



# Nature du Gard

La revue de l'Observatoire  
du Patrimoine Naturel du Gard

A propos des observations naturalistes  
mises à disposition par l'INPN  
(Inventaire National du Patrimoine Naturel).

Proposé par Jean-Laurent Hentz

Document de discussion rédigé dans le cadre du projet  
d'Atlas des papillons de jour et zygènes du Gard

N°9 - novembre 2024

# Nature du Gard

La revue de l'Observatoire du Patrimoine Naturel du Gard, projet participatif et collaboratif initié, développé et animé par l'association Gard Nature.

La revue Nature du Gard a pour objectif la diffusion de savoirs relatifs à la faune, à la flore, à l'écologie des espèces et des espaces, et aux sciences appliquées à leur connaissance. Le secteur géographique ciblé est essentiellement le département du Gard.

**Directeur de la rédaction :** Charlotte Herry.

**Comité de rédaction :** collectif.

**Maquette :** Jean-Laurent Hentz.

**Secrétariat de rédaction :** Jean-Laurent Hentz.

**Ont collaboré à ce numéro :** .

**Relecture :** David Delon, Lou-Baptiste Reboul-Thibert, Solène Robert.

**Date de publication :** N°9 - novembre 2024.

Ce document est librement diffusé au format .pdf : il peut à ce titre être repris, en intégralité ou partiellement, pour tout usage privé ou public.

La citation recommandée est la suivante :

Hentz, J.-L. (2024) : A propos des observations naturalistes mises à disposition par l'INPN (Inventaire National du Patrimoine Naturel). Nature du Gard N°9 - novembre 2024.

Critiques, questions et remarques :

Gard Nature  
Mas du Boschet Neuf  
1059E, chemin du Mas du Consul  
30300 Beaucaire (France)  
Mail : [contact@gard-nature.com](mailto:contact@gard-nature.com)  
Téléphone : 04 66 02 42 67

# Sommaire

Préambule .....	1
A propos des champs et de la forme de la donnée .....	2
Exploitation de l'INPN .....	4
A propos des doublons .....	5
Inter-opérabilité des bases de données .....	6
A propos de l'observation originelle .....	7
A propos de l'observation originelle - validation .....	12
Identifiant unique et paternité de la donnée .....	14
Identifiant unique : bis ! .....	16
A propos des coordonnées géographiques .....	17
A propos des noms d'observateurs .....	18
Conclusion .....	19

# Préambule

L'association Gard Nature a lancé, officiellement en 2018, un projet d'Atlas des papillons de jour et zygènes du Gard.

Ce projet participatif (chacun peut participer individuellement), collectif (l'animation et l'organisation sont imaginées et mises en oeuvre par un groupe ouvert à toutes les bonnes volontés) et collaboratif (toute structure est invitée à participer en tant que gestionnaire d'une base de données ou avec une capacité de prospection) s'appuie notamment sur l'aboutissement d'un autre projet : le Guide photographique des papillons de jour et zygènes de France, initié en 2014, abouti en octobre 2022.

Dès sa création en 2003 Gard Nature a prôné la libre circulation des données naturalistes.

Cette idée est renforcée en 2005 par la création de l'ONEM (Observatoire Naturaliste des Ecosystèmes Méditerranéens), en pleine période de balbutiement des outils de saisie dynamique sur Internet.

L'Observatoire du Patrimoine Naturel du Gard (OPNG) est lancé en 2006, avec un module de saisie cartographique arrivant en 2010. Il s'affiche comme un projet de sciences participatives.

Cette même année, Charlotte Herry et Jean-Laurent Hentz, tous deux animateurs de l'association Gard Nature, vont rencontrer de nombreux partenaires potentiels pour leur proposer de collaborer à l'Observatoire. Conservatoire des Espaces Naturel, Parc National des Cévennes, Conservatoire Botanique Méditerranéen de Porquerolles, Office National des Forêts, certains bureaux d'études sont ainsi sollicités après présentation détaillée du projet. Sans la moindre décision de la part de ces structures.

Plus tard, des échanges ponctuels ou plus durables ont pu être mis en place avec le Syndicat Mixte des Hautes Vallées Cévenoles, l'EPTB Gardons, le Syndicat Mixte de la Camargue Gardoise ainsi que Nîmes Métropole.

