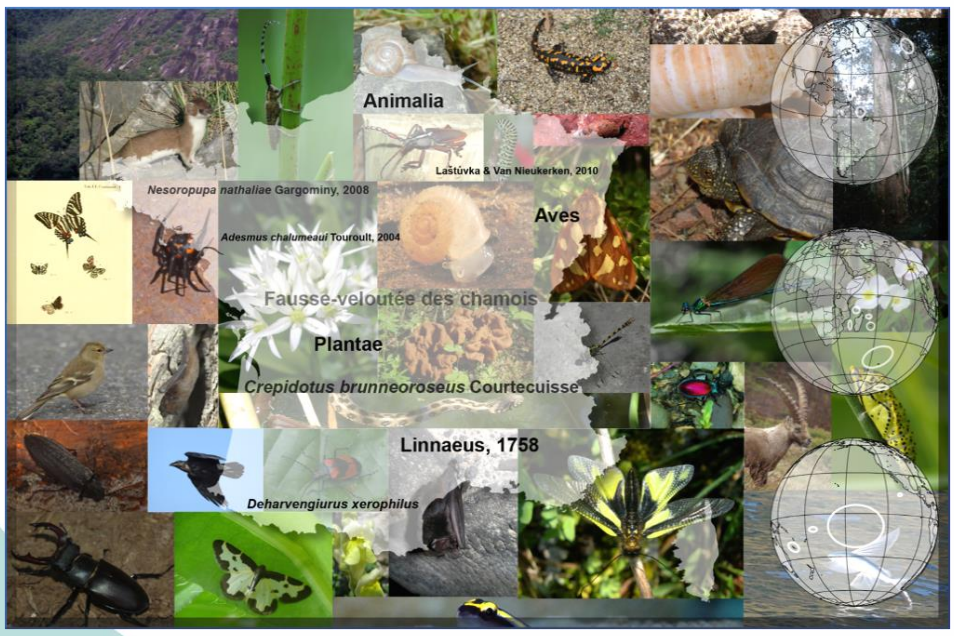
PATRINAT
CENTRE D'EXPERTISE ET DE DONNÉES
OFB•MNHN•CNRS•IRD

TAXREF, référentiel taxonomique pour la France : méthodologie, mise en œuvre et diffusion

O. Gargominy, S. Tercerie, C. Régnier, T. Ramage,

P. Dupont, P. Daszkiewicz & L. Poncet



Janvier 2025

PATRINAT

Centre d'expertise et de données sur le patrimoine naturel

Un service commun
de l'Office français de la biodiversité,
du Muséum national d'Histoire naturelle,
du Centre national de la recherche scientifique
et de l'Institut pour la recherche et le développement

Nom du Programme/Projet : TAXREF, référentiel taxonomique pour la France

Chef de projet : Olivier Gargominy

Chargé.e.s de mission : P. Antonetti, P. Daszkiewicz, O. Fogliani, C. Régnier, G. Simian, S. Tercerie

Référence du rapport conseillée : Gargominy, O., Tercerie, S., Régnier, C., Ramage, T., Dupont, P., Daszkiewicz, P. & Poncet, L. 2025. *TAXREF, référentiel taxonomique pour la France : méthodologie, mise en œuvre et diffusion.* Rapport PatriNat (OFB-CNRS-MNHN), Muséum national d'Histoire naturelle, Paris. 47 pp.

Photos de couverture : Yan BUSKE, Olivier GARGOMINY, Arnaud HORELLOU, Philippe GOURDAIN, Renaud PUISSAUVE, Jean-Philippe SIBLET, Julien TOUROULT.

PatriNat

Centre d'expertise et de données sur le patrimoine naturel

Dans une unité scientifique associant des ingénieurs, des experts et des spécialistes de la donnée, PatriNat rapproche les compétences et les moyens de ses quatre tutelles que sont l'OFB, le MNHN, le CNRS et l'IRD.

PatriNat coordonne des programmes nationaux d'acquisition de connaissance pour cartographier les écosystèmes, les espèces et les aires protégées, surveiller les tendances de la biodiversité terrestre et marine, répertorier les zones clefs pour la conservation de la nature (Znieff), et produire des référentiels scientifiques et techniques (TaxRef, HabRef, etc.). Ces programmes associent de nombreux partenaires et fédèrent les citoyens à travers des observatoires de sciences participatives (tels que Vigie-Nature, INPN espèces ou Vigie-terre).

PatriNat développe des systèmes d'information permettant de standardiser, partager, découvrir, synthétiser et archiver les données aussi bien pour les politiques publiques (SIB, SINP) que pour la recherche (PNDB) en assurant le lien avec les systèmes internationaux (GBIF, CDDA, etc.)

PatriNat apporte son expertise dans l'interprétation des données pour accompagner les acteurs et aider les décideurs à orienter leurs politiques : production d'indicateurs, notamment pour l'Observatoire national de la biodiversité (ONB) et des livrets de chiffres clés, élaboration des Listes rouges des espèces et écosystèmes menacés, revues systématiques, préparation des rapportages pour les directives européennes, élaboration d'outils de diagnostic de la biodiversité pour les acteurs des territoires, ou encore évaluation de l'efficacité des mesures de restauration. PatriNat organise également l'autorité scientifique CITES pour la France.

L'ensemble des informations (de la donnée brute à la donnée de synthèse) est rendu publique dans les portails NatureFrance, INPN et Compteur BIOM.

En savoir plus : www.patrinat.fr

Direction : Laurent PONCET et Julien TOUROULT

Naturefrance

Le service public d'information sur la biodiversité

Naturefrance représente le service public d'information sur les politiques publiques de biodiversité en France. Il se décline dans plusieurs portails d'information, dont le portail général naturefrance.fr. Destiné à un public aussi large que possible, il propose des clés de lecture des grands enjeux liés à la biodiversité et à son évolution, aux pressions qu'elle subit, et aux réponses de la société. Naturefrance présente des chiffres clés, des indicateurs développés dans le cadre de l'ONB (Observatoire national de la biodiversité), des articles et des publications, issus de l'analyse scientifique des données provenant des politiques publiques de conservation ou d'activités socio-économiques favorables ou défavorables à la biodiversité.

Dans le cadre de cette mission confiée par l'OFB, PatriNat gère ce portail et participe au traitement, à l'analyse et à l'interprétation d'une partie des données versées sur Naturefrance : par exemple, celles provenant du Système d'information de l'inventaire du patrimoine naturel (SINP) ou encore du Système d'information de la Convention sur le commerce international des espèces de faune et de flore sauvages menacées d'extinction (SI CITES).

En savoir plus : naturefrance.fr

Inventaire national du patrimoine naturel

Le portail de la biodiversité et de la géodiversité françaises, de métropole et d'outre-mer

Dans le cadre de Naturefrance, l'Inventaire national du patrimoine naturel (INPN) est le portail de la biodiversité et de la géodiversité françaises, de métropole et d'outre-mer (www.inpn.fr). Il regroupe et diffuse les informations sur l'état et les tendances du patrimoine naturel français terrestre et marin (espèces animales, végétales, fongiques et microbiennes actuelles et anciennes, habitats naturels, espaces protégés et géologie) en France métropolitaine et ultramarine.

Les données proviennent du Système d'information de l'inventaire du patrimoine naturel (SINP) et de l'ensemble des réseaux associés. PatriNat organise au niveau national la gestion, la validation, la centralisation et la diffusion de ces informations. L'inventaire consolidé qui en résulte est l'aboutissement d'un travail associant scientifiques, collectivités territoriales, naturalistes et associations de protection de la nature, en vue d'établir une synthèse régulièrement mise à jour du patrimoine naturel en France.

L'INPN est un dispositif de référence français pour la connaissance naturaliste, l'expertise, la recherche en macroécologie et l'élaboration de stratégies de conservation efficaces du patrimoine naturel. L'ensemble de ces informations sont mises à la disposition de tous, professionnels, amateurs et citoyens.

En savoir plus : www.inpn.fr

Compteur Biodiversité Outre-mer

Le portail des indicateurs, des enjeux et des initiatives sur la biodiversité en outre-mer

Dans le cadre de Naturefrance, le Compteur de la biodiversité Outre-mer (BiOM) développe une entrée dédiée aux territoires ultramarins français qui abritent une part importante de la biodiversité mondiale. Portail accessible, actualisé et pérenne, il favorise la rencontre des citoyens et des acteurs de la biodiversité, autour de trois objectifs : partager la connaissance scientifique, valoriser les actions des territoires ultramarins, et encourager chacun à agir. Cette démarche vise à relater les contextes culturels et mettre en avant des enjeux spécifiques de chaque territoire, pour répondre à un engagement du Livre bleu des Outre-mer.

Des études auprès des citoyens viennent compléter l'initiative : par exemple le premier panorama des programmes de sciences participatives dans les territoires, et une enquête sur la perception de la nature et l'utilisation des outils numériques.

PatriNat assure la mise en œuvre du projet et avec la participation des acteurs des outre-mer, suivant trois axes : production d'indicateurs de biodiversité (connaissances, espèces menacées, espaces protégés, etc.), relai des actions de mobilisation et de sciences participatives (écogestes, inventaires participatifs, etc.) et gestion technique du portail

En savoir plus : biodiversite-outre-mer.fr

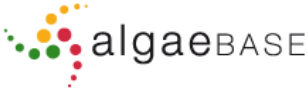
PARTENAIRES (liste non exhaustive)





















































Pour toutes remarques ou demandes relatives à TAXREF,

utilisez la page « Signalements » de TAXREF-Web :

https://taxref.mnhn.fr/taxref-web/issues

SOMMAIRE

AVANT-PROPOS **9**

PARTIE 1 : ACQUISITION **11**

1. Contexte et objectifs **11**

1.1. Objectifs **11**

1.2. Portée du référentiel **12**

1.2.1. Portée géographique **12**

1.2.2. Portée taxonomique **12**

1.2.3. Portée temporelle **12**

1.3. Périmètre du référentiel **12**

1.4. Coordination et consolidation, rôle de PatriNat **13**

2. Gestion des noms scientifiques **13**

2.1. Principes généraux **13**

2.2. Pérennité des noms diffusés **14**

2.3. Règles pour la création de nouveaux noms **14**

2.4. Recommandations nomenclaturales **15**

3. Sources de l'information **16**

3.1. Sources utilisées **16**

3.2. Global Species Database (GSD) **16**

3.3. Consolidations nationale et internationale **18**

3.3.1. Cas des listes de taxons liés à une GSD 18

3.3.2. Cas des listes de taxons non liés à une GSD 19

3.4. Références bibliographiques 20

3.4.1. Pour ajout de noms scientifiques 20

3.4.2. Pour ajout de noms vernaculaires 20

3.4.3. Pour occurrences nouvelles dans les territoires français 20

3.5. Choix de nouvelles sources 20

4. Biogéographie des taxons 21

4.1. Logique référentielle et logique d'inventaire 21

4.2. Notion de statuts biogéographiques 21

4.2.1. Définition des statuts biogéographiques 21

4.2.2. Attribution des statuts biogéographiques 24

4.2.3. Synonymies et gestion des héritages 25

5. Structure du référentiel 25

5.1. Structure de la table TAXREF 25

5.1.1. Liste des champs 26

5.1.2. Clé primaire 29

5.1.3. Contraintes d'intégrité 29

5.2. Gestion de la synonymie 29

5.2.1. Cas particulier de la synonymie d'usage 29

5.3. Gestion de la classification 30

5.4. Liens vers les bases de données sources 30

5.5. Références bibliographiques 30

5.6. Noms vernaculaires 31

5.6.1. Structure de la table TAXVERN 31

5.6.2. Gestion des héritages de noms vernaculaires 32

6. Gestion du référentiel 32

6.1. Historique et suivi des modifications 32

6.2. Gestion des ressources bibliographiques (DOCS-Web) 33

6.3. Application TAXREF-Web 34

6.4. Contrôles effectués sur le référentiel 37

PARTIE 2 : DIFFUSION 39

7. Téléchargement 39

8. Structure du référentiel diffusé 40

9. Modifications depuis la version précédente 42

10. Cas des enregistrements (CD_NOM) qui disparaissent 42

11. Référentiel en ligne sur l'INPN 43

12. API 43

13. Historique des différentes versions de TAXREF diffusées 44

PARTIE 3 : UTILISATION 45

14. Utilisation du référentiel dans la gestion de données 45

14.1. Principes généraux 45

14.1.1. Possibilité de gestion multi-référentielle 46

14.2. Suivi des modifications 46

15. Recherche de correspondance (réconciliation) 47

15.1. Directives 47

15.2. Pièges 47

15.3. Outil de réconciliation TAXREF_MATCH 47

15.4. Stockage des informations dans les tables de données 47

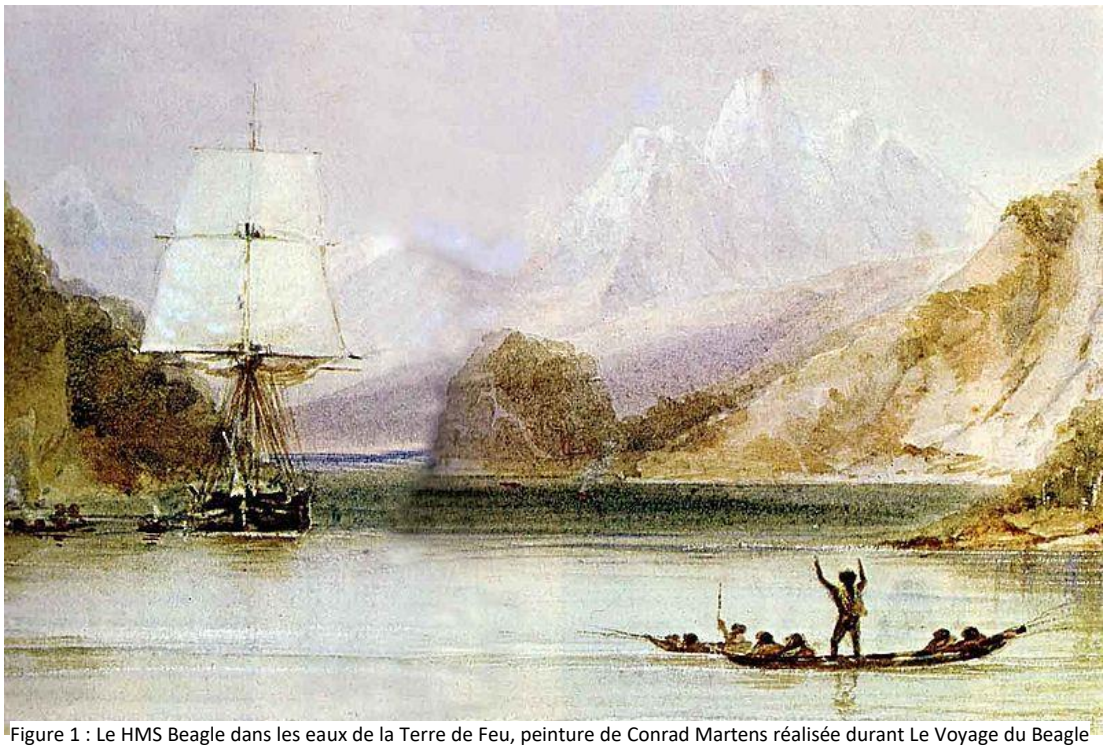


Figure 1 : Le HMS Beagle dans les eaux de la Terre de Feu, peinture de Conrad Martens réalisée durant Le Voyage du Beagle (1831-1836), provenant de *The Illustrated Origin of Species* de Charles Darwin.

