

Observatoire National des Mammifères

Base de données d'occurrence



Guide méthodologique pour la validation scientifique des données d'occurrence de Mammifères

Version finale 1

Septembre 2022

SFEPM
19 allée René Ménard
18000 Bourges
contact@sfepm.org
02.48.70.40.03



Versions de travail non diffusées	Auteurs	Relu par	Date	Commentaires
1.0	CP Arthur	F Paperin, L Girard	16/04/2018	Version initiale
1.1	F Paperin	CP Arthur	30/04/2018	Poursuite structuration
1.1	CP Arthur	F Paperin, L Girard	12/05/2018	Poursuite structuration
1.2	F Paperin	CP Arthur, L Girard	29/05/2018	Poursuite structuration, réorganisation du document
		Julie Marmet, JF Julien	19/02/2019	Relecture du document et propositions d'améliorations
2.0	F Paperin	R. Baghi	15/08/2019	Refonte du document
		A. Savouré-Soubelet	20/02/2020	Relecture du document et propositions d'améliorations
3 & 4.0	A. Savouré-Soubelet	Christian Arthur, Pascal Dupont, Patrick Haffner, Julie Marmet Solène Robert, Thomas Ruys, Eva Rodinson,	05/07/2022	Refonte du document. Puis réunion le 2 septembre pour apporter des précisions par rapport au lien avec le SINP
5.0	A. Savouré-Soubelet	Christian Arthur, Romain Baghi Fabrice Darinot, Patrick Haffner, Julie Marmet, Thomas Ruys, Nathalie de Lacoste	08/09/2022	Réunion pour présenter la méthode. Quelques corrections apportées à la suite de cette réunion

Coordinatrices du document

Fanny PAPERIN-CARILLO, Chargée de mission « *Chiroptères* » et Coordinatrice pour l'Observatoire National des Mammifères à la SFPEM.

Puis Audrey SAVOURÉ-SOUBELET, Chargée de mission « *Atlas des Mammifères de France* ».

Citation recommandée

Paperin-Carillo F. & Savouré-Soubelet A. (coords.) 2022. – *Guide méthodologique pour la validation scientifique des données d'occurrence de Mammifères. Version finale 1*. Observatoire National des Mammifères, base de données d'occurrence. Société Française pour l'Etude et la Protection des Mammifères (SFPEM). Bourges, 16 p.

Relecteurs du document

Christian Arthur (SFPEM)	Et
Hélène Chauvin (SFPEM / CCN)	
Romain Baghi (SFPEM)	Patrick Haffner (PatriNat)
Fabrice Darinot (SFPEM / Groupe Petits et moyens Mammifères)	Eva Rodinson (PatriNat)
Nathalie De Lacoste (SFPEM / Groupe Petits et moyens Carnivores)	Julie Marmet (MNHN)
Thomas Ruys (SFPEM)	

Document connexe

Robert S., Barneix M., Body G., Castanet J., Caze G., Cellier P., Desse A., de Mazières J., Fromage P., Gourvil J., Jomier R., Juste A., Landry P., Lebeau Y., Lecoq M.E., Lescure J., Marage D., Meyer D., Pamerlon S., Papacotsia A., Poncet L., Quintenne G., Saltré A. & Touroult J. 2016. *Guide méthodologique pour la conformité, la cohérence et la validation scientifique des données et des métadonnées du SINP – Volet 1 : occurrences de taxons, Version 1*. Rapport pour le SINP, rapport MNHN-SPN 2016-77, 63 p.

SFPEM - SOCIETE FRANCAISE POUR L'ETUDE ET LA PROTECTION DES MAMMIFERES

Association loi de 1901, agréée au titre de la protection de l'environnement -article L.252.1 du code rural- Siret : 348 398 371 000 32
C/o Muséum d'Histoire Naturelle, 19 allée René MENARD,
18000 BOURGES – Tél : 02 48 70 40 03 - E-mail : contact@sfepm.org

TABLE DES MATIERES

Table des matières	1
I. Préambule	2
II. Protocole de validation scientifique des données de Mammifères.....	7
1. Validation automatique.....	7
a. Vérification relative à l'identification du taxon.....	8
b. Vérification relative à l'aire de présence connue.....	8
c. Vérification relative à la période d'observation	10
2. Validation manuelle	10
III. Attribution des niveaux de validité	11
1. A partir de la validation automatique	11
2. A partir de la validation manuelle	11
Bibliographie.....	14
Annexe 1 : Article L411-1 A - Code de l'environnement.	15
Annexe 2 : version de travail du référentiel de détermination.....	16

I. PREAMBULE

L'Observatoire National des Mammifères (ONM) est un projet fédérateur porté par la SFEPM dont l'objectif prioritaire est l'agrégation de données de Mammifères sauvages issues des différentes bases de données régionales ou locales. Il peut s'agir de données d'occurrence (présence ou absence vérifiée) ou de données de synthèse (par exemple données de comptage en gîte pour les Chiroptères) produites par tous acteurs collectant des données sur les Mammifères sauvages, institutionnels, privés ou autres acceptant de mettre leurs données en commun et à la disposition de tous.