A partir de 2010-2011 se met en place la politique publique de la France en matière de mise en commun et d'accès aux données naturalistes, à travers les plateformes régionales du SINP (Système d'Information sur la Nature et le Paysage) et l'INPN : l'Inventaire National du Patrimoine Naturel<sup>1</sup>. Si le développement du SINP du Languedoc-Roussillon, devenu SINP de l'Occitanie, est particulièrement opaque et non conforme à l'attendu (un citoyen lambda ou une collectivité publique n'ayant pas accès aux informations collectées), l'INPN, en revanche, se présente comme un site Internet accessible à tous. L'oeuvre est de grande ampleur et il faudra un peu de temps pour caler le projet, tant du point de vue technique (travail des informaticiens...) que dans la mise en oeuvre que nous osons appeler «pédagogique» : ce qui est donné à voir pour l'utilisateur, en particulier le non spécialiste.

Arrive enfin, en 2020, le module d'accès public aux données naturalistes : OpenObs (<https://openobs.mnhn.fr>), dans la lignée des projets portés au niveau mondial par le GBIF (Global Biodiversity Information Facility).

Cet outil est remarquable d'un point de vue technique : facilité de prise en main, rapidité de traitement et de visualisation cartographique des résultats, capacité d'export. **C'est, disons-le, l'outil idéal de diffusion des informations naturalistes pour la France, base qui devrait devenir incontournable dans tous les travaux nécessitant une compilation d'informations naturalistes.**

Dans le cadre de notre projet d'Atlas départemental, nous avons donc eu grand plaisir à récupérer, une première fois début 2023, puis début 2024, facilement et gracieusement, un jeu de données d'observations des Lépidoptères du Gard, venant compléter avec bonheur les observations déjà à notre disposition.

Le présent document est donc une critique, qui se veut constructive, de l'utilisation de ces informations.

Nous remercions très chaleureusement Solène Robert, Cheffe de l'équipe «Diffusion et médiation» au sein de PatriNat, qui a accepté ce jeu de questions réponses pas toujours faciles ; en effet, certains choix sont assumés par l'INPN, mais la teneur de certaines informations est parfois liée à la politique de mise à disposition propre au gestionnaire initial.

Insistons lourdement dès le début : nous basons notre critique sur des éléments constatés, relatifs à la teneur ou à la forme de l'information diffusée par l'INPN, dans le but de questionner ces faits, parfois en proposant d'autres façons de faire, issues de notre propre expérience en matière de bases de données naturalistes collaboratives, ou issues d'autres réseaux, discussions et rencontres...

<sup>1</sup> - Pour en savoir plus sur le SINP : [https://www.bulletin-officiel.developpement-durable.gouv.fr/documents/Bulletinofficiel-0032602/TREL2224513S\\_Arrete%2030-09-2022\\_SINP\\_Schema\\_metier\\_complet3.pdf;jsessionid=04BA8D8AEA6EDD319609E98B8AE48574](https://www.bulletin-officiel.developpement-durable.gouv.fr/documents/Bulletinofficiel-0032602/TREL2224513S_Arrete%2030-09-2022_SINP_Schema_metier_complet3.pdf;jsessionid=04BA8D8AEA6EDD319609E98B8AE48574)

# A propos des champs et de la forme de la donnée

Un rappel s'impose : de quoi parlons-nous ?

Une *donnée naturaliste* est la transcription écrite, donc partageable, d'une observation réalisée par une personne physique.

Les éléments constitutifs de cette donnée sont les suivants :

- une information sur le sujet, généralement un nom d'espèce (par exemple : Merle noir *Turdus merula*) ;
- une localisation, si possible précise, qui se traduit généralement par un nom de commune, un nom de site, des coordonnées géographiques X et Y (longitude et latitude) ;
- une date d'observation ;
- un nom d'observateur.

Cette donnée naturaliste peut être accompagnée par des métadonnées, qui sont de deux sortes :

- des métadonnées encadrant l'appréciation de la donnée, par exemple le fait qu'un effectif soit proposé (1 Merle noir) dans le cadre d'un protocole (point d'écoute de 10 minutes) ;
- des métadonnées utiles à la gestion administrative des données, par exemple le programme dans lequel cette donnée précise a été collectée.

Pour résumer :

- la donnée constitue le socle de l'information ;
- les métadonnées d'accompagnement permettent, le cas échéant, le bon usage de tout ou partie de la donnée ;
- les métadonnées de gestion sont essentiellement des filtres permettant de regrouper et de trier les jeux de données.

L'export des données de l'INPN via OpenObs se présente sous la forme d'un tableur constitué de 119 champs !

Nous les mentionnons ci-après, selon leur ordonnancement, avec un exemple concret. Nous attribuons à chacun sa «famille» informative, **donnée**, **métadonnée d'accompagnement** ou **métadonnée de gestion**. Puis nous proposerons une réflexion quant à l'intérêt de ces champs et de cet ordonnancement.

