

RAPPORT

 **Conservatoire
d'espaces naturels
Provence-Alpes-Côte d'Azur**



Protocole de validation des données naturalistes faunistiques dans Silene - 2021

Mars 2021



Rapport

Protocole de validation des données naturalistes faunistiques dans Silene - 2021

Document réalisé par :

Conservatoire d'espaces naturels de Provence-Alpes-Côte d'Azur

Coordination :

Julie DELAUGE – Adjointe à la Direction - Responsable Connaissance et Programmes

Rédaction :

Géraldine KAPFER – Responsable du Pôle Biodiversité Régional

Paul Honoré – Chargé de mission – base de données

Fanny Guillaud – Chargée de mission assistance technique et scientifique

Relecture :

Julie DELAUGE – Adjointe à la Direction - Responsable Connaissance et Programmes

Géraldine KAPFER – Responsable du Pôle Biodiversité Régional

Participation :

Patrick BAYLE

Stéphane BENCE

Gilles CHEYLAN

Paul HONORE

Françoise POITEVIN

Sonia RICHAUD

Cédric ROY

Nicolas VINCENT-MARTIN

Date de réalisation : mars 2021

Citation recommandée :

Delaugé J., Kapfer, G., Honoré, P., Guillaud, F., 2021. Protocole de validation des données naturalistes faunistiques dans Silene. Conservatoire d'espaces naturels de Provence-Alpes-Côte d'Azur – Janvier 2021. Sisteron, 13

Table des matières

Préambule	4
1. Objectifs et principe du protocole de validation scientifique	4
2. Qualification automatique	6
2.1. Processus de Qualification automatique	6
2.2. Détail des contrôles	7
a. Contrôle du programme d'acquisition	7
b. Contrôle des espèces à valider obligatoirement par département	7
c. Contrôle des données au rang de genre	7
d. Contrôle phénologique - étape 1 pour les oiseaux	7
e. Contrôle des espèces communes par département	8
f. Statut biogéographique	8
g. Contrôle orthodromique	8
h. Contrôle phénologique - étape 2 pour les oiseaux	9
i. Contrôle du statut du déterminateur/observateur	9
2.3. Listes de référence	9
3. Validation manuelle	10
4. Mise à jour et validation des données	11
Bibliographie	12

Préambule

Pour la circulation et la mise à disposition de la donnée naturaliste, celle-ci doit être techniquement conforme et surtout scientifiquement pertinente. Pour cela, toute base de données doit faire l'objet d'une vérification constante.

La démarche première a été de responsabiliser les fournisseurs de données vis-à-vis de la qualité de la données fournies en s'engageant à en assurer la validation à leur échelle. Force est de constater que les procédures de qualification sont plus ou moins robustes et que Silene doit pouvoir assurer la validation globale des données fournies ou, du moins, auditer une base de données, dite fiable.

Le CEN PACA a élaboré en 2012 un protocole de validation scientifique pour sa propre base intégrant une qualification automatique des données en complément de l'analyse technique existante. Ce travail a ensuite été mis à disposition de tous pour avis et complément.

La vérification manuelle de plusieurs centaines de milliers de données n'étant pas envisageable, le CEN PACA, administrateur des données faunistiques de SILENE, a la volonté de faciliter la validation des données acquises et centralisées de manière régulière dans Silene. En 2013, après une phase de test de la version beta de ce protocole, ce dernier a été soumis aux référents thématiques Silene afin de le mettre en place sur les données alimentant Silene. Ce protocole est actualisé régulièrement.

En 2017, le protocole de validation a été audité par le national et a été par la suite amélioré à la faveur de la procédure nationale de validation scientifique des données élémentaire d'échange du SINP pour les occurrences de taxons (Robert *et.al*, 2017).

Jusqu'en 2019, la validation par expertise dit validation manuelle n'est encore que partielle en raison de la prise en charge financière ponctuelle de cette mission. Elle était donc réalisée à la faveur de programmes d'amélioration de la connaissance tels que les Inventaires Régionaux, l'animation de PNA ou l'évaluation des Listes Rouges Régionales. Depuis 2019, des moyens financiers plus importants permettent la prise en charge d'une plus grande partie de la validation manuelle.

Certains groupes taxonomiques (poissons, certaines familles d'arthropodes par exemple) ne bénéficient pour le moment pas de validation manuelle satisfaisante du fait de l'absence de spécialistes en région ou de leur disponibilité.