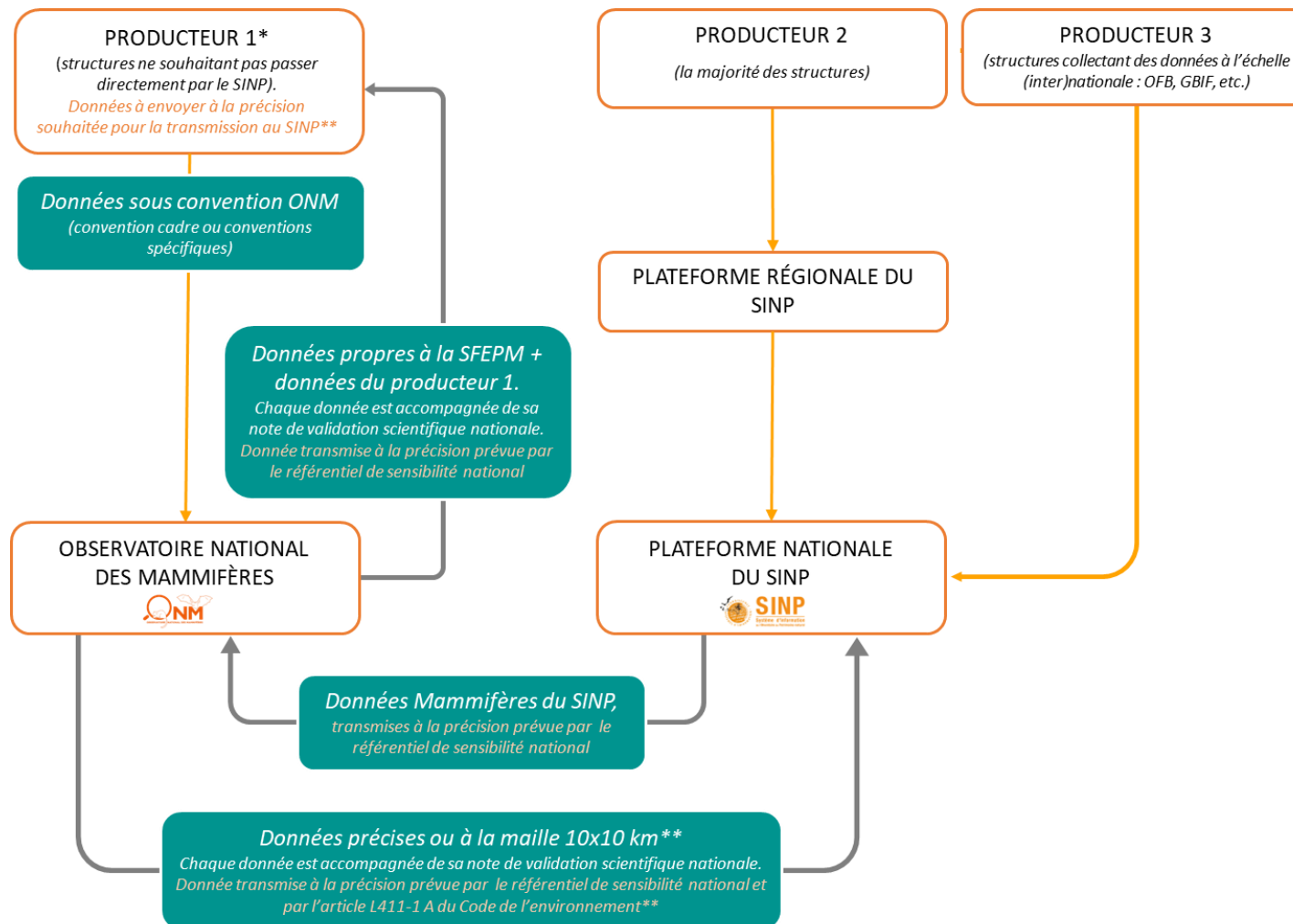
Les données arrivant à l'ONM ont toutes pour vocation à être transmises au SINP (figure 1). Seules les données collectées sur fonds privés **ET** transmises dans le cadre d'une mission ne relevant pas d'un partenariat avec le Service public pourront être envoyées floutées à la maille 10 x 10 km au SINP. Pour toutes les autres données, si le producteur ne souhaite pas que la localisation précise soit transmise au SINP, le floutage des données à la maille 10 x 10 km sera alors à sa charge. Dans ce cas, l'ONM pourra apporter si besoin un appui technique aux différents producteurs pour leur faciliter ce travail de floutage des données.

De fait, et comme le montre l'exemple ci-après, dans le cadre de l'atlas des Mammifères sauvages de France bénéficiant d'un partenariat avec le Service public (MNHN), la SFEPM a l'obligation (Article L411-1 A - Code de l'environnement, disponible en annexe de ce document) d'envoyer toutes les données transmises pour ce projet à la précision à laquelle elle les a reçues. Pour ce projet, l'ONM n'a pas le droit de les recevoir à la localisation précise et de les envoyer floutées au SINP.

Toutes les données transmises au SINP suivent alors les règles de diffusion et de mise à disposition adoptées par le SINP, c'est-à-dire qu'elles deviennent *de facto* des données publiques pouvant être diffusées suivant le niveau de floutage défini dans les listes de sensibilité régionales et/ou nationale.

La donnée élémentaire d'échange (ou DEE selon le code SINP) agrégée dans cet Observatoire est une donnée de présence d'une espèce à une date donnée et localisée le plus précisément possible (localisation précise XY ou *a minima* la maille 10 x 10 km voire par défaut à la commune). Toutes les données sont également assorties d'indications sur leur provenance (organisme ayant transmis l'observation, observateur ayant procédé à l'observation, identification du jeu de données d'où provient la donnée, etc.) et d'un fichier de métadonnées, l'ensemble permettant la traçabilité de chacune des données de l'ONM. Cette traçabilité permet notamment d'identifier les doublons et de retourner à la source en cas de question. En outre, des informations diverses peuvent également être renseignées afin de permettre des analyses ou vérifications plus poussées (dénombrement, méthode d'observation, programme dans lequel la donnée est récoltée, etc.). L'ensemble de ces informations est présenté dans le format standard de l'ONM¹ (format pensé au plus près de celui préconisé dans le standard d'occurrence du SINP).

¹ SFEPM (Paperin F., coord.) (2019) – *Format standard des données d'occurrence de Mammifères. Version 2.* Observatoire National des Mammifères, base de données d'occurrence. SFEPM, Bourges, 25p. Disponible sur : http://www.observatoire-mammiferes.fr/atlas/static/docs/ONM_Format_standard_2019.pdf



* En fonction des jeux de données et cadres d'acquisition associés, certaines structures peuvent envoyer une partie de leurs données directement à l'ONM (données collectées sur fonds privés) et une autre partie directement au SINP (données collectées sur fonds publics). Il faudra faire preuve de vigilance sur ce point afin qu'un même jeu de données ne soit pas envoyé aux deux (ONM & SINP). Il est indispensable de s'assurer que le producteur a bien fourni un identifiant unique par donnée.

△ Seules les données collectées sur fonds privés **ET transmises dans le cadre d'une mission ne relevant pas d'un partenariat avec le Service public pourront être envoyées au SINP floutées à la maille 10 x 10 km. Pour toutes les autres données, si le producteur ne souhaite pas les transmettre à la localisation précise au SINP, il sera de son devoir de les envoyer floutées à l'ONM. △

Figure 1. Schéma de synthèse du flux de données entre les différents producteurs, l'ONM et les différentes plateformes du SINP.