Recherche : *Hyponephele lupina* dans le département du Gard entre 2020 et 2024, en date du 14 janvier 2024. Une seule mention connue.

1. **idSINPOccTax** : CBDCAEAF-C7AD-1735-E053-0514A8C036F2
2. **libelleCadreAcquisition** : CardObs : Observations naturalistes issues de l'outil CardObs
3. **idCadreAcquisition** : 4A9DDA1F-B693-3E13-E053-2614A8C02B7C
4. **descriptionCadreAcquisition** : Les données issues de l'outil de gestion CardObs et de son application mobile Carnat (hors programme relevant d'un cadre d'acquisition spécifique) regroupent les observations de naturalistes expérimentés. Ces observations concernent la faune, la flore et la fonge tant continentales que marines. L'emprise géographique des saisies couvre la France métropolitaine et l'Outre-mer. Les données peuvent être opportunistes aussi bien que protocolées en fonction des utilisateurs, et peuvent disposer de champs complémentaires propres à chaque jeu de données. L'outil CardObs a été mis en place en 2007 par le Service du Patrimoine Naturel du MNHN. Il est actuellement géré par PatriNat (OFB, MNHN, CNRS, IRD)
5. **objectifCadreAcquisition** : Multiple ou autres
6. **motsClefsCadreAcquisition** : CardObs | Donnée Naturaliste | Faune | Flore | Fonge | Métropole | Outre-Mer
7. **referenceBiblioCadreAcquisition**
8. **maitreOuvrage** :
9. **maitreOeuvre** : PATRINAT (OFB-MNHN-CNRS-IRD)
10. **financeur** :
11. **contactPrincipal** : PATRINAT (OFB-MNHN-CNRS-IRD)
12. **typeFinancement** : Mixte
13. **libelleJeuDonnees** : Données naturalistes de MATIGNON Théo
14. **idJeuDonnees** : C32307AF-DB57-5D61-E053-0514A8C09FD8
15. **descriptionJeuDonnees** :
16. **objectifJeuDonnees** : Observations naturalistes opportunistes
17. **jsonProtocole** :
18. **libelleProtocole** :
19. **idCampanuleProtocole** :
20. **motsClefsJeuDonnees** :
21. **territoireJeuDonnees** :
22. **fournisseurJeuDonnees** :
23. **producteurJeuDonnees** :
24. **typeDonneesJeuDonnees** : Occurrence de taxon
25. **idOrigine** : 3824152
26. **statutSource** : Ne Sait Pas
27. **statutObservation** : Présent
28. **observateur** : MATIGNON Théo
29. **determinateur** : MATIGNON Théo
30. **nomScientifiqueRef** : *Hyponephele lupina*
31. **nomCite** : *Hyponephele lupina*
32. **nomVernaculaire** : Louvet (Le)
33. **cdNom** : 53678
34. **cdRef** : 53678
35. **rangTaxo** : species
36. **regne** : Animalia
37. **classe** : Insecta
38. **ordre** : Lepidoptera
39. **famille** : Nymphalidae
40. **genre** : *Hyponephele*
41. **espece** : *Hyponephele lupina*
42. **groupeTaxoSimple** : Insectes et araignées