AVANT-PROPOS

Au début du 18ème siècle, notamment avec l'accélération de l'exploration outre-mer et du commerce des ressources naturelles, les espèces de plantes et d'animaux qui arrivent sur les comptoirs européens et les paillasses des scientifiques deviennent tellement nombreuses que les noms pour les désigner se multiplient et se complexifient à outrance. Trouver un standard de nommage devient urgent, et le système de nomenclature binominale de Carl von Linné s'impose, notamment avec la publication en 1753[1] de son *Species plantarum* puis de la dixième édition de son *Systema Naturæ*, en 1758[2], qui généralise le système de nomenclature binominale.

Au 21ème siècle, les urgences en matière de gestion de la biodiversité pressent l'ensemble des acteurs de réunir la connaissance sur l'état et l'évolution de cette biodiversité. Les quantités

[1] Linnaeus, C. 1753. *Species plantarum, exhibentes plantas rite cognitas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas*. Holmiæ. (L. Salvii). 1200 pp.

[2] Linnaeus, C. 1758. *Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis*. Holmiæ. (Salvius). 1384 pp.

d'informations collectées sur la nature par trois siècles de naturalistes parcourant la planète entière ou explorant le moindre recoin de leur jardin sont tout simplement énormes, mais complètement disséminées : collections publiques ou privées, littérature volumineuse et éparpillée, carnets de terrains ou fichiers informatiques forment un vaste monde où l'accessibilité n'est pas toujours de mise. Avec l'avènement des technologies de l'information, les bases de données se sont révélées être une clé pour l'accès à l'information. Mais les bases de données du 21ème siècle sont comme les hommes du 18ème et ceux de maintenant : sans références, difficile d'organiser l'information et de dialoguer.

Dans le cadre du Système d'Information sur la Nature et les Paysages (SINP), le Muséum national d'Histoire naturelle est en charge de l'élaboration du référentiel taxonomique des organismes vivant sur l'ensemble des territoires de la République française, qui représente la pierre angulaire du système d'information sur les espèces. Le but du présent document est d'exposer la méthodologie d'élaboration, de diffusion et d'utilisation de ce référentiel électronique TAXREF.



Figure 2 : Planche d'herbier de l'herbier de Tournefort, Muséum national d'Histoire naturelle de Paris. En bas à gauche, un polynom antérieur à la nomenclature binominale(source :**http://science.mnhn.fr/institution/mnhn/collection/p/item/p05361881**)

PARTIE 1 : ACQUISITION

1. Contexte et objectifs

TAXREF est le référentiel taxonomique national des organismes vivant dans les territoires de la République française, élaboré et diffusé par le Muséum national d'Histoire naturelle (MNHN) dans le cadre de la mise en œuvre du Système d'Information de l'iNventaire du Patrimoine naturel (SINP). Ce référentiel est unique pour la France ; il a pour but de lister et d'organiser les noms scientifiques de l'ensemble des êtres vivants recensés sur le territoire. Le référentiel TAXREF est consultable et téléchargeable sur le site internet de l'Inventaire national du Patrimoine naturel (INPN) coordonné par le MNHN.

1.1. Objectifs

Les objectifs du référentiel taxonomique sont de :

- donner un nom scientifique unique non ambigu pour chacune des espèces de France qui soit consensuel aux niveaux national et international ;
- permettre une interopérabilité entre les jeux et bases de données ;
- gérer les évolutions taxonomiques et nomenclaturales dans les données concernant ces espèces (suivi et gestion de la synonymie et de la hiérarchie taxonomique).

TAXREF est une pièce essentielle du dispositif d'information sur la nature en France, qui permet de communiquer tant au niveau des personnes que des ordinateurs grâce à l'utilisation de normes communes de références. Il permet la communication entre taxonomistes et utilisateurs de noms d'espèces (gestionnaires de données naturalistes, gestionnaires d'espaces naturels, écologues, généticiens, muséologues, enseignants et législateurs) et rend les bases de données interopérables.

Le référentiel doit permettre de répondre rapidement, par simple requête, à des questions telles que : combien y a-t-il d'espèces d'insectes en France, quelles sont les données qui concernent tel taxon, etc.

1.2. Portée du référentiel

1.2.1. Portée géographique

Le référentiel TAXREF couvre l'ensemble des organismes vivant dans les territoires de la République française :

- la France hexagonale et la Corse,
- la Guyane,
- la Martinique,
- la Guadeloupe et ses dépendances (Marie-Galante, l'archipel des Saintes et la Désirade),
- Saint-Martin,
- Saint-Barthélemy,
- Saint-Pierre-et-Miquelon,
- Mayotte,
- la Réunion,
- les îles Éparses,
- les îles subantarctiques (archipel des Kerguélen, archipel de Crozet, îles Saint-Paul et Amsterdam),
- la Terre Adélie,
- la Nouvelle-Calédonie,
- Wallis-et-Futuna,
- la Polynésie française,
- l'île de Clipperton.

Le référentiel couvre les milieux continentaux et marins.

Le référentiel inclut des espèces citées par erreur de ces territoires ainsi que des espèces non encore présentes mais pour lesquelles une surveillance est mise en place (cas des espèces envahissantes).

1.2.2. Portée taxonomique

La portée taxonomique du référentiel s'étend sur l'ensemble du domaine vivant et les virus, avec en priorité les plantes, les animaux et les champignons.

1.2.3. Portée temporelle

Le référentiel TAXREF prend en compte l'ensemble des êtres vivants actuels et ceux qui constituent l'histoire naturelle proche, c'est-à-dire depuis le paléolithique. On y trouvera ainsi le mammouth ou l'ours de caverne.

1.3. Périmètre du référentiel

TAXREF est un **référentiel des espèces sauvages.** Les taxons domestiques ou cultivés (races ou cultivars) ne sont *a priori* pas traités, sauf si la gestion de certaines données l'exige ou qu'on les trouve dans le milieu naturel. Quoi qu'il en soit, l'exhaustivité n'est pas recherchée, à la différence des espèces sauvages.

La priorité du référentiel est de mettre à disposition autant de noms scientifiques que possible, en attendant de pouvoir prétendre à l'exhaustivité, sachant que la complétude de l'information reste hétérogène selon les groupes et les espaces géographiques, en fonction de l'expertise disponible (handicap taxonomique).

1.4. Coordination et consolidation, rôle de PatriNat

De nombreuses listes de référence existent déjà selon les groupes taxonomiques et les différents territoires de la République française, souvent avec une expertise taxonomique et nomenclaturale variable. Il s'agit donc de coordonner l'ensemble de ces connaissances et de faire valoir les compétences locales dans une taxonomie générale. Cette coordination implique une **consolidation nationale**, exprimant la nécessité d'utiliser un seul et même nom pour une espèce, qu'il s'agisse de la Martinique, la Réunion ou la Polynésie française. Cette réconciliation est notamment indispensable dans le contexte des espèces exotiques envahissantes. Dans la mesure de l'existant, cette consolidation est portée au niveau mondial par l'utilisation de Global Species Database en liens uniques (voir 3.2).

Le rôle de PatriNat est donc :

- d'assurer la synthèse, la coordination et la consolidation nationale, régionale ou globale du référentiel ;
- de produire de nouveaux référentiels lorsque ceux-ci n'existent pas ;
- d'assurer la diffusion du référentiel, notamment au travers du SINP, vers l'ensemble de la communauté.

2. Gestion des noms scientifiques

2.1. Principes généraux

TAXREF est un référentiel taxonomique et nomenclatural. Il fournit une liste de noms scientifiques valides et leurs synonymes reflétant les connaissances taxonomiques à un temps donné. Seuls les noms disponibles au sens des codes de nomenclature entrent dans le cadre de TAXREF :

- pour les animaux : code international de nomenclature zoologique (ICZN[3]) ;
- pour les algues, les champignons et les plantes : code international de nomenclature des algues, de la fonge et des plantes (Code de Shenzhen[4]) ;
- pour les bactéries : code international de nomenclature des procaryotes[5] ;
- pour les virus : code international de classification et de nomenclature des virus[6].

TAXREF n'est pas un espace de publication : aucune combinaison nouvelle, c'est-à-dire non publiée, n'y est établie.

En conséquence, TAXREF refuse (liste non exhaustive) :

- les groupes dits fonctionnels, c'est-à-dire des groupes pratiques pour identifier un ensemble d'espèces mais pas nécessairement clairement définis taxonomiquement et non publiés validement (par exemple « papillons de nuit » ou « petits poissons »). Ces groupes doivent être gérés aux niveaux des données et non dans le référentiel taxonomique lui-même. Seuls les

3 https://www.iczn.org/the-code/the-code-online/
4 https://www.iapt-taxon.org/nomen/main.php
5 https://www.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.000778
6 https://ictv.global/about/code

groupes qui ont fait l'objet d'une publication peuvent être intégrés à TAXREF (par exemple la section *Ruderalia* Kirschner, H. Øllg. et Stepanek, définie dans le genre *Taraxacum*).

- les noms comportant des mentions telles que « sp. », « cf. », « aff. », « prox. », « ? », etc. qui sont liés à des problèmes d'identification et doivent être gérés au niveau des données et non au niveau du référentiel.
- les noms « *sensu* », c'est-à-dire les noms utilisés dans un sens autre que celui de l'auteur du taxon, ne peuvent pas être considérés comme de référence. En revanche, TAXREF peut prendre en compte les noms « *sensu* » qui sont synonymes d'un autre nom, à condition que les auteurs concernés soient clairement mentionnés et que la correspondance ait été établie par un spécialiste ayant une très bonne connaissance des usages au cours du temps.

Pour ces deux derniers points, il existe des exceptions discutées ou en cours de discussion.

2.2. Pérennité des noms diffusés

Les noms scientifiques diffusés dans le référentiel TAXREF sont pérennes.

Les changements de nom d'autorité sont gérés par la création d'un nouvel enregistrement, à l'exception des cas suivants :

- ajout de l'auteur (et/ou de la date) lorsque celui-ci n'était pas renseigné dans les versions précédentes du référentiel ;
- ajout ou retrait des parenthèses à l'autorité (cas de la faune) ;
- faute de frappe évidente par exemple « Hartret, 1921 » corrigé en « Hartert, 1921 » ;
- synonymie d'usage : afin de respecter la continuité des usages dans les jeux de données, il est possible de transformer l'autorité en « auct. non » suivi de l'auteur du nom valide, à condition que la synonymie d'usage s'applique à l'ensemble du champ géographique de TAXREF (territoires de la République française). Voir 5.2

Le nom scientifique peut être exceptionnellement modifié dans le cas d'émendations orthographiques prévues par les codes de nomenclature. Toutefois, dans le cas où l'orthographe injustifiée est la plus utilisée, alors il y a création d'un nouvel enregistrement.

Par ailleurs, un enregistrement peut ne plus être diffusé dans TAXREF : il s'agit de cas exceptionnels, lorsqu'il s'avère qu'un nom n'est rattachable à aucun autre nom du référentiel (ou si jamais l'enregistrement correspond à un doublon d'un autre enregistrement). Ces noms sont alors retirés de la diffusion mais restent dans la table de travail (table TAXREF). De cette façon, il reste possible de ré-inclure ultérieurement ces noms dans la diffusion si cela s'avère justifié. Par la suite, ils pourront être supprimés, mais toute délétion d'un nom de la table TAXREF s'accompagne de la copie de ce nom dans une table de structure similaire : aucun nom n'est jamais vraiment effacé (voir 10. Cas des enregistrements (CD_NOM) qui disparaissent).

En aucun cas ces rares changements ne doivent induire une quelconque ambigüité taxonomique.

2.3. Règles pour la création de nouveaux noms

Le référentiel TAXREF a pour but de lister les noms de taxons figurant dans la littérature taxonomique. Les noms non corrects mais qui sont malgré tout couramment utilisés et qui figurent dans une publication de référence ou un texte officiel (par exemple un texte juridique) sont acceptés dans le référentiel, par exemple :

- *Gymnophthalmus pleei* est couramment utilisé alors que le nom correct est *Gymnophthalmus pleii*. Ces deux noms figurent dans TAXREF, le premier étant rattaché au deuxième.

- *Potimirim glabra* (Müller, 1881) est mentionné à tort dans une publication scientifique. Ce nom est intégré à TAXREF et rattaché à *Potimirim glabra* (Kingsley, 1878) qui est l'autorité correcte.

Malgré tout, la liste des taxons de TAXREF est contrainte par l'utilisation des référentiels sources et les règles des codes de nomenclature.

2.4. Recommandations nomenclaturales

Dans le cas où plusieurs solutions nomenclaturales sont autorisées par ces codes, TAXREF applique les recommandations suivantes :

NOM SCIENTIFIQUE

Le nom de l'espèce doit suivre la nomenclature binominale : l'écriture sans le sous-genre entre parenthèses est considérée par défaut comme le nom de référence. Des exceptions peuvent apparaître dans TAXREF en raison de la liaison avec WoRMS qui ne suit pas toujours cette règle.

Pour les rangs infra-spécifiques, la nomenclature trinominale est adoptée :

- dans le cas de la flore, avec indication systématique du rang (« subsp. » pour les sous-espèces, « var. » pour les variétés, « f. » pour les formes) entre l'épithète spécifique et le nom terminal (voir article 24.1 (ICBN)), par exemple : *Euphorbia flavicoma* subsp. *costeana*.
- dans le cas de la faune, sans aucun ajout pour les sous-espèces (ex : *Acmaeodera bipunctata obenbergeriana*); avec indication du rang dans les autres cas (*Ascidia sydneiensis* var. *samea*).

CITATION DES AUTEURS

- S'il y a 2 auteurs, ils sont séparés par une esperluette (signe '&') encadrée d'espaces, par exemple : Linden & André (cf. Recommandation 46C1 (ICBN)).
- les synonymies d'usage sont notées sous la forme « auct. non » suivi de l'auteur du nom valide (e.g. *Crocodilurus lacertinus* auct. non (Daudin, 1802) ou *Alchemilla hybrida* auct. non L. (recommandation 50D (ICBN)). « auct. » peut être remplacé par « sensu » suivi du nom de l'auteur employant le nom à tort.

Pour les algues, champignons et plantes :

- TAXREF adopte la recommandation 46A du Code qui préconise l'utilisation des abréviations standardisées de Brummitt & Powell (1992)[7] pour les noms d'auteurs (travail d'homogénéisation des abréviations en cours pour l'ensemble du référentiel). En conséquence, un même auteur peut apparaître sous deux orthographes différentes s'il s'agit d'un nom de plante ou d'animal (cas de Carl Reinhold Sahlberg : « C. R. Sahlberg » en zoologie et « Sahlb. » en botanique).
- TAXREF suit rigoureusement l'IPNI et ne met aucun espace dans les abréviations d'auteur, que ce soit entre l'initiale(s) et le nom ou dans l'abréviation du nom, par exemple :

  Écrire « W.D.J.Koch» et non « W. D. J. Koch» ;

  Écrire « Balf.f. » et non « Balf. f. ».

Pour les animaux :

- s'il existe plusieurs auteurs portant le même nom, le plus ancien s'écrit sans initiale et les suivants s'écrivent avec les initiales afin d'éviter toute homonymie d'auteur. Dans ce cas, il faut

[7] Brummitt, R. K. & Powell, C. E. 1992. *Authors of plant names: a list of authors of scientific names of plants*. Royal Botanic Gardens, Kew London, 732 pp.

mettre un espace uniquement entre la dernière initiale et le nom (pas d'espace entre les initiales). De nombreux cas sont encore à traiter dans TAXREF.

- par ailleurs, « Linnaeus » est utilisé pour Linné, Fuessly préférentiellement à Fuesslin et Füssli, et Motschulsky préférentiellement à Motschoulsky. Les caractères diacritiques sont respectés au maximum (Strøm utilisé plutôt que Ström). Cette liste est complétée au fur et à mesure de la découverte de cas litigieux.

3. Sources de l'information

3.1. Sources utilisées

Trois types de sources sont utilisés pour alimenter TAXREF :

- des bases de données validées et officielles (WoRMS, Index Fungorum, FishBase, Fauna Europaea, base Nadeaud, etc.)[8] ;
- des publications scientifiques ;
- les données de groupes d'experts : Société Herpétologique de France, Société Française d'Arachnologie, IFRECOR, etc.

