau de fiabilité globale	Le niveau de fiabilité est une note qui synthétise les différentes qualifications apportées aux observations lors des phases de prévalidation chronologique, géographique et nomenclaturale, lesquelles étant confirmées ou modifiées, par les référents régionaux, lors du processus de validation.
FLORA	Validation pour diffusion	Validation experte par les botanistes.
	Validation retour de terrain	Relecture des bordereaux papier.
	Validation post saisie	Validation experte par les botanistes.
GBIF/GBIF France	Vérification Qualification Endorsement (approbation)	Ensemble de contrôles comprenant la complétude (ensemble des champs nécessaires renseignés, métadonnées), la cohérence entre les informations fournies et la recherche d'erreur (taxonomique, géoréférencement...). Vérification de la conformité du format et des données (doublons, identifiants, etc.). Chaque nouveau fournisseur de données doit être approuvé par un point nodal national ou thématique.
INPN	Vérification	Processus technique de contrôle de conformité qui vise à détecter les données qui ne sont pas conformes à des critères pré-établis. Cette vérification ne préjuge pas de l'usage, elle se fait par rapport à des critères génériques et des bases de connaissance.
	Validation pour diffusion	Valeur liée aux résultats des vérifications précédentes permettant de définir de la diffusion ou non des données pour la diffusion.
	Qualification	La qualification est le résultat final de la sélection de données pour un usage défini (distribution, répartition, etc). Seules les données validées sont qualifiées.
NBN Record Cleaner	Validation	La validation est le processus qui consiste à vérifier l'adéquation aux critères établis (ensemble intégré de règles). Comprend un ensemble de test. Résultat binaire.
	Vérification	Consiste à vérifier la crédibilité des données par rapport à une expertise.
Observatoire de la Flore Sud-Atlantique	Donnée sous forme qualifiée	Donnée la plus fiable possible, techniquement et scientifiquement.
	Qualification du lot de données	Renseignement et vérification des métadonnées.
	Normalisation technique	Mise en conformité au format standard de données et contrôles géographiques.
	Normalisation scientifique	Mise en conformité au référentiel taxonomique officiel (Taxref).
	Validation scientifique des observations	Contrôle de fiabilité, détection des erreurs éventuelles de détermination...
ODIN	Validation	Consiste à apporter une caution scientifique par la qualification des données stockées dans la plate-

		forme après intégration.
	Qualification	Résultat de la validation (attribution du statut O_VALI)
Pêches électriques (SIE)	Niveau de validation (Statut)	Etat d'avancement de la validation de l'ensemble des données biologiques produites au cours de l'opération de prélèvement. Cette information est complétée par la qualification des résultats biologiques. Le statut des résultats de l'opération de prélèvement biologique repose sur la liste des valeurs possibles définies dans la nomenclature n°446. (Cf. dictionnaire Sandre « Processus d'acquisition des données biologiques » version 4.2)
	Niveau de qualification	Etat d'avancement de la validation de l'ensemble des données biologiques produites au cours de l'opération de prélèvement. Cette information est complétée par le statut de l'opération de prélèvement biologique. <i>Par exemple, pour différencier une donnée « validée par un processus automatisée » d'une donnée « validée par un expert », il n'est pas créé deux niveaux de validation. Au contraire, dans les deux cas, la donnée est qualifiée « Correcte » mais elle est accompagnée d'une information complémentaire, le statut de la donnée, qui indique l'état d'avancement de cette validation : dans le premier cas, on indique le couple (CORRECTE, CONTROLE AUTOMATIQUE). Dans le second, il s'agit du couple (CORRECTE, EXPERT). Toute validation est toujours indiquée par ce couple.</i> La qualification des résultats biologiques repose sur la liste des valeurs possibles définies dans la nomenclature n°414 (Cf. dictionnaire Sandre « Processus d'acquisition des données biologiques » version 4.2)
	Validation	La validation s'effectue par le couple (qualification, statut).
Sigogne	Validation scientifique	-
SILENE FAUNE	Analyse technique	Validation technique de la donnée consistant à vérifier la présence valide des informations : espèces, observateur, localisation et date.
	Validation scientifique	S'appuie sur une série de listes (tableurs) interrogées par scripts informatiques et une validation manuelle (faisant intervenir de l'expertise).
Si Quadrige ²	Qualification	Le niveau de qualification correspond au niveau de confiance des données. Il conditionne les modalités de diffusion de ces données et l'utilisation dans le cadre de traitement spécifique.
	Qualification automatique	Consiste à rechercher des erreurs « grossières » et facilement identifiables pouvant être décelées informatiquement en définissant des règles de contrôle simples.
	Qualification experte	Consiste à mettre en évidence les données statistiquement aberrantes via des méthodes adaptées (séries temporelles, tests statistiques...).

	Contrôle/validation	Consiste à éditer les données saisies pour vérifier leur cohérence vis-à-vis du cahier de paillasse (ou des fiches terrain). La validation est le résultat du contrôle.
VerNet	Qualité spatiale	Ensemble de contrôles comprenant la complétude (ensemble des champs nécessaires renseignés), l'inconsistance (la cohérence entre les informations fournies) et la recherche d'erreur (erreur de système de coordonnées par exemple).

**Liste des résultats (valeurs en sortie) : Cf. 2.6*

Les définitions actuelles du glossaire SINP sont les suivantes :

Qualification : C'est un ensemble de procédures permettant d'apprécier la qualité technique et scientifique d'un lot de données. Cette qualification est effectuée initialement par le producteur, puis par les plate-formes régionales ou thématiques enfin par le MNHN au niveau de la plateforme nationale.

Qualité (de la donnée) : C'est l'étude des caractéristiques d'exactitude, de complétude, de précision, de cohérence interne et externe des données par rapport à leur méthode de production et/ou à leur utilisations.

Validation : La validation correspond à l'attribution d'un niveau de qualité prédéfini (norme, standard, fiabilité, exploitabilité, accessibilité...).

Bilan :

Force est de constater l'hétérogénéité des terminologies utilisées par les différents systèmes. Ce constat abonde la nécessité de définir des terminologies communes dans le cadre de la qualification des données du SINP.

On constate, qu'au-delà du vocabulaire utilisé, se dessine plusieurs grands types de « contrôle ». Notamment un contrôle « automatique informatique » (formats, types, etc.), un contrôle « semi-automatique » lié à de l'expertise en amont (faisant appel à des bases de connaissance, des référentiels préexistants) et un contrôle « manuel » par expertise (dire d'expert).

Certains systèmes ne mentionnent pas directement l'usage de la donnée dans les définitions transmises. Cependant, dans certains cas, l'usage est implicite, comme la diffusion ou l'usage en carte de synthèse.

Par ailleurs, certaines définitions (statut) font référence à un état dans le processus de contrôle.

Proposition de définition suite aux réunions du groupe de travail sur la validation:

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.

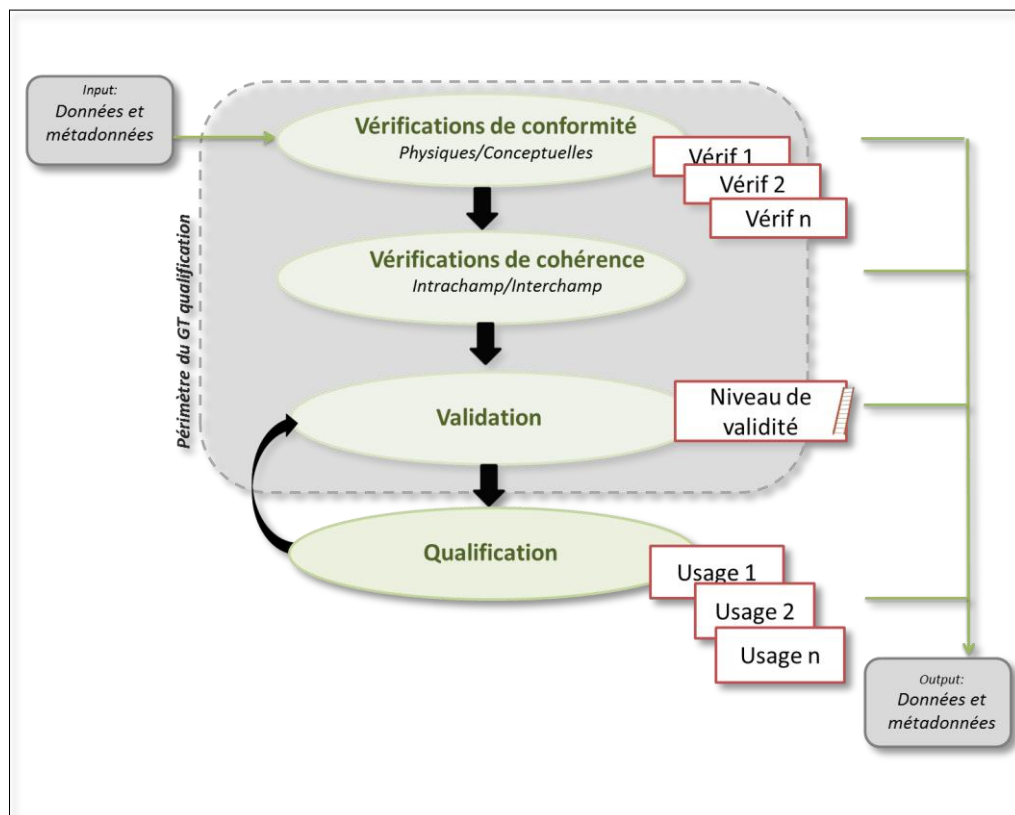
Validation :

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).

Schéma général du processus de contrôle



2. Types de contrôle

«X» = contrôle réalisé

« » = contrôle non réalisé

«?»= information inconnue

SYSTEME	Type de « contrôle » mis en place			
	Contrôle de conformité au regard d'un standard (format, champs obligatoires, cohérence inter-champs)	Contrôle semi-automatique (en général par rapport à des référentiels ou des bases de connaissance existantes)	Expertise (avis d'expert sur la conformité de l'information)	Expertise visant à permettre l'adéquation à l'usage
Atlas floristique de la FCBN	X Conformité/standardisation	X	X	X (pour la production des distributions de référence notamment)
BDN	X	X	X	
BRAHMS	X (standard interne pour les spécimens de collection)	X (notamment pour les pays cités et les taxons cités)	?	?
CBN PMP	X	X	X	X (impact la prise en compte dans le cadre de l'évaluation liste rouge)
CSCF	X	X	X	
Digitale2	X	X	X	X (cartes de répartition, exploitation à la commune, exploitation à la localisation citée (la validation est réalisée à la région, à la commune ou à la localisation citée))
FLORA	X	X	X	
GBIF/GBIF France	X	X		
INPN	X	X	+/- variable selon les programmes	X (impacte la diffusion de la donnée)
NBN	X	X	X	
Observatoire de la Flore Sud-Atlantique	X (mise en conformité au Format Standard de Données (FSD) & référentiel taxonomique)	X (pour les espèces communes et d'identification aisée)	X	Normalisation scientifique (rattachement à Taxref) : Automatique pour les données saisies

				sous OFSA/ Semi-automatique pour les données transmises par fichier/manuelle pour les données bibliographiques anciennes.
ODIN	X	X	X	
Pêches électriques (SIE)	X (contrôle de conformité aux formats et référentiels Sandre)	X (contrôle de cohérence par rapport aux protocoles)	X	
Sigogne	X (renseignement par le producteur d'une fiche de description des opérations de validation et des dispositifs de collecte)	?	?	?
SILENE	X	X	X	X (impacte la diffusion de la donnée)
Si Quadrige ²	X	X	X	X (impacte la diffusion de la donnée)
VerNet	X	X		

Bilan :

La plupart des systèmes analysés met en place une procédure « multicritères » en associant des paramètres automatiques, semi-automatiques et de l'expertise manuelle dans l'objectif d'aboutir à une caractérisation finale des informations (SIE, INPN, Q², ODIN, FLORA, etc.). Les différentes étapes de contrôle suivent généralement une démarche chronologique. Les contrôles « techniques » de format, de type de champs, etc. consistent en la première étape mise en œuvre. Les contrôles semi-automatiques faisant appel à des référentiels sensu stricto ou à des expertises standardisées (liste de référence ou profil type défini) arrivent ensuite. La sollicitation d'experts dédiés pour intervenir directement sur la donnée arrive en général en fin de chaîne de traitement.