43. **groupeTaxoAvance** : Insectes  
44. **dateObservation** : 2021-08-31T00:00:00Z  
45. **datePrecision** : DAY  
46. **decennie** : 2020  
47. **annee** : 2021  
48. **mois** : 8  
49. **dateDeTermination** : 13/09/21  
50. **latitude** : 43.92110  
51. **longitude** : 3.50678  
52. **precisionGeometrieMetres** : 4249  
53. **systemeCoordonnees** : WGS84  
54. **precisionLocalisation** : XY centroïde commune  
55. **typeObjetSource** : Maille 1km  
56. **toponyme** :  
57. **commune** : Blandas  
58. **codeInseeCommune** : 30040  
59. **EPCI** : CC du Pays Viganais  
60. **codeInseeEPCI** : 243000270  
61. **departement** : Gard  
62. **codeInseeDepartement** : 30  
63. **region** : Occitanie  
64. **codeInseeRegion** : 76  
65. **altitudeMin** : 634.0  
66. **altitudeMax** : 634.0  
67. **profondeurMin** :  
68. **profondeurMax** :  
69. **codeMaille10Km** : 10kmL93E074N631  
70. **objetGeoWKT** :  
71. **denombrementMinMax** : (6-6)  
72. **objetDenombrement** : NSP  
73. **occEtatBiologique** :  
74. **occSexe** :  
75. **occStadeDeVie** :  
76. **occComportement** :  
77. **occStatutBiologique** :  
78. **urlPreuveNumerique** :  
79. **referenceBiblio** :  
80. **commentaire** :  
81. **jsonAttributsAdditionnels** :  
82. **idRegroupement** :  
83. **typeRegroupement** :  
84. **methodeRegroupement** :  
85. **jsonRegroupement** :  
86. **idSINPRegroupement** :  
87. **typeRegroupement** :  
88. **codeNivValNationale** : 2  
89. **nivValNationale** : Probable  
90. **validateurNationale** : PatriNat  
91. **dateCtrlNationale** : 09/03/23  
92. **typeValNationale** : Automatique  
93. **protocoleValNationale** : <https://inpn.mnhn.fr/docs-web/docs/download/247445>  
94. **commValNationale** :  
95. **codeNivValRegionale** :  
96. **nivValRegionale** :  
97. **validateurRegionale** :  
98. **dateCtrlRegionale** :  
99. **typeValRegionale** :  
100. **protocoleValRegionale** :  
101. **commValRegional** :  
102. **nivValProducteur** :  
103. **validateurProducteur** :  
104. **dateCtrlProducteur** :  
105. **protocoleValProducteur** :  
106. **especeProtegee** :  
107. **especeReglementee** :  
108. **especeDirectiveEuropeenne** :  
109. **especeEvalueeLR** : true  
110. **especeMenaceeLRNat** : true  
111. **especeMenaceeLRReg** : true  
112. **especeDeterminanteZnieff** : true  
113. **statutBiogeoEspeceTaxref** : Présent (indigène ou indéterminé)  
114. **habitatEspeceTaxref** : Terrestre  
115. **sensible** : Non  
116. **sensiNiveauCode** :  
117. **sensiNiveau** :  
118. **sensiDateAttribution** :  
119. **sensiReferentiel** :

Quelque peu surpris, pour ne pas dire, assommés, par cette quantité d'information relative à l'observation d'un papillon quelque part dans le Gard, voyons ce que nous pouvons en dire, étant entendu que notre perception des choses est en lien avec notre usage attendu de l'information. Il est tout à fait possible qu'un autre lecteur aura une vision différente. Nous revendiquons ici une lecture naturaliste.

### Où se trouve l'information constitutive de l'observation ?

Il faut attendre le champ n°28 pour connaître l'observateur, le champ n°30 pour la citation de l'espèce, le champ n°44 pour connaître la date d'observation ; latitude et longitude arrivent en 50 et 51<sup>ème</sup> position; l'information communale en n°57 ; l'altitude en 65 et 66. Enfin, une information numérique est donnée au champ n°71. La réécriture de ces seuls champs nous donne donc :

- MATIGNON Théo (28), *Hyponephele lupina* (30), 2021-08-31 (44), 43.92110/3.50678 (50 et 51), Blandas (57), 634 m (65), 6-6 (71).

Que nous aurions tendance à lire comme ceci, pour répondre à la question «qu'est-ce qui a été observé ?» :

- 6 *Hyponephele lupina* le 31/08/2021 à Blandas, 43.92110/3.50678/634 m, par Théo MATIGNON.

Ainsi parmi les 119 champs présentés, il nous semble que :

- 16 champs décrivent réellement l'observation, avec des doublons (4 champs différents pour la dénomination de l'espèce, par exemple + un champ supplémentaire pour le genre) ;
- 48 champs participent à la bonne appréciation de l'observation, mais là encore avec des champs peut-être superflus ?
- 55 champs ont pour seule utilité la gestion administrative de la donnée.

# Exploitation de l'INPN

Ce qui nous intéresse ici est de tester, grandeur nature oserions-nous dire, l'utilisation des données proposées dans l'INPN.

Notre cadre imparti : nous voulons que les informations participent à l'appréciation

- de la répartition géographique des espèces ;
- de la répartition phénologique des espèces (périodes de vol, cycles biologiques) ;
- d'une éventuelle dynamique des populations, pouvant conduire à une appréciation de la fragilité des populations et au statut de conservation selon différentes échelles.

Le tout dans notre projet d'Atlas des papillons de jour et zygènes du Gard, ayant vocation à être publié et à servir d'outil de réflexion pour la conservation de ces espèces à l'échelle de ce département.

La recherche faite sur le module OpenObs de l'INPN, en date du 14 janvier 2024, donne 120 569 résultats.

De ce premier jeu nous excluons :

- 1 351 observation en provenance d'iNaturalist ; à la même date nous intégrons en direct 5 459 observations d'iNaturalist ;
- 6 379 observations en provenance d'Observado (observation.org) ; à la même date nous intégrons 31 341 observations d'Observado ;
- 226 données rapportées au département, sans intérêt pour un atlas départemental.