Comme toute base de données, celle de l'ONM doit se doter d'un processus d'expertise permettant d'évaluer la qualité et fiabilité des données intégrées. Ce processus se découpe en plusieurs phases (figure 2) :

- **le contrôle de conformité**, c'est-à-dire le respect des règles définies dans le format standard (renseignement des champs obligatoires, format), basées tant sur la conformité numérique de la saisie (numérique, alphanumérique, date... selon les modalités définies : nombre de caractères, typographie des caractères, ...) que sur l'utilisation d'un nombre restreint, et défini au départ, de possibilités de termes descriptifs (nomenclatures : TAXREF, code mailles, etc.) ;
- **le contrôle de cohérence**, c'est-à-dire le respect de la logique combinatoire des informations transmises dans la donnée. Cette étape permet de vérifier la cohérence de chaque champ obligatoire (par exemple, la date d'observation est antérieure à la date du jour, c'est-à-dire qu'il n'est pas possible en 2022 d'avoir une donnée datée de 2023), la cohérence inter-champs (par exemple, si la localisation a été transmise au point avec notification dans un autre champ de la commune, il faut vérifier que les coordonnées géographiques sont bien comprises dans la commune citée) et la cohérence avec les métadonnées (par exemple les données concernent bien le groupe taxonomique et le(s) territoire(s) mentionnés dans les métadonnées).
- **la validation scientifique** (automatique et manuelle) qui consiste en des processus d'expertises visant à renseigner sur le degré de confiance que l'on peut accorder à la donnée, en fonction notamment des modalités (méthode, programme, etc.) ayant présidé à sa récolte. Cette validation permet l'attribution d'un niveau de validité selon le type de validation réalisée (automatique ou manuelle ou combinatoire, cf. tableau 1) et de définir les possibilités de diffusion et utilisation ultérieures de la donnée.

Seule l'absence de conformité ou de cohérence d'une donnée peut empêcher l'entrée (la saisie et son acceptation) de cette donnée dans la base de données de l'ONM. La validation scientifique n'induit quant à elle pas d'obstruction à la circulation d'une donnée une fois celle-ci saisie ; elle peut cependant en limiter l'utilisation.

Nota important : la donnée intégrée à l'ONM peut déjà avoir fait l'objet d'une validation (locale, régionale, institutionnelle, etc.) par la personne ou l'organisme ayant fourni l'observation. Cette information, quand elle existe, est conservée au sein de la donnée mais la mise en œuvre de la validation nationale est réalisée de manière indépendante du niveau de validation locale.

Le présent document a pour objectif de définir les modalités de cette validation scientifique au sein de la base de données de l'ONM, sur les données de présence uniquement (la validation des données d'absence étant un processus totalement différent). Le protocole national de validation scientifique présenté ci-après a vocation à être applicable à l'ensemble des Mammifères sauvages et à l'ensemble des données de Mammifères du SINP, tous producteurs confondus. En effet, s'inscrivant dans le cadre de la Convention de partenariat entre le MNHN et la SFEPM, la SFEPM a été mandatée pour prendre en charge la validation scientifique nationale des données de Mammifères du SINP de tous les producteurs. Des référentiels de validation par groupe taxonomique seront ensuite rédigés en concertation avec les experts afin de pouvoir mettre en œuvre ce protocole.

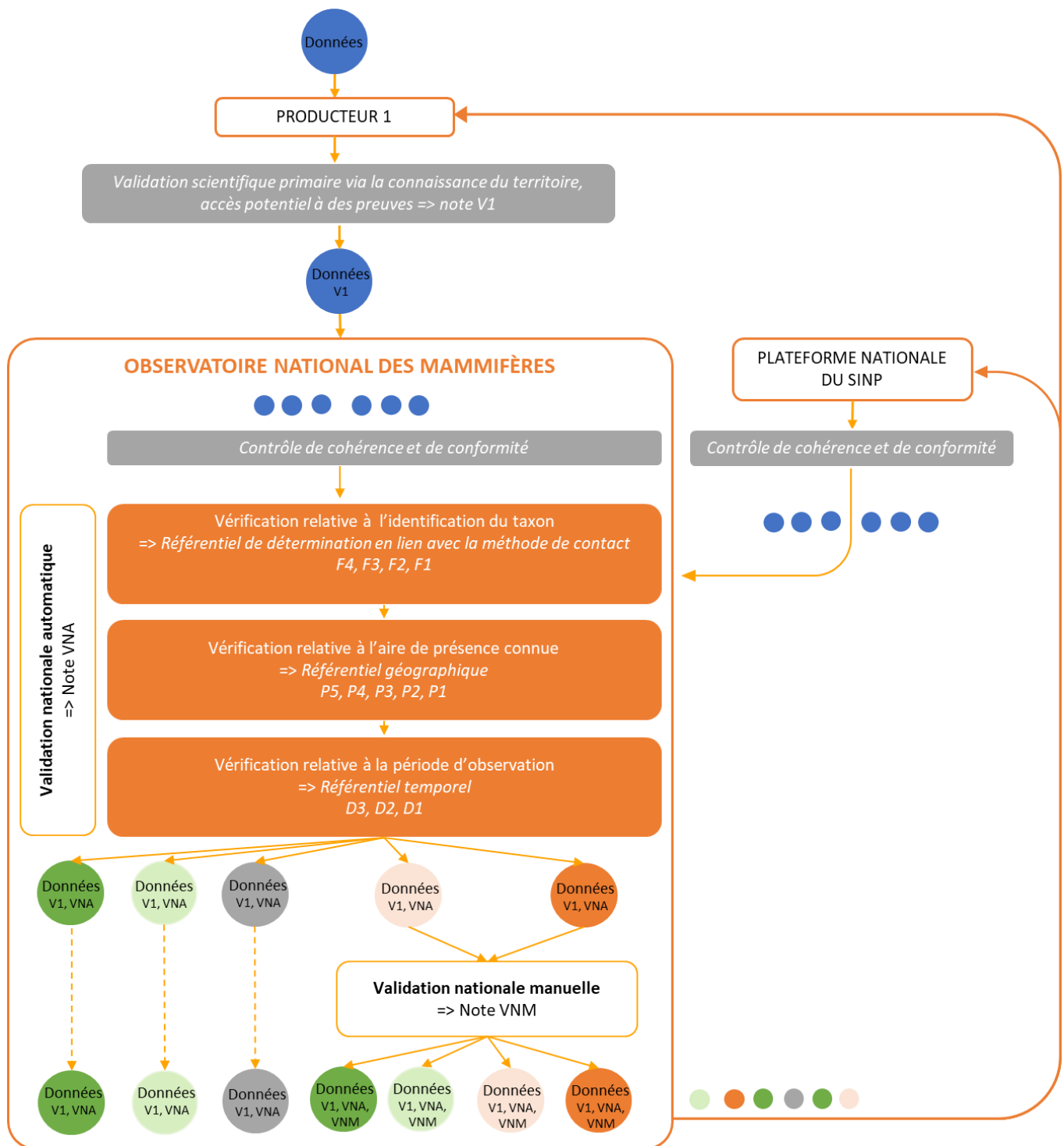


Figure 2. Schéma global des différentes phases de contrôle et de validation des données d'occurrence. VNA = validation nationale automatique, VNM = validation nationale manuelle. Cf. le tableau 1 pour la définition des niveaux de validité.