Certains systèmes (NBN Record Cleaner par exemple) n'ont pas vocation à appliquer une procédure de contrôle stricte sur les données mais à fournir des solutions à l'utilisateur qui souhaiterait en améliorer la qualité. C'est ainsi l'utilisateur de l'outil qui est maître de la méthodologie qu'il souhaite appliquer.

Pour les systèmes analysés, il est rare qu'une expertise explicite de l'adéquation à l'usage soit mise en place. Cette analyse tend à être déportée au niveau de l'utilisateur ou du programme, qui définit sa méthodologie et, en fonction, détermine les requêtes (notion de filtres) pour sélectionner les données pertinentes pour répondre à la question posée.

Cependant, même si cette qualification par rapport à l'usage n'est pas toujours explicite, elle est néanmoins souvent sous-jacente dès lors qu'il y a sélection avec un objectif de diffusion (par exemple de couches de référence).

Concernant le niveau d'intervention des différents contrôles dans la procédure de diffusion, il semble que les contrôles de structure, de format ou de cohérence interchamps tendent à être réalisés au préalable à l'intégration dans le système (cas de Brahms par exemple). Les contrôles semi-

automatiques tendent à être réalisés avant la diffusion des données (cas de l'INPN par exemple). Les contrôles de qualité faisant appel à une démarche d'expertise sont réalisés soit avant la diffusion soit après.

3. Niveau de mise en œuvre des procédures de contrôle

L'analyse porte ici sur les niveaux de mise en œuvre des procédures de contrôles dans le processus de partage de la données, depuis la production jusqu'à la diffusion nationale.

«X» = contrôle réalisé

« » = contrôle non réalisé

«?» = information inconnue

SYSTEME	Niveau de mise en œuvre			Commentaire
	Producteur	« Regroupeur » 1-n	« Regroupeur » National (ou supra-national)	
Atlas floristique de la FCBN			X	Mise en place de seuils de rejet
BDN			X	Réseaux naturalistes comprenant des experts nationaux et des experts plus locaux.
BRAHMS			X	
CBN PMP		X	X	Pour validation de la liste rouge régionale et nationale
CSCF			X	Une confirmation de l'information concernée peut être demandée à l'observateur
Digitale2		X		
FLORA	X	X		Au retour de terrain, les bordereaux papier sont relus, validés puis saisis (validation producteur).
GBIF/GBIF France	X		X	La responsabilité des données publiées revient aux producteurs de données ; un contrôle avec retour au fournisseur est également mis en place au niveau nodal (le fournisseur n'étant pas nécessairement le producteur).
INPN	X	X	X	Des échelles producteur ou regroupeur infra-national peuvent être mises en place dans le cadre de programmes spécifiques.
NBN Record Cleaner	X	X	X	Le logiciel peut être utilisé à différentes étapes (dépend de l'utilisateur : individu, organisme, etc.)
Observatoire de la Flore	X	X		Validation manuelle par des botanistes au plus proche du terrain

Sud-Atlantique				(« référents départementaux »)
ODIN	X	X		Le niveau producteur peut être contacté en cas de doute.
Pêches électriques (SIE)	X			Le responsable de la procédure de contrôle peut cumuler les deux niveaux.
Sigogne	X	X		Le questionnaire producteur inclus le renseignement des procédures de validation producteur et des moyens mis en œuvre pour y aboutir.
SILENE FAUNE	X	X		La démarche première consiste en la responsabilisation des fournisseurs de données vis-à-vis de qualité des données fournies. Un retour au fournisseur est également mis en place.
Si Quadrige ²	X	X	X	Les contrôles sont d'abord producteur, puis effectués par le centralisateur national. A chaque doute sur une donnée, le qualificateur national contacte un représentant local des producteurs de données.
VerNet			X (International)	

Bilan :

L'échelle d'application de la procédure de contrôle varie fortement selon les structures consultées. Certains acteurs au sein d'une structure peuvent parfois cumuler plusieurs rôles (plusieurs niveaux d'implication dans la procédure avec par exemple un contrôle en tant que producteur tout d'abord, puis un contrôle en tant que tête de réseau). A l'inverse, un même rôle peut être tenu par des experts de plusieurs niveaux. C'est-à-dire que des acteurs intervenant à des échelles géographiques différentes (experts nationaux, régionaux, locaux, etc.) peuvent intervenir sur le même niveau de contrôle et réaliser la même opération de validation.

Lorsque des procédures de contrôle sont mises en place sur plusieurs niveaux (une validation producteur et « regroupeur » par exemple), elles sont en général complémentaires et associées. C'est-à-dire que le résultat des contrôles au niveau producteur est utilisé pour mettre en œuvre ceux réalisés au niveau regroupeur, servant de filtre pour orienter la procédure du niveau supra vers les données à vérifier prioritairement.

La validation des données dans certains cas (SIE notamment) est sous la responsabilité des producteurs des données. Des contrôles supplémentaires (conformité et cohérence principalement) sont réalisés par les systèmes d'informations récepteurs de ces données, pouvant aboutir à des rejets de données ou à de simples alertes, les producteurs en étant informés et devant corriger leurs données.

On constate également la mise en place de réseaux de contrôles par spécialité (géographique ou thématique).

On note par ailleurs la responsabilisation du niveau producteur quant à la pré-validation de ses données. Le retour, en cas de doute, a généralement lieu en prenant contact avec le partenaire, au plus près de la production de la donnée. Il a été cependant remonté que ce niveau de contrôle n'est pas nécessairement fiable car la procédure de contrôle n'est bien souvent pas documentée ni

partagée. De plus, les échanges retours du « regroupueur » avec les producteurs ne sont pas nécessairement pris en compte et répercutés dans les bases sources.

Cependant, ce fonctionnement pluri-échelles apparaît intéressant de par la dynamique et les liens qu'il génère et maintient entre les différents partenaires.

4. Sources utilisées dans les processus de contrôle

Nous traitons ici de la provenance des informations utilisées pour réaliser les différents contrôles.

«X» = contrôle réalisé

« » = contrôle non réalisé

«?» = information inconnue

SYSTEME	Sources utilisées pour l'analyse		Commentaire
	Données	Métadonnées	
Atlas floristique de la FCBN	X	?	
BDN	X		
BRAHMS	X		Pas de gestion de métadonnées pour les spécimens d'herbier
CBN PMP	X	?	
CSCF	X	?	
Digitale2	X	*	* La référence bibliographique de la source des données est saisie et vérifiée dans un logiciel documentaire PMB.
FLORA	X		
GBIF/GBIF France	X	X	Contrôle de fiabilité du fournisseur+ contrôle sur les données et les métadonnées
INPN	X	X	Vérification notamment de la cohérence entre les métadonnées et les données.
NBN Record Cleaner	X	X	
Observatoire de la Flore Sud-Atlantique	X	X	
ODIN	X	X	Les métadonnées sont utilisées uniquement en cas de validation manuelle (expert).
Pêches électriques (SIE)	X		
Sigogne		X	La documentation transmise concerne uniquement les dispositifs de collecte.
SILENE FAUNE	X	X	Prise en compte du programme d'acquisition source ; donc des métadonnées.
Si Quadriges ²	X	X	Chaque niveau de métadonnée peut être qualifié. L'analyse de tout l'historique permet

			plus facilement de discriminer les données qui sortent des valeurs classiques.
VerNet	X		

Bilan :

Tous les jeux de données traités par les différentes structures ne sont pas nécessairement accompagnés d'un fichier de métadonnées. Cependant, lorsqu'un fichier de métadonnées est associé au jeu de données, il est fréquent que ce fichier soit également soumis à des contrôles (qui consistent en général à vérifier la présence des informations « obligatoires » pour la structure).

Des contrôles de cohérence entre métadonnées et données transmises peuvent parfois être mis en place. Cela consiste par exemple à vérifier que, lorsque les métadonnées mentionnent que le jeu concerne la flore, les données portent bien uniquement sur des taxons flore ou encore qu'un jeu de données de suivi temporel entre 2000 et 2010 ne contient pas de données qui outrepassent les bornes annoncées.

Même si la majorité des procédures ne mentionne pas explicitement la prise en compte des métadonnées, les analyses les intègrent pour autant en général de manière implicite. La connaissance de l'origine des données, du but pour lequel elles ont été collectées, du protocole utilisé, de la structure productrice, etc. sont des informations qui vont être susceptibles d'être prises en compte lors de l'intervention des experts dans la procédure de validation.

On constate également que certains système (SIE notamment) n'appliquent pas de procédure de contrôle complète lorsque les données ont été fournies par une structure reconnue et que celle-ci a son propre processus de validation. La procédure réalisée par la structure receveuse, dans le cas de « reprise de données » est donc bien modifiée en fonction de la provenance, et donc des métadonnées du jeu.

Les informations portées au niveau du jeu de données (métadonnées) sont cependant rarement utilisées dans les opérations automatiques probablement en lien avec le manque de standardisation des formats d'échange pour les métadonnées.

Les informations relatives au protocole de collecte de données, à l'objectif, au degré de complétude (taxonomique, géographique ou temporelle), au maître d'œuvre sont tout particulièrement utilisées pour définir de l'aptitude à l'usage (cas de l'INPN notamment).

5. Objet portant le résultat de l'analyse

Nous analysons ici l'objet (donnée ou métadonnée) portant le résultat suite à la mise en place des procédures de contrôle.

«X» = contrôle réalisé

« » = contrôle non réalisé

«?» = information inconnue

SYSTEME	Résolution du résultat de l'analyse		Commentaire
	Données	Métadonnées	
Atlas floristique de la FCBN	X	?	
BDN	X		
BRAHMS	X		
CBN PMP	X		
CSCF	X		
Digitale2	X	X*	Le résultat de la procédure abouti à l'attribution d'une valeur de « fiabilité » pour l'observation. * La référence bibliographique de la source des données est saisie et vérifiée dans un logiciel documentaire PMB.
FLORA	X		
GBIF/GBIF France	X		Le résultat du contrôle sur les données est porté au niveau de la donnée elle-même et affiché sur le site.
INPN	X	X	Un premier résultat aux contrôles est porté sur la donnée. Une expertise par jeu de donnée permet d'informer sur la complétude géographique du jeu.
NBN Record Cleaner	X		
Observatoire de la Flore Sud-Atlantique	X		
ODIN	X		
Pêches électriques (SIE)	X		Le résultat est porté sur la donnée mais peut être attribué aussi par lot de données.
Sigogne		X	La documentation transmise concerne uniquement les dispositifs de collecte.
SILENE	X		
Si Quadrige ²	X	X	Chaque niveau de métadonnées porte son propre niveau de qualité (l'information de localisation spatio-temporelle peut être

			qualifiée « Bon », tandis que le résultat d'une analyse correspondante peut être « Douteux »). Chaque niveau de métadonnées fait l'objet de contrôles spécifiques.
VerNet	X		

Bilan :

Dans la majorité des systèmes étudiés, les résultats de contrôle sont portés au niveau de la donnée avec une possibilité d'attribution du résultat par « lot de données » dans l'objectif de faciliter le renseignement par l'opérateur. Il faut ici bien distinguer ce qui concerne les outils techniques permettant d'appliquer un résultat d'analyse à un ensemble ou groupe de données de la résolution finale qui porte le résultat de l'analyse (en général la donnée).

Une information complémentaire liée à la complétude des données d'un jeu ou en lien avec la notion de distribution ou de répartition peut être définie à l'échelle du jeu de données, et s'ajoute alors à l'analyse de résolution « données ». La notion de complétude est une notion relative car variable selon l'usage visé.

Les contrôles relatifs à la « validation » portent donc sur la donnée tandis que les analyses réalisées en fonction d'un usage spécifique s'appliquent à des ensembles de données et se véhiculent ainsi via les métadonnées.