Ensuite nous recherchons les observations saisies, lors de leur première mise en base de données, dans l'Observatoire du Patrimoine Naturel du Gard et qui auraient pu arriver dans l'INPN via les diffusions réalisées auprès du CEN Languedoc-Roussillon dans le cadre de la mise en place des ZNIEFF en Languedoc-Roussillon, dans le cadre du projet d'Atlas des papillons de jour et libellules de la région Languedoc-Roussillon, ou encore lors de la mise à disposition d'informations auprès du CEN-PACA.

Dès la première mention, nous voilà plongés au coeur des problèmes de respect des informations...

L'observation originelle, que chacun peu consulter (<http://www.naturedugard.org/obs.php?obs=3829>), fait état de l'observation d'une Diane *Zerynthia polyxena* le 14/05/2008 à Aiguèze par Benjamin Adam et Nicolas Forestier, cette information étant liée à la structure ONEM (Observatoire Naturaliste des Ecosystèmes Méditerranéens).

Dans l'INPN, la structure ONEM a disparu au profit de la structure Gard Nature ; on retrouvera le même traitement pour l'ensemble des informations issues de l'ONEM. Nous reviendrons plus loin sur la notion de « paternité » de la donnée, de sensibilité - réelle - des observateurs au respect des informations diffusées, et l'importance du respect apporté par les gestionnaires de bases de données à ce sujet particulier.

Un peu plus loin, on notera :

- *Z. polyxena*, par Frédéric Andrieu (Gard Nature) le 12/05/2007, avec un idOrigine 53823 ;
  - *Z. polyxena*, par Frédéric Andrieu (INCONNU) le 12/05/2007, avec un idOrigine 53823-opng-3845 ;
- pour une donnée qui aurait du être estampillée ainsi :
- *Z. polyxena*, par Frédéric Andrieu (ONEM) le 12/05/2007, avec un idOrigine opng-3845.

L'ONEM est donc maudit ! Et les idOrigine vont bon train, situation idéale pour créer des doublons d'informations non décelable de façon automatisée...

Après force de recherche... nous excluons :

- 13 818 observations avec mention Gard Nature associée aux observateurs (observations déjà existantes dans l'OPNG) ;
- 1 358 observations de Christian Jourdan qui envoie annuellement un export de ses observations saisies sous faune-france.org dans l'OPNG ;
- 522 observations estampillées opng dans leur idOrigine.

Ces dernières 522 observations sont toutes des doublons. On notera que le champ producteurJeuDonnees donne « Association Lozérienne pour l'Etude et la Protection de l'Environnement | Aude Nature | CEFÉ-CNRS Montpellier | Centre Ornithologique du Gard | Fédération des Réserves Naturelles Catalanes | Groupe Ornithologique du Roussillon | Ligue pour la Protection des Oiseaux de l'Aude | Ligue pour la Protection des Oiseaux de l'Hérault | ONCFS-DR-LR | Syndicat Mixte du bassin de l'On », c'est-à-dire sans la moindre mention de la structure réellement productrice de la donnée, et avec mention de nombreuses structures qui n'ont pas concouru à la création de cette donnée...

# A propos des doublons

On parle de doublon lorsque deux lignes de la table de données relatent la même observation, à savoir :

- la même date ;
- les mêmes coordonnées géographiques X et Y ;
- la même espèce et informations associées (sexe, âge...) ;
- le même observateur.

Les doublons informatiques ne sont pas graves en eux-même : simplement, ils prennent une place - informatique - pour rien, et faussent les valeurs statistiques de base comme par exemple le nombre d'observations pour une espèce donnée, le nombre d'observations dans une commune, etc.

Au niveau technique (tableur), nous recherchons les doublons de la manière suivante :

- concaténation des champs visés pour créer une valeur de dédoublonnage ;
- tri sur cette valeur ;
- affichage des doublons.

Nous obtenons ainsi :

- 4 924 doublons probablement créés par le formatage des données, les coordonnées étant ramenées à une maille (donc différentes observations peuvent être ramenées à un même jeu X et Y exploitable) ;
- 3 844 vrais doublons.

Après avoir codé proprement les observateurs, lors d'une nouvelle session de recherche des doublons, ce sont encore 7 267 lignes qui sortent ! Cette erreur dans une base de données ayant vocation à la mise en commun de sources diverses est tout à fait attendue : certains gestionnaires vont par exemple écrire DUPONT Henry quand d'autre noteront Henry DUPONT (inversion nom-prénom), ou DUPONT H. (partie tronquée)... On aura aussi des divergences entre André et Andre, Saint-Aignan et St-Aignan, etc. Nous verrons tout ceci dans un paragraphe dédié.