Une attention particulière est portée aux sources de données selon qu'elles représentent de l'expertise taxonomique ou une source pour l'occurrence des différents taxons sur les territoires concernés (« liste de référence » ou « checklist ») qui ne suit pas nécessairement la taxonomie et la nomenclature la plus à jour. En conséquence, une méthodologie est mise en place afin d'utiliser ces différentes sources dans le contexte de consolidation nationale.

Les sources utilisées dans le cadre de TAXREF sont consultables sur le site de l'INPN (par espèce) et sur le site TAXREF-web (par taxon (onglet « Notes ») ou par groupe opérationnel (onglet « Biblio »)).

3.2. Global Species Database (GSD)

Les GSD sont des référentiels taxonomiques qui traitent d'un groupe taxonomique dans sa globalité, ce qui est l'objectif même de la taxonomie. Ce sont des sources de données *a priori* fiables, puisque ce sont les taxonomistes spécialistes du groupe eux-mêmes qui assurent la consolidation à l'échelle mondiale. En conséquence, **TAXREF favorise l'utilisation de référentiels mondiaux (GSD) et participe à leur amélioration**.

Les GSD sont logiquement prioritaires sur les référentiels régionaux (RSD, Regional Species Database) ou locaux (LSD, Local Species Database). Néanmoins, il est évident que les référentiels locaux évoluent généralement plus rapidement que les GSD. En conséquence, la méthodologie TAXREF vise à faire remonter les informations nouvelles du niveau local au niveau mondial.

[8] https://taxref.mnhn.fr/taxref-web/contexts

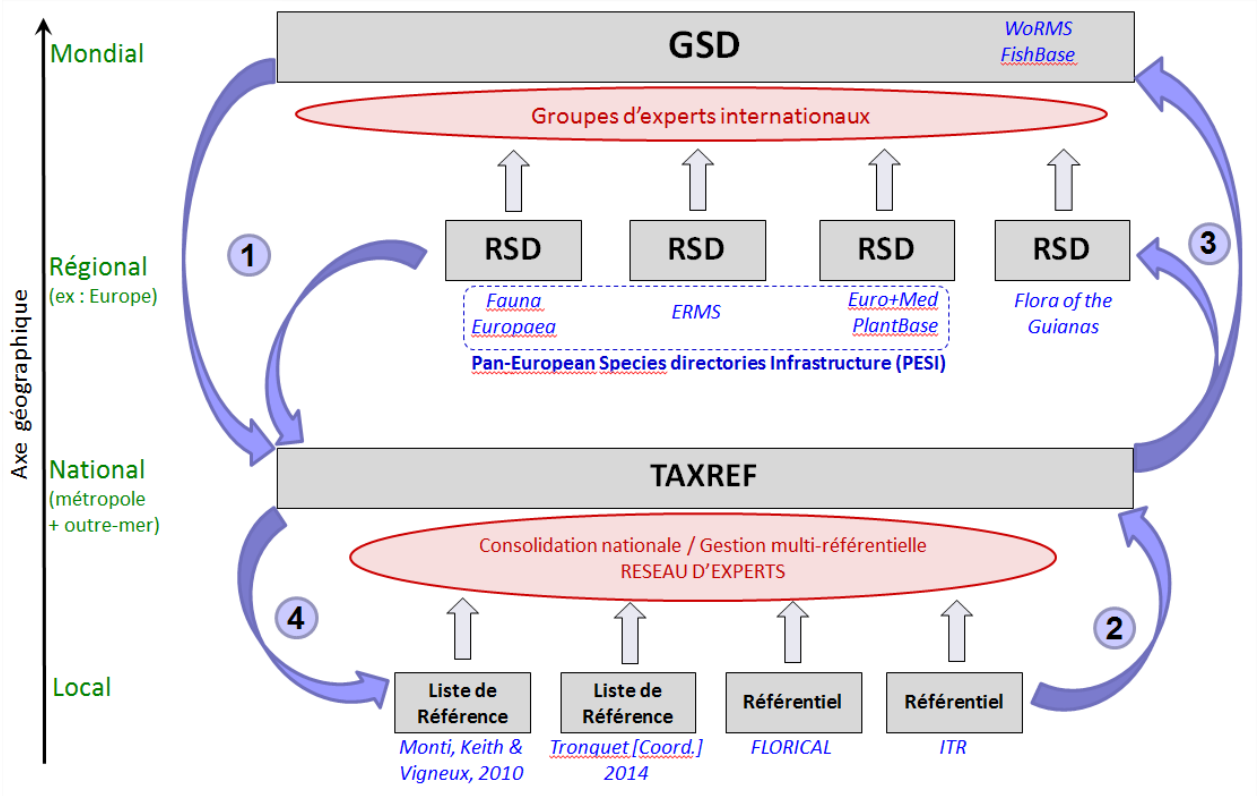


Figure 3: Interaction entre TAXREF et les référentiels existants (GSD : Global species database ; RSD : Regional species database).

Lorsqu'une base mondiale (GSD) existe pour un groupe taxonomique donné et qu'elle est suffisamment à jour et avancée pour être reconnue par la communauté scientifique, TAXREF l'utilise donc prioritairement comme source (**Erreur ! Source du renvoi introuvable.Erreur ! Source du renvoi introuvable.Erreur ! Source du renvoi introuvable.**Figure 3, flèche 1). La procédure consiste alors à **identifier les taxons présents en France**. Des extractions de la GSD sont effectuées et une correspondance unique (lien 1-1) est établie entre les identifiants de la GSD et les identifiants de TAXREF (CD_NOM), ce qui assure une équivalence stricte entre les deux bases. TAXREF utilise à l'heure actuelle plusieurs GSD, notamment Index Fungorum, Algaebase ou le World Register of Marine Species (WoRMS).

Si pour un groupe taxonomique donné il n'y pas de GSD mais qu'il existe en revanche une RSD, alors celle-ci est utilisée comme source et mise à profit de la même façon qu'une GSD.

TAXREF utilise par exemple Fauna Europaea (faune continentale d'Europe).

Si pour un groupe donné, il existe une GSD et une RSD (e.g. WoRMS (World Register of Marine Species) et ERMS (European Register of Marine Species)), c'est naturellement la GSD qui a priorité.

Enfin, si une GSD ou une RSD n'est pas reconnue par la communauté scientifique parce qu'elle ne s'appuie pas sur un réseau d'experts suffisants et pratique une logique de compilation sans contrôle de noms existants ou sans références aux codes de nomenclatures internationaux, alors elle n'est pas utilisée.

3.3. Consolidations nationale et internationale

Au niveau local, TAXREF intègre des listes validées et publiées qui ne couvrent qu'une partie du territoire français (« listes de référence », **Erreur ! Source du renvoi introuvable.Erreur ! Source du renvoi introuvable.Erreur ! Source du renvoi introuvable.**Figure 3, flèche 2). La problématique consiste alors à trouver un consensus entre les différentes sources lorsque des désaccords existent : c'est le travail de consolidation nationale, qui est effectué en concertation avec les experts taxonomistes des groupes et territoires concernés.

3.3.1. Cas des listes de taxons liés à une GSD

Dans le cas où la liste de référence traite d'un groupe dont il existe une GSD, elle est mise en correspondance avec celle-ci (Figure 4**Erreur ! Source du renvoi introuvable.Erreur ! Source du renvoi introuvable.**). Cette comparaison entre le niveau local (listes de référence) et le niveau global (GSD) est doublement bénéfique :

- d'un côté, la liste de référence permet de repérer les manques ou les erreurs de la GSD qui en est alors informée (**Erreur ! Source du renvoi introuvable.Erreur ! Source du renvoi introuvable.**Figure 3, flèche 3);

- de l'autre côté la GSD, grâce à sa conception globale du groupe taxonomique, apporte la cohérence nécessaire entre les différentes listes de référence (Figure 3, flèche 4).

Dans le cas de données nouvelles, il est possible de disposer des nouveaux noms immédiatement sans attendre le retour de la GSD, à condition que ce nom désigne une espèce nouvellement décrite qu'il est normal de ne pas trouver dans la GSD (ou une espèce récemment arrivée sur le territoire dans le cas de l'utilisation de RSD).

Dans le cas où la GSD refuse de corriger ou d'ajouter le nom dans sa base :

- soit le refus est justifié d'un point de vue nomenclatural / taxonomique, auquel cas ce nom doit être reconsidéré au niveau de la liste de référence ;
- soit il s'agit d'un nom qui est hors cadre méthodologique de la GSD (type de synonyme qui n'est pas pris en compte dans la GSD par exemple). Si ce nom entre dans la méthodologie de TAXREF, alors il y sera intégré.

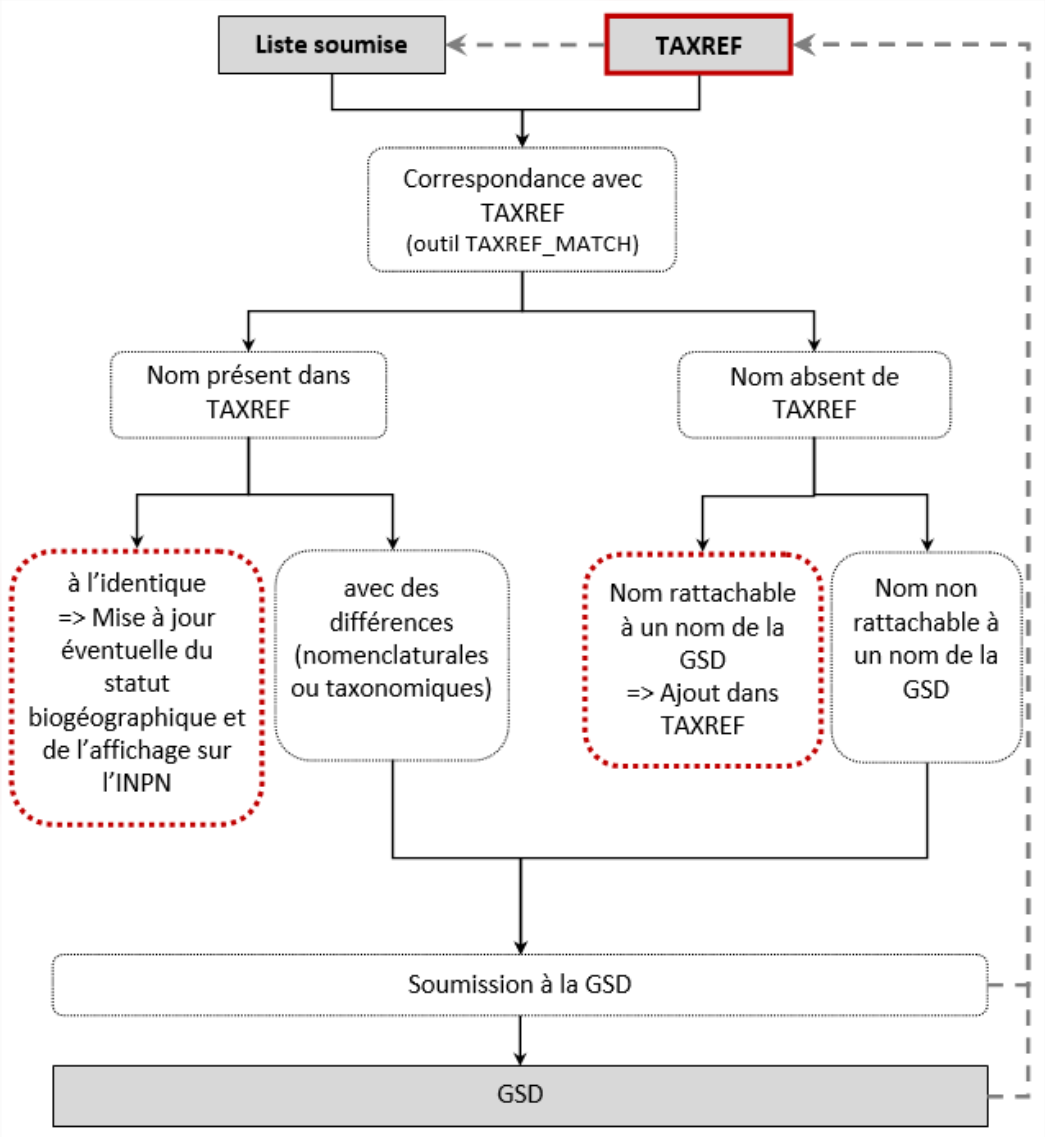


Figure 4 : Méthodologie mise en place entre TAXREF et les Global Species Database (GSD). En gris : référentiels taxonomiques ou « checklists » ; cerclé de rouge : ce qui concerne TAXREF.

3.3.2. Cas des listes de taxons non liés à une GSD

Si la liste concerne un groupe taxonomique non encore traité dans TAXREF, elle est incluse telle quelle à condition qu'il s'agisse d'une liste publiée et reconnue comme référence pour le groupe en question (cas de la plupart des insectes d'Outre-mer par exemple).

Si la liste concerne un groupe taxonomique déjà traité dans TAXREF :

- soit la liste à intégrer est une version actualisée de la liste de TAXREF, auquel cas la mise à jour est effectuée (e.g. Catzeflis (2008) et Catzeflis (2011)) ;
- soit il s'agit d'une liste nouvelle pour TAXREF, auquel cas elle doit être comparée aux données déjà dans TAXREF (en portant attention aux dates de publication, certaines incohérences pouvant provenir de sources trop anciennes). En cas de désaccords entre les deux listes, il peut être nécessaire de contacter les experts taxonomistes du groupe pour décision.

3.4. Références bibliographiques

En fonction du type d'ajouts ou de modifications effectués sur le référentiel, différentes ressources bibliographiques sont nécessaires pour TAXREF.

3.4.1. Pour ajout de noms scientifiques

- Pour les espèces nouvelles : **publication originale** respectant les codes de nomenclature.

ex : Roux, C., Boulanger, M. & Malle, N. 2009. *Caloplaca itiana* sp. nov., espèce nouvelle de lichen du littoral de la Manche (France). *Bulletin de l'Association Française de Lichénologie*, 34(1): 1-10.

- Pour des changements de synonymie ou de classification : **publication scientifique** nécessaire.

ex : Ahrens, D. 2007. Taxonomic changes and an updated catalogue for the Palaearctic Sericini (Coleoptera : Scarabaeidae : Melolonthinae). *Zootaxa*, 1504: 1-51.

3.4.2. Pour ajout de noms vernaculaires

Tout type de littérature peut être pris en compte, y compris les guides naturalistes ou les sites web spécialisés (ex : site DORIS sur la faune et de la flore subaquatiques de la Fédération française d'études et de sports sous-marins).

3.4.3. Pour occurrences nouvelles dans les territoires français

- Pour les statuts biogéographiques (indigénat, introduction, endémisme, invasion, etc.) : **publication scientifique** nécessaire.

ex : Dusoulier, F., Lupoli, R., Aberlenc, H. P. & Streito, J.-C. 2007. L'invasion orientale de *Leptoglossus occidentalis* en France: bilan de son extension biogéographique en 2007 (Hemiptera: Coreidae). *L'Entomologiste*, 63(6): 303–308.

- Pour la présence / absence d'une espèce dans un territoire :

Tout type de littérature peut être pris en compte, y compris les nouvelles occurrences mentionnées sur des forums naturalistes, sous réserve que la détermination de l'espèce soit accompagnée d'une photo, de la date de détermination et du nom du déterminateur.