6. Valeurs attribuées dans les processus de contrôle

Atlas floristique de la FCBN
<i>Valeurs sous forme de note pour 9 critères:</i>
Conformité
Complétude
Précision
Cohérence
Exactitude
Fidélité
Interprétabilité
Opportunité
Duplication
<i><u>Puis traduction selon 3 échelles :</u></i>
<i><u>Echelle de conséquence</u></i>
Inutilisable
Partiellement utilisable
Utilisable
<i><u>Echelle de gravité</u></i>
Aucune
Mineur
Modéré
Majeur
Grave
<i><u>Echelle de priorité de traitement</u></i>

Très élevée Elevée Moyenne Faible Nulle
BDN
<u>Valeurs pour la validation :</u> Donnée valide Donnée plausible Donnée refusée
BRAHMS
Concaténation de plusieurs résultats de contrôle au sein d'un même champ porté au niveau enregistrement.
CBN-PMP
<u>Valeurs pour la validation des données élémentaires:</u> Valide indigène Valide manifestement introduit Valide subspontané Valide origine inconnue Exclue Douteuse A confirmer A réindexer
<u>Valeurs pour la validation départementale</u> Valide indigène Valide manifestement introduit Valide subspontané Valide origine inconnue Exclue Douteuse A confirmer
CSCF
VALID = Null donnée non encore traitée VALID = 1 donnée qualifiée de sûre sans procédure particulière VALID = 2 donnée pour laquelle une validation a été demandée et qui s'est révélée exacte VALID = 3 donnée considérée comme incertaine après validation VALID = 4 donnée validée qui correspond à un individu dont l'identification est correcte mais la provenance est erronée en fonction de son aire de distribution VALID = 5 donnée validée qui correspond à un individu dont l'identification est correcte mais accidentellement ou volontairement introduite dans le site correspondant VALID= 6 donnée validée d'individus dont l'identification est correcte mais dont l'origine est ancienne (issus de sédiments par ex.)
Digitale2

Valeurs par type de pré-validation :

1: Prévalidation géographique

Localisation avérée

Localisation interprétée

2: Prévalidation chronologique (action de validation, niveau de précision)

Action de validation :

Inconnu

A analyser

Aucune (à ne pas analyser)

Plus aucune action

Niveau de précision :

Vague/extrapolé

Imprécis

Précis

3: Prévalidation taxonomique ou habitat (action de validation, niveau de validité, fiabilité)

Action de validation :

Aucune action

Taxon inconnu

Taxon présent

Taxon hypothétique

Taxon absent

Taxon douteux

Taxon cité par erreur

Taxon nouveau

Taxon à présence inconnue

A ne pas valider

Valeurs de validité (par échelle : région/commune/localisation citée ou retenue):

Plausible à la région

Plausible à ma commune

Plausible au polygone

Douteux, à confirmer

Cité par erreur

Non traité

Traduction en niveau de validité :

Aucun

Plausible dans la région

Fiabilité :

Observation non validée

Observation à analyser, à vérifier

Valeurs pour la « fiabilité » de l'observation :

Observation fiable au polygone

Observation plausible au polygone

Observation fiable à la commune

Observation plausible à la commune

- Observation plausible à la région
- Observation à analyser, à vérifier
- Observation douteuse ou à confirmer
- Observation erronée
- Observation non validée

Niveau de validité taxon. ou syntax.	Précision chronologique + Action de validation	Nature de localisation									
		1 : Localisation toponymique					2 : Localisation cartographique				
		Cité à la région /départ (T)	Cité à la commune (C)	Cité au site (S)	Cité au lieu-dit (L)	Cité à la maille (M) etc.	Secteur approximatif (Lo) ou indéterminé (?)	Zone de prospection (Zo)	Occurrence (Oc)	Occupation (In)	
Plausible au polygone (5)	Précis (2 ; 2.1 ; 2.2) + Aucune(2) ou Validation réalisée(#)	/	1a	1a	1a	1a	1a	1a	1a	1	1
	Imprécis (1 ; 1.1 ; 1.2) + Validation réalisée(#)	/	1a	1a	1a	1a	1a	1a	1a	1a	1a
	Imprécis (1 ; 1.1 ; 1.2) + A analyser (1)	/	2c	2c	2c	2c	2c	2c	2c	2c	2c
	Vague/extrapolé (0 ; 0.1 ; 0.2) + Validation réalisée(#)	/	1a	1a	1a	1a	1a	1a	1a	1a	1a
	Vague/extrapolé (0 ; 0.1 ; 0.2) + A analyser (1)	/	2c	2c	2c	2c	2c	2c	2c	2c	2c
Non traité (null aucune date) + Inconnu, à déterminer (0)	/	0	0	0	0	0	0	0	0	0	
Plausible dans com (3 et 4)	Précis (2 ; 2.1 ; 2.2) + Aucune(2) ou Validation réalisée(#)	/	2a	2a	2a	2a	2a	2a	2a	2	2
	Imprécis (1 ; 1.1 ; 1.2) + Validation réalisée(#)	/	2a	2a	2a	2a	2a	2a	2a	2a	2a
	Imprécis (1 ; 1.1 ; 1.2) + A analyser (1)	/	2c	2c	2c	2c	2c	2c	2c	2c	2c
	Vague/extrapolé (0 ; 0.1 ; 0.2) + Validation réalisée(#)	/	2a	2a	2a	2a	2a	2a	2a	2a	2a
	Vague/extrapolé (0 ; 0.1 ; 0.2) + A analyser (1)	/	2c	2c	2c	2c	2c	2c	2c	2c	2c
Non traité (null aucune date) + Inconnu, à déterminer (0)	/	0	0	0	0	0	0	0	0	0	
Plausible dans rég/dép (2 et 1)	Précis (2 ; 2.1 ; 2.2) + Aucune(2) ou Validation réalisée(#)	2b	2c	2c	2c	2c	2c	2c	2c	2c	2c
	Imprécis (1 ; 1.1 ; 1.2) + Validation réalisée(#)	2b	2c	2c	2c	2c	2c	2c	2c	2c	2c
	Imprécis (1 ; 1.1 ; 1.2) + A analyser (1)	2c	2c	2c	2c	2c	2c	2c	2c	2c	2c
	Vague/extrapolé (idem) + Validation réalisée(#)	2b	2c	2c	2c	2c	2c	2c	2c	2c	2c
	Vague/extrapolé (idem) + A analyser (1)	2c	2c	2c	2c	2c	2c	2c	2c	2c	2c
Non traité (null aucune date) + Inconnu, à déterminer (0)	0	0	0	0	0	0	0	0	0	0	
Douteux (-1)	Précis (2 ; 2.1 ; 2.2) + Aucune(2) ou Validation réalisée(#)	3	3	3	3	3	3	3	3	3	3
	Imprécis (1 ; 1.1 ; 1.2) + Validation réalisée(#)	3	3	3	3	3	3	3	3	3	3
	Imprécis (1 ; 1.1 ; 1.2) + A analyser (1)	3	3	3	3	3	3	3	3	3	3
	Vague/extrapolé (idem) + Validation réalisée(#) ou A analyser (1)	3	3	3	3	3	3	3	3	3	3
	Non traité (null aucune date) + Inconnu, à déterminer (0)	0	0	0	0	0	0	0	0	0	0
Erreur (-2)	Précis (2 ; 2.1 ; 2.2) + Aucune(2) ou Validation réalisée(#)	4	4	4	4	4	4	4	4	4	4
	Imprécis (1 ; 1.1 ; 1.2) + Validation réalisée(#)	4	4	4	4	4	4	4	4	4	4
	Imprécis (1 ; 1.1 ; 1.2) + A analyser (1)	4	4	4	4	4	4	4	4	4	4
	Vague/extrapolé (idem) + Validation réalisée(#) ou A analyser (1)	4	4	4	4	4	4	4	4	4	4
	Non traité (null aucune date) + Inconnu, à déterminer (0)	0	0	0	0	0	0	0	0	0	0
non traité (0)	Précis (2 ; 2.1 ; 2.2) + Aucune(2) ou Validation réalisée(#)	0	0	0	0	0	0	0	0	0	0
	Imprécis (1 ; 1.1 ; 1.2) + Validation réalisée(#)	0	0	0	0	0	0	0	0	0	0
	Imprécis (1 ; 1.1 ; 1.2) + A analyser (1)	0	0	0	0	0	0	0	0	0	0
	Vague/extrapolé (idem) + Validation réalisée(#) ou A analyser (1)	0	0	0	0	0	0	0	0	0	0
	Non traité (null aucune date) + Inconnu, à déterminer (0)	0	0	0	0	0	0	0	0	0	0

FLORA

Valeurs pour l'état de vérification du relevé :

- A vérifier
- Correction demandée
- (les relevés relus sont ceux sans flag)

Valeurs pour la validation de la saisie :

- La saisie est conforme au bordereau
- La saisie n'est pas conforme au bordereau

Description	Liste des espèces	Cartographie Scan 25 ©	Cartographie Google ©	Espèces non prises en compte
STATION				
Géomorphologie (échelle kilométrique) :				
Topographie :				
Pente du versant :		Pente de la station :		Exposition du versant :
Roche mère (carte géologique) :				
Sol (Texture dominante de surface) :				
Commentaire :				
Milieu				
Code corine		Libellé		Pourcentage
Commentaire :				
Etat de la Station				
Etat général :				
Facteur d'influence :				
Menace :	Actif ou potentiel (A ou P) :		Positif ou Négatif (+ ou -) :	

La saisie est conforme au bordereau

La saisie n'est pas conforme au bordereau

Le relevé est considéré comme valide (niveau x) et peut être visible sur le site si on considère que ce niveau de validation est suffisant.

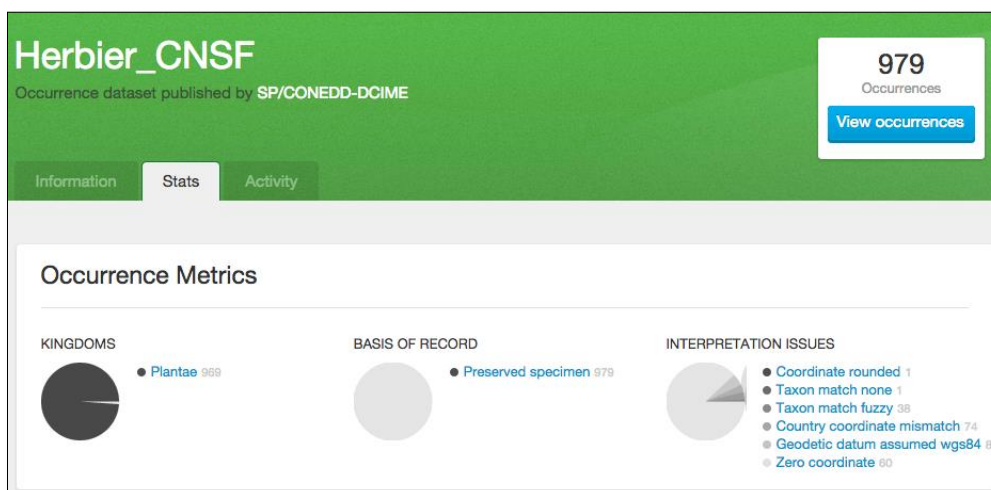
On ouvre la page de demande de modification (exemple ci-dessous)

Valeurs pour la validation des données élémentaires:

- Valide indigène
- Valide Naturalisés
- Valide accidentelle
- Valide plantée/cultivée
- Valide spontané
- Douteuse
- A définir

GBIF/GBIF France

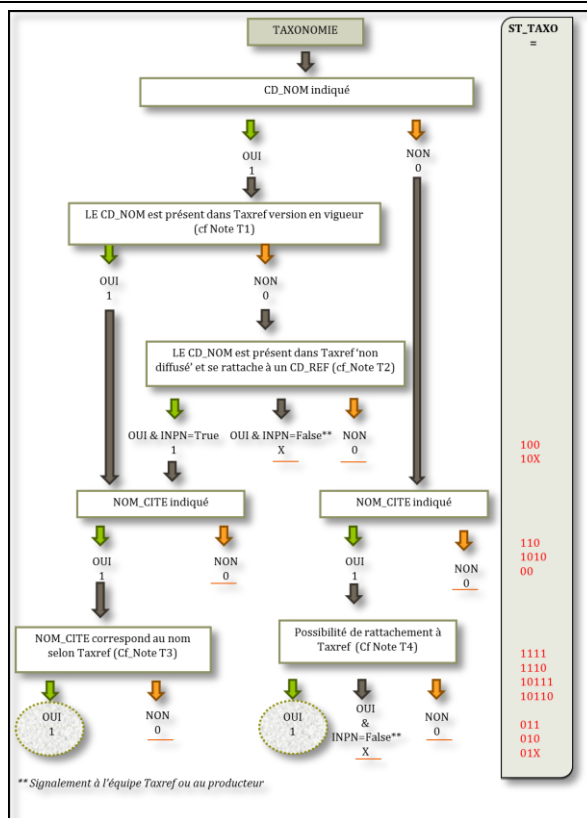
Pages statistiques pour chaque jeu de données avec les « interprétations issues »