Au final, sur les 96 915 données que nous envisageons pouvoir utiliser, il en reste 80 880 seulement (16% de doublons). Ce qui est tout à fait intéressant néanmoins pour notre projet !

Quelques broutilles nous amènent à éliminer quelques lignes encore.

Puis 3 890 observations rapportées à 2 à 5 communes, dans une géométrie «XY Centroïde ligne/polygone», doivent être éliminées : on ne sait pas ce que signifient les coordonnées géographiques indiquées, sur quelle commune et parfois dans quel département a été faite l'observation...

Quelques surprises nous attendent encore : 281 observations de Raymond Gaillard datées de 1758 ? Cet observateur arpenterait le Gard au milieu du XX<sup>ème</sup> siècle... C'est là une erreur facile à détecter... Plus aisément en tout cas que les dizaines de mentions de Gaillard datées de décembre 2011 ! Tout aussi erronées (pour la date en tout cas).

Au final, nous pouvons jouer avec un jeu conséquent de 76 955 données.

Nous allons pouvoir entrer dans les détails, en prenant les thèmes les uns après les autres, sans aucune forme de hiérarchisation : les questions posées sont toutes légitimes, selon nous, et toutes aussi importantes.

# Inter-opérabilité des bases de données

Pour faire bref : l'outil de base de données, développé avec l'informatique, est précieux pour le naturaliste souhaitant agréger de la donnée (c'est le terme imparfait...).

Faire un atlas, dessiner les contours d'une aire de répartition, imaginer un graphique de présence décadaire, vérifier, le cas échéant, une observation hors norme... tout ceci a été réalisé de tout temps par les naturalistes avec des bouts de papier, des collages, des gommages, des coups de crayon...

Une solide base de données associée à un logiciel de cartographie et un tableur sont les outils d'aujourd'hui, qui augmentent considérablement notre capacité individuelle et collective à traiter des informations multiples.

A chaque besoin, son outil :

- la base de données sert à homogénéiser l'information (on la fait rentrer dans des champs) et à la conserver ;
- l'outil SIG sert à représenter sur carte les informations géographiques ;
- le tableur va nous aider pour produire des graphiques et des calculs.

Qu'appelle-t-on donc l'inter-opérabilité des bases de données ?

Dans la pratique il existe une multitude d'outils développés par les uns et les autres au fil du temps, en fonction des besoins spécifiques (groupe considéré, une étude particulière, une structure durable ou éphémère...). Mais, en matière naturaliste, nous l'avons écrit dès le début de ce document, il nous importe de pouvoir enregistrer et utiliser les informations relatives à l'objet (espèce), au lieu (coordonnées, nom de commune, etc.), à la date et à l'observateur. S'ajoutent toutes informations utiles à la bonne compréhension de l'information (les données complémentaires et les métadonnées de gestion). Toutes nos bases de données utilisent donc, de fait, ces types de champs.

Donner un nom d'espèce... En français ? En Catalan ? En scientifique ? Avec ou sans l'auteur et la date ?

L'ordinateur est un outil d'une logique infaillible. Aussi considérera-t-il ces différents termes comme des informations différentes :

- le Damier athalie
- *Mellicta athalia*
- la Mélitée du Mélampyre
- *Melitaea athalia*
- *Melitaea athalia* Roth.
- *M. athalia*...

Et que fait-on avec *Melitaea athalia* ssp. *celadussa*, *M. helvetica*, *M. nevadensis* et *M. celadussa* ?

A multitude de possibilités de décrire la même information, multiplication du temps passé pour faire une synthèse des connaissances !

L'Index Synonymique de la Flore de France de Michel Kerguelen dans les années 1990 a donné la voie qui a abouti, depuis une dizaine d'années, au Référentiel Taxonomique des espèces de France. Par un simple jeu de code unique associé à chaque dénomination unique, on peut arriver au nom dit valide du moment. Si celui-ci change (ce qui arrive assez souvent...), on pourra se plonger dans le référentiel pour trouver, avec notre code initial, la nouvelle mention valide.

Ainsi, en reprenant mon exemple de *Melitaea athalia* avec, mettons, 6 entrées et codes initiaux différents, on va pouvoir tout agglomérer pour travailler sur les données selon le nom valide, sans perte d'information, ni perte de temps pour aller à la pêche aux données.