Dans ce cas, une référence bibliographique est créée avec l'url et une copie de la page web sous forme de zip contenant la page html et les photos associées (« page web complète »).

ex : http://mersociety.wordpress.com/2012/11/07/extraordinarily-rare-whale-sighting-omuras-whale-balaenoptera-omurai/ référencé comme :

Krajewsky, H. 2012. Extraordinarily Rare Whale Sighting – Omura's Whale (*Balaenoptera omurai*). *Marine Education and Research Society Blog*. Posted on November 7, 2012, downloaded on May 5, 2014. http://mersociety.wordpress.com/2012/11/07/extraordinarily-rare-whale-sighting-omuras-whale-balaenoptera-omurai/

3.5. Choix de nouvelles sources

Si pour un groupe taxonomique non traité dans TAXREF, il existe un référentiel reconnu par la communauté (publications récentes faisant référence, e.g. flores, monographies, articles scientifiques), alors il peut être intégré dans TAXREF.

S'il existe plusieurs référentiels qui se valent (cas des oiseaux), le ou les experts du groupe concerné sont consultés pour définir le référentiel à suivre, ou du moins pour discuter des choix à adopter en cas de divergence entre référentiels.

Si aucun référentiel n'existe, la consultation d'experts peut être l'occasion d'élaborer un nouveau référentiel.

Étant donné la grande portée du référentiel taxonomique, le choix de nouvelles sources pour TAXREF se fait en fonction de deux grands types de priorités, à savoir :

- apporter le support référentiel aux différents programmes ayant besoin de nommer les espèces : ZNIEFF, SIE-Sandre, inventaires nationaux, textes juridiques, listes rouges, etc. ;
- rester à l'écoute de l'actualité taxonomique, d'une part en assurant un suivi des publications majeures ou bien des listes de référence, d'autre part en jouant l'opportunisme pour les groupes où des taxonomistes se déclarent comme collaborateurs, ou bien pour les groupes où la connaissance disparaît qu'il est donc urgent de conserver.

Au-delà de ces priorités, le premier objectif reste la complétude du référentiel.

4. Biogéographie des taxons

4.1. Logique référentielle et logique d'inventaire

TAXREF est le référentiel taxonomique national : il liste l'ensemble des êtres vivants recensés sur le territoire et est donc directement lié à une logique d'inventaire. TAXREF gère la répartition des espèces via l'utilisation de statuts biogéographiques par territoire administratif ou biogéographique français. Ces statuts permettent d'avoir une vision synthétique de la répartition des espèces dans les différents territoires français. Ils permettent aussi d'estimer la proportion d'espèces indigènes par rapport aux espèces introduites, ou bien le nombre d'espèces endémiques ou éteintes et sont donc de bons indicateurs.

4.2. Notion de statuts biogéographiques

Les statuts biogéographiques définis se veulent suffisamment généraux pour être transversaux à l'ensemble des êtres vivants. Ils couvrent trois notions différentes : notion de présence (présence/absence), d'origine (indigénat ou introduction) et de surface d'aire d'occupation (endémisme).

4.2.1. Définition des statuts biogéographiques

Tableau 1 : Définitions des statuts biogéographiques utilisés dans TAXREF (valeurs de la table TAXREF_STATUTS).

STATUT	DESCRIPTION	DEFINITION
P	Présent (indigène ou indéterminé)	Taxon présent au sens large dans la zone géographique considérée, c'est-à-dire taxon indigène ou taxon dont on ne sait pas s'il appartient à l'une des autres catégories. Le défaut de connaissance profite donc à l'indigénat. Par **indigène** on entend : taxon qui est issu de la zone géographique considérée et qui s'y est naturellement développé sans contribution humaine, ou taxon

STATUT	DESCRIPTION	DEFINITION
		qui est arrivé là sans intervention humaine (intentionnelle ou non) à partir d'une zone dans laquelle il est indigène[9]. (NB : sont exclus les hybrides dont l'un des parents au moins est introduit dans la zone considérée) Sont regroupés sous ce statut tous les taxons catégorisés « natif » ou « autochtone ». Les taxons hivernant quelques mois de l'année entrent dans cette catégorie.
E	Endémique	Taxon naturellement restreint à la zone géographique considérée.
S	Subendémique	Taxon naturellement restreint à une zone non intégralement incluse dans la zone géographique considérée mais dont les principales populations se situent dans celle-ci (par ex distribution à cheval sur les Pyrénées françaises et espagnoles, sur les Jura français et suisse, endémique corso-sarde, etc.). Pour l'Outre-mer, on applique ce statut à l'endémisme régional : - pour la Guyane= endémique du plateau des Guyanes, - pour les Antilles françaises = endémique des petites Antilles, Trinidad et Tobago non incluses, - pour Mayotte = endémique de l'archipel des Comores, - pour la Réunion = endémique des Mascareignes, - pour les îles subantarctiques = endémique de la South Indian Ocean Province[10]. - pour Wallis-et-Futuna (WF) = endémique Fidgi / WF ou endémique WF / Samoa / Tonga - pour la Polynésie française (PF) = endémique PF / Iles Cook
C	Cryptogène	Taxon dont l'aire d'origine est inconnue et dont on ne peut donc pas dire s'il est indigène ou introduit[11].
I	Introduit	Taxon introduit (établi ou possiblement établi) dans la zone géographique considérée. Par **introduit** on entend : taxon dont la présence dans la zone géographique considérée est due à une intervention humaine, intentionnelle ou non, ou taxon qui est arrivé dans la zone sans intervention humaine mais à partir d'une zone dans laquelle il est introduit[9]. Par **établi** (terme pour la faune, = **naturalisé** pour la flore) on entend : taxon introduit qui forme des populations viables (se reproduisant) et durables qui se maintiennent dans le milieu naturel sans besoin d'intervention humaine[9]. Sont regroupés sous ce statut tous les taxons catégorisés « naturalisé », « établi », et *a priori* ceux catégorisés « non-indigène », « exotique », « exogène », « allogène », « allochtone », « non-natif », (en anglais : *alien*) dans une publication scientifique, sous réserve qu'ils soient établis et non envahissants.
J	Introduit envahissant	Taxon introduit dans la zone géographique considérée, qui produit des descendants fertiles souvent en grand nombre, et qui a le potentiel pour

[9] Pyšek, P., Richardson, D. M., Rejmánek, M., Webster, G. L., Williamson, M. & Kirschner, J. 2004. Alien plants in checklists and floras: towards better communication between taxonomists and ecologists. *Taxon*, 53(1): 131-143.

[10] Marion et Prince Edward Islands, Heard et McDonald Islands, Iles Crozet et Kerguelen au sens de Lewis-Smith (1984), étendu aux Iles Saint-Paul et Amsterdam en raison de leur rattachement administratif aux TAAF [Lewis-Smith, R. I. 1984. Terrestrial plant biology. *In*: Laws, R.M. [Ed] *Antarctic Ecology*. Academic Press, London. 1: 61-162.].

[11] Carlton, J. T. 1996. Biological invasions and cryptogenic species. *Ecology*, 77(6): 1653-1655.

STATUT	DESCRIPTION	DEFINITION
		s'étendre de façon exponentielle sur une grande aire, augmentant ainsi rapidement son aire de répartition[9]. Cela induit souvent des conséquences écologiques, économiques ou sanitaires négatives (IUCN, 2000). Sont regroupés sous ce statut tous les taxons catégorisés « introduit envahissant », « exotique envahissant » ou « invasif » (*invasive* en anglais) dans une publication scientifique.
M	Introduit non établi (dont domestique)	Taxon introduit qui peut occasionnellement se reproduire en dehors de son aire de culture ou de captivité, mais qui ne peut se maintenir à l'état sauvage car ne pouvant former de populations viables sans intervention humaine, et qui dépend donc d'introductions répétées pour se maintenir dans la nature[9] Sont regroupés sous ce statut tous les taxons catégorisés « introduit occasionnel », « subspontané », « échappé de culture ou de captivité » (en anglais : *casual alien*). Ce statut inclut les taxons strictement domestiques (faune), cultivés (flore) ainsi que les espèces commercialisées (espèces d'aquariophilie par exemple).
B	Occasionnel	Taxon occasionnel, non nicheur, accidentel dans la zone géographique considérée (par exemple migrateur de passage).
D	Douteux	Taxon dont la présence dans la zone géographique considérée n'est pas avérée (en attente de confirmation).
A	Absent	Taxon non présent dans la zone géographique considérée.
W	Disparu	Taxon qui n'est plus présent à l'état sauvage dans la zone géographique considérée mais qui n'est pas globalement éteint. Rq : en cas de doute sur la présence ancienne ou non du taxon à l'état sauvage, utiliser le statut absent (A).
X	Éteint	Taxon globalement éteint (= ayant totalement disparu de la surface du globe terrestre).
Y	Introduit éteint / disparu	Taxon introduit naturalisé par le passé mais aujourd'hui disparu de la zone géographique considérée (W) ou éteint (X).
Z	Endémique éteint	Taxon endémique et aujourd'hui disparu, donc globalement éteint (X).
Q	Mentionné par erreur	Taxon mentionné par erreur comme présent sur le territoire considéré.
G	ADNe uniquement	Taxon détecté uniquement par du matériel moléculaire, non confirmé morphologiquement (nouveauté v18).

Ces notions sont contingentes à chacun des territoires administratifs de la République française, à savoir :

- la France hexagonale et la Corse,
- la Guyane,
- la Martinique,
- la Guadeloupe et ses dépendances (Marie-Galante, l'archipel des Saintes et la Désirade),
- Saint-Martin,
- Saint-Barthélemy,
- Saint-Pierre-et-Miquelon,
- Mayotte,
- la Réunion,
- les îles Éparses,
- les îles subantarctiques (archipel des Kerguélen, archipel de Crozet, îles Saint-Paul et Amsterdam),

- la Terre Adélie,
- la Nouvelle-Calédonie,
- Wallis-et-Futuna,
- la Polynésie française,
- l'île de Clipperton.

Ces territoires sont à considérer dans leur intégralité. Ainsi, une espèce qui est indigène d'au moins un site du territoire considéré est à prendre en compte comme indigène du territoire entier, même si elle s'avère introduite dans d'autres parties de ce territoire.

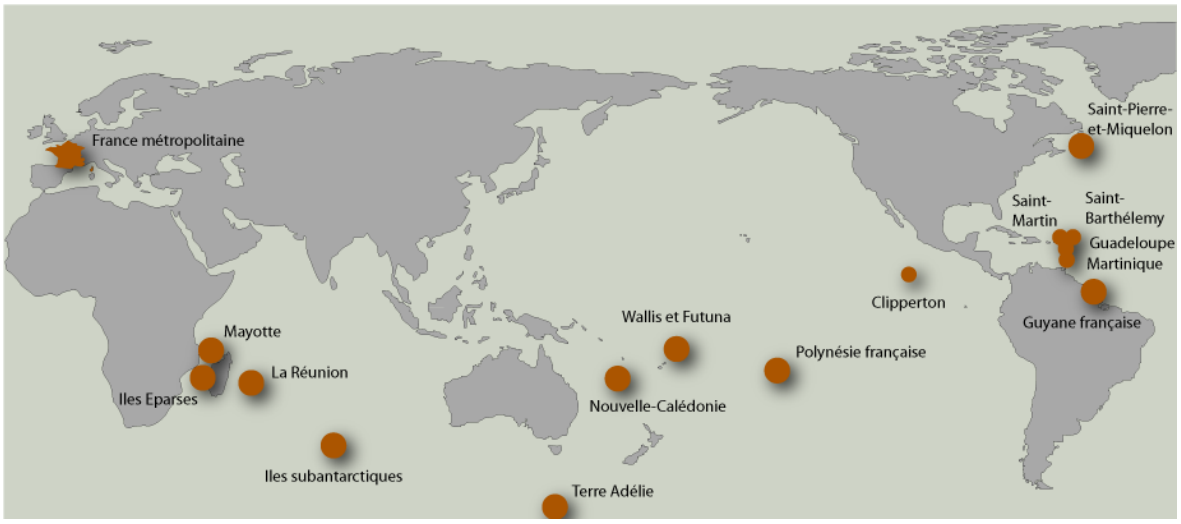


Figure 5 : Localisation des territoires français couverts par TAXREF.

4.2.2. Attribution des statuts biogéographiques

L'attribution des statuts biogéographiques aux taxons repose sur des publications scientifiques, et éventuellement sur les données d'experts dans le cadre de programmes nationaux (inventaires, listes rouges, etc.). La Figure 6 propose un diagramme pour l'attribution d'un statut biogéographique à un taxon. Ce diagramme se parcourt de haut en bas : chaque niveau parcouru permet d'attribuer un statut de plus en plus précis. Ainsi dans tous les cas où une espèce est présente sur le territoire considéré (statut P) mais qu'on peut caractériser cette présence (présence occasionnelle ou non), préciser l'origine du taxon (indigénat ou introduction) ou son endémisme, alors on le précise. En cas de doute entre deux statuts, il est plus judicieux de conserver le statut du niveau précédent sur le diagramme, voire de remonter au statut P (en cas de présence) ou au statut A (en cas d'absence), lesquels seront moins précis mais corrects.

Étant donné que TAXREF recherche l'**exhaustivité des espèces sauvages**, une espèce qui serait à la fois à l'état domestique/cultivé et à l'état sauvage est considérée comme une espèce sauvage. Ainsi, une espèce qui est à la fois cultivée et naturalisée dans le milieu naturel est considérée comme une espèce naturalisée (statut I). De même, il est plus important pour TAXREF de savoir qu'une espèce est disparue à l'état sauvage (statut W) que de savoir qu'elle est toujours cultivée sur le territoire.

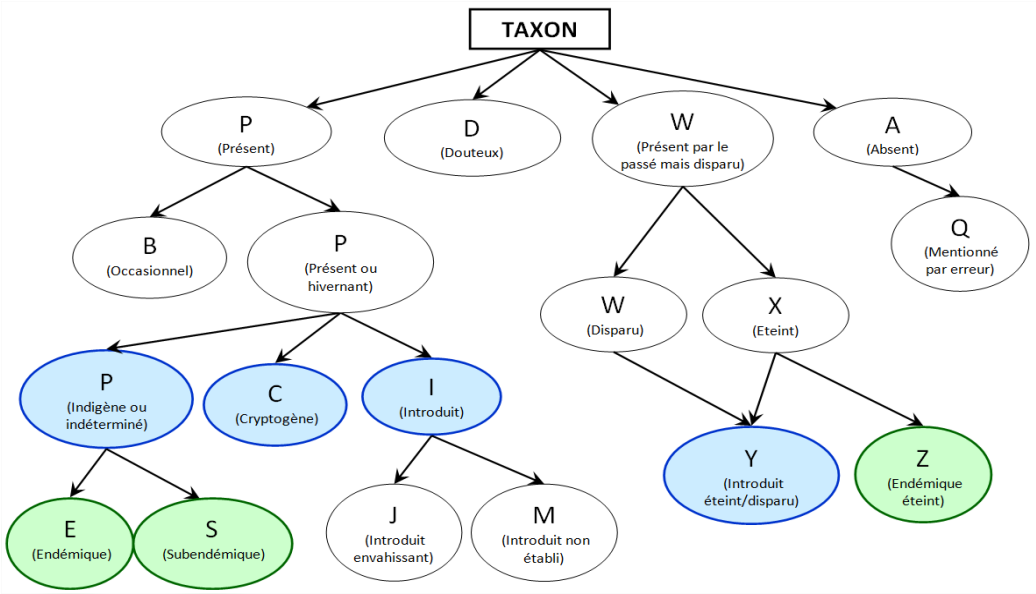


Figure 6: Diagramme pour l'attribution d'un statut biogéographique à un taxon (voir le Tableau 1 pour les définitions des statuts). En bleu : notion d'origine du taxon (indigénat ou introduction) ; en vert : notion de surface de l'aire d'occupation du taxon (endémisme).

4.2.3. Synonymies et gestion des héritages

Lorsqu'un nom devient synonyme d'un autre, et lorsque ce nom possédait un statut pour un territoire donné, la transmission de ce statut vers le nouveau nom de référence n'est pas automatique, de façon à ne pas inclure d'erreurs éventuelles. Par contre, il existe des tests de contrôle pour repérer d'éventuelles incohérences dans les statuts d'occurrence.

