



UNION REGIONALE
PAYS DE LA LOIRE

Note sur le protocole de validation en cours de fonctionnement pour les données de la base de données de l'URCPIE des Pays de la Loire

(<https://cpie.kollect.fr/index.php>)

Contexte

Suite à la mise en ligne à l'échelle régionale de la base « CPIE Kollect » en 2019 un protocole de validation manuelle provisoire a été mis en place dans l'attente de l'avancée de la mise en place de la plateforme SINP Régionale et l'implication des pôles thématiques dans les protocoles de validation associés. Cette note détaille le protocole de validation en cours depuis 2019.

La procédure de validation appliquée du « CPIE Kollect » a pour objectif d'attribuer un niveau de fiabilité à chaque donnée intégrée à l'outil dans le respect du cadre SINP. L'outil a été développé par le CEN Nouvelle-Aquitaine qui en assure la maintenance technique et est propulsé par Obs'Nat.

1. Vérification relative à la conformité SINP/TAXREF

Le développement de « CPIE Kollect » a été réalisé en visant une compatibilité maximale avec le SINP, ainsi sa conception a pris en compte :

- le respect du standard "Occurrences taxons" en vigueur
- l'utilisation du référentiel taxonomique TAXREF le plus récent
- la création et la gestion des métadonnées de cadres d'acquisition et jeux de données [développement en cours, encore à finaliser/renseigner de façon optimale]
- l'autogénération de l'identifiant unique permanent national (UUID)
- la possibilité de transfert automatique des métadonnées et données vers la plateforme régionale SINP à terme.

À ce titre, la conformité du CDNom est directe (sans préjuger à ce stade de la qualité de détermination de la donnée). Autrement dit, il n'est pas possible de saisir une espèce non présente au sein de Taxref.

2. Validation manuelle / automatique

Aucun protocole de validation automatique n'est aujourd'hui en place dans la base « CPIE Kollect ».

Ainsi, aucun processus automatisé, lié à des critères écologiques établis à partir de bases de connaissances (contrôles sur la date, la localisation, l'écologie, la difficulté d'identification...), n'est appliqué sur les données saisies. Ceci, tant faute de développement informatique dédié (nécessité de moyens) que de bases de connaissances disponibles (réflexions en cours au niveau des pôles thématiques).

La validation en cours dans la base « CPIE Kollect » est donc 100% manuelle et appliquée à toutes les données saisies.

La validation manuelle est réalisée principalement par :

- des salariés des CPIE des Pays de la Loire ayant des compétences taxonomiques dans certains groupes,
- des bénévoles du réseau des CPIE, spécialistes de groupes taxonomiques,
- des salariés ou bénévoles de structures partenaires acceptant d'aider à la qualité des données de la base « CPIE Kollect ».

La clé de validation est présentée ci-après.

Il est à noter que le niveau de connaissance naturaliste des observateurs est aussi estimé et pris en compte en amont de l'application du protocole via une prise de contact auprès de celui-ci lors de son inscription sur la base ou lors de l'application du processus de validation de ses premières données (dans l'objectif de savoir si la personne possède une notion de la diversité du vivant et des difficultés d'identifications inhérentes à la grande majorité des espèces).

3. Visibilité des espèces sur le site

Sur la base « CPIE Kollect », les fiches des espèces n'affichent que les espèces validées « *Probable* » à « *Certain - très probable* ».

Sur la page listant les observations, toutes les données présentées le sont avec le niveau de validation appliqué.