**Le TaxRef (référentiel taxonomique) du Muséum National d'Histoire Naturelle, quelles que soient les critiques que l'on peut faire à son endroit, est un outil indispensable et d'un très haut intérêt.**

C'est une première étape importante pour pouvoir échanger facilement entre différentes bases.

Généralement, les champs de date et de coordonnées sont compréhensibles.

Les noms des observateurs et les champs associés sont plus difficiles à traiter : nous y reviendrons plus loin dans un paragraphe dédié.

# A propos de l'observation originelle

Lorsque l'on fait circuler des données naturalistes, il est parfois intéressant de pouvoir accéder à l'observation originelle. Cette idée est incluse dans la plupart des bases de données, mais donne bien du fil à retordre aux informaticiens... Pourtant, une solution simple existe, proposée par le GBIF et Tela-Botanica depuis une quinzaine d'années...

Un exemple pour illustrer la question : je dispose dans l'export de l'INPN d'une observation de *Charaxes jasius* faite à Saint-Just-et-Vacquières le 23/08/2023 par un anonyme. Je souhaite accéder à la fiche de cette observation sur le site de l'INPN. Je dispose, dans l'information de la donnée, d'un identifiant SINP : 60ebdcba-d5ec-4f88-9f85-b418de83b86c. Mais je ne peux rien en faire puisque l'identifiant de la fiche INPN est 1448076e-9a76-4235-b287-56c7253bf278, ouvrant la page unique <https://openobs.mnhn.fr/openobs-hub/occurrences/1448076e-9a76-4235-b287-56c7253bf278>

## Pourquoi revenir à la source ?

D'un premier abord, on ne voit pas forcément très bien l'intérêt de remonter à la source, dès lors que l'on dispose de la teneur de l'information naturaliste : objet, lieu, date et observateur. Cet intérêt, devenant nécessité parfois, voit le jour lorsque l'on s'amuse à compiler de nombreuses informations pour en faire quelque chose : une synthèse, une carte, un graphique, une liste d'espèces, d'observateurs, de structures...

Dans une base de données, toute observation est dotée d'un identifiant unique, un numéro qui s'incrémente tout seul, par exemple.

Ainsi, 2 exuvies de Gomphe à pattes jaunes *Stylurus flavipes* observées le 23/06/2014 en pied de berge du Rhône, juste en aval du barrage de Vallabrègues, par Philippe Baldi et Jean-Laurent Hentz, est affublé du numéro d'identifiant 174060.

Ce numéro est unique dans la base de données de l'Observatoire du Patrimoine Naturel du Gard.

Il permet de visualiser l'intégralité de l'information sur une page web, si la base de données est liée à un serveur Internet : <http://www.naturedugard.org/obs.php?obs=174060> (fig. 1).

Cette information a été collectée dans le cadre d'un programme particulier : l'Atlas des papillons de jour et libellules du Languedoc-Roussillon, financé notamment par l'argent public via la DREAL Languedoc-Roussillon.

L'information diffusée sur cette page est précise, au niveau géographique.

Conformément à la commande et à notre engagement moral auprès des animateurs du programme régional, cette observation a été transférée, *in extenso*, au CEN Languedoc-Roussillon, qui l'a transmise à la DREAL Languedoc-Roussillon dans le cadre du SINP régional.

Ainsi cette observation est-elle normalement rendue accessible à tout un chacun dès le 26/07/2014.

Or, la base de données du CEN Languedoc-Roussillon n'est pas rendue publique ; ni celle du SINP Languedoc-Roussillon !

Par contre, dès 2014, les informations sont mises à disposition via un site web dédié à l'Atlas des papillons de jour et libellules du Languedoc-Roussillon, via l'espèce par exemple (<https://atlas.libellules-et-papillons-lr.org>). La figure 2 présente la même observation, incomplète mais bien localisée.

Et enfin, en 2023, on trouve une page d'information de l'espèce ciblée au titre du SINP Occitanie (<https://sinp-occitanie.fr/atlas/>) (fig. 3)...



## Consultation de l'observation n° 174060.



Photo :

Observateurs /1/811/ : Hentz Jean-Laurent, Baldi Philippe

Structure /2/: Gard Nature

Programme 6 : Recherche - Atlas LR Papillons-Libellules

Commanditaire : CEN-LR

Saisi par : Hentz Jean-Laurent

Espèce (fr) : Gomphe à pattes jaunes (Le)

**Espèce (sc) : Stylurus flavipes (Charpentier, 1825)**

Gomphus flavipes

TAXREF : 65234

Effectif : 2 mâle ou femelle exuvie

Nature 0 :

Protocole 25 : Observation occasionnelle

**Date : 23/06/2014**

**Commune : Vallabrègues (30336)**

Site : non précisé (id : 303360099000)