5. Structure du référentiel

Le référentiel se compose de 4 tables principales (voir Figure 7) :

- la table TAXREF qui constitue le référentiel *sensu stricto* (noms scientifiques) ;
- la table TAXVERN qui gère les noms vernaculaires ;
- la table CONTEXT qui référence les bases de données source utilisées dans le cadre de TAXREF et fait le lien vers celles-ci ;
- la table DOCS qui gère les références bibliographiques.

Les autres tables sont des tables explicatives (dictionnaires) ou des tables de correspondance.

5.1. Structure de la table TAXREF

TAXREF est un référentiel unique regroupant plantes, champignons, animaux, bactéries, etc., des territoires de la République française. Il n'est pas composé d'une multitude de référentiels juxtaposés.

En conséquence, la structure de l'information disponible doit être simple et souple pour répondre aux exigences de la taxonomie et de la nomenclature des êtres vivants. Afin de simplifier la complexité inhérente à l'exercice, le choix est le suivant :

- une seule table gère l'ensemble des noms scientifiques, les autres tables n'étant que des tables explicatives (dictionnaires) ;

- quel que soit leur rang taxonomique, tous les noms sont traités de façon équivalente (dans le même champ). Un autre champ contient le rang taxonomique.

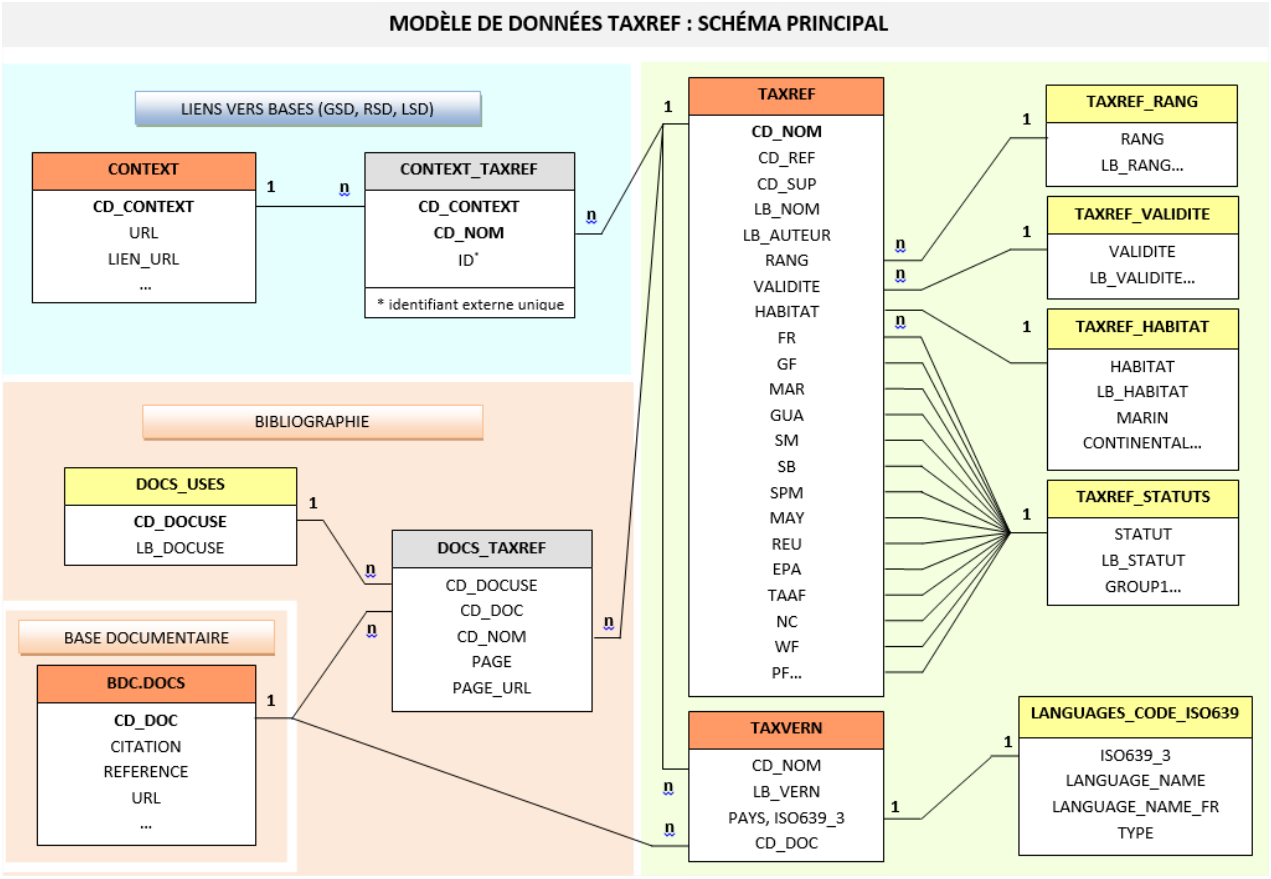


Figure 7: Modèle de données TAXREF. La base documentaire est accessible sous forme de web-service (DOCS-Web). En orange : références ; en jaune : dictionnaires ; en gris : correspondances.

5.1.1. Liste des champs

La liste des champs de la table TAXREF est fixée dans le Tableau 2.

Tableau 2 : Liste des champs de la table TAXREF (en gras la clé unique).

NOM	TYPE	COMMENTAIRES
CD_NOM	**Entier long**	**Identifiant unique du nom scientifique**
CD_SUP	Entier long	Identifiant (CD_NOM) du taxon supérieur
CD_REF	Entier long	Identifiant (CD_NOM) du taxon de référence (nom retenu)
CD_BA	Entier long	Identifiant (CD_NOM) du nom original (basionyme / protonyme)
LB_NOM	Texte	Nom scientifique du taxon (sans l'autorité)
LB_AUTEUR	Texte	Autorité du taxon (Auteur, année, gestion des parenthèses)
NOMENCLATURAL_COMMENT	Texte	Commentaires nomenclaturaux (inclus dans LB_AUTEUR avant la v18)
NOM_COMPLET	Texte	Combinaison des champs pour donner le nom complet (~LB_NOM+" " +LB_AUTEUR)
RANG	Varchar2(4)	Rang taxonomique (clé vers TAXREF_RANG)
VALIDITE	Varchar2(4)	Type de validité du nom scientifique complet (clé vers TAXREF_VALIDITE)
HABITAT	Entier court	Code de l'habitat (clé vers TAXREF_HABITATS)

NOM	TYPE	COMMENTAIRES
FR	Varchar2(1)	Statut biogéographique en France hexagonale et de Corse (clé vers TAXREF_STATUTS)
GF	Varchar2(1)	Statut biogéographique en Guyane (clé vers TAXREF_STATUTS)
MAR	Varchar2(1)	Statut biogéographique à la Martinique (clé vers TAXREF_STATUTS)
GUA	Varchar2(1)	Statut biogéographique à la Guadeloupe (clé vers TAXREF_STATUTS)
SM	Varchar2(1)	Statut biogéographique à Saint-Martin (clé vers TAXREF_STATUTS)
SB	Varchar2(1)	Statut biogéographique à Saint-Barthélemy (clé vers TAXREF_STATUTS)
SPM	Varchar2(1)	Statut biogéographique à Saint-Pierre-et-Miquelon (clé vers TAXREF_STATUTS)
MAY	Varchar2(1)	Statut biogéographique à Mayotte (clé vers TAXREF_STATUTS)
EPA	Varchar2(1)	Statut biogéographique aux Îles Éparses (clé vers TAXREF_STATUTS)
REU	Varchar2(1)	Statut biogéographique à la Réunion (clé vers TAXREF_STATUTS)
SA	Varchar2(1)	Statut biogéographique aux îles subantarctiques (clé vers TAXREF_STATUTS)
TA	Varchar2(1)	Statut biogéographique en Terre Adélie (clé vers TAXREF_STATUTS)
TAAF	Varchar2(1)	Statut biogéographique aux TAAF, calculé à partir des champs SA et TA (clé vers TAXREF_STATUTS)
PF	Varchar2(1)	Statut biogéographique en Polynésie française (clé vers TAXREF_STATUTS)
NC	Varchar2(1)	Statut biogéographique en Nouvelle-Calédonie (clé vers TAXREF_STATUTS)
WF	Varchar2(1)	Statut biogéographique à Wallis-et-Futuna (clé vers TAXREF_STATUTS)
CLI	Varchar2(1)	Statut biogéographique à Clipperton (clé vers TAXREF_STATUTS)
INPN	Booléen	Diffusion publique sur l'INPN (true/false)

Les valeurs prises par le champ RANG sont fixées par la table TAXREF_RANG (Tableau 3).

Les valeurs prises par le champ VALIDITE sont fixées par la table TAXREF_VALIDITE, non décrite ici car en construction.

Les valeurs prises par le champ HABITAT sont fixées par la table TAXREF_HABITATS (

Tableau 4).

Les valeurs prises par les champs FR, GF, MAR, GUA, SM, SB, SPM, MAY, EPA, REU, TAAF, SA, TA, PF, NC, WF et CLI sont fixées par la table TAXREF_STATUTS (voir Tableau 1 page 21).



Figure 8 : Situation géographique des champs d'occurrence biogéographique.

Tableau 3: Valeurs de la table TAXREF_RANG. L'ensemble des rangs disponibles sont dorénavant diffusés sur l'INPN. Les rangs en gris correspondent aux rangs pris en compte par le CD_TAXSUP afin de disposer d'une classification simplifiée (ajouts de la tribu et de la sous-famille depuis la version 11).

RANG	DETAIL	RANG	DETAIL	RANG	DETAIL
Dumm	Domaine	OR	Ordre	AGES	Agrégat
SPRG	Super-Règne	SBOR	Sous-Ordre	ES	Espèce
KD	Règne	IFOR	Infra-Ordre	SSES	Sous-Espèce
SSRG	Sous-Règne	PVOR	Parv-Ordre	NAT	Natio
IFRG	Infra-Règne	SCO	Section	VAR	Variété
PH	Phylum/Embranchement	SSCO	Sous-Section	SVAR	Sous-Variété
SBPH	Sous-Phylum	SPFM	Super-Famille	FO	Forme
IFPH	Infra-Phylum	FM	Famille	SSFO	Sous-Forme
DV	Division	SBFM	Sous-Famille	RACE	Race
SBDV	Sous-Division	SPTR	Super-Tribu	CAR	Cultivar
SPCL	Super-Classe	TR	Tribu	AB	Abberatio
CLAD	Clade	SSTR	Sous-Tribu		
CL	Classe	GN	Genre		
SBCL	Sous-Classe	SSGN	Sous-Genre		
IFCL	Infra-Classe	SC	Section		
PVCL	Parv-Classe	SBSC	Sous-Section		
SPOR	Super-Ordre	SER	Série		
COH	Cohorte	SSER	Sous-Série		

Tableau 4 : Valeurs de la table TAXREF_HABITATS.

HABITAT	DESCRIPTION	REMARQUES
1	Marin	Espèces effectuant l'intégralité de leur cycle de vie en milieu marin. Les espèces vivant en mer mais pouvant occasionnellement supporter les eaux douces entrent dans cette catégorie (exemple de la sardine).
2	Eau douce	Espèces effectuant l'intégralité de leur cycle de vie en eau douce. Les espèces vivant en eau douce mais pouvant occasionnellement supporter les eaux saumâtres entrent dans cette catégorie.
3	Terrestre	Espèces vivant uniquement en milieu terrestre.
4	Marin & Eau douce	Espèces pouvant être présentes en eau douce et en mer de par leur cycle de vie diadrome pour les organismes amphihalins, ou par tolérance aux fortes variations de salinités pour les organismes euryhalins.
5	Marin & Terrestre	Espèces effectuant une partie de leur cycle de vie en mer et l'autre partie à terre (cas des pinnipèdes, des tortues et des oiseaux marins par exemple).
6	Eau saumâtre	Espèces vivant exclusivement en eau saumâtre.
7	Continental (terrestre et/ou eau douce)	Espèces continentales (non marines) dont on ne sait pas si elles sont terrestres et/ou d'eau douce (taxons provenant de Fauna Europaea).
8	Continental (terrestre ET eau douce)	Espèces terrestres effectuant une partie de leur cycle en eau douce (odonates par exemple), ou fortement liées au milieu aquatique (loutre par exemple).

5.1.2. Clé primaire

L'idéal serait de faire une clé unique sur les champs LB_NOM et LB_AUTEUR. Les codes de nomenclature existent justement pour qu'il y ait une relation non ambiguë, de 1 à 1, entre un nom scientifique et une espèce biologique.

Or cela n'est possible car :

- TAXREF fait appel à deux codes de nomenclature distincts, celui de botanique et celui de zoologie, et qu'en conséquence des doublons sont possibles (et ils existent) ;
- il existe des problèmes nomenclaturaux d'homonymie non encore résolus par les systématiciens ;
- compte-tenu de l'existence de données, et de la nécessité de pouvoir réattribuer ces données à des entités biologiques évoluant au gré des travaux taxonomiques, il est parfois nécessaire de dupliquer un nom.

En conséquence, une clé primaire numérique a été créée : CD_NOM.

Afin de coller au plus près du monde parfait où le nom scientifique serait la clé primaire, il est institué une pérennité du couple identifiant unique / nom scientifique.

Cette méthode de gestion représente un avantage considérable quant au suivi des modifications.

Rares exceptions : il existe quelques exceptions qui sont opérées afin que les changements nomenclaturaux n'affectent pas les concepts taxonomiques visés dans les données (voir 2.2).

5.1.3. Contraintes d'intégrité

Le champ CD_REF ne doit pas être *null* et le champ CD_SUP ne doit pas être *null* pour les taxons de référence. Ils doivent désigner un CD_NOM qui existe (clé étrangère).

5.2. Gestion de la synonymie

La gestion de la synonymie s'opère via le champ CD_REF qui correspond au CD_NOM du nom de référence. Si le taxon considéré fait référence, alors CD_REF = CD_NOM. Le CD_REF doit correspondre à un taxon de référence (synonymie sur un niveau).

5.2.1. Cas particulier de la synonymie d'usage

Compte tenu de l'objectif premier de TAXREF qui est la gestion de données, le lien entre CD_NOM et CD_REF peut amener à effectuer des changements relativement conséquents dans TAXREF du moment où l'on veut que le CD_NOM continue de désigner la même entité biologique. Par exemple, le taxon que l'on désignait autrefois par *Lacerta viridis* (Laurenti, 1768) (CD_NOM = 77686) s'est vu redistribué en plusieurs espèces, et toutes les populations françaises doivent être maintenant nommées *Lacerta bilineata* Daudin, 1802. En conséquence, et compte tenu du fait qu'il s'agit d'un référentiel national, le CD_NOM 77686 s'est vu réattribuer le nom *Lacerta viridis* auct. non (Laurenti, 1768) et renvoie vers *Lacerta bilineata*, tandis que *Lacerta viridis* (Laurenti, 1768) s'est vu réattribuer un nouveau CD_NOM (non diffusé puisque non présent en France).

Cette gestion n'est plus possible dès lors que plusieurs espèces issues de la scission sont présentes sur le territoire, auquel cas le référentiel ne peut rien apporter : soit sa précision géographique est insuffisante, soit il faut de toute manière retourner au spécimen pour l'identification. Dans cette situation, on rajoute au référentiel les nouveaux noms et on laisse le CD_NOM préexistant dans son état. C'est notamment le cas des goujons (*Gobio* spp.) et des chabots (*Cottus* spp.) de France.

5.3. Gestion de la classification

La gestion de la hiérarchie taxonomique s'opère via le champ CD_SUP qui correspond au CD_NOM du taxon parent (par exemple le genre ou le sous-genre pour l'espèce).