Clé de validation manuelle appliquée

*	Aucune compétence dans le groupe taxonomique considéré au sein des salariés du réseau des CPIE des Pays de la Loire ou de ses partenaires, <i>de facto</i>	« cas 6 - Non évalué »
**	Des compétences dans le groupe taxonomique considéré dans le réseau des CPIE ou ses partenaires	1
1	Une photo est disponible pour la donnée	2
1'	Aucune photo n'est disponible pour la donnée	8
2	Une photo est disponible pour la donnée et peut correspondre à l'espèce	3
2'	Une photo est disponible pour la donnée et ne correspond pas à l'espèce	« cas 4 - Invalide »
3	Une photo est disponible pour la donnée et <u>pouvant</u> correspondre à l'espèce avec une date en accord avec sa phénologie et une localisation correspondant aux connaissances biogéographiques	4
3'	Une photo est disponible pour la donnée et <u>pouvant</u> correspondre à l'espèce, mais avec une date ou une localisation ne correspondant pas aux connaissances locales. Demande de précision à l'auteur de la donnée et redémarrage au point 1 après retour de l'observateur	dans l'attente « cas 3- Douteux »
3''	Une photo est disponible pour la donnée et <u>pouvant</u> correspondre à l'espèce, mais avec une date ou une localisation ne correspondant pas aux connaissances locales. <u>L'auteur confirme la date/localisation</u>	4
4	Aucune espèce proche ne peut être confondue/pas de nécessité de passage sous loupe binoculaire ou microscope	« cas 1 - Certain - très probable ».
4'	Des espèces proches existent et /ou un examen sous loupe binoculaire est nécessaire pour une détermination rigoureuse	5
5	La détermination a été assurée par un spécialiste du groupe concerné	« cas 1 - Certain - très probable ».
5'	La détermination n'a pas été assurée par un spécialiste du groupe concerné	6
6	Des taxons proches sont connus des Pays de la Loire	« cas 5 - Non réalisable »
6'	Aucun taxon proche n'est connu des Pays de la Loire	7
7	Un ou des taxons proches sont connus sur les régions périphériques ou taxon(s) étendant régulièrement son(leurs) aire(s) géographique(s)	« cas 5 - Non réalisable »
7'	Aucun taxon proche n'est connu des régions périphériques et taxon(s) restreint(s) à une aire géographique éloignée sans dynamique particulière témoignée	« cas 2 - Probable »
8	La détermination a été assurée par un spécialiste du groupe concerné	« cas 1 - Certain - très probable »
8'	La détermination n'a pas été assurée par un spécialiste du groupe concerné	9
9	La détermination concerne une espèce connue des Pays de la Loire	11
9'	La détermination concerne une espèce inconnue des Pays de la Loire - Demande de précision à l'auteur de la donnée et redémarrage au point 1 après retour de l'observateur, dans l'attente « cas 3- Douteux »	10
10	La détermination concerne une espèce inconnue des Pays de la Loire, après retour de l'observateur, obtention d'une preuve de l'observation (photo, spécimen) et détermination par un spécialiste	« cas 1 - Certain - très probable »
10'	La détermination concerne une espèce inconnue des Pays de la Loire, après retour de l'observateur, absence de preuve de l'observation, espèce connue sur les régions périphériques ou taxon étendant régulièrement son aire géographique	« cas 5 - Non réalisable »
10''	La détermination concerne une espèce inconnue des Pays de la Loire, après retour de l'observateur, absence de preuve de l'observation, espèce inconnue des régions périphériques et taxon restreint à une aire géographique éloignée sans dynamique particulière témoignée	« cas 4 - Invalide »
11	La date d'observation est cohérente par rapport à la phénologie connue de l'espèce en Pays de la Loire	14
11'	La date d'observation n'est pas cohérente par rapport à la phénologie connue de l'espèce en Pays de la Loire Demande de précision à l'auteur de la donnée et redémarrage au point 1 après retour de l'observateur, dans l'attente « cas 3- Douteux » - Si confirmation par l'observateur :	12
12	La date d'observation n'est pas cohérente par rapport à la phénologie connue de l'espèce en Pays de la Loire, après retour de l'observateur, obtention d'une preuve de l'observation (photo, spécimen) et détermination par un spécialiste	« cas 1 - Certain - très probable ».
12'	La date d'observation n'est pas cohérente par rapport à la phénologie connue de l'espèce en Pays de la Loire, absence de preuve de l'observation, date proche	13
12''	La date d'observation n'est pas cohérente par rapport à la phénologie connue de l'espèce en Pays de la Loire, absence de preuve de l'observation, date très éloignée	14

13	Date proche et espèce aisée d'identification <u>et</u> commune en Pays de la Loire	« cas 2 - Probable »
13'	Date proche et espèce difficile d'identification ou date proche et espèce aisée d'identification, <u>mais</u> peu commune à rare en Pays de la Loire [<i>Principe : pas de « prise de risque » de création d'une donnée erronée pour une espèce rare</i>]	« cas 3 - Douteux »
14	Date très éloignée et espèce aisée d'identification <u>et</u> commune en Pays de la Loire	« cas 3 - Douteux »
14'	Date très éloignée et espèce difficile d'identification	« cas 4 - Invalide »
15	Identification aisée sur le terrain sans risque de confusion manifeste	« cas 1 - Certain - très probable »
15'	Identification difficile/délicate sur le terrain ou nécessité d'un examen sous loupe binoculaire/analyse de son	16
16	Le spécimen a été collecté et examiné sous loupe binoculaire/analyse de son réalisé	17
16'	Le spécimen n'a pas été collecté	18
17	L'observateur n'est pas aguerri à l'étude du groupe taxonomique concerné	19
17'	L'observateur est aguerri à l'étude de ce groupe taxonomique / a déjà vu les différentes espèces pouvant être confondues	« cas 1 - Certain - très probable »
18	L'observateur n'est pas aguerri à l'étude du groupe taxonomique concerné	« cas 3 - Douteux »
18'	L'observateur est aguerri à l'étude de ce groupe taxonomique / a déjà vu les différentes espèces pouvant être confondues	« cas 2 - Probable »
19	Une double-détermination par un spécialiste a été réalisée	« cas 1 - Certain - très probable »
19'	Pas de double détermination par un spécialiste	20
20	Espèce « commune » en Pays de la Loire	« cas 2 - Probable »
20'	Espèce « peu commune » à « Rare » en Pays de la Loire	« cas 5 - Non réalisable »

Rappel des niveaux de validité (SINP) utilisés

- 1 - Certain - très probable.

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 notamment 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.

- 2 – Probable

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.

- 3 – Douteux

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.

- 4 – 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.

- 5 - 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 :

- État des connaissances du taxon insuffisant
- Informations insuffisantes sur l'observation

- 6 - 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 de validation.