1. Objectifs et principe du protocole de validation scientifique

La validation scientifique ne se fait que sur de la donnée techniquement conforme, c'est-à-dire une donnée présentant a minima les informations suivantes reconnues et standardisée (Delauge J., Kapfer, G. Honoré P., Guillaud, F., 2021. Organisation de l'intégration des données dans Silene Faune. Conservatoire d'espaces naturels de Provence-Alpes-Côte d'Azur – Mars 2021. Sisteron, 26 p) : un taxon, un observateur (ou observateur non mentionné suivi du nom du fournisseur de données le cas échéant), une localisation et une date.

L'objectif du protocole de la validation scientifique des données de Silene est de qualifier chaque donnée par rapport à sa fiabilité, c'est-à-dire au degré de confiance que l'on peut lui accorder.

La problématique posée est celle-ci : telle espèce observée, dans tel lieu, à telle date, à telle altitude, par tel observateur nécessite-t-elle une validation manuelle ?

La validation ne porte donc pas sur l'ensemble des éléments de l'observation mais uniquement sur les champs contenant les valeurs : taxons, lieu, date.

La validation scientifique se déroule en deux temps :

- une qualification automatique : série de contrôles par rapport à des référentiels d'expertises.
- une validation scientifique manuelle : vérification de la donnée par un expert du groupe concerné.

La qualification automatique consiste en une série de contrôles, interrogeant successivement la donnée par rapport à des référentiels d'expertises. Ces référentiels ont été élaborés puis validés par les référents SILENE. Ils ne sont pas figés et peuvent être améliorés et complétés continuellement à la faveur des remarques, suggestion des experts et du jeu de données de référence (données validées) disponible. Aucune donnée n'est validée de façon automatique.

A l'issue de la qualification automatique, une donnée peut être qualifiée :

Probable : retenue informatiquement

A confirmer : à valider manuellement

Dans Silene, la validation scientifique manuelle consiste à expertiser les données manuellement pour leur affilier un des statuts suivant :

validée

retenue

à confirmer (donnée plausible mais retour d'information en attente)

douteuse (donnée aberrante).

La correspondance entre les attributs du niveau de validité avec le protocole national de validation des données est proposée ci-après :

PROTOCOLE NATIONAL	PROTOCOLE REGIONAL
CERTAIN / TRES PROBABLE	Validée
PROBABLE	Retenue
DOUTEUX	A confirmer/Douteuse
INVALIDE	Douteuse
NON REALISABLE	Douteuse ou A valider

Par ailleurs, les données « non évaluées » sont qualifiées « à valider ».

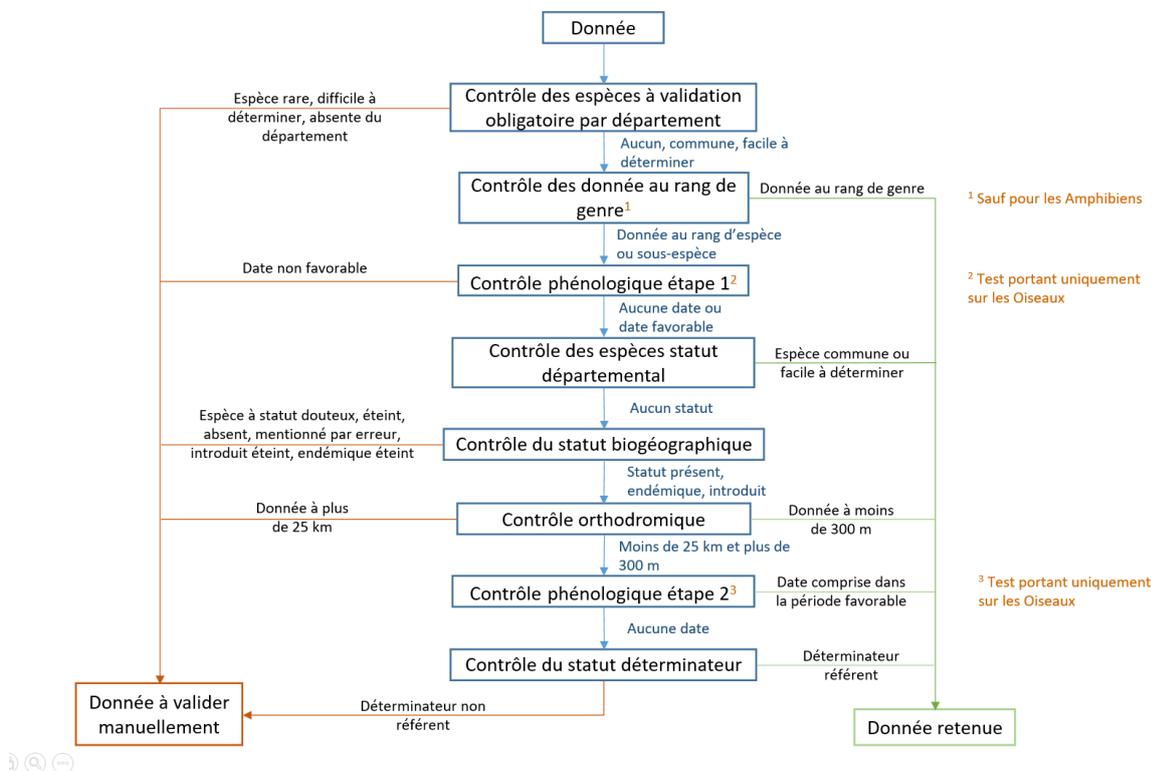
Dans Silene, seules les données qualifiées comme « retenues » ou « validées » seront diffusées. Lorsque les données sont qualifiées de « douteuses » ou « à confirmer », elles nécessitent un complément d'information et ne sont pas diffusées dans SILENE en ligne, elles restent disponibles sur demande.