En toute logique, le CD_SUP doit pointer vers un taxon de référence. Il existe quelques rares exceptions, comme le cas d'un genre qui tomberait dans la synonymie d'un autre mais pour lequel certaines espèces initialement classées dans ce genre n'auraient pas été révisées, qui conservent donc le même nom et pointent en conséquence vers un synonyme.

En toute logique également, le CD_SUP d'une espèce devrait pointer vers le genre qui compose son nom (avec le sous-genre en intermédiaire ou non). Il existe encore quelques rares exceptions, comme par exemple des escargots du Pacifique décrits dans le genre « *Helix* » mais dont aucune nouvelle combinaison n'a pour l'instant été publiée. Dans ce cas, on fait pointer le CD_SUP vers la famille à laquelle l'espèce se rattache, en l'occurrence les Charopidae, sans passer par le genre qui de toute évidence est faux.

5.4. Liens vers les bases de données sources

La table CONTEXT référence les bases de données source utilisées pour TAXREF et permet de faire le lien vers celles-ci (Tableau 5).

Tableau 5 : Liste des champs de la table CONTEXT (en gras la clé unique).

NOM	TYPE	COMMENTAIRES
CD_CT	**Entier long**	**Identifiant unique de l'enregistrement**
CT_NAME	Texte	Acronyme de la base
CT_URL	Texte	Adresse internet (URL) de la base
CT_TYPE	Texte	Portée de la base (mondiale, régionale, locale)
CT_TITLE	Texte	Nom de la base
CT_AUTHORS	Texte	Auteurs de la base
URL_SP	Texte	URL pour accéder aux pages des taxons (à compléter avec l'identifiant du taxon)

Les tables CONTEXT et TAXREF sont liées via la table de correspondance CONTEXT_TAXREF (Tableau 6) par un lien 1-n. Cette table permet de stocker l'identifiant unique du nom scientifique dans la base source. Chaque identifiant ne peut être attribué qu'une seule fois à un CD_NOM de TAXREF (index unique), ce qui assure une **équivalence stricte entre TAXREF et la base source**.

Tableau 6 : Liste des champs de la table CONTEXT_TAXREF (en gras la clé unique).

NOM	TYPE	COMMENTAIRES
CD_CONTEXT_TAXREF	**Entier long**	**Identifiant unique de l'enregistrement**
CD_NOM	Entier long	Identifiant unique du nom scientifique dans TAXREF (clé vers TAXREF)
CD_CT	Entier long	Identifiant unique de la base source (clé vers CONTEXT)
CT_SP_ID	Texte	Identifiant unique du nom scientifique dans la base source

5.5. Références bibliographiques

La table DOCS constitue la base documentaire de l'INPN où sont stockées toutes les références bibliographiques utilisées dans le système d'information (Tableau 7).

Tableau 7 : Liste des principaux champs de la table DOCS (en gras la clé unique).

NOM	TYPE	COMMENTAIRES
CD_DOC	**Entier long**	**Identifiant unique de l'enregistrement**
CITATION	Texte	Citation du document (auteur(s) et année de publication)
REFERENCE	Texte	Référencement complet du document
URL	Texte	URL d'accès au document, de préférence un permalien
DIFFUSION	Booléen	Type de diffusion (aucune, limitée, publique)

Les références bibliographiques de TAXREF sont utilisées à différentes fins : description originale du nom, illustration originale, source de la nouvelle combinaison, source de la synonymie, statut biogéographique du taxon pour un territoire donné, etc. Ces différentes utilisations sont stockées dans la table DOCS_USES, et appelées dans la table DOCS_TAXREF qui lie les références bibliographiques aux noms de TAXREF (Tableau 8).

Tableau 8 : Liste des champs de la table DOCS_TAXREF (en gras la clé unique).

NOM	TYPE	COMMENTAIRES
CD_DOC_TAXREF	**Entier long**	**Identifiant unique de l'enregistrement**
CD_NOM	Entier long	Identifiant unique du nom scientifique (clé vers TAXREF)
CD_DOC	Entier long	Identifiant unique de la référence bibliographique (clé vers DOCS)
CD_DOCUSE	Entier long	Utilisation qui est faite de la référence (clé vers DOCS_USES)
PAGE	Entier court	Page où le nom scientifique apparaît
URL_PAGE	Texte	URL de la page où le nom scientifique apparaît

5.6. Noms vernaculaires

5.6.1. Structure de la table TAXVERN

Tableau 9 : Liste des champs de la table TAXVERN (en gras la clé unique).

NOM	TYPE	COMMENTAIRES
CD_VERN	**Entier long**	**Identifiant unique de l'enregistrement**
CD_NOM	Entier long	Identifiant unique du nom scientifique (clé vers TAXREF)
CD_REF	Entier long	Identifiant (CD_NOM) du taxon de référence (nom retenu)
NOM_VERN	Texte	Nom(s) vernaculaire(s) du taxon
ISO639-3	Texte	Code de la langue dans la norme internationale ISO 639-3 (clé vers LANGUAGES_ISO639-3)
PAYS	Texte	Pays d'utilisation de l'enregistrement
CD_DOC	Entier long	Identifiant unique de la référence bibliographique (clé vers DOCS)
NOM_CITE	Texte	Nom scientifique associé au(x) nom(s) vernaculaire(s)
COMPTE	Entier court	Nombre de noms vernaculaires de l'enregistrement (champ calculé)

La table TAXVERN (Tableau 9) permet la gestion des noms vernaculaires dans différentes langues et dialectes (nouveauté version 12), l'objectif étant le français, l'anglais ainsi que l'ensemble des langues et dialectes locaux parlés dans les différents territoires français. Ces langues suivent la norme internationale ISO 639-3 qui établit des codes à 3 lettres pour l'ensemble des langues vivantes, mortes et anciennes (http://www.iso.org/iso/fr/home/standards/language_codes.htm). Depuis la v12, **71**

langues et dialectes différents sont utilisés pour traduire les noms vernaculaires au sein du territoire français.

Un pays d'utilisation est associé à la langue, ce qui permet de gérer les noms locaux. Ainsi, TAXVERN permet de distinguer le fait que le nom français pour *Ficus nymphaeifolia* est « Figuier grande feuille » à la Guadeloupe et « Figuier blanc » à la Martinique.

Le champ NOM_VERN peut contenir plusieurs noms vernaculaires (séparés par des virgules). Le premier nom vernaculaire de la liste est considéré comme le nom de référence pour la langue donnée. Les noms vernaculaires ne doivent pas être précédés d'un déterminant (à mettre entre parenthèses après le nom si nécessaire). Dans le cas où il existe un nom vernaculaire spécifique pour le mâle et la femelle, il est précisé entre parenthèses après le nom (exemple de *Lucanus cervus* (Linnaeus, 1758) : « Cerf-volant (mâle), Biche (femelle) »).

5.6.2. Gestion des héritages de noms vernaculaires

Lorsqu'un nom devient synonyme d'un autre, et lorsque ce nom possédait un(des) nom(s) vernaculaire(s), la transmission de ce(s) nom(s) vernaculaire(s) vers le nouveau nom de référence n'est pas automatique, de façon à ne pas inclure d'erreurs éventuelles. Par contre, il existe des tests de contrôle pour repérer d'éventuelles incohérences dans les noms vernaculaires.

6. Gestion du référentiel

6.1. Historique et suivi des modifications

Toutes les modifications effectuées sur le référentiel (tables TAXREF et TAXVERN) sont enregistrées dans une table d'audit au travers de déclencheurs informatiques avec les informations suivantes (Figure 9):

- le nom de la table et du champ concernés (TABLE_NAME et COLUMN_NAME)
- l'ancienne valeur du champ (OLD)
- sa nouvelle valeur (NEW)
- par qui et quand la modification a été effectuée (ID_UTILISATEUR et DATE)
- pour quel CD_NOM de TAXREF.

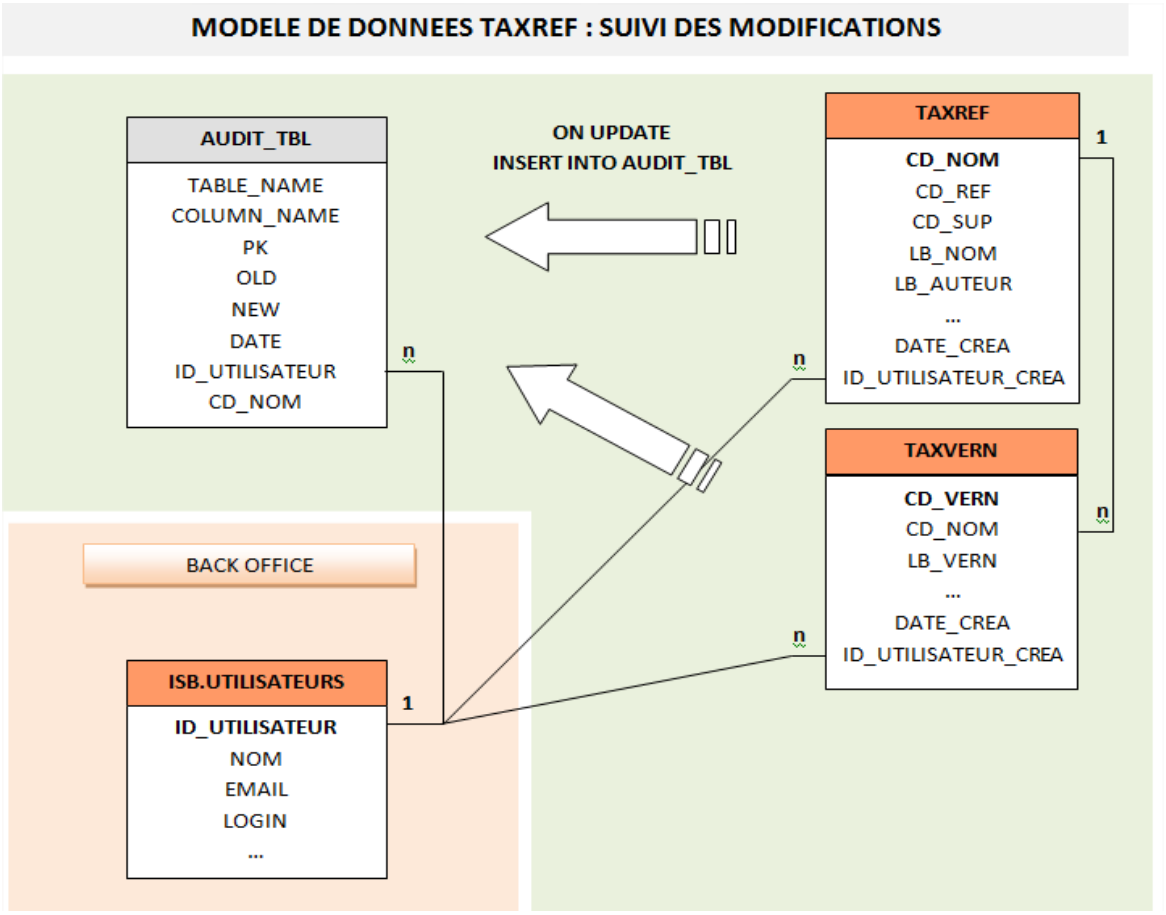


Figure 9 : Modèle de données de TAXREF : schéma du suivi des modifications.

6.2. Gestion des ressources bibliographiques (DOCS-Web)

Tous les ajouts ou modifications effectués sur le référentiel (tables TAXREF et TAXVERN) sont, dans la mesure du possible, justifiés par des ressources bibliographiques. Ces ressources sont stockées et gérées dans la base documentaire de l'INPN (tables DOCS, Figure 7), accessible à travers l'application DOCS-Web (https://inpn.mnhn.fr/docs-web) laquelle permet en outre le partage des ressources (Figure 10).

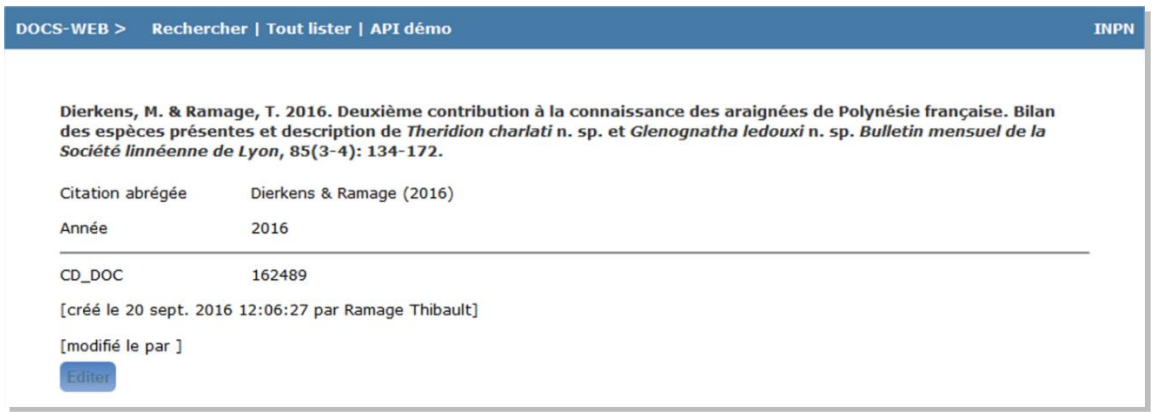


Figure 10 : Référence bibliographique dans DOCS-Web.

L'implémentation du flux RSS de Web of Knowledge (Thomson Reuters) au niveau de DOCS-web permet un contrôle des connaissances et une intégration aux grandes bases de données internationales (Figure 11).



Figure 11 : Flux RSS de Web of Knowledge mis en œuvre dans DOCS-Web.

6.3. Application TAXREF-Web

TAXREF est issu de la collaboration entre de nombreux scientifiques, qu'il faut faire travailler sur un seul et même objet, la table référentielle. À cette fin, une application Web a été mise en ligne afin que chaque expert en taxonomie travaille et bénéficie du travail des autres experts en temps réel (Figure 12).

L'accès en mode « édition » à cette application est contrôlé et se fait par identification avec mot de passe. Chaque expert possède un domaine d'expertise identifié par un taxon et tous ses descendants. Ainsi, un spécialiste des papillons de Guyane aura accès à l'ensemble des papillons de TAXREF, ceci dans le souci du respect de la consolidation nationale.

Afin de faciliter les échanges entre experts, une fonctionnalité « email » permet de générer automatiquement un email pré-rempli avec le nom du taxon consulté dans TAXREF-web, les liens vers les différentes bases source concernées et de le destiner à l'ensemble des experts / coordinateurs de ces bases conjointement pour discussion.

Depuis la version 16 de TAXREF, l'application est **librement accessible en consultation** afin que l'ensemble de la communauté puisse disposer des nouveaux noms en temps réel, sans attendre le versionnage annuel, et ait accès à l'ensemble des informations concernant les taxons (notamment les notes, historique de suivi du taxon, etc.).