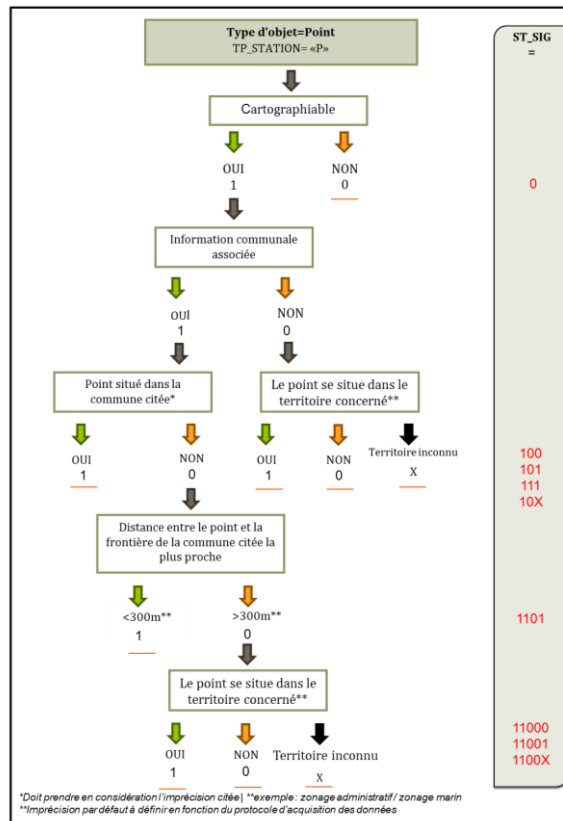
INPN

Valeurs issues des vérifications (les résultats sont exprimés sous forme de code-barres):

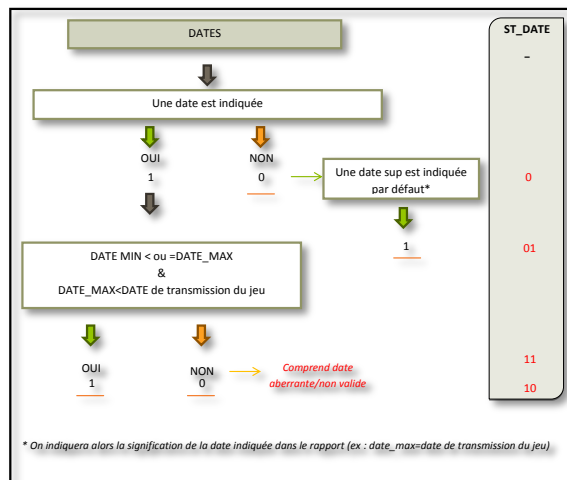
-> Statut taxonomique (ST_TAXO) :



-Statut géographique (ST_SIG)



-Statut date (ST_DATE)



-Statut habitat (ST_HABITAT)

- 0 : L'habitat associé au taxon dans Taxref n'est pas cohérent avec l'emprise géographique du jeu
- 1 : L'habitat associé au taxon dans Taxref est cohérent avec l'emprise géographique du jeu

-Statut de présence

- 0 : Le statut (présent, absent, disparu, etc) associé au taxon dans Taxref n'est pas cohérent avec l'emprise géographique du jeu
- 1 : Le statut associé au taxon dans Taxref est cohérent avec l'emprise géographique du jeu

Valeurs pour la diffusion des données (validation):

- Donnée diffusable pour l'INPN
- Donnée non diffusable pour l'INPN

Valeurs pour la « qualification » des jeux en fonction d'un usage « cartes de répartition et distribution » :

- Distribution de référence
- Données de référence
- Données expertes
- Non Défini
- Données de zonages/programmes espaces

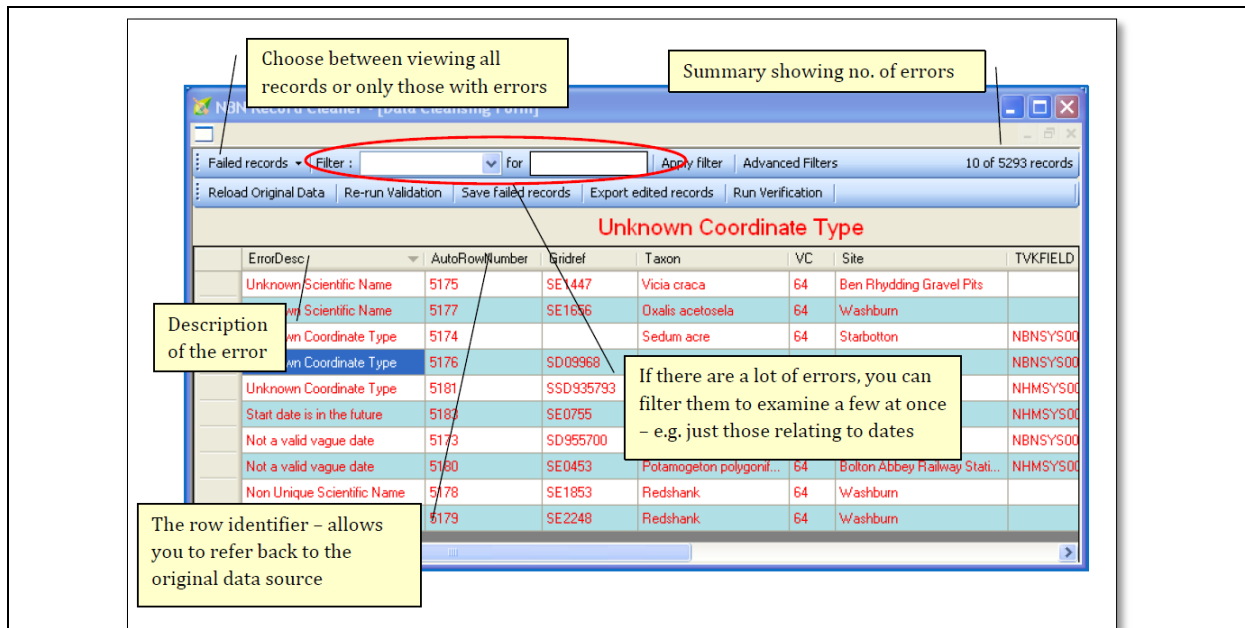
NBN Record Cleaner

Validation :

OK

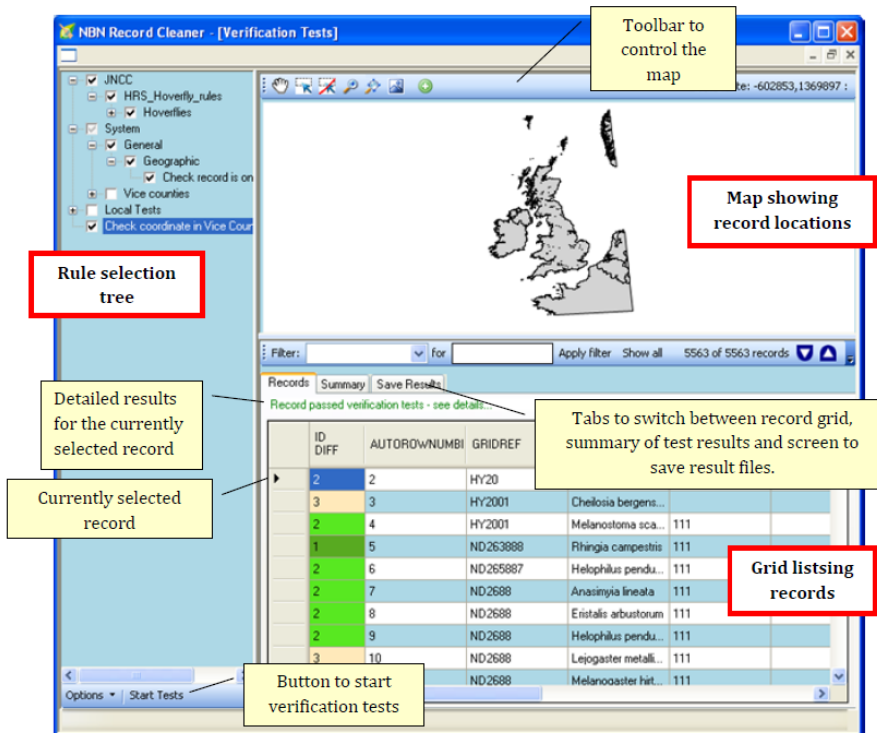


Erreur (avec description de l'erreur) :

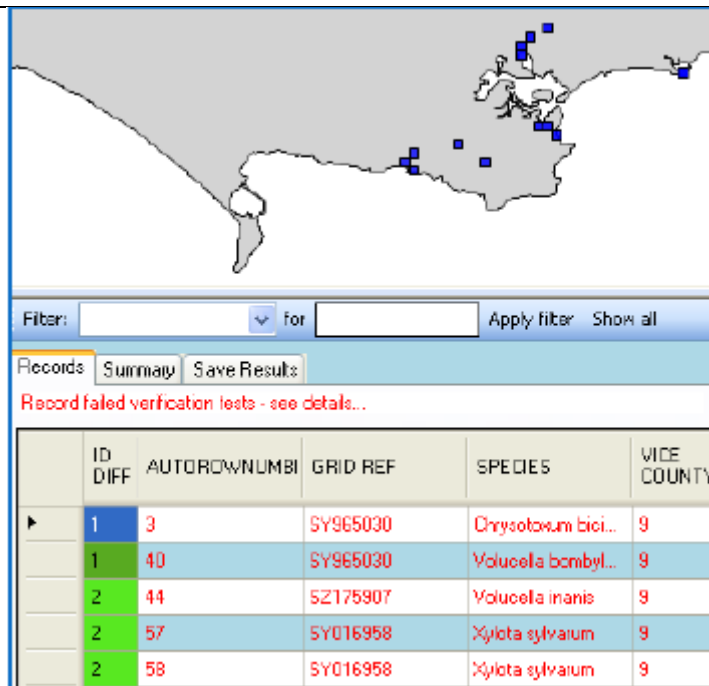


Vérifications :

Les vérifications appliquées sont à l'appréciation de l'utilisateur.



Code couleur : rouge, enregistrements rejetés par le test



Filter: [] for [] Apply filter Show all 5132 of 5563 records

Records Summary Save Results

Test Name	Records Tested	Records Failed	
▶ Melanostoma dubium: check grid is in range	250	241	F
Check coordinate in Vice County	5563	165	
Dasytyrphus hilaris: check grid is in range	17	17	F
Meliscaeva auricollis: check date is during flight period	16	16	F
Melanostoma dubium: check date is during flight period	250	3	F
Eupeodes corollae: check grid is in range	80	2	F
Dasytyrphus hilaris: check date is during flight period	17	1	F
Epistrophe diaphana: check date is during flight period	3	1	F
Melanostoma scalare: check date is during flight period	203	1	F
Anasimyia contracta: check date is during flight period	15	0	F
Anasimyia contracta: check grid is in range	15	0	F
Anasimyia interpuncta: check date is during flight period	1	0	F
Anasimyia interpuncta: check grid is in range	1	0	F
Anasimyia lineata: check date is during flight period	28	0	F
Anasimyia lineata: check grid is in range	28	0	F

Observatoire de la Flore Sud-Atlantique

Validation scientifique des observations :

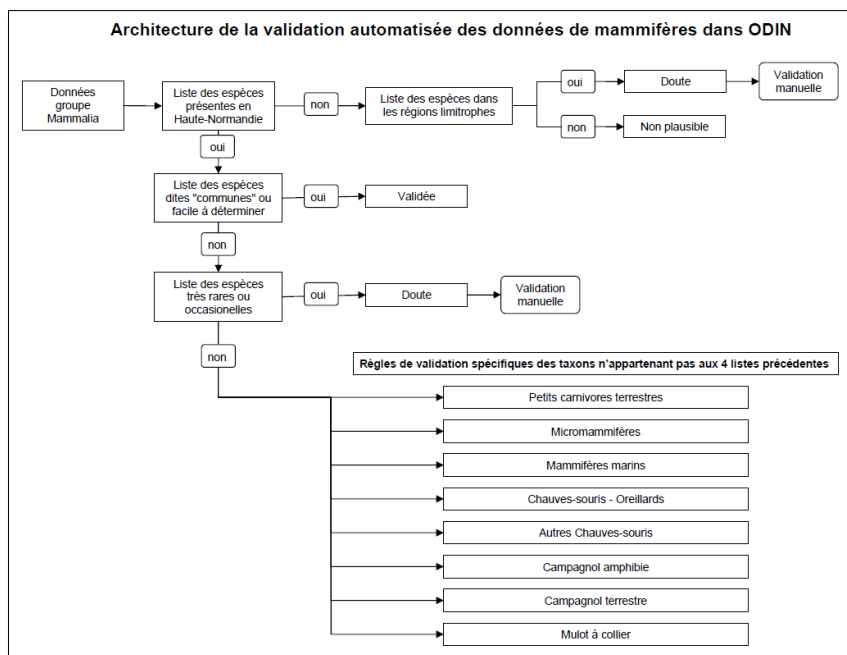
- Valide
- Valide provisoire
- Invalide provisoire
- Invalide
- Non traité

Indicateur	Etat	Note
	VALIDE	4
	VALIDE Provisoire	3
	INVALIDE Provisoire	2
	INVALIDE	1
	NON TRAITE	0

ODIN

Valeurs de la validation :

A valider
Validée
Plausible
Doute
Non plausible



Pêches électriques (SIE)

Pour le contrôle standard (format, champs obligatoires, ...)