NIVEAU DE VALIDITE	DIFFUSION SILENE
VALIDE	OUI
RETENUE	OUI
A CONFIRMER	NON
DOUTEUSE	NON

2. Qualification automatique

2.1. Processus de Qualification automatique

Le schéma ci-dessous représente le parcours de chaque donnée conforme dans le processus automatique de qualification des données. Il indique pour chaque contrôle les alternatives existantes permettant de qualifier la donnée comme « retenue » (flèche verte à droite) ou comme « à valider manuellement » (flèche rouge à gauche). Lorsque le contrôle ne peut qualifier la donnée dans l'une de ses deux alternatives, la donnée poursuit le processus de qualification jusqu'au contrôle suivant. Par principe, si la donnée ne peut être retenue par l'un des contrôles, elle est qualifiée « à valider manuellement ».



Le processus est différent selon le groupe ou la précision de localisation de la donnée.

2.2. Détail des contrôles

a. Contrôle du programme d'acquisition

Les statuts de validation fournis par les fournisseurs ne sont jamais pris en compte lors du processus de validation à l'exception de rares fichiers très spécifiques (aujourd'hui un unique fournisseur de données spécialiste du groupe taxonomique concerné dont le protocole de validation est connu). Les données sont donc validées par le SINP régional.

b. Contrôle des espèces à valider obligatoirement par département

Portée : le contrôle porte sur toutes les données d'occurrences mentionnant une espèce dans un département de PACA.

Description : le contrôle vérifie l'existence d'un statut départemental pour chaque espèce. Si l'espèce (ou la sous-espèce) est identifiée comme étant rare, absente ou difficile à déterminer, la donnée est à valider manuellement.

Règles :

Statut départemental	Résultat
Rare	Donnée à valider manuellement
Absente	
Difficile à déterminer	
Aucun	Poursuite du processus

c. Contrôle des données au rang de genre

Portée : la contrôle concerne toutes les données de tous groupes renseignés au rang de genre.

Description : il consiste à vérifier les données au rang de genre.

Règle : si la donnée est au rang de genre, elle est retenue sauf exception pour les genres Rana ou Pelophylax qui poursuit le processus comme toutes les autres données au rang d'espèce ou de sous-espèce.

Statut rang de genre	Résultat
Donnée au rang de genre (hors Rana ou Pelophylax)	Donnée retenue
Autre	Poursuite du processus

d. Contrôle phénologique - étape 1 pour les oiseaux

Portée : le contrôle concerne toutes les données d'occurrence d'oiseaux renseignées.

Description : il consiste à vérifier la cohérence entre la date d'observation de l'espèce et la période phénologique favorable à l'observation.

Règle : si la date n'est pas comprise dans la période phénologique favorable à l'espèce, la donnée est à valider manuellement sinon elle poursuit le processus de validation.

Phénologie	Résultat
Période non favorable	Donnée à valider manuellement

Autre période	Poursuite du processus
---------------	------------------------

Actuellement, la précision de date n'est pas prise en compte (le contrôle peut donc être amélioré) mais cela a peu de conséquences sur les oiseaux.