Pour consulter TAXREF-Web : https://taxref.mnhn.fr

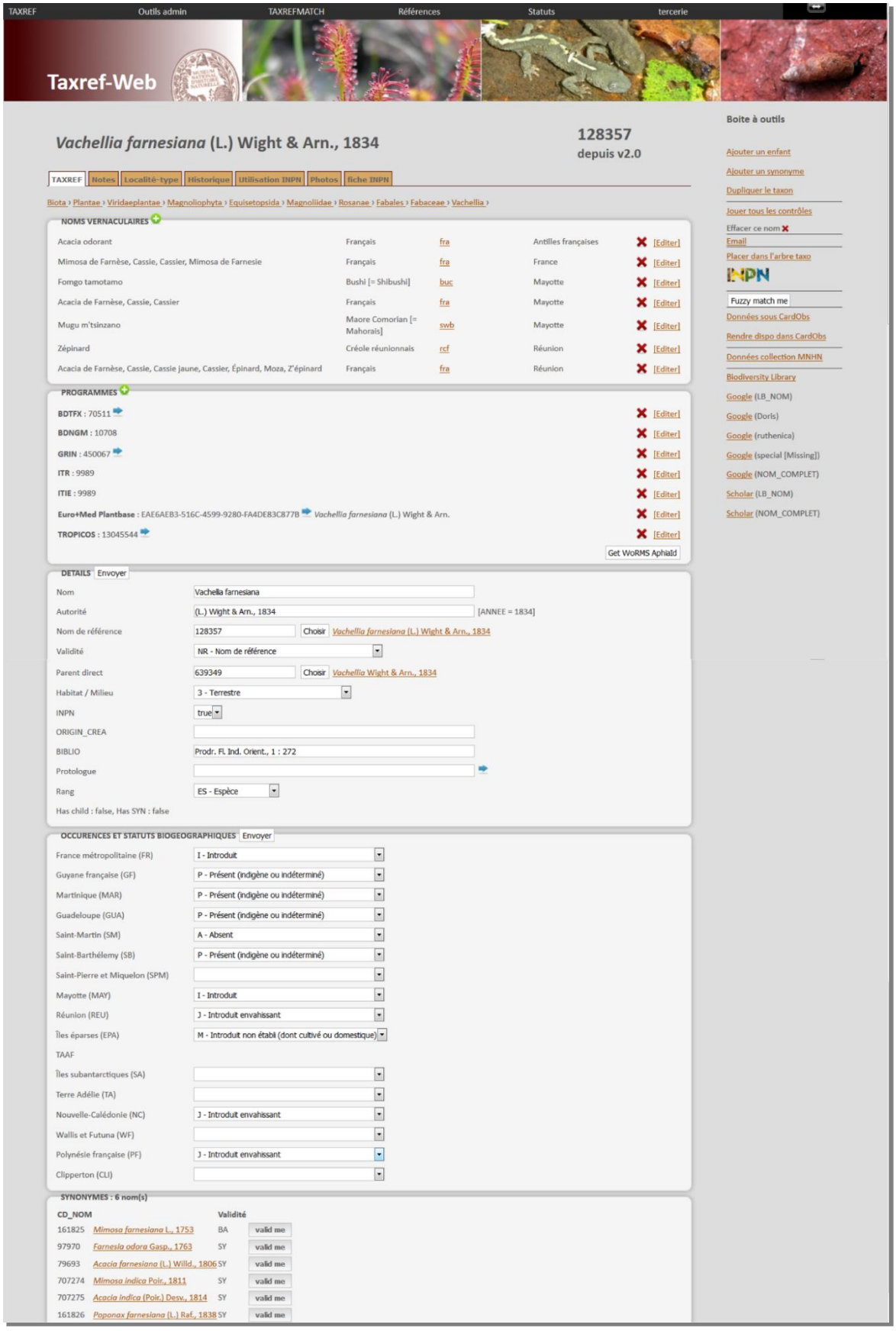


Figure 12 : Détails d'un taxon dans TAXREF-Web (onglet principal « TAXREF »).

Le suivi des modifications est visible dans l'application : pour chaque nom du référentiel, un historique complet est fourni (date de création du nom, par qui, quelles modifications y ont été apportées, quand, etc.). Chaque expert peut ainsi suivre les changements réalisés par les autres experts travaillant sur TAXREF-Web, et éventuellement l'invalider (Figure 13).

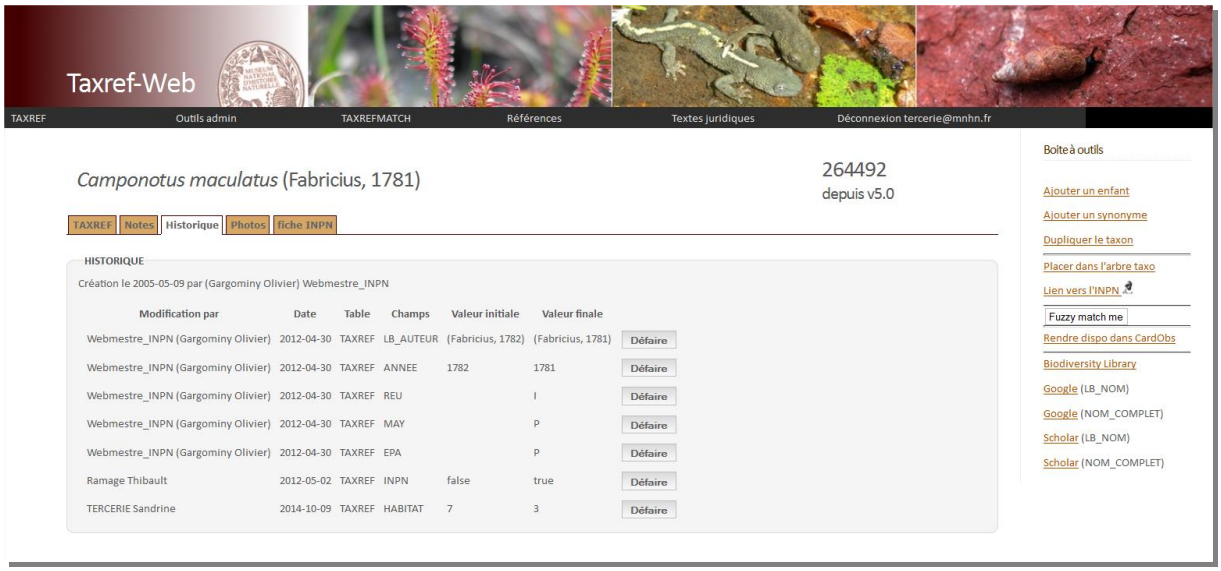


Figure 13 : Historique et suivi des modifications dans TAXREF-Web (onglet « Historique »).

TAXREF-Web permet également d'associer des références bibliographiques à chaque nom du référentiel, en indiquant l'utilisation qui est faite de chaque référence (description originale du nom, source de la synonymie, source pour le statut biogéographique à Mayotte, etc.) (Figure 14). L'ajout de notes, associées ou non à une référence bibliographique, permet d'apporter des informations complémentaires et de justifier les choix adoptés dans TAXREF.

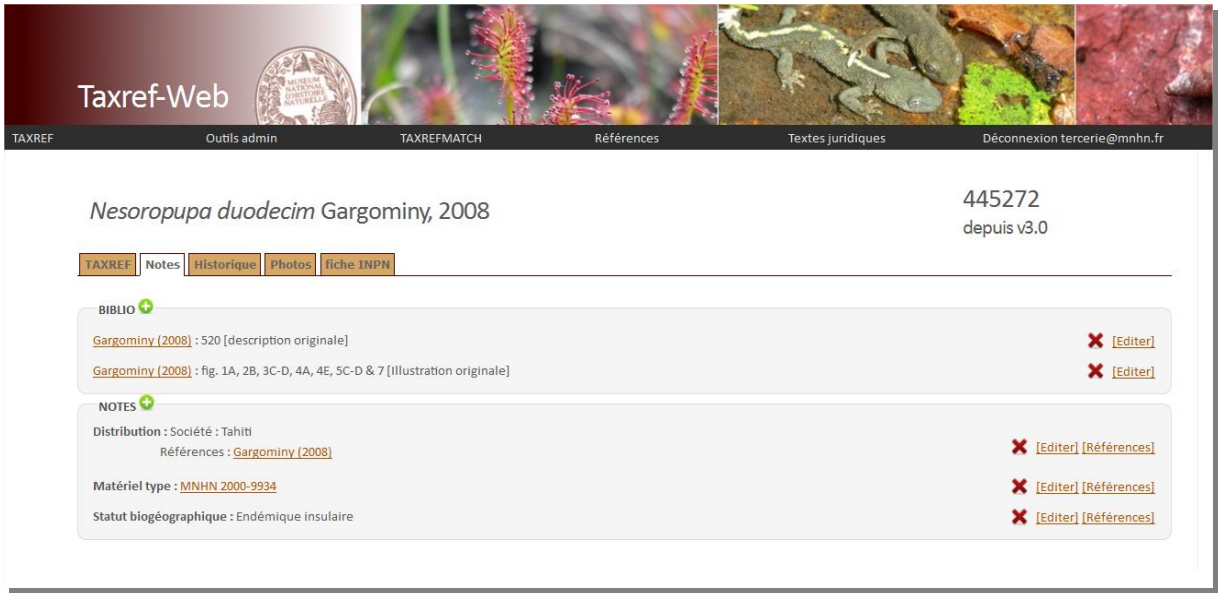


Figure 14 : Notes et gestion des ressources bibliographiques dans TAXREF-Web (onglet « Notes »).

6.4. Contrôles effectués sur le référentiel

Plus de 200 contrôles sont effectués de façon à assurer la cohérence et l'intégrité du référentiel : écriture correcte des noms scientifiques par rapport aux codes de nomenclature, respect de la logique de la classification, recherche de combinaisons génériques non décelées, etc.

Ces contrôles sont effectués à la volée lors de la création ou la modification de noms dans TAXREF-Web, mais ils font surtout partie d'un travail spécifique avant chaque nouvelle version.

Mettant en évidence certains problèmes nomenclaturaux non encore résolus par les systématiciens, ils ne peuvent pas tous recevoir un traitement définitif dans TAXREF.

PARTIE 2 : DIFFUSION

7. Téléchargement

Le référentiel taxonomique TAXREF est publié sur le site de l'INPN en libre accès. Il est disponible intégralement depuis la version 2.0.

https://inpn.mnhn.fr/telechargement/referentielEspece/referentielTaxo

L'utilisateur a le choix entre :

- le téléchargement du référentiel intégral, sous forme de fichier texte zippé ;
- le téléchargement de parties du référentiel avec la possibilité de filtrer taxonomiquement (par embranchement, classe, ordre ou famille), géographiquement (chacun des territoires de la République française) ou écologiquement (marin, continental).

Le fichier archive Zip comporte huit fichiers :

- TAXREFvXX.txt (UTF-8), le référentiel en lui-même,
- TAXVERNvXX.txt (UTF-8), le référentiel des noms vernaculaires,
- TAXREF_CHANGES.txt (UTF-8), tableau décrivant l'ensemble des modifications depuis la version précédente,
- TAXREF_LIENS.txt (UTF-8), tableau permettant un lien unique vers les bases de données sources,
- CDNOM_DISPARUS.txt, tableau listant les CD_NOM de niveau spécifique anciennement diffusés mais qui ne le sont plus dans la dernière version, avec les causes de leur disparition ou les CD_NOM de remplacement dans le cas des doublons.
- trois fichiers (UTF-8) habitats_note.csv, rangs_note.csv et statuts_note.csv qui permettent d'interpréter respectivement la colonne Habitat, la colonne Rang et les colonnes de statuts biogéographiques dans les territoires (FR, GF, etc.).

8. Structure du référentiel diffusé

Le référentiel TAXREF tel qu'il est diffusé (fichier TAXREFvXX.txt) est décrit dans le Tableau 10. Il reprend la table de travail telle que décrite précédemment (voir 5.1) avec les conditions suivantes :

- INPN = 'true' (filtre des taxons français[12]) ;
- Les noms vernaculaires français et anglais sont regroupés dans la table référentielle ;
- Le champ CD_TAXSUP propose une classification supérieure simplifiée aux rangs suivants : règne, embranchement, classe, ordre, famille, sous-famille et tribu. Le CD_TAXSUP est une valeur calculée à partir du CD_SUP. De cette façon, l'utilisateur a le choix d'utiliser une classification complète ou simplifiée.

Par ailleurs, afin d'avoir une lisibilité plus grande et des temps de requête plus courts, la classification supérieure est mise à plat dans les champs correspondants.

Tableau 10 : Liste des champs de la table TAXREF diffusée (en gras la clé unique).

NOM	TYPE	COMMENTAIRES
REGNE	Texte	Règne auquel le taxon appartient
PHYLUM	Texte	Embranchement auquel le taxon appartient
CLASSE	Texte	Classe à laquelle le taxon appartient
ORDRE	Texte	Ordre auquel le taxon appartient
FAMILLE	Texte	Famille à laquelle le taxon appartient
SOUS_FAMILLE	Texte	Sous-famille à laquelle le taxon appartient
TRIBU	Texte	Tribu à laquelle le taxon appartient
GROUP1_INPN	Texte	Regroupement vernaculaire (niveau 1)
GROUP2_INPN	Texte	Regroupement vernaculaire (niveau 2)
GROUP3_INPN	Texte	Regroupement vernaculaire (niveau 3)
CD_NOM	**Entier long**	**Identifiant unique du nom scientifique**
CD_TAXSUP	Entier long	Identifiant (CD_NOM) du taxon supérieur calculé dans la classification simplifiée (voir Tableau 3 pour les valeurs de rangs pris en compte)
CD_SUP	Entier long	Identifiant (CD_NOM) du taxon directement supérieur
CD_REF	Entier long	Identifiant (CD_NOM) du taxon de référence (nom retenu)
CD_BA	Entier long	Identifiant (CD_NOM) du nom original (basionyme / protonyme)
RANG	Texte	Rang taxonomique (lien vers TAXREF_RANG)
LB_NOM	Texte	Nom scientifique du taxon (sans l'autorité)
LB_AUTEUR	Texte	Autorité du taxon (Auteur, année, gestion des parenthèses)
NOMENCLATURAL_COMMENT	Texte	Commentaires nomenclaturaux (inclus dans LB_AUTEUR avant la v18)
NOM_COMPLET	Texte	Combinaison des champs précédents pour donner le nom complet (~LB_NOM+" " +LB_AUTEUR)
NOM_COMPLET_HTML	Texte	NOM_COMPLET avec balises HTML pour gérer les italiques dans le nom
NOM_VALIDE	Texte	Le NOM_COMPLET du CD_REF
NOM_VERN	Texte	Noms vernaculaires français
NOM_VERN_ENG	Texte	Noms vernaculaires anglais
HABITAT	Entier court	Code de l'habitat (clé vers TAXREF_HABITATS)

[12] Quelques espèces non présentes en France sont diffusées ; il s'agit d'espèces potentiellement envahissantes, présentes en pays frontaliers mais pas encore en France, qui demandent une surveillance.

NOM	TYPE	COMMENTAIRES
FR	Texte	Statut biogéographique en France hexagonale et Corse (clé vers TAXREF_STATUTS)
GF	Texte	Statut biogéographique en Guyane (clé vers TAXREF_STATUTS)
MAR	Texte	Statut biogéographique à la Martinique (clé vers TAXREF_STATUTS)
GUA	Texte	Statut biogéographique à la Guadeloupe (clé vers TAXREF_STATUTS)
SM	Texte	Statut biogéographique à Saint-Martin (clé vers TAXREF_STATUTS)
SB	Texte	Statut biogéographique à Saint-Barthélemy (clé vers TAXREF_STATUTS)
SPM	Texte	Statut biogéographique à Saint-Pierre-et-Miquelon (clé vers TAXREF_STATUTS)
MAY	Texte	Statut biogéographique à Mayotte (clé vers TAXREF_STATUTS)
EPA	Texte	Statut biogéographique aux Îles Éparses (clé vers TAXREF_STATUTS)
REU	Texte	Statut biogéographique à la Réunion (clé vers TAXREF_STATUTS)
SA	Texte	Statut biogéographique aux îles subantarctiques (clé vers TAXREF_STATUTS)
TA	Texte	Statut biogéographique en Terre Adélie (clé vers TAXREF_STATUTS)
TAAF	Texte	Statut biogéographique aux TAAF, calculé à partir des champs SA et TA (clé vers TAXREF_STATUTS)
PF	Texte	Statut biogéographique en Polynésie française (clé vers TAXREF_STATUTS)
NC	Texte	Statut biogéographique en Nouvelle-Calédonie (clé vers TAXREF_STATUTS)
WF	Texte	Statut biogéographique à Wallis-et-Futuna (clé vers TAXREF_STATUTS)
CLI	Texte	Statut biogéographique à Clipperton (clé vers TAXREF_STATUTS)
URL	Texte	Permalien TAXREF-Web = 'https://taxref.mnhn.fr/taxref-web/taxa/' + CD_NOM
URL_INPN	Texte	Permalien INPN = 'https://inpn.mnhn.fr/espece/cd_nom/' + CD_NOM. Uniquement pour les taxons de rangs spécifiques, renvoie vers le CD_REF

L'ensemble des noms vernaculaires associés à TAXREF sont disponibles dans le fichier TAXVERNvXX.txt dans le format décrit Tableau 11.