Rond bleu = données n'ayant pas été soumises à la validation scientifique nationale,
 Rond vert foncé = données dont la note de validation est « certain, très probable »,
 Rond vert clair = données dont la note de validation est « probable »,
 Rond rose clair = données dont la note de validation est « douteux »,
 Rond orange foncé = données dont la note de validation est « invalide »,
 Rond gris = données dont la validation n'est pas réalisable à l'instant t.

Tableau 1 : Niveaux de validité pouvant être attribués à une donnée. Source : Robert *et al.* 2016

Libellés	Définitions	
	Processus automatique	Processus manuel ou combinatoire
Certain - très probable	La donnée présente un haut niveau de vraisemblance (très majoritairement cohérente) selon le protocole automatique appliqué. Le résultat de la procédure correspond à la définition optimale de satisfaction de l'ensemble des critères du protocole automatique, par exemple, lorsque la localité correspond à la distribution déjà connue et que les autres paramètres écologiques (date de visibilité, altitude, etc.) sont dans la gamme habituelle de valeurs et que le paramètre méthodologique est cohérent.	La donnée est exacte. Il n'y a pas de doute notable et significatif quant à l'exactitude de l'observation ou de la détermination du taxon. La validation a été réalisée <u>notamment</u> à partir d'une preuve de l'observation qui confirme la détermination du producteur ou après vérification auprès de l'observateur et/ou du déterminateur.
Probable	La donnée est cohérente et plausible selon la procédure automatique appliquée mais ne satisfait pas complètement (intégralement) l'ensemble des critères automatiques appliqués. La donnée présente une forte probabilité d'être juste. Elle ne présente aucune discordance majeure sur les critères jugés les plus importants mais elle satisfait seulement à un niveau intermédiaire, ou à un ou plusieurs des critères automatiques appliqués, soit sur des considérations biologiques ou éthologiques, voire spatiales, mais aussi parfois méthodologiques.	La donnée présente un bon niveau de fiabilité. Elle est vraisemblable et crédible. Il n'y a, a priori, aucune raison de douter de l'exactitude de la donnée mais il n'y a pas d'éléments complémentaires suffisants disponibles ou évalués (notamment la présence d'une preuve ou la possibilité de revenir à la donnée source) permettant d'attribuer un plus haut niveau de certitude.
Douteux	La donnée concorde peu selon la procédure automatique appliquée. La donnée est peu cohérente ou incongrue. Elle ne satisfait pas ou peu un ou plusieurs des critères automatiques appliqués, biologiques et/ou éthologiques et/ou spatiaux, voire méthodologiques. Elle ne présente cependant pas de discordance majeure sur les critères jugés les plus importants qui permettraient d'attribuer le plus faible niveau de validité (invalide).	La donnée est peu vraisemblable ou surprenante mais on ne dispose pas d'éléments suffisants pour attester d'une erreur manifeste. La donnée est considérée comme douteuse.
Invalide	La donnée ne concorde pas selon le protocole automatique appliqué. Elle présente au moins une discordance majeure sur un des critères, jugés les plus importants ou la majorité des critères déterminants sont discordants. Elle est considérée comme trop improbable (aberrante notamment au regard de l'aire de répartition connue, des paramètres biotiques et abiotiques de la niche écologique du taxon, ou de la méthode utilisée). Elle est considérée comme invalide.	La donnée a été infirmée (erreur manifeste/avérée) ou présente un trop bas niveau de fiabilité. Elle est considérée comme trop improbable (aberrante notamment au regard de l'aire de répartition connue, des paramètres biotiques et abiotiques de la niche écologique du taxon, la preuve révèle une erreur de détermination). Elle est considérée comme invalide.
Non réalisable	La donnée a été soumise à l'ensemble du processus de validation mais l'opérateur (humain ou machine) n'a pas pu statuer sur le niveau de fiabilité, notamment du fait : -d'un état des connaissances du taxon insuffisant ; -d'informations insuffisantes sur l'observation (champs non renseignés ou de façon trop imprécise).	
Non évalué	Niveau initial ou temporaire. La donnée n'a pas été soumise à l'opération de validation ou l'opération n'est pas encore terminée (validation en cours). Elle n'est donc pas évaluée à un temps précis défini par la date d'exécution du processus de validation.	

II. PROCOLE DE VALIDATION SCIENTIFIQUE DES DONNEES DE MAMMIFERES

Une fois les contrôles de conformité et de cohérence effectués, la validation scientifique des données au sein de la base de données de l'ONM se déroule en deux phases. Dans un premier temps, la donnée fait l'objet d'un examen dans le cadre d'un processus informatisé de validation automatique qui consiste à évaluer la congruence des informations de la donnée par rapport à des référentiels et bases de connaissances préétablis par un réseau d'experts.

Cette automatisation de la validation permet de traiter un nombre important de données sans les examiner une par une à l'aide d'un réseau d'experts, et de ne s'intéresser dans un premier temps qu'aux données « posant problème » qui seront alors ensuite validées manuellement par des experts.

L'expertise manuelle peut en outre être réalisée quel que soit le statut de validation et également à tout moment, y compris après la diffusion de la donnée.

1. Validation automatique

Une donnée d'occurrence est récoltée par une **personne** ayant observé un **taxon** (espèce, groupe d'espèces...) pouvant présenter un certain **comportement** (en vol, en chasse...), selon une **méthode** particulière (opportuniste, capture, enregistrement audio...), en un **lieu** et un **habitat** donné et à un **moment** précis. La validation de la donnée doit donc s'appuyer sur les informations associées à cette donnée (méthode d'observation, maille, commune, etc.) et sur les métadonnées associées au jeu de données (protocole, programme, etc.). Dans l'absolu, ces informations doivent permettre de répondre aux questions suivantes :

- **Le lieu géographique d'observation de ce taxon est-il cohérent avec ce que l'on connaît de la répartition de ce taxon ?**
- **Le taxon pouvait-il être observé, et surtout déterminé, à l'aide de la méthode indiquée ?**
- **La date d'observation de ce taxon est-elle cohérente avec la phénologie de ce taxon (hibernation, migration...)?**
- L'habitat naturel dans lequel ce taxon a été observé est-il cohérent avec ce que l'on connaît de l'écologie de ce taxon et de son utilisation de l'espace (en fonction du jour, des saisons...)?
- Le comportement démontré par ce taxon en ce lieu et à cet instant est-il cohérent avec ce que l'on connaît de l'éthologie de ce taxon (en fonction du jour, de la saison, du lieu...)?

La précision de la validation automatique d'une donnée va donc dépendre de la quantité et de la qualité des informations accompagnant cette donnée et des métadonnées associées.