Remarque : si ce contrôle est mis en place pour d'autres groupes (entomologie), le stade devra être pris en compte.

e. Contrôle des espèces communes par département

Portée : le contrôle porte sur toutes les données d'occurrences mentionnant une espèce dans un département de PACA.

Description : le contrôle vérifie l'existence d'un statut départemental pour chaque espèce. Si l'espèce (ou la sous-espèce) est identifiée comme commune et simple d'identification la donnée est retenue.

Règles :

Statut départemental	Résultat
commune	Donnée retenue
Aucun	Poursuite du processus

f. Statut biogéographique

Portée : le contrôle porte sur toutes les données d'occurrence

Description : il consiste à vérifier le statut biogéographique MNHN du taxon dans le référentiel.

Règle : si le statut biogéographique du taxon est douteux, introduit éteint, absent, mentionné par erreur, éteint, endémique éteint, la donnée est à valider manuellement sinon la validation se poursuit.

Statut biogéographique	Résultat
Vide, non renseigné, douteux, introduit éteint, absent, mentionné par erreur, éteint, endémique éteint	Donnée à valider manuellement
Autre	Poursuite du processus

g. Contrôle orthodromique

Portée : le contrôle concerne toutes les données d'occurrence.

Description : le contrôle consiste à permettre la qualification automatique selon sa distance orthodromique (à vol d'oiseau) avec une donnée de la même espèce. Le contrôle amène à trois alternatives possibles : soit la donnée est taguée à valider manuellement (si la donnée la plus proche est à plus de 25 km) et quitte le processus de qualification automatique, soit la donnée est retenue (si la donnée est située à moins de 300 m), soit la donnée poursuit le processus de validation.

Règle :

Distance orthodromique	Résultat
------------------------	----------

Donnée à plus de 25 km (toutes précisions)	Donnée à valider manuellement
Donnée à moins de 300 m (données précises)	Donnée retenue
Autres	Poursuite du processus

h. Contrôle phénologique - étape 2 pour les oiseaux

Portée : le contrôle concerne toutes les données d'occurrence d'oiseaux renseignées.

Description : il consiste tout comme pour l'étape 1 à vérifier la cohérence entre la date d'observation de l'espèce et la période phénologique favorable à l'observation de celle-ci.

Règle : si cette espèce dispose d'une date de période phénologique favorable et que la date d'observation est comprise dans cette dernière : la donnée est retenue. Si l'espèce ne dispose pas de période favorable : la donnée est retenue

Phénologie	Résultat
Donnée d'une espèce disposant d'une période favorable observée dans cette période	Donnée retenue
Donnée d'une espèce qui ne dispose pas de période favorable.	Donnée retenue
Autre cas	Poursuite du processus

i. Contrôle du statut du déterminateur/observateur

Portée : le contrôle porte sur toutes les données d'occurrence (hors oiseaux).

Description : il consiste à vérifier si le déterminateur ou un des observateurs (en l'absence de déterminateur renseigné) fait partie d'une liste de référence pour le groupe concerné.

Règle :

Déterminateur/observateur référent	Résultat
OUI	Donnée retenue
NON	Poursuite du processus

2.3. Listes de référence

Ces contrôles automatiques reposent sur des listes spécifiques qui ont été créées :

- La liste du statut des espèces par département

Elle statue pour toutes les espèces si elles sont :

- rares ou absentes par département (ex : un Apollon dans les Bouches-du-Rhône nécessitera une validation manuelle car il n'est pas présent dans ce département, alors qu'un Apollon dans les Hautes-Alpes ne nécessitera pas de validation manuelle à ce stade car il n'est pas rare dans ce département) → validation manuelle obligatoire
- à détermination difficile (ex : certains micromammifères) → validation manuelle obligatoire

- commune et facile de détermination → retenue
- **La liste des déterminateurs ou des observateurs référents**
Elle contient les observateurs expérimentés pour chaque groupe faunistique. Cette liste est actualisée en continu.
- **La liste des périodes favorables d'observations**
Elle contient la liste des espèces d'oiseaux migrateurs avec leur période(s) favorable(s) d'observation.
- **Le statut biogéographique**
Le statut associé actuellement aux taxons est hérité du référentiel TAXREFv15.

3. Validation manuelle

La qualification automatique n'est que la première étape de validation qui vise à faciliter la phase nécessaire de validation scientifique manuelle par un expert pour juger de la fiabilité de l'observation.