Tableau 11: Liste des champs de la table TAXVERN diffusée (en gras la clé primaire).

NOM	TYPE	COMMENTAIRES
CD_VERN	**Entier long**	**Identifiant unique de l'enregistrement dans TAXVERN**
CD_NOM	Entier long	Identifiant unique du nom scientifique (clé vers TAXREF)
LB_VERN	Texte	Liste de noms vernaculaires du taxon
NOM_VERN_SOURCE	Texte	Référence bibliographique associée à l'enregistrement (contient quelques balises html pour gérer les italiques)
LANGUE	Texte	Langue de l'enregistrement
ISO639_3	Texte	Code de la langue dans la norme internationale ISO 639-3 (clé vers LANGUAGES_ISO639-3)
PAYS	Texte	Pays d'utilisation de l'enregistrement

La correspondance entre les identifiants de TAXREF (CD_NOM) et les identifiants des bases externes est diffusée via le fichier TAXREF_LIENS.txt dans le format décrit

Tableau 12.

Tableau 12: Liste des champs de la table TAXREF_LIENS diffusée.

NOM	TYPE	COMMENTAIRES
CT_NAME	Texte	Acronyme de la base source
CT_URL	Texte	Adresse internet (URL) de la base source
CT_TYPE	Texte	Portée de la base source (mondiale (GSD), régionale (RSD), locale (LSD))
CT_TITLE	Texte	Nom de la base source
CT_AUTHORS	Texte	Auteurs de la base source
CD_NOM	Entier long	Identifiant unique du nom scientifique dans TAXREF
CT_SP_ID	Entier long	Identifiant unique du nom scientifique dans la base source
URL_SP	Texte	URL pour accéder aux pages des taxons dans la base source (facultatif)

9. Modifications depuis la version précédente

Une table détaillant les ajouts, retraits et modifications opérés entre une version et sa précédente est également fournie. Sa structure est détaillée dans le Tableau 13.

Les modifications sont fournies pour l'ensemble des noms diffusés, y compris la classification supérieure. Le champ CD_SUP (diffusé depuis la version 9.0) et les champs SA et TA (diffusés depuis la version 10.0) sont également pris en compte dans cette table.

Tableau 13 : Liste des champs de la table TAXREF_CHANGES.

NOM	COMMENTAIRES
CD_NOM	Code du nom (numéro de la ligne auquel le changement s'applique)
NUM_VERSION_INIT	Numéro de la version initiale
NUM_VERSION_FINAL	Numéro de la version finale
CHAMP	Champ auquel le changement s'applique
VALEUR_INIT	Valeur initiale du champ
VALEUR_FINAL	Valeur finale du champ
TYPE_CHANGE	Type du changement effectué ; peut prendre 3 valeurs {AJOUT, RETRAIT, MODIFICATION}

10. Cas des enregistrements (CD_NOM) qui disparaissent

Il est possible qu'un CD_NOM diffusé dans une version ne le soit plus ultérieurement. Si ce phénomène peut poser des problèmes d'intégrité des données, il s'agit cependant d'un mal nécessaire ; les raisons de non-diffusion sont regroupées en trois catégories :

Cas n°1. Doublons orthographiques, combinaisons erronées qui proviennent principalement d'erreurs lors de la consolidation de listes locales. Pour les doublons, il est jugé nécessaire de ne plus diffuser

ces noms afin de ne pas surcharger le référentiel, notamment pour les opérations de réconciliation. Dans ce cas, le CD_NOM est à supprimer définitivement et un CD_NOM de remplacement est proposé.

Cas n°2. Taxons diffusés à tort, ne respectant pas la méthodologie TAXREF (non présent et jamais signalé de France par exemple). Le CD_NOM est toujours potentiellement valable et peut éventuellement être rediffusé dans une version à venir.

Cas n°3. Noms ambigus dont le contenu taxonomique est indéfini. Il peut s'agir de *nomen nudum* ou *nomen dubium* au sens des codes de nomenclature, mais également d'erreurs faites lors de la constitution du référentiel TAXREF. Aucune solution ne peut être proposée, le CD_NOM est à supprimer définitivement et il faut retourner à la donnée.

L'ensemble de ces enregistrements est livré sous la forme d'une table CDNOMS_DISPARUS (Tableau 14) dépendant de la version en cours.

Tableau 14: Liste des champs de la table CDNOM_DISPARUS.

NOM	COMMENTAIRES
CD_NOM	Identifiant unique du nom scientifique dans TAXREF
PLUS_RECENTE_DIFFUSION	Numéro de la dernière version dans laquelle le CD_NOM a été diffusé
CDNOM_REMPLACEMENT	CD_NOM de remplacement pour les cas de doublons supprimés
CD_RAISON_SUPPRESSION	Code de la raison de suppression ; peut prendre 3 valeurs (voir les 3 cas ci-dessus)
RAISON_SUPPRESSION	Raison de la suppression ; peut prendre 3 valeurs (voir les 3 cas ci-dessus)

11. Référentiel en ligne sur l'INPN

LA VERSION UTILISÉE EN LIGNE SUR L'INPN CORRESPOND À LA VERSION EN COURS DIFFUSÉE DANS LES PAGES DE TÉLÉCHARGEMENT.

Les références bibliographiques associées aux taxons spécifiques et infra-spécifiques sont disponibles sur l'INPN au niveau des fiches espèces, dans l'onglet « Taxonomie ».

Sauf cas exceptionnel, le versionnage est fait annuellement en fin d'année.

12. API

Depuis la v12, TAXREF est également accessible via un web-service dont la documentation est disponible ici : https://taxref.mnhn.fr/taxref-web/api/doc.

Cette API regroupe les informations provenant du référentiel national **TAXREF** et de la **Base de connaissance sur les espèces** qui lui est associée. L'API fournit des réponses sur la nomenclature, la taxonomie, la répartition, les statuts (protections, menaces, etc.) et les interactions biologiques entre espèces. Elle permet également d'accèder à l'ensemble des sources bibliographiques utilisées dans le cadre de TAXREF et de l'INPN. Cette API est **restFul** autant que possible. Elle exprime les réponses au format **HAL** (**JSON Hypertext Application Language**, https://tools.ietf.org/html/draft-kelly-json-hal-08).

13. Historique des différentes versions de TAXREF diffusées

Voir https://taxref.mnhn.fr/taxref-web/stat/versions et https://taxref.mnhn.fr/taxref-web/stat/biodiv.

PARTIE 3 : UTILISATION

14. Utilisation du référentiel dans la gestion de données

14.1. Principes généraux

Le référentiel TAXREF est notamment conçu pour gérer l'évolution du statut taxonomique ou nomenclatural d'un nom dans les jeux de données. Afin que les mises à jour des différentes versions soient le plus facile possible, il faut que les gestionnaires de données prévoient l'architecture de leur base de façon à autoriser le remplacement complet d'une version du référentiel par une autre version plus récente. Il pourra alors être nécessaire de créer une table d'indexation des noms entre les jeux de données et TAXREF, afin d'éviter toute modification de la table TAXREF d'origine.

Un CD_NOM qui a été transmis une fois n'est jamais effacé. Par contre, il se peut que des noms diffusés dans une version antérieure de TAXREF ne le soient plus, suite à une impossibilité de lier ce nom à une espèce biologique. De tels cas sont rares et l'utilisateur pourra discuter de ces points particuliers en prenant contact avec l'équipe TAXREF, soit par mail (taxref@mnhn.fr), soit sur la page dédiée du site de l'INPN https://inpn.mnhn.fr/contact/contacteznous (Thématique : "Espèces, Référentiel taxonomique").

En ce qui concerne le stockage du nom d'espèce, il est hautement recommandé de :

1) stocker la chaine de caractère exacte à l'origine de la donnée
2) stocker le CD_NOM
3) stocker le CD_REF

En effet, le couple CD_NOM / CD_REF étant évolutif au cours de l'évolution de la taxonomie et des référentiels qui s'en font l'écho, il est indispensable de conserver la signification taxonomique du CD_NOM (le CD_REF qui, par construction désigne un nom de référence) au moment où il a été saisi.

Il est également possible de stocker la version de TAXREF sous laquelle la donnée a été saisie.

14.1.1. Possibilité de gestion multi-référentielle

Compte tenu du fait que TAXREF est en liaison unique avec d'autres référentiels, préférentiellement des GSD (WoRMS, FishBase, AlgaeBase, etc.), tout jeu de données géré avec ce référentiel est directement transposable en références sous TAXREF (Figure 15**Erreur ! Source du renvoi introuvable.Erreur ! Source du renvoi introuvable.**).

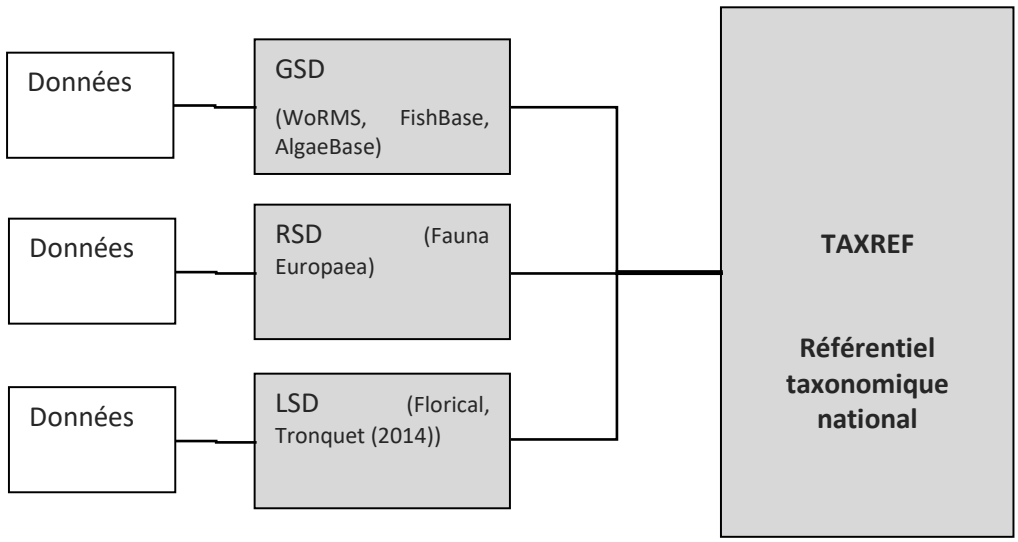


Figure 15: Possibilité de gestion multi-référentielle dans TAXREF.

14.2. Suivi des modifications

Lors de la mise à jour du référentiel, celui-ci gère automatiquement les changements taxonomiques pour peu que l'on opère les traitements sur le CD_REF.

Il peut cependant s'avérer indispensable de contrôler les changements entre le CD_REF stocké dans la table de données et la nouvelle version de TAXREF. À cette fin, chaque version du référentiel est livrée avec une table des changements opérés par rapport aux versions précédentes.

Le cas du « splittage » (scission) d'un taxon (l'espèce *Aa* est reconnue comme deux espèces distinctes *Aa* et *Ab*) doit retenir l'attention des gestionnaires de données, car ce cas n'est pas entièrement soluble via un référentiel taxonomique (voir 5.2 Gestion de la synonymie) : il faut réattribuer le bon nom aux spécimens identifiés, c'est-à-dire pour certains cas retourner dans la collection !

15. Recherche de correspondance (réconciliation)

Dans le cas où les noms d'espèces concernées par le jeu de données ne sont liés à aucun référentiel avec lequel il existe un rapport 1 à 1 avec TAXREF (WoRMS, AlgaeBase, Index Fungorum, etc.), il est nécessaire que le gestionnaire des données établisse une correspondance (réconciliation) entre le nom utilisé dans TAXREF et le nom utilisé dans son jeu de données.

15.1. Directives

La gestion des taxons dans une table de données est délicate : elle définit un lien entre une détermination par un individu X à un temps T, avec un nom dans un référentiel taxonomique établi à un temps T. En conséquence, la première recommandation consiste à rester au plus proche de l'auteur de la détermination (seul lui sait ce qu'il a voulu dire) et de stocker l'information selon les quelques directives suivantes :

- se procurer le plus d'informations possibles sur le nom utilisé : famille ou autre groupe parent, autorité, synonymie éventuelle ;
- établir la correspondance sur le maximum de champs possible : en plus du nom avec le LB_NOM, le nom d'auteur (LB_AUTEUR) est souvent indispensable pour séparer des synonymes ou des homonymes et tout groupe taxonomique supérieur doit être pris en compte (famille, ordre, classe, etc.) ;
- bien vérifier l'écriture des noms dans les tables de données. Avant toute correction à effectuer, dupliquer le champ du nom dans la table de données afin d'en conserver l'origine exacte (voir paragraphe suivant).

15.2. Pièges

Les erreurs suivantes doivent être rigoureusement évitées :

- une correspondance sur le genre uniquement : en nomenclature binominale ou trinominale, le nom cité le plus à droite est le plus précis pour désigner l'entité biologique que l'on observe ;
- une correspondance sur le nom d'espèce (deuxième partie du binom) sans disposer de la famille et de l'autorité : les homonymes de ce type sont légions.

15.3. Outil de réconciliation TAXREF_MATCH

L'outil TAXREF_MATCH a pour but d'aider à la mise en correspondance de listes taxonomiques avec TAXREF (réconciliation) ; il est accessible en ligne sur la plateforme TAXREF-web. Il permet le chargement de fichiers aux formats csv ou txt et retourne les résultats par mail.

https://taxref.mnhn.fr/taxref-match

15.4. Stockage des informations dans les tables de données

Un nom d'espèce attaché à une donnée (observation ou collection) résulte d'une détermination d'un spécimen faite par un individu X à un temps T. Il est donc fortement recommandé de stocker ces deux informations. On peut être amené à stocker plus d'informations pour préciser l'identification (guide ou référence bibliographique utilisée) soit par ligne de donnée, soit en métadonnées rapportées au jeu de données.

RÉSUMÉ

Après avoir rempli la planète internet d'informations, après avoir rempli nos carnets et nos ordinateurs de données de terrain, l'envie ou le besoin de partager la connaissance fait rapidement face à la nécessité de parler un langage commun et de disposer de référentiels. En matière de taxons biologiques, ces référentiels sont cruciaux tant le foisonnement des noms est à l'image de la diversité de ce qu'ils représentent.

Dans le cadre du Système d'Information sur la Nature et les Paysages (SINP), le Muséum national d'Histoire naturelle est en charge de la réalisation du référentiel national sur l'ensemble des organismes vivants dans les territoires de la République française. L'objectif du présent document est d'exposer le cadre méthodologique global, de la mise en œuvre à la diffusion, qui fonde ce référentiel TAXREF. En particulier, une composante essentielle réside dans le lien unique vers les référentiels globaux, respectant ainsi l'exigence de consolidation de la connaissance qui est le fondement de la nomenclature et la taxonomie. La gestion bibliographique, le suivi des modifications apportées au référentiel, les problématiques de gestion de données sont aussi des aspects fondamentaux mis en œuvre pour assurer la cohérence et la rigueur du référentiel et de l'utilisation qui en est faite.

Mots-clés : TAXREF, référentiel, nomenclature, taxonomie, espèces, méthodologie, diffusion.