La validation au sein de la base de données de l'ONM concerne les trois premiers points listés ci-dessus, à savoir le lieu, la validation du taxon (en lien avec la méthode d'observation utilisée) et la période d'observation. Par la suite, si le niveau de précision des données intégrées le permet, une validation secondaire plus élargie pourra être envisagée, en prenant en compte des champs connexes associés à la donnée (ex. habitat, comportement) permettant des analyses plus poussées. Pour vérifier ces trois points constituant cette validation primaire, des référentiels doivent être rédigés par groupe taxonomique : référentiel géographique, référentiel de détermination, référentiel temporel. Ces référentiels représentent donc des filtres par lesquelles les données vont circuler afin de retenir les données qui nécessiteront ensuite une validation manuelle.

Par principe, tout observateur faisant remonter une donnée à l'ONM, via les structures locales, sera considéré comme compétent, la compétence d'un observateur devant être évaluée au plus près du terrain (et donc par les structures locales qui auront en charge cette responsabilité). La crédibilité de l'observateur, fonction de la compétence qui lui est attribuée (cette compétence peut avoir été

certifiée, par exemple dans le cadre d'une validation à la capture), n'est donc pas un critère intervenant dans le processus de validation au niveau de l'ONM.

a. Vérification relative à l'identification du taxon.

Afin de transcrire la difficulté de détermination d'une espèce et ainsi identifier les données nécessitant un degré d'expertise plus élevé, un degré de fiabilité de la donnée sera attribué en lien avec la méthode utilisée pour répondre à la question « *quel est le degré de fiabilité de détermination de ce taxon par cette méthode ?* ». L'information concernant la méthode utilisée n'est malheureusement pas systématiquement renseignée. Il est donc nécessaire d'avoir un échelon permettant de poursuivre la validation même si cette information n'est pas disponible (= « F0 »). Par la suite, afin d'avoir des bases de données les plus fiables possible, il est important que les observateurs pensent à renseigner ce champs (ObsMeth dans la V3 du protocole du SINP) et que les producteurs acceptent de partager cette information.

Cette vérification s'appuie sur un référentiel (référentiel de détermination) construit sur la base de connaissances d'experts. Ce référentiel reprend l'ensemble des méthodes listées dans la nomenclature « *Méthodes d'observation* »² et attribue à chaque combinaison espèce/méthode un degré de fiabilité. Afin de préciser cette combinaison, les informations sur le comportement disponibles dans le champ « *OccurrenceComportementValue* » (V3 du protocole du SINP) ou dans les métadonnées peuvent être utilisées (cf. annexe 2). Les niveaux de fiabilités retenus sont :

- **F0. « Méthode non renseignée ».** Dans la base de données reçue, nous ne disposons pas d'informations sur la méthode utilisée.
- **F1. « Fiabilité de la méthode non connue ».** Dans l'état actuel des connaissances, nous ne pouvons pas nous prononcer sur le niveau de fiabilité de la méthode de détermination. *Exemple : acoustique pour les petits Mammifères.*
- **F2. « Fiabilité basse ».** Ce statut est attribué lorsque la détermination de cette espèce est considérée comme difficile et source d'erreur, et lorsque la méthode employée pour déterminer le taxon n'apparaît pas comme suffisamment discriminante et robuste au vu de la difficulté de détermination liée à ce taxon : « cette méthode ne permet en aucune façon d'être formel sur la détermination de ce taxon ». *Exemple : identification du Murin cryptique à partir de l'acoustique.*
- **F3. « Fiabilité moyenne ».** La probabilité de confusion avec un autre taxon n'est pas nulle, et la méthode de contact utilisée permet de déterminer le taxon mais avec un risque d'erreur réel : « cette méthode ne permet pas la discrimination fiable et certaine de ce taxon avec un autre taxon proche ». *Exemple : identification d'une Martre à partir de fèces.*
- **F4. « Fiabilité élevée ».** L'espèce est facilement identifiable, et la méthode de contact utilisée ne permet aucune confusion quant au taxon déterminé. *Exemple : observation directe d'un Écureuil roux.*

b. Vérification relative à l'aire de présence connue

La vérification de la localisation du taxon s'appuie sur les informations de référence concernant l'aire de présence connue pour cette espèce. Le référentiel utilisé ici est constitué par l'enveloppe englobant l'ensemble des mailles géographiques au sein desquelles la présence (ou l'absence) de l'espèce a été confirmée « récemment ». C'est donc un référentiel « glissant », c'est-à-dire qu'il considère une période en remontant dans le temps à partir de l'année considérée. Pour la carte de distribution initiale, il est préconisé d'utiliser les distributions de référence réalisées et validées dans

² <http://standards-sinp.mnhn.fr/nomenclature/14-methodes-dobservation-observationmethode-2018-05-14/>

le cadre de l'atlas des Mammifères sauvages de France lorsqu'elles sont disponibles (actuellement : Mammifères marins, ongulés, Lagomorphes, Carnivores et Primates) ou en cas d'absence les enquêtes spécifiques réalisées au niveau national ou les cartes réalisées et publiées dans le cadre d'atlas régionaux récents.

Pour les statuts géographiques, nous nous inspirons de la méthodologie nationale du MNHN pour l'élaboration de l'ABDSM³. Les statuts renseignent sur la présence ou l'absence de l'espèce dans la maille considérée et indiquent également le niveau de certitude de cette information lorsque l'espèce est considérée comme présente. La distribution initiale réalisée à partir de ces statuts a un objectif de validation des données. Cette carte peut donc être différente de la carte affichant les données d'occurrence. Par exemple, pour le Loup gris, une occurrence peut être validée en Ile-de-France mais dans le cadre du protocole de la validation le statut géographique reste P2 « Absence probable ou certaine ou données occasionnelle », la SFEPM souhaitant valider manuellement les données occasionnelles.