Non exploitable-non qualifiable
Qualifiable avec avertissement (warning)
Qualifiable

Pour le contrôle producteur, expert :

Valeurs pour la qualification :

Non qualifié
Correcte
Incorrecte
Incertaine
Non définissable

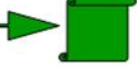
Valeur pour le niveau de validation (statut) :

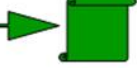
Donnée brute
Donnée contrôlée niveau 1

Donnée contrôlée niveau 2
Donnée interprétée

Notion de Validation

La validation s'évalue par le couple (qualification, statut)

Phase de test	Qualification	Statut	
Contrôle automatique	Non qualifié	Donnée brute	 Jeu de données validé
Contrôle manuel	Correcte	Donnée contrôlée de niveau 1	
Expertise	Incorrecte	Donnée contrôlée de niveau 2	
	Incertaine	Donnée contrôlée de niveau 2	
	Non définissable	Donnée interprétée	


Sandre = standardisé

La validation peut très bien conclure à une qualification incorrecte

SIGOGNE

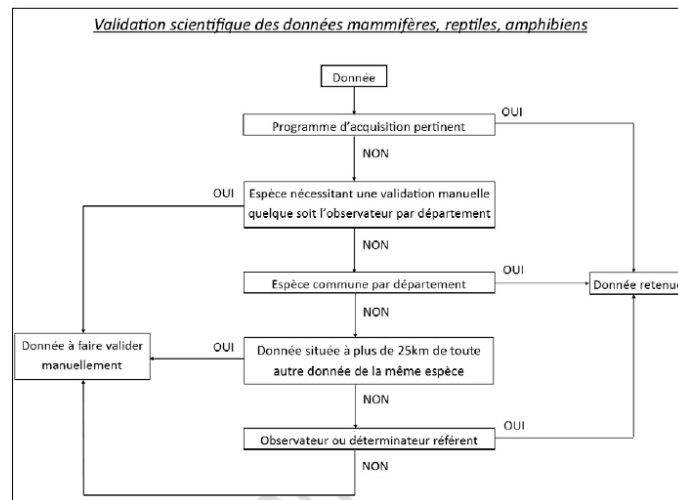
Non renseigné

SILENE FAUNE

Valeurs pour la validation automatique :

Donnée retenue

Donnée à faire valider manuellement



Valeurs pour la validation finale :

Validée par un expert

Retenue informatiquement

A confirmer

Douteuse

Valeurs pour la diffusion :

Diffusée


Non diffusée

Valeurs pour la qualification experte et automatique:

Bon
Douteux
Faux

A l'issue de cette qualification experte, les données sont qualifiées de la façon suivante :

Niveau initial \ Niveau final		Qualification experte		
		BON	DOUTEUX	FAUX
Qualification automatique	BON	X	X	X
	DOUTEUX	Possible, mais rare (cas des données dont l'analyse statistique peut lever le doute)	X	X
	FAUX			

 cas non concerné

Valeurs pour la qualification finale :

Non qualifié
Bonne
Douteuse
Fausse

VerNet

Spatial Quality :

Complétude

Présence des coordonnées : Oui/NON
Présence du pays : OUI/NON
Coordonnées=0 : OUI/NON
Coordonnées avec trois décimales ou plus : OUI/NON
Présence d'un Datum Géodésique (Ellipsoïde + centroïde) : OUI/NON

Inconsistance

Cohérence entre coordonnées et pays cité* : OUI/NON
Ecart (en degrés) entre coordonnées et pays cité* : Nombre
Ecart (en degrés) entre coordonnées et la zone de répartition connue pour l'espèce* : Nombre/Non évalué (lorsque carte de référence inexistante)

Erreurs :

Latitude comprise entre 90 et -90 : OUI/NON
Longitude entre 180 et -180 : OUI/NON
Transposition des coordonnées* : OUI/NON
Latitude correcte* : OUI/NON
Longitude correcte* : OUI/NON

Data completeness +		
Are coordinates present?	✓	Yes
Is the country value present?	✓	Yes
Are both coordinates 0 (zero)?	✓	No
Do coordinates have three or more decimal figures?	✓	Yes
Do coordinates have datum?	✓	Yes
Data inconsistencies +		
Are coordinates within specified country? ¹	✓	Yes
Distance outside of specified country (in degrees) ¹	✓	0
Distance outside of species range map (in degrees) ¹	✓	0
Data Errors +		
Is latitude between 90 and -90?	✓	Yes
Is longitude between 180 and -180?	✓	Yes
Are coordinates transposed? ¹	✓	No
Is latitude hemisphere correct? ¹	✓	Yes
Is longitude hemisphere correct? ¹	✓	Yes

¹Assessed with [Map Of Life](http://map.mol.org/) validation tools

*Évalué à partir l'outil de validation de Map Of Life (<http://map.mol.org/>). Les cartes de répartition sont issues de l'UICN (Union for Conservation of Nature)

Bilan :

Les valeurs en sortie diffèrent entre systèmes de par le nombre d'informations portées par les valeurs de la liste. En fonction des systèmes, une ou plusieurs listes de valeurs sont utilisées selon le nombre d'étapes dans la procédure de contrôle en place. On observe ainsi des résultats « intermédiaires ».

Certaines listes contiennent ainsi des valeurs qui intègrent la traçabilité des modifications (donnée corrigée, etc.), l'état dans le traitement (à valider, donnée en cours de validation, etc.), ou encore l'usage (diffusée pour une carte de distribution, etc.).

La notion de doute, de plausibilité peut être nuancée sur plusieurs valeurs (douteuse, hors statistique, etc.).

On remarque que les contrôles de validité donnent lieu à des résultats graduels (notion de seuil).

7. Sectorisation des contrôles

Est analysé ici si la procédure utilisée est semblable pour l'ensemble des données traitées ou si celle-ci est adaptée pour les types de données particuliers ou pour des groupes taxonomiques spécifiques.

SYSTEME	Emprise thématique du système	Sectorisation pour l'analyse
Atlas floristique de la FCBN	Flore/Habitat	Non
BDN	Faune, Flore	Oui : par groupe taxonomique
BRAHMS	Flore	Oui : les contrôles mis en place sont propres à chaque projet. Au sein d'un même projet, les contrôles sont identiques.
CBN PMP	Flore/Habitat	Non
CSCF	Faune	Oui pour la validation scientifique (distinction entre rhopalocères, odonates, orthoptères, coléoptères du bois, etc.)
Digitale2	Flore/Habitat	Oui : par plante, par habitat, par source, par lieu, permettant au validateur d'avoir plusieurs angles d'analyse et même si les « statuts » de validation sont les mêmes.
FLORA	Flore/Habitat	Possible : par rareté
GBIF/GBIF France	Biodiversité en général	Non
INPN	Faune, Flore, Géologie	Oui : les règles de validation dépendant des programmes et inventaires (méthodologie propre).
NBN Record Cleaner	Faune/Flore	Règles automatiques développées pour les oiseaux, les plantes vasculaires, les papillons et les espèces marines.
Observatoire de la Flore Sud-Atlantique	Flore	Possible (par niveau de difficulté d'identification)
ODIN	Faune, Flore/Habitats (via Digitale2)	Oui : par groupe taxonomique
Pêches électriques (SIE)	Faune (poissons)	Non
Sigogne	Flore/Faune/Habitats	Non renseigné
SILENE	Faune/ Flore	Oui : par groupe taxonomique (oiseaux, mammifères, amphibiens et reptiles, insectes, etc.)
Si Quadrige ²	Hydrologie, microbiologie, biogéochimie, biologie	Oui : les règles de qualifications automatiques et d'expertise diffèrent selon les thématiques des données (hydrologie, microbiologie, biodiversité, etc.).
VerNet	Vertébrés	Non

Bilan :

Les analyses faisant intervenir de l'expertise sont systématiquement sectorisées par thématique.

Les contrôles techniques sont en général réalisés sans sectorisation.

Les analyses semi-automatiques (faisant appel à une expertise anté-traitement) peuvent être sectorisées notamment lorsque les contrôles font appel à des référentiels différents selon la thématique ou le groupe taxonomique. Un travail d'expertise par groupe est alors réalisé en amont.

8. Critères/Paramètres pris en compte dans le processus de contrôle

« » = contrôle non réalisé

« ? » = information inconnue

SYSTEME	Paramètres intégrés pour :			
	Contrôle de structure (format, champs obligatoires, cohérence inter-champs)	Contrôle « semi-automatique » (en général par rapport à des référentiels)	Expertise d'expert sur la conformité de l'information) (avis sur la de	Adéquation à l'usage
Atlas floristique de la FCBN	Conformité/standardisation : Conforme au format défini	Cohérence taxonomique, géographique et temporelle, Précision géographique et temporelle, Fidélité taxonomique, temporelle, Opportunité, Duplication	Cohérence taxonomique, Exactitude taxonomique, Précision taxonomique, Interprétabilité, Duplication	Complétude taxonomique, géographique et temporelle
BDN		Aires de répartitions connues Liste de référence (Taxref, etc.).	Compétences de l'auteur Facilité de détermination Chorologie et congruence des données Saison et type d'habitat	
BRAHMS	Conformité au standard Kew	Contrôle par rapport à des listes de valeurs contrôlées		Non explicite
CBN PMP		En fonction de la connaissance actuelle (détermination, écologie, répartition générale). Congruence avec d'autres données.	En fonction de la connaissance actuelle (détermination, écologie, répartition générale)	Usage Liste rouge : Date, rang taxonomique, statut de présence (espèces introduites, erratiques, etc.)
CSCF	Coordonnées, encodages, code de l'espèce, source de l'information.	Difficulté de détermination, méthode de détermination.	Systématique pour les taxons nouveaux dans une région biogéographique. Preuve (photo, etc)	
Digitale2	Conformité aux standards régionaux et celui de Digitale2 et aux concepts métier	Prévalidation géographique (notamment en lien avec les procédures de	Critères écologiques, chronologiques, d'indigénat + cohérence des	

	associés.	croisements géographiques), chronologique, taxonomique ou habitat Définition de règles de calcul pour l'attribution d'un degré de fiabilité pour l'observation à partir des résultats de prévalidation.	informations de l'observation entre elles. Degré d'expertise de l'observateur. Un niveau de validité est alors affecté.	
FLORA	Standardisation : conforme au format défini.	Cohérence taxonomique, liste d'espèces nécessitant une vérification manuelle (prise en compte de la rareté et de la difficulté de détermination).	Plusieurs paramètres rentrent en ligne de compte et sont plutôt intuitifs faisant partie de l'expérience du contrôleur qui sont l'expertise de l'observateur, sa proportion à faire des erreurs, le cortège de plante observée, la zone géographique habituel de présence, le risque d'erreur sur certaine espèce; l'ambiguïté taxonomique etc....	
GBIF/GBIF France	Présence des champs obligatoires, formats	Contrôle par rapport à des listes de valeurs contrôlées. Exemple : Catalogue of Life, Taxref, pour la taxonomie. Cohérence géographique, duplication.		Information taxonomique, géographique et temporelle.
INPN	Informations obligatoires présentes, cohérence de valeur, cohérence entre les champs (cohérence interne).	Rattachement à Taxref et cohérence de répartition Chorologie par rapport à une base de connaissance sur la répartition ABDSM (Atlas de la Biodiversité Départementales et des Secteurs Marins) Congruence	Variable en fonction des programmes et des inventaires. Interviennent notamment : -Statut biologique -Chorologie -Phénologie -Date	Usage répartition et distribution de référence : Complétude, niveau d'expertise du fournisseur, objectifs de l'acquisition du jeu de données
NBN	Format, type, doublons,	Habitat, aire de	A la charge de	