Cette expertise passe par l'utilisation d'une plateforme dédiée aux validateurs sur l'outil « SILENE Admin », une interface de Silene permettant de voir toutes les données d'occurrences à la suite de la qualification scientifique automatique.

Un réseau de validateurs, volontaires et compétents sur un ou plusieurs groupes concernés, peuvent, grâce à cet outil de validation, affecter un statut aux données : retenue, validée, à confirmer ou douteuse.

Ces experts peuvent évaluer l'ensemble des données, quel que soit le niveau de qualification attribué automatiquement à la donnée. Des modifications peuvent avoir lieu après diffusion des données lors des utilisations et des questionnements ou autres retours d'utilisateurs. La validation manuelle fait foi par rapport à la qualification automatique.

Les validateurs ont accès à l'ensemble des informations portées au niveau des données nécessaires à assurer leur expertise. En aucun cas, le validateur ne peut corriger ou modifier les données (sauf en cas de synonymie) sans en avoir eu l'autorisation de l'observateur. Ils ont la possibilité de recontacter l'observateur, le déterminateur ou l'organisme producteur de manière à demander des éléments complémentaires à leur analyse. En cas de problème identifié, le validateur invalidera la donnée. Celle-ci sera soit retournée vers le fournisseur qui effectuera les modifications avec l'observateur, soit modifiée par le validateur après obtention d'une autorisation.

Comme pour le processus de qualification automatique de la donnée, l'expertise porte sur les valeurs suivantes de la donnée d'occurrence : taxon /date /lieu. Autrement dit, le validateur valide la probabilité de la présence d'un taxon en un lieu ou un secteur (en fonction de la précision de localisation), à une date ou sur une période donnée (en fonction de la précision de la date).

Les autres informations constituant la données ne sont pas validées (statut biologique, altitude,...) mais elles participent au processus de qualification.

Il est donc demandé au validateur de regarder la probabilité de présence du taxon au lieu et à la date mentionnés en fonction de :

- Sa propre connaissance du taxon, de sa répartition et de sa biologie
- Une redétermination est possible sur photos si présente ou demandée

- De la difficulté de détermination ou du risque de confusion du taxon en ce lieu à cette date
- De la provenance des données : contributeurs considérés comme fiables pour le taxon en question.

Sur cette base le validateur peut :

- Valider la donnée, qualifiée « Validée », s'il considère cette présence CERTAINE ou TRES PROBABLE
- Retenir la donnée, qualifiée « Retenue », s'il considère cette présence PROBABLE mais que son expertise n'est pas plus ni moins fiable qu'une qualification automatique se basant sur les mêmes processus.
- Invalider la données temporairement, qualifiée « A confirmer », s'il ne dispose pas des informations nécessaires à cette validation, et a besoin de contacter l'observateur ou soumettre cette donnée à un autre expert.
- Invalider la donnée jusqu'à correction ou suppression, qualifiée « Douteuse », s'il considère cette occurrence comme INVALIDE ou TRES PEU PROBABLE

4. Mise à jour et validation des données

La validation des données s'effectue une fois par semaine, à partir du vendredi soir et se termine le samedi en début d'après-midi. Un journal « log » permet de savoir à quel moment la validation s'est achevée.

Elle concerne toutes les données qui ont un statut « A valider ». C'est le cas des données nouvellement intégrées.

Bibliographie

Robert S., Dupont P., de Mazières J., Poncet L., Touroult J., 2017. *Procédure nationale de validation scientifique des données élémentaires d'échanges du SINP pour les occurrences de taxons. Version 1*. Service du patrimoine naturel, Muséum national d'histoire naturelle, Paris. Rapport SPN 2017 – 2. 16 p.



**Conservatoire
d'espaces naturels
Provence-Alpes-Côte d'Azur**

Siège :

4, avenue Marcel Pagnol

Immeuble Atrium Bât B.

13 100 Aix-en-Provence

Tél : 04 42 20 03 83

Fax : 04 42 20 05 98

Email : contact@cen-paca.org

www.cen-paca.org

Pôle Biodiversité Régionale

18 avenue du Gand

04200 SISTERON

Tél : 04 92 34 40 10

Le Conservatoire d'espaces naturels
de Provence-Alpes-Côte d'Azur
est membre de la Fédération
des Conservatoires d'espaces naturels



Ce travail a été réalisé grâce au soutien financier des partenaires suivants :