Les différents statuts géographiques sont :

- **P1. « Pas d'information ».** Dans l'état actuel de la connaissance, on ne peut pas se prononcer sur la présence ou l'absence du taxon dans cette maille. *Exemple de la Taupe d'Europe dans certaines mailles du sud-ouest à la suite du splittage de l'espèce.*
- **P2. « Absence probable ou certaine ou présence occasionnelle ».** Ce point recouvre l'absence (probable ou certaine), par nature plus difficile à démontrer que la présence. Il se base sur l'absence de données concernant ce taxon dans cette maille depuis toujours (jamais d'observation) et/ou avec un ou plusieurs des éléments suivants :
 - recherche ciblée de l'espèce exhaustive mais infructueuse ;
 - absence de milieux adéquats ;
 - espèce non observée alors que sa présence est facile à détecter ;
 - présence peu vraisemblable pour des raisons historiques ou biogéographiques.Ce statut doit également être attribué à une maille dans laquelle la présence de l'espèce est occasionnelle, car une observation réalisée en dehors de la répartition « classique et actuelle » d'une espèce doit être validée manuellement et non automatiquement. *Exemple de l'observation d'un Loup en Ile-de-France.*
- **P3. « Disparition avérée ».** Cas particulier : absence liée à une disparition avérée (on connaît la dernière date d'observation du taxon) depuis moins d'un demi-siècle (si disparition ancienne, utiliser alors le code P2). *Exemple de la Loutre d'Europe en Alsace.*
- **P4. « Présence probable ».** Ce statut se base sur un ou plusieurs des éléments suivants :
 - la dernière observation validée a plus de 12 ans par rapport à la date de référence (période correspondant à deux « rapportages Natura 2000 »), aucune recherche spécifique menée et aucune preuve ou présomption de disparition depuis cette date ;
 - la présence certaine de l'espèce dans des mailles contiguës s'il existe dans la maille considérée des milieux favorables.
- **P5. « Présence certaine ».** Correspond à un signalement sur la base d'au moins une observation avérée dans une période de 12 ans précédant l'année de référence, et aucune preuve ou présomption de disparition depuis l'obtention de la dernière donnée.

³ Haffner P., Poncet L., Da Costa H. & Touroult J. 2012. *Atlas de la biodiversité départementale et des secteurs marins. Rapport méthodologique – version 1.* Rapport SPN 2012-36, 9 p.

c. Vérification relative à la période d'observation

Cette vérification consiste à contrôler la cohérence entre la période favorable à l'observation du taxon et la date de l'observation renseignée dans la donnée. Elle s'appuie sur un référentiel (référentiel temporel) construit sur la base de connaissances d'experts et suivant les niveaux ci-dessous :

- **D1. Date d'observation trop imprécise.** La date renseignée dans la donnée ne permet pas de savoir si le taxon a été observé dans une période favorable à son observation. *Exemple : donnée d'une espèce hibernante pour laquelle la date est renseignée sur une fourchette annuelle (du 1 janvier au 31 décembre).* Cela n'invalide pas la donnée mais peut restreindre son utilisation
- **D2. Date d'observation hors de la période favorable d'observation.** *Exemple : observation en hiver d'une Marmotte.*
- **D3. Date d'observation en période d'observation favorable.** Cet échelon s'applique aussi aux espèces pour lesquelles il n'y a pas de période d'observation préférentielle, c'est-à-dire qui peuvent être observées toute l'année. *Exemple : observation d'un Cerf élaphe en octobre, observation d'un Chevreuil à n'importe quelle période de l'année.*

A l'issue de ces vérifications, un niveau de validité est attribué à chaque donnée (cf. §III). Celles ayant obtenu un statut « douteux » ou « invalide » sont soumises à une validation manuelle. Les autres peuvent d'ores et déjà être utilisées en fonction des besoins.

2. Validation manuelle

L'expertise manuelle concerne prioritairement les données ayant obtenu un statut « douteux » ou « invalide » via la validation automatique mais elle peut aussi être réalisée quel que soit le statut de validation (« probable » notamment) et également à tout moment, y compris après la diffusion de la donnée.

Les experts s'appuient sur l'ensemble des informations liées à la donnée et peuvent également solliciter l'organisme ayant transmis cette dernière afin d'obtenir des informations supplémentaires, notamment auprès du découvreur.

Ils statuent alors sur le niveau de validation de la donnée, qui prévaut par rapport au niveau obtenu à la suite de la validation automatique.

L'identité du ou des experts, ainsi que la date et le processus appliqué sont alors renseignés au sein même de la donnée. De manière à éviter les sollicitations multiples, s'il y a lieu, le contact avec le producteur (nom, prénom, organisme et date de contact) devra également être retranscrit au sein des données. Dans le cadre du règlement général de protection des données (RGPD), soit l'observateur a d'ores et déjà donné son accord au niveau local pour que ses informations personnelles soient retranscrites dans les bases de données, et cela vaudra aussi pour accord au niveau national, soit ces informations manquent et il faudra alors les récolter avec l'accord de stockage de la part de l'observateur.

III. ATTRIBUTION DES NIVEAUX DE VALIDITE

Les niveaux de validité retenus dans la base de données de l'ONM pour le résultat du processus de validation scientifique sont ceux préconisés par le SINP⁴ (Tableau 1).

A noter que le niveau de validité correspondant à un état des connaissances à un moment donné, il est important de renseigner la date à laquelle la validation a eu lieu.

1. A partir de la validation automatique

La donnée intégrée à la base de données de l'ONM est par défaut considérée à son entrée comme « *Non évaluée* » au niveau national. Elle est ensuite décrite par les résultats des trois vérifications automatiques présentées précédemment (vérification relative à l'aire de présence connue de l'espèce, vérification relative à l'identification du taxon en lien avec la méthode d'observation utilisée et vérification relative à la période d'observation). C'est la combinaison de ces trois résultats qui permet d'aboutir au niveau de validité déterminant le degré de confiance attribué à cette donnée. Il n'y a pas de hiérarchie entre le référentiel géographique et le référentiel de détermination. En revanche, le référentiel relatif à la période d'observation permet uniquement de diminuer le niveau de confiance attribué à une donnée lorsque le taxon a été observé hors de la période d'observation favorable.

La logique d'attribution des niveaux de validité est présentée dans le schéma ci-dessous (figure 3) et les résultats par niveau de validité sont synthétisés dans le tableau qui suit (tableau 2).

2. A partir de la validation manuelle

Les niveaux de validité issus de la validation manuelle suivent les définitions présentées dans le tableau 1. Comme précisé précédemment, le niveau obtenu à la suite d'une validation manuelle prévaut par rapport au niveau obtenu à la suite d'une validation automatique.

Dans tous les cas, le ou les niveaux de validité (automatique et/ou manuelle) sont renseignés pour chaque donnée et diffusés aux producteurs des données et au SINP. Un historique des différents statuts de validation est conservé pour chaque donnée. Cet historique contient le niveau de validité, la date de la validation, le type ou nom du validateur et le commentaire associé à la décision.

⁴ Robert S., Barneix M., Body G., Castanet J., Caze G., Cellier P., Desse A., de Mazières J., Fromage P., Gourvil J., Jomier R., Juste A., Landry P., Lebeau Y., Lecoq M.E., Lescure J., Marage D., Meyer D., Pamerlon S., Papacotsia A., Poncet L., Quaintenne G., Saltré A. & Touroult J. 2016. *Guide méthodologique pour la conformité, la cohérence et la validation scientifique des données et des métadonnées du SINP –Volet 1 : occurrences de taxons, Version 1*. Rapport pour le SINP, rapport MNHN-SPN 2016-77, 63 p.