record Cleaner	références spatiales, etc.	répartition connue, période, listes de référence	l'utilisateur qui sélectionne les règles qu'il souhaite appliquer	
Observatoire de la Flore Sud-Atlantique	Conformité par rapport au FSD. Rattachement au référentiel taxonomique .	Difficulté d'identification du taxon, rareté du taxon, niveau de l'observateur.	Difficulté d'identification du taxon (possible confusions), rareté du taxon, niveau de l'observateur.	
ODIN	Conformité au standard régional.	Comparaison par rapport à des données déjà validées. Usage des référentiels.	Connaissances scientifiques et bibliographie existante. Preuves.	
Pêches électriques (SIE)	Conformité par rapport au format Sandre.	Conformité par rapport aux référentiels du SIE.	Complément de validation par les producteurs lors des valorisations.	
Sigogne	Analyse de la complétude : Présence des informations fondamentales (Où, quand, qui, quoi)	Analyse de la complétude : Taxon : au moins un élément de taxonomie selon un référentiel taxonomique standard Analyse de la cohérence : taxon vraisemblable, phénologie, distribution géographique, écologie	Identification/expertise : photo, ADN, audio, description, expérience de l'observateur	
SILENE FAUNE	Date, lieu, espèce et observateur	Liste d'espèces nécessitant une vérification manuelle (prise en compte de la rareté et de la difficulté de détermination) Liste d'espèces communes par département Liste de programmes d'acquisition pertinents Liste d'observateurs référents par groupe taxonomique Comparaison par rapport à des données déjà validées (distance entre les localités)		
Si Quadrigé ²	Cohérence par rapport aux fiches de terrain ou cahier de paillasse. Champs obligatoires Incohérence inter-champs	Définition de bornes/seuils Variable en fonction de la thématique Profondeur, salinité, pH, concentration en	Séries temporelles Tests statistiques Saisonnalité Analyse des outliers Utilisation de	

	Formats	éléments, etc. Listes de valeurs possibles (référentiels) pour de nombreux champs (niveau d'immersion, engin de prélèvement utilisé, paramètre, support d'analyse, méthode d'analyse, laboratoires impliqués, etc.)	modèles statistiques Variable en fonction de la thématique Profondeur, salinité, pH, concentration en éléments, etc.	
VerNet	Utilisation de l'IPT (ETL développé par le GBIF). Présence de certains champs obligatoire. Contrôle intra-champ (coordonnées, pays, etc.) Cohérence inter-champ (coordonnées/pays).	Confrontation avec des cartes de répartition IUCN		

Bilan :

Les critères de contrôle les plus fréquemment rencontrés sont en tout premier lieu la présence des informations fondamentales (où, quand, qui et quoi). Cette vérification est en général associée à une vérification automatique de format (très fréquent pour le contrôle des dates et des coordonnées) et de contenu (référence à des listes normalisées ou à des référentiels).

Des vérifications de cohérence interchamps peuvent être mises en place de manière automatique (ex : cohérence entre les coordonnées transmises et la commune citée). Certains systèmes font intervenir des procédures de comparaison par rapport à d'autres données déjà validées (INPN, ODIN,...) -on parle de congruence entre les données-, par rapport à des couches de référence de la distribution des taxons ou par rapport à des profils types permettant l'extraction des données aberrantes (outliers).

On constate globalement que toutes les vérifications préliminaires automatiques ont en général pour but de filtrer les données soumises à expertise dans l'objectif d'optimiser les efforts de contrôle qui reposent sur les experts.

L'expertise « manuelle » repose en général sur la présence de preuves (photos, vidéos, enregistrements audio, spécimen, etc.), sur le niveau d'expertise de l'observateur, sur la détectabilité des taxons, sur la chorologie, la phénologie et la biologie du taxon...

Les données validées lors des contrôles servent bien souvent à produire des couches de références ou à alimenter des référentiels.

Les contrôles automatiques se basent, dans la majorité des cas, sur des couches de référence (distribution de référence par exemple) et des référentiels (expertise amont).

On constate donc une rétroaction entre validation et qualification. 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.

Le contrôle par rapport à des référentiels ou couches de référence implique que le contrôle va « rejeter » les nouvelles mentions d'un taxon sur une zone. La validation de ces nouvelles données

implique alors l'intervention d'un expert, qui, en général, demandera des preuves (sous forme de photos, d'enregistrements audio, etc.) pour statuer.

9. Gestion des données inexploitablees ou douteuses pour la diffusion

«?»= *information inconnue*

SYSTEME	
Atlas floristique de la FCBN	Pas de diffusion des données n'ayant pas répondu favorablement à la procédure de contrôle. Les données non conformes au standard, incohérentes d'un point de vue taxonomique ou géographique sont retournées au CBN concerné pour correction.
BDN	Données plausibles diffusées (données pour lesquelles il manque des éléments pour valider mais pour lesquels il y a un intérêt de porter l'information à connaissance pour la gestion et la conservation). Données refusées stockées mais non visibles = non diffusées
BRAHMS	Pas de diffusion des données n'ayant pas répondu favorablement à la procédure de contrôle.
CBN PMP	Exclusion de certaines données par rapport à l'usage prévu : l'évaluation LR (taxons introduits, erratiques, adventices, apomictiques et fonction du rang taxonomique)
CSCF	Seules les données dont le code VALID=1 ou 2 sont diffusés.
Digitale2	Les données erronées ou douteuses sont identifiées en tant que telles, mais elles ne sont pas accessibles par défaut dans Digitale2. C'est l'utilisateur qui doit lui-même modifier les critères de restitution pour les visualiser. Un jeu de données considéré comme ayant beaucoup de données douteuses ne sera pas diffusé sans une validation de ces données (cas rare).
FLORA	Pas de diffusion des données n'ayant pas répondu favorablement à la procédure de contrôle.
GBIF/GBIF France	Pas d'exclusion. Hormis si absence de champs obligatoires pour l'IPT.
INPN	Exclusion pour la diffusion des données non validées pour l'usage prévu
NBN Record Cleaner	?
Observatoire de la Flore Sud-Atlantique	Diffusion uniquement des données « fiables » (notes 3 et 4)
ODIN	Contrôle selon le profil utilisateur (données validées uniquement pour les « non experts »).
Pêches électriques (SIE)	Si les données sont inexploitablees elles ne rentrent pas dans la base de données et les données sont renvoyées au producteur. Les données sont contrôlées par les producteurs et corrigées si besoin.

Sigogne	Mise en quarantaine des données inexploitablees ou douteuses
SILENE	Exclusion pour la diffusion des données non validées. Catalogage et archivage des données douteuses.
SI Quadrige ²	Pas de diffusion des données fausses (ni via Surval – ni grand public, ni pour les évaluations DCE). Données « Douteuses » diffusées avec commentaire expliquant le caractère « Douteux ». Le caractère « Douteux » est pris en compte dans les traitements statistiques des données.
VerNet	Pas d'exclusion (hormis si présence des champs obligatoires pour l'IPT non présents)

Bilan :

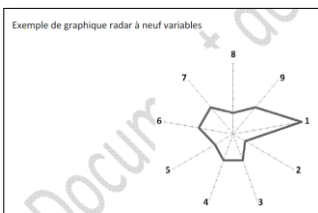
La gestion des données douteuses ou erronées est très variable. Cependant, sans surprise, on constate une exclusion quasi-systématique des données invalidées ou douteuses dès lors qu'il s'agit de répondre à un usage de type couche de référence.

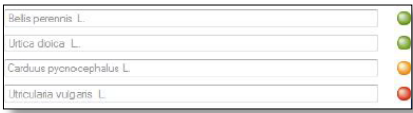
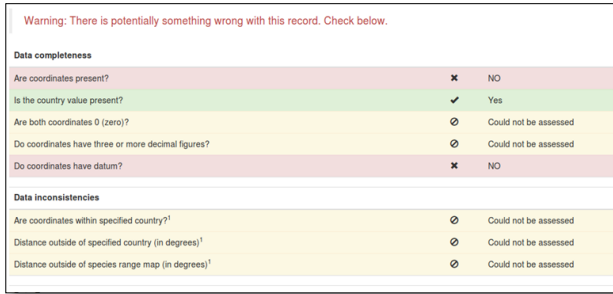
Lorsque toutes les données sont diffusées, sans sélection en fonction du niveau de validation, le niveau de validation est associé à la donnée et donc visible pour l'utilisateur.

10. Retour aux utilisateurs ou aux producteurs du résultat de contrôle

« » = pas de retour

« ? » = information inconnue

SYSTEME	Informations aux producteurs	Information aux utilisateurs
Atlas floristique de la FCBN	Oui : via les indicateurs de qualité. Le contrôle de la conformité au standard fait l'objet de la production d'un rapport d'audit qui est transmis au CBN concerné. Les données expertisées comme douteuses (avis expert FCBN) sont soumises à l'avis du CBN qui a transmis la donnée.	Oui : via les indicateurs de qualité 
BDN	Oui : en cas de doute, contact avec l'auteur.	A priori non, seules les données validées ou plausibles sont diffusées.
BRAHMS	Oui : les producteurs ont accès à la base de données qui portent le résultat des contrôles.	Pas directement. Cela dépend ensuite du site de diffusion (Herbat, GBIF, etc.)
CBN PMP	?	?
CSCF	Oui : toute personne ayant fourni de l'information, qu'elle soit spécialiste ou non, peut être appelée à apporter la preuve tangible de l'identification de l'une ou l'autre des espèces qu'elle a observée. Dans tous les cas, l'observateur à qui une demande de validation a été adressée, est informé de la décision prise.	La documentation concernant la validation est accessible directement sur le site. Seules les données VALID=1 & 2 diffusées. <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 10px auto;">Seules les données dont le code VALID = 1 ou 2 sont diffusées par le CSCF.</div>
Digitale2	Oui par export de leurs données complétés avec les attributs de validation, mais aussi par échange avec le producteur (direct ou par courriel).	Oui, le niveau de fiabilité est visualisable dans Digitale2. L'utilisateur peut même modifier les niveaux de fiabilité et de validité diffusés par défaut.
FLORA	Non systématique : en cas de doute, retour possible vers l'auteur.	Mise en place d'une plate-forme d'échange sur le site flora concernant la validation des données.
GBIF/GBIF France	Oui : sur les problèmes de format et sur la présence des informations obligatoires	Oui : affichage via le site internet du GBIF Ex : http://www.gbif.org/occurrence/910529573
INPN	Oui : retour au producteur via rapport d'audit	Non : seule les données validées sont diffusées
NBN Record Cleaner	Dépend des structures et de si producteur = utilisateur	Oui : sous forme de messages d'erreur détaillés
Observatoire de la Flore Sud-Atlantique	Oui, via l'outil de saisie (<i>pastilles colorées</i>)	Oui : <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 10px auto;">Les données diffusées sont des données validées, c'est-à-dire normalisées et cohérentes, réputées fiables. Si toutefois vous repérez des erreurs, nous vous remercions de bien vouloir nous les signaler en contactant le référent botaniste départemental de l'OFSA ou en vous adressant à : contact[at]ofsa.fr</div>

		
ODIN	Oui : en cas de doute, contact avec l'auteur	A partir d'une donnée lien vers la fiche de métadonnées IDCNP correspondante.
Pêches électriques (SIE)	Les résultats des contrôles d'intégration dans BDMAP ou postérieurs sont retournés aux producteurs ; par ailleurs ces derniers disposent d'outil d'autocontrôle (cohérence) lors des alimentations.	Seuls les champs de validation internes à la donnée indiquent à l'utilisateur l'état de la validation. Aucune autre information n'est actuellement disponible.
Sigogne	?	?
SILENE	Oui : Flore : en cas de doute, contact avec l'auteur Faune : id° + transmission d'un bilan technique par lot de données	Seules les données validées sont diffusées
Si Quadrige ²	Oui : retour format csv aux producteurs	Oui : niveau de qualité et commentaire de qualification diffusés dans les jeux de données (via l'application Q ² , ou via l'outil Internet Surval pour le grand public).
VerNet	Oui : à minima via le portail	Oui : via le portail 

Bilan :

En cas de doute ou d'informations manquantes nécessaires pour la procédure de contrôle, il est très courant que la structure reprenne contact avec le fournisseur ou le producteur des données concernées. Ce contact a généralement lieu pendant la phase d'intervention d'un expert.