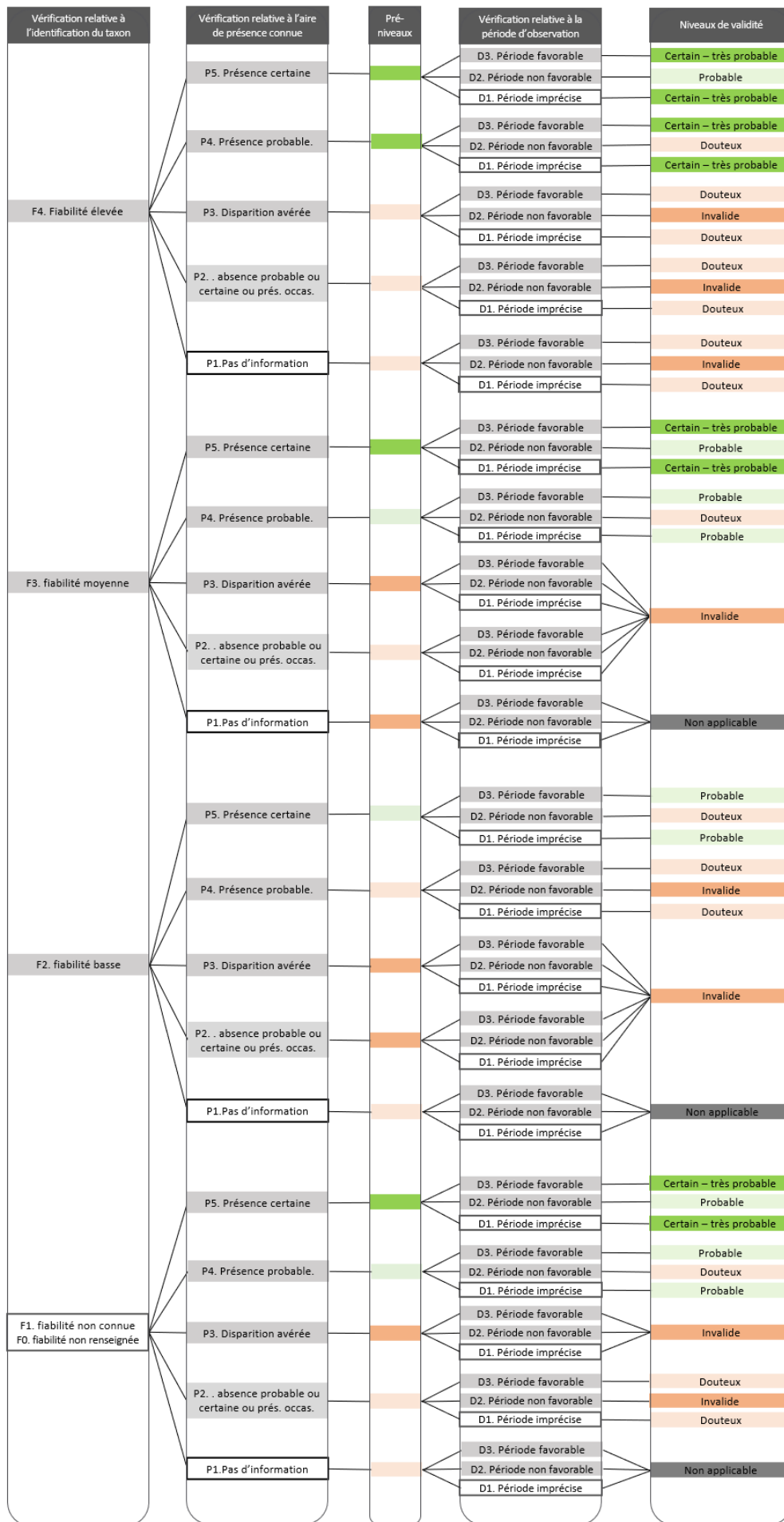


Figure 3 : Schéma d'attribution des niveaux de validité à l'issue de la validation automatique

Tableau 3. Synthèses des résultats obtenus à la suite de la validation automatique.

Fiabilité	Présence	Période
Certain - très probable		
Elevée	Certaine	Favorable Imprécise
	Probable	Favorable Imprécise
Moyenne	Certaine	Favorable Imprécise
Non connue / Non renseignée	Certaine	Favorable Imprécise
Probable		
Elevée	Certaine	Non favorable
Moyenne	Certaine	Non favorable
	Probable	Favorable Imprécise
Basse	Certaine	Favorable Imprécise
Non connue / Non renseignée	Certaine	Non favorable
	Probable	Favorable Imprécise
Douteux		
Elevée	Probable	Non favorable
	Disparition avérée	Favorable Imprécise
	Absence certaine / probable / Présence occasionnelle	Favorable Imprécise
	Pas d'information	Favorable Imprécise
Moyenne	Probable	Non favorable
Basse	Certaine	Non favorable
	Probable	Favorable Imprécise
Non connue / Non renseignée	Probable	Non favorable
	Absence certaine / probable / Présence occasionnelle	Favorable Imprécise
Invalide		
Elevée	Disparition avérée	Non favorable
	Absence certaine / probable / Présence occasionnelle	Non favorable
	Pas d'information	Non favorable
Moyenne	Disparition avérée	Favorable Non favorable Imprécise
	Absence certaine / probable / Présence occasionnelle	Favorable Non favorable Imprécise
Basse	Probable	Non favorable
	Disparition avérée	Favorable Non favorable Imprécise
	Absence certaine / probable / Présence occasionnelle	Favorable Non favorable Imprécise
Non connue / Non renseignée	Disparition avérée	Favorable Non favorable Imprécise
	Absence certaine / probable / Présence occasionnelle	Non favorable
Non réalisable		
Moyenne	Pas d'information	Favorable
Basse		Non favorable
Non connue / Non renseignée		Imprécise

BIBLIOGRAPHIE

Haffner P., Poncet L., Da Costa H. & Touroult J. 2012. *Atlas de la biodiversité départementale et des secteurs marins. Rapport méthodologique – version 1*. Rapport SPN 2012-36, 9 p.

Robert S., Barneix M., Body G., Castanet J., Caze G., Cellier P., Desse A., de Mazières J., Fromage P., Gourvil J., Jomier R., Juste A., Landry P., Lebeau Y., Lecoq M.E., Lescure J., Marage D., Meyer D., Pamerlon S., Papacotsia A., Poncet L., Quaintenne G., Saltré A. & Touroult J. 2016. *Guide méthodologique pour la conformité, la cohérence et la validation scientifique des données et des métadonnées du SINP –Volet 1 : occurrences de taxons, Version 1*. Rapport pour le SINP, rapport MNHN-SPN 2016-77, 63 p.

SFEPM (Paperin F., coord.) (2019) – *Format standard des données d’occurrence de Mammifères. Version 2*. Observatoire National des Mammifères, base de données d’occurrence. SFEPM, Bourges, 25p.

ANNEXE 1 : ARTICLE L411-1 A - CODE DE L'ENVIRONNEMENT.

Version en vigueur depuis le 25 février 2022. Modifié par LOI n°2021-1104 du 22 août 2021 - art. 228 (V).

I. – (...) Les maîtres d'ouvrage, publics ou privés, des projets, plans et programmes mentionnés aux articles L. 122-1 et L. 122-4 ou bénéficiant d'une dérogation prévue à l'article L. 411-2 contribuent à [l'inventaire du patrimoine naturel] par la saisie ou, à défaut, par le versement des données brutes de biodiversité acquises à l'occasion des études d'évaluation réalisées préalablement à la décision d'autorisation, d'approbation ou de dérogation appliquée à leur projet, plan ou programme et à l'occasion des mesures de suivi des impacts environnementaux, notamment celles relevant des mesures d'évitement, de réduction ou de compensation définies en application de l'article L. 110-1, réalisées après cette même décision.

On entend par données brutes de biodiversité les données d'observation de taxons, d'habitats d'espèces ou d'habitats naturels, recueillies par observation directe, par bibliographie ou par acquisition de données auprès d'organismes détenant des données existantes.

Les modalités de saisie ou de versement des données sont fixées par décret, pris après concertation avec les organisations représentatives des maîtres d'ouvrage, des bureaux d'études concernés et des associations contribuant ou susceptibles de contribuer à l'inventaire du patrimoine naturel. La saisie ou le versement de données s'effectue au moyen d'une application informatique mise gratuitement à la disposition des maîtres d'ouvrage par l'Etat.

II. – En complément de l'inventaire du patrimoine naturel, les collectivités territoriales, les associations ayant pour objet l'étude ou la protection de la nature et leurs fédérations, les associations naturalistes et les fédérations de chasseurs et de pêcheurs peuvent contribuer à la connaissance du patrimoine naturel par la réalisation d'inventaires locaux ou territoriaux ou d'atlas de la biodiversité, ayant notamment pour objet de réunir les connaissances nécessaires à l'élaboration du schéma régional de cohérence écologique mentionné à l'article L. 371-3 ou à la mise en œuvre des articles L. 412-7 à L. 412-9 lorsque l'assemblée délibérante concernée a adopté la délibération prévue à l'article L. 412-15.

Le représentant de l'Etat dans la région ou le département et les autres collectivités territoriales concernées sont informés de ces réalisations. (...).

IV. – (...). Les données brutes contenues dans les inventaires mentionnés au présent article sont diffusées comme des données publiques, gratuites et librement réutilisables, sauf si leur diffusion porte atteinte aux intérêts mentionnés aux 1° à 4° du I de l'article L. 124-4. Les conditions dans lesquelles la diffusion des données prévue au présent alinéa peut être restreinte pour des motifs de protection de l'environnement sont précisées par décret. (...).

ANNEXE 2 : VERSION DE TRAVAIL DU REFERENTIEL DE DETERMINATION

ESPECE	CD_REF	0.Vu	0. Vu (ni capture ni en vol)	0. Vu (en vol)	0. Vu (et manipulé)	1. Entendu	3. Ultrasons	4. Empreintes	(...)
Espèce 1	XX1	F4. Élevé	F4. Élevé	F3. Moyenne	F4. Élevé	F2. Basse	F3. Moyenne	F2. Basse	
Espèce 2	XX2	F4. Élevé	F4. Élevé	F4. Élevé	F4. Élevé	F2. Basse	F2. Basse	F4. Élevé	
Espèce 3	XX3	F3. Moyenne	F3. Moyenne	F3. Moyenne	F4. Élevé	F2. Basse	F2. Basse	F3. Moyenne	

Libellé « <i>ObsMeth</i> » + (précision)	Définition
0. Vu	Observation directe d'un individu vivant.
0. Vu (ni capture ni en vol)	Observation directe d'un individu vivant sans qu'il ait été capturé ou vu en vol. Cette information peut être disponible dans des champs du JDD [OccurrenceComportementValue = 15 ou dans commentaire], ou dans les métadonnées du JDD.
0. Vu (en vol)	Observation directe d'un individu vivant en vol. Cette information peut être disponible dans d'autres champs du JDD comme [OccurrenceComportementValue = 10 ou 16] ou dans les métadonnées du JDD
0. Vu (et manipulé)	Observation directe d'un individu vivant capturé. Prise de mesure et observation minutieuse. Cette information peut être disponible dans les métadonnées du JDD ou dans d'autres champs
1. Entendu	Observation acoustique d'un individu vivant.
3. Ultrasons	Observation acoustique indirecte d'un individu vivant avec matériel spécifique permettant de transduire des ultrasons en sons perceptibles par un humain.
4. Empreintes	Observation indirecte via empreintes.
6. Fèces/Guano/Epreintes	Observation indirecte par les excréments
7. Mues	Observation indirecte par des plumes, poils, phanères, peau, bois... issus d'une mue.
8. Nid/Gîte	Observation indirecte par présence d'un nid ou d'un gîte non occupé au moment de l'observation.
10. Restes dans pelote de réjection	Identifie l'espèce à laquelle appartiennent les restes retrouvés dans la pelote de réjection (os ou exosquelettes, par exemple).
11. Poils/phanères	Observation indirecte de l'espèce par ses poils, plumes ou phanères, non nécessairement issus d'une mue.
12. Restes de repas	Observation indirecte par le biais de restes de l'alimentation de l'individu.
19. ADN environnemental	Séquence ADN trouvée dans un prélèvement environnemental (eau ou sol).
20. Autre	Pour tout cas qui ne rentrerait pas dans la présente liste. Le nombre d'apparitions permettra de faire évoluer la nomenclature.
21. Inconnu	Inconnu : La méthode n'est pas mentionnée dans les documents de l'observateur (bibliographie par exemple).
23. Galerie/terrier	Observation indirecte : Galerie forée dans le bois, les racines ou les tiges, par des larves (Lépidoptères, Coléoptères, Diptères) ou creusée dans la terre (micromammifères, mammifères...).
25. Vu et entendu	Vu et entendu : l'occurrence a à la fois été vue et entendue.
26. Contact olfactif	Contact olfactif : l'occurrence a été sentie sur le lieu d'observation
27. Empreintes et fèces	Empreintes et fèces