Cependant, le contact avec le producteur dans le but d'obtenir une correction ou un complément d'information (photos, description fine, etc.) est conditionné par le délai entre l'acquisition de la donnée et le processus de contrôle. Il semble souvent complexe d'obtenir ces compléments lorsque la donnée est ancienne.

On notera parallèlement qu'il est peu courant que le fournisseur ou le producteur soit recontacté pour l'informer de la bonne validation des données transmises.

Le résultat de la procédure de contrôle n'est pas toujours clairement visualisable pour les utilisateurs. Il peut être implicite (cas où seules les données validées sont diffusées) ou incomplet (seules les données non validées sont flagguées, l'utilisateur déduisant ainsi, en l'absence de flag que la donnée est validée). Le caractère incomplet de l'information fournie aux usagers est en mesure de générer des erreurs d'interprétation. Une donnée non flagguée pourrait par exemple ne pas être validée mais n'avoir tout simplement pas encore subi la procédure complète de validation.

11. Gestion de la traçabilité des corrections

« » = pas de gestion de la traçabilité

« ? » = information inconnue

SYSTEME	De la procédure de contrôle De la mise en œuvre des corrections
Atlas floristique de la FCBN	Données originales remplacées ou corrigées en fonction du problème détecté.
BDN	Scripts de correction archivés avec date de mise à jour mais pas de conservation de la donnée initiale en base.
BRAHMS	Oui tag sur les données.
CBN PMP	?
CSCF	?
Digitale2	Données originales conservées dans des champs spécifiques. Lors de la validation, en plus des attributs de validation, la date, le nom de l'expert et éventuellement un commentaire de validation sont affectés à la donnée.
FLORA	Données originales conservées, champs spécifiques de corrections (date, nom de l'expert si expertise, raison de la correction).
GBIF/GBIF France	?
INPN	Données originales conservées, champs spécifiques de corrections (date, nom de l'expert si expertise, raison de la correction)
NBN Record Cleaner	
Observatoire de la Flore Sud-Atlantique	Oui : données originales conservées, suivi des modifications, validateurs, argumentaire, dates, etc.
ODIN	?
Pêches électriques (SIE)	Non
Sigogne	?
SILENE FAUNE	Traçabilité de la date de validation et de l'expert si expertise. Traçabilité de la date du retour au producteur mais aucune modification de la donnée n'est faite au niveau de SILENE-Faune, cette dernière incombe au producteur.
Si Quadrige ²	Oui : « historique de qualification » des données = enregistrement de tous les statuts successifs des données, avec les dates et la personne responsable des changements. Dans le cas des changements de valeur de résultat, l'ancienne valeur est conservée. Dans le cas des changements de taxon, l'ancien taxon est conservé dans un commentaire.
VerNet	

Bilan :

Tous les systèmes consultés ne bancarisent pas la donnée initiale lorsque celle-ci a subi une correction. Cependant, la traçabilité de la procédure par le renseignement de champs complémentaires dédiés ou « tag » (date de la modification, raison de la correction, nom de l'expert, etc.) est très fréquente. Chez les « centralisateurs de données » (SILENE-FAUNE, INPN, etc.), la

correction doit être réalisée au préalable ou parallèlement sur la base source (ou de gestion), d'où l'importance du contact avec le producteur ou le gestionnaire de cette base.

III. ELEMENTS COMPLEMENTAIRES ISSUS DE LA BIBLIOGRAPHIE

1. Identification des besoins

La détermination des besoins des utilisateurs n'est pas chose facile, et il est aussi difficile d'élaborer des spécifications détaillées que de concevoir ensuite la structure de données adaptée à ces spécifications. Il est en revanche important d'identifier des utilisateurs clés avec qui travailler sur l'analyse des besoins. (1)

Il est nécessaire d'identifier également les besoins des validateurs (de quelles informations ils ont besoin pour réaliser une validation). (18)

Il faut garder à l'esprit que certains usages ne nécessitent pas la même qualité en termes d'identification taxonomique, dès lors par exemple qu'on cherche à dégager de grandes tendances à partir d'un grand jeu de données (5).

Une des attentes exprimées consiste en la nécessité d'apporter une caution scientifique aux données qui sont diffusées. (10)

Les enjeux liés à la diffusion de données normalisées et fiables sont :

- scientifiques (fiabilité des cartes de répartition, des coefficients de rareté, des référentiels d'évaluation de vulnérabilité, des indicateurs d'état et de suivi, etc.),
- juridiques (diffusion des données visant à nourrir les politiques publiques de préservation de la biodiversité),
- de valorisation (disponibilité pérenne des données sur le long terme, valorisation des productions à des fins diverses de connaissance et de préservation de la nature). (16)

Disposer d'une Base de Données fiable est une nécessité et un atout compte tenu des nombreuses possibilités d'utilisation et d'analyse qu'elle permet :

- Réalisation de cartes de distribution réelles voire potentielles
- Analyses de la phénologie ou de la répartition altitudinale des espèces
- Calcul de tendances pour la révision de leur statut liste rouge
- Localisation de sites méritant des mesures de conservation particulières ou abritant des espèces menacées par exemple (19)

La procédure de validation experte des données est facilitée par la possibilité de validation en masse, réalisable par le biais de filtres sur les données. (22)

Une démarche de type forum (participation de l'ensemble d'une communauté pour la validation) permet d'augmenter l'exécution de la procédure et d'aider à la mise en œuvre d'une validation experte. (22)

2. Eléments de terminologie

D'un point de vue technique, la qualification des données est un processus permettant de connaître la qualité des données sans jugement de valeur, de connaître l'état d'une donnée et d'évaluer la qualité de ce qui est produit. (3)

La qualification en droit est une opération consistant à rattacher un acte, un fait, une chose, une situation à une catégorie juridique déjà établie.(3)

La notion d'utilisabilité (norme ISO 19157 et Inspire) est directement liée à une notion de qualité externe : degré d'adéquation à un besoin, une application ou à une exigence particulière. (3)

La vérification est un processus technique de contrôle de conformité qui vise à détecter les données qui ne sont pas conformes à des critères pré-établis. En théorie, cette vérification ne préjuge pas de l'usage, elle se fait par rapport à des critères génériques. (5)

La validation est l'attribution d'un jugement d'expert sur la donnée, prenant en compte les résultats de la vérification, une connaissance approfondie du sujet (critères implicites) et l'usage prévu.

La qualification est un processus de classement des jeux de données en fonction de critères. Ce classement peut être selon l'usage ou la question posée et selon l'existence et l'organisation des vérifications et/ou validations appliqués aux données qui le composent. (5)

3. Traçabilité de la procédure

Il est important de garder la trace des opérations de vérification : quelles données ont été vérifiées et quand. Ceci permet d'éviter les redondances et d'empêcher que des données ne passent à la trappe. Le meilleur moyen de réaliser cela est de tenir un journal des opérations de validation. Il est conseillé d'utiliser des champs additionnels dédiés à la qualité et à la précision (1)

Les traitements doivent préserver l'intégrité des données (3)

4. Délais

Les objectifs de performance sont un bon moyen pour un organisme de maintenir un niveau cohérent de vérification et validation de qualité. Par exemple, que 95 % des enregistrements soient documentés et validés dans les six mois suivant leur réception. (1)

Il faut en particulier être en mesure de valider la donnée avec un délai « raisonnable » qui permette à la personne contributrice de voir sa donnée accepter (ou pas) rapidement. (2)

Il faut mettre en place un système de mesure qui permette de surveiller les évolutions dans le temps. La publication des indicateurs de qualité, leur cible et leur évolution permettent de définir les plans d'actions éventuels à mettre en œuvre pour corriger une situation (15).

5. Types d'erreurs

(5) Partant des définitions de Chapman (2005), erreurs possibles dans la réalisation d'un inventaire:

- les doublons d'attributs:

Il s'agit du cas où deux ou plusieurs codes ou noms représentent la même donnée. Ils sont dus à des évolutions de codification ou à la consolidation de données de source hétérogène utilisant des référentiels différents. Les doublons n'affectent pas nécessairement la qualité de l'information mais ils peuvent être gênants pour certaines valorisations (par exemple ceci peut faire croire à une station importante d'une espèce rare par la présence de plusieurs observations qui n'en sont qu'une en réalité).

- les données incorrectes :

Portant une ou plusieurs informations erronées, les données incorrectes comprennent surtout les erreurs d'identification. Ce sont les plus courantes et elles peuvent avoir plusieurs causes. Les principales sources sont l'erreur humaine de détermination (liée à la compétence/formation/information de l'observateur ou aux conditions d'observation ...), l'erreur de saisie, ou l'erreur de conversion (lors de la consolidation par exemple, quand les référentiels initiaux ne sont pas les mêmes).

- les données incomplètes :

Une ou plusieurs informations de la donnée indispensables à l'identification de l'espèce sont manquantes. Exemple : non-identification de l'observateur.

- les incohérences :
Des informations sont discordantes entre elles.

Selon Robertson et al. (2010) l'erreur la plus grave est la mauvaise identification de l'espèce, qui aurait tendance à être plus importante lors d'un inventaire avec participation des non-scientifiques. De la même façon, Chapman (2005) affirme qu'une détermination fiable ne peut être se faire qu'avec une implication constante des spécialistes puisque l'identification automatisée ne semble pas opérationnelle. (5)

6. Types de contrôles

La consolidation est un processus responsable de mettre dans le même format et la même logique (référentiel taxonomique, géographique...) toutes les données d'un inventaire. Ce point est particulièrement critique quand les sources n'ont pas toutes le même référentiel, voire quand certaines sources n'ont pas réellement de référentiel. (5)

Certains outils permettent de convertir des données textuelles, développe des aides à la réconciliation. Ils peuvent identifier des problèmes d'encodages, des doublons ou une absence d'identifiant (ID unique) (8).

Les experts sont indispensables dans un processus de validation car ils sont la principale garantie de fiabilité des informations. Selon Oliver et al. (2000), le nombre de taxonomistes professionnels décline depuis des années et le nombre de taxonomistes actif pour un groupe d'espèces ne correspond pas à la richesse spécifique du groupe. C'est un point critique pour l'efficacité des méthodologies de validation. (5)

C'est pour cette raison que les méthodes mises en place doivent simplifier et optimiser le travail des spécialistes pour valider les données d'inventaires qui le nécessitent. Il s'agit donc de concentrer le travail des experts sur les activités où leur valeur ajoutée est la plus importante. Une des piste pour cela est de trier les données de façon automatisée, selon des critères définis par les experts, pour ne soumettre à la vérification de l'expert qu'une faible proportion des données, celles délicates qui ne sont ni manifestement à rejeter, ni très vraisemblables. (5)

Editer des règles basiques sur les espèces lorsque cela fait consensus (i.e le statut telle espèce dans telle aire est *valide*) permet de créer à moyen terme des validations automatiques. (2)

Le but de la vérification semi-automatique est ainsi d'alléger l'implication des experts. (5)

7. Critères et paramètres

La variabilité dans la compétence et l'expertise des amateurs pose des questions en ce qui concerne l'exploitation scientifique des données (Levrel *et al.*, 2010, Wiggins *et al.*, 2011). D'autre part, les difficultés d'identification de certains groupes taxonomiques et d'accès de certains milieux (tels que le milieu marin) sont des freins à prendre en compte dans le processus de bancarisation des données (4).

D'autres processus sont utilisés pour valider les données comme l'apport de preuves visuelles ou de demande de fiche de description des espèces, le filtrage des données aberrantes ou inhabituelles, le contrôle par rapport à d'autres sources de données ou encore l'attribution d'un niveau d'expertise à l'observateur. (4)

Peuvent être vérifiés, la cohérence de la distribution, la taxonomie, les dates, les coordonnées, les codes ISO (7) et la présence d'un minimum de métadonnées (8).

Certaines informations peuvent être vérifiées en faisant intervenir la notion de référentiel (checklists) (8).

La preuve photographique est un élément majeur pour la validation experte des données. (22)

La congruence entre les données (présence d'une observation du taxon dans un rayon de 10km autour de la donnée à valider) peut être utilisée de manière automatique pour pré-valider les données. (23)

8. Codages

Lorsqu'un jeu de données arrive aux gestionnaires, il faut classer le potentiel de validité de chaque donnée. Cette classification peut être faite avec l'attribution des codes de qualité. Il est évident que c'est un processus long et laborieux de devoir traiter toutes les données une par une. Pour cette raison l'attribution de codes de qualité est l'aspect majeur abordé lors de l'automatisation de la validation. (5)

9. Feedback

Etablir des canaux efficaces pour le retour d'information de la part des utilisateurs et des fournisseurs est un mécanisme aisé et productif pour améliorer la qualité des données. (1)

L'outil doit faciliter le retour à la communauté et aider les producteurs de données à réaliser les corrections (9).

Chaque critère peut-être mesuré soit de manière subjective en recueillant la perception des producteurs, des fournisseurs ou des utilisateurs, soit de manière objective au travers de suivis automatisés des indicateurs spécifiques (15).

10. Augmentation de la qualité des données

Les principes de la qualité au sens de la norme ISO 9001 sont de limiter les erreurs en amont, les détecter au plus près de la source, dès leurs premières constatations, ce qui évite l'émergence des futures confusions et inexactitudes et diminue l'effort et les coûts de détection et de nettoyage dans les étapes finales de validation (Dalcin, 2004 ; Chapman, 2005). (5)

Des programmes fournissent des guides d'identification des espèces concernées ou des formations afin d'améliorer le niveau de compétence des observateurs (4).

Certains programmes ont décidé d'agir directement au cours de la collecte des données en encadrant scientifiquement les observateurs sur le terrain. (4)

L'amélioration de la qualité des données par leur validation en aval de la collecte de données est le cas le plus fréquemment observé. (4)

Le contrôle de cohérence peut s'appuyer sur un processus informatique. Un outil de saisi en ligne peut permettre de diminuer les erreurs quant aux données recueillies. Cet outil peut porter quatre procédures visant réduire les erreurs:

- Obligation de renseigner certains champs ce qui permet de réduire les données manquantes
- Diminuer les champs libres et favoriser les listes déroulantes, les cases à cocher ou les champs formatés afin d'éviter les fautes de frappe.
- Mise en place un système de géolocalisation, par exemple avec GoogleMap, permettant ainsi au contributeur de localiser facilement ses observations
- Production d'un récapitulatif des informations saisies afin que l'observateur puisse vérifier ses données avant de les soumettre. (4)

Des techniques de détermination semi-automatique ont été testées et se sont avérées efficaces dans certains cas. Motta *et al.* (2001) (5)

L'outil développé pour la saisie de données doit être générique et facilement utilisable par les utilisateurs. (7)

Le principe est de se baser sur des formats et des valeurs standards (liste de choix, etc) qui permettent, dès la saisie, d'éviter les erreurs. (2)

Une approche proactive peut-être mise en place, incluant les bonnes pratiques à appliquer quand de nouvelles données sont intégrées, des projets utilisant les données sont lancés ou des actions de qualité sont effectuées (15).

L'accès, lors de l'acquisition, lors de la saisie tout comme lors de la validation experte à des clefs de détermination permet de diminuer le risque d'erreur de détermination. (22)

11. Mises à jour

Les contrôles effectués en temps réel permettent de prendre en considération les évolutions, notamment concernant les couches de répartition de référence. Cependant, le calcul en temps réel induit des problématiques de vitesse d'exécution et le contrôle est réalisé enregistrement par enregistrement. (17)

IV. BIBLIOGRAPHIE

- (1) Chapman, A. D. (2005). *Les principes de qualité des données, version 1.0*. Trad. Chenin, N. Copenhague : Global Biodiversity Information Facility, 76 p.
Disponible sur http://links.gbif.org/gbif_qualite_donnees_manual_fr_v1.pdf
- (2) ONF (2009). *Validation des données de la Base de Données Naturalistes*. Rapport ONF non publié, 5p.
- (3) Verdier-Maillot, A. & Troispoux, G. (2014). *Les enjeux de la qualité des données géographiques: Eléments juridiques et techniques*. Présentation au Géo-séminaire 2014 – Données géographiques : qualité et conditions d’usage.
- (4) Marchand, I., de Mazières J., Aish A., Poncet L. (2013). *Rapport préparatoire pour l’élaboration d’une note méthodologique : Intégration des données issues des programmes marins de sciences participatives dans l’INPN*. Rapport SPN 2013-17, Service du Patrimoine Naturel, Muséum national d’Histoire naturelle, Paris, 60 p.
Disponible sur <http://spn.mnhn.fr>
- (5) Bogéa Soares, L. V., Touroult, J., Poncet, L. (2013). *Réflexions sur la validation des données naturalistes : le cas des erreurs d’occurrence dans la distribution des espèces*.
Disponible sur <http://spn.mnhn.fr>
- (6) Savouré-Soubelet, A., Haffner, P., Aulagnier, S., Moutou, F. (2013). *Atlas des Mammifères marins de France. Notice méthodologique et organisationnelle*. Rapport SPN-2013-05, 22p. Disponible sur <http://spn.mnhn.fr>
- (7) Güntsch, A. (2013). *The OpenUp Data Quality Toolkit*. Présentation à la Conférence Annuelle TDWG 2013, Italie.
- (8) Ó Tuama, E., Hahn, A., Döring, M. (2013). *Documenting data quality in a global network: the challenge for GBIF*. Présentation à la Conférence Annuelle TDWG 2013, Italie.
- (9) Gendreau, C., Shorthouse, D., Desmet, P. (2013). *Data quality challenges in the Canadensys network of occurrence records: examples, tools, and solutions*. Présentation à la Conférence Annuelle TDWG 2013, Italie.
- (10) Observatoire de la Biodiversité de Haute-Normandie. (2013). *Plate-forme partagée pour la diffusion des données naturalistes de Haute-Normandie*. Cahier des Clauses Techniques Particulière (version de travail), 62 p.
- (11) Chapman, A. D. (2005). *Principes et méthodes de nettoyage de données, version 1.0*. Trad. Roosen, A. Copenhague: Global Biodiversity Information Facility, 90 p.
Disponible sur http://links.gbif.org/gbif_nettoyage_donnees_manual_fr_v1.pdf
- (12) Ball, S. & French, G. (2012). *NBN Record Cleaner user guide. Version 1.0.8.3*. Joint Nature Conservation Committee and National Biodiversity Network, 49 p.
- (13) National Biodiversity Network (2011a). *Sharing Information about Wildlife: Running a Biological Recording Scheme or Survey*. NBN Trust, Nottingham, 32 p.

- (14) National Biodiversity Network (2011b). *Sharing Information about Wildlife: Improving Wildlife Data Quality*. NBN Trust, Nottingham, 16 p.
- (15) Gourvil, J. (2014). *Projet de notice sur la méthodologie de qualification de la complétude et de la cohérence des données de présence. Partie : Métiers*. Atlas floristique de la FCBN. Fédération des Conservatoires botaniques nationaux, Paris, 6 p.
- (16) Caze, G. (2014). *Validation technique et scientifique des observations naturalistes. Cas de l'Observatoire de la Flore Sud-Atlantique (www.ofsa.fr)*. Présentation au Séminaire national des observatoires de biodiversité 2014, Rochefort.
- (17) Otegui, J., Guralnick, R., Wieczorek, J., Bloom, D., Russell, L. (2014). *The Spatio-Taxonomic Data Quality API. Rationale, Development and Examples*. Présentation à la Conférence Annuelle TDWG 2014, Jönköping.
- (18) Gendreau, C., Shorthouse, D., Lecoq, M-E., Roberston, T. (2014). *Darwin Core Archive (DwC-A) validation: A New Collaborative Effort*. Présentation à la Conférence Annuelle TDWG 2014, Jönköping.
- (19) Anonyme. [2014]. *Validation des données*. Centre Suisse de Cartographie de la Faune (CSCF). http://www.cscf.ch/cscf/welcome_to_www_cscf_ch/validation. Consulté le 10/12/2014.
- (20) Gauthier, E., Huguet, A., Belin, C., Claisse, D., Duval M. (2014). *Qualification des données du système d'information Quadrige²*. Rapport interne Ifremer, 12p. http://www.ifremer.fr/quadrige2_support/content/download/79613/1012251/file/SINP_Qualification_PorcessusQ2_140924_V5.pdf
- (21) Delauge, J., Meyer, D., Noble, V., Chondroyannis, P. (2013). Le portail des données naturalistes SILENE en région Provence-Alpes-Côte d'Azur. *Networks and Communication Studies, NETCOM*, 27(1-2): 245-253.
Disponible sur: <http://www.netcom-journal.com/volumes/articlesV271/Netcom245-253.pdf>
- (22) Anonyme. [2014]. *Suivi photographique des insectes pollinisateurs (SPIPOLL)*. <http://www.spipoll.org>. Consulté le 23/01/2014.
- (23) Anonyme. [2014]. *Système d'information régionale sur la faune (SIRF)*. <http://www.sirf.eu/index.php?cont=common&tpl=accueil>. Consulté le 23/01/2014.
- (24) Centre régional de phytosociologie agréé Conservatoire botanique national de Bailleul. [1994-2015]. *Digitale2 (Système d'information sur la flore et la végétation sauvage du Nord-Ouest de la France)*. Bailleul. <http://digitale.cbnbl.org>. Consulté le 25/03/2015
- (25) Desse, A. (2014). *Validation des observations flore et habitats dans Digitale*. Centre régional de phytosociologie agréé Conservatoire botanique national de Bailleul, Bailleul, 44 p.
- (26) Observatoire Batrachologique Herpétologique de Normandie (2014). *Protocole de validation automatique et manuelle des données relatives aux amphibiens et reptiles de Haute-Normandie*. 6p.
- (27) Conservatoire d'Espaces Naturels de Haute-Normandie, Stalleguer, P. (2014). *Protocole de validation automatique et manuelle des données relatives aux invertébrés de Haute-Normandie*. 12p.

(28) Groupe Mammalogique Normand (2014). *Protocole de validation automatique et manuelle des données relatives aux mammifères de Haute-Normandie*. 23p.

(29) Ligue de Protection des oiseaux de Haute-Normandie (2014). *Protocole de validation automatique et manuelle des données relatives aux oiseaux de Haute-Normandie*. 18p.

(30) Fédération de Seine-Maritime pour la Pêche et la Protection du Milieu Aquatique (2014). *Protocole de validation automatique et manuelle des données relatives aux poissons de Haute-Normandie*. 13p.



SINP

Systeme d'Information
sur la Nature et le Paysage



Une multitude de jeux de données sur la nature et les paysages est produite chaque année par de nombreux acteurs sur les territoires français. Ces données sont collectées selon des méthodes et protocoles différents et pour des objectifs hétérogènes ; elles concernent également des périmètres taxonomiques et géographiques divers ; les compétences et moyens des observateurs sont tout aussi variables.

Dans le cadre du SINP, les différents utilisateurs des données doivent être en mesure de connaître le niveau de confiance associé à celles-ci. Cela permet notamment une bonne utilisation des données au regard des usages.

Le SINP doit mettre en place des dispositifs puissants d'échanges, de communication et d'intégration des informations. Cela nécessite notamment le partage de principes scientifiques et règles de contrôle de la qualité et de la validation des données.

Ce document est produit dans le cadre du lancement du Groupe de Travail sur la validation des données d'occurrence du SINP. Le recensement de l'existant est le produit de la phase 1 définie dans le mandat.

Suite à l'étape préalable de recensement non exhaustif des travaux existants en matière de contrôle, de validation ou de qualification de données, au niveau régional, national et international, ce document présente une synthèse des éléments analysés de manière à alimenter les discussions et les décisions du GT Validation.



MUSÉUM
NATIONAL D'HISTOIRE NATURELLE