X (longitude) : 4.62102

Y (latitude) : 43.84008

Altitude : 5 m

Précision : point

Commentaire : TB, prospection exuvies, tronçon 13

Validation : Observation validée

Saisie le : 26/07/2014

Dernière modification : 09/12/2014 à 15:40:11

Numéro unique de l'Observatoire :  
<http://www.naturedugard.org/obs.php?obs=174060>

Des questions, des remarques ? - COPIL

Figure 1 : copie d'écran de la page web <http://www.naturedugard.org/obs.php?obs=174060>

Atlas des papillons de jours et des libellules du Languedoc-Roussillon

Rechercher une espèce

obs. n° 66224 © Sannier David

## Gomphe à pattes jaunes (Le)

### *Gomphus flavipes* (Charpentier, 1825)

Ordre : *Odonata* > Famille : *Gomphidae*  
> Genre : *Gomphus*

INPN Inventaire National du Patrimoine Naturel

Informations sur  
Description En s

espèce eurosibérien sibérie.  
En France l'essentie de la Loire et de ses sud-ouest.  
La découverte de se les observations se :

198 observations

12 communes

32 observateurs

Première observation 2010

Dernière observation 2016

12 communes 32 observateurs

Date: 2014-06-23  
Altitude : 4  
Observateurs : Baldi Philippe, Hentz Jean  
Validation : valide  
Effectif: 2

Nombre d'observation(s): 198

Observations

200  
150  
100  
50  
0

0-100 100-200 200

Observations

200  
150  
100  
50  
0

Janvier Février Mars

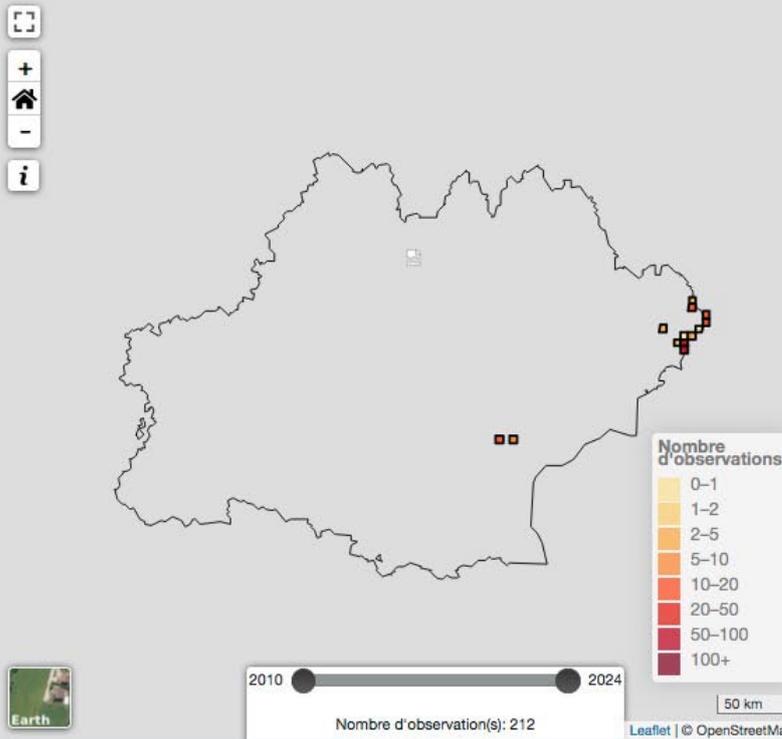
Figure 2 : copie d'écran de la page web <https://atlas.libellules-et-papillons-lr.org/atlas/espece/65236>, en janvier 2024.



### Gomphe à pattes jaunes (Le)

*Stylurus flavipes*  
(Charpentier, 1825)

Classe : *Insecta* > Ordre : *Odonata* >  
Famille : *Gomphidae* > Genre : *Stylurus*



212 observations

17 communes

34 observateurs

Première observation 2010

Dernière observation 2022

- 17 communes 34 observateurs
- Aramon (30) - Beaucaire (30) - Codolet (30) - Comps (30) - Coursan (11) - Cuxac-d'Aude (11) - Fleury (11) - Laudun-l'Ardoise (30) - Les Angles (30) - Lespignan (34) - Montfaucon (30) - Montfrin (30) - Salles-d'Aude (11) - Sanilhac-Sagriès (30) - Sauveterre (30) - Vallabrègues (30) - Villeneuve-lès-Avignon (30)

Figure 3 : copie d'écran de la page web <https://sinp-occitanie.fr/atlas/espece/65234>, en janvier 2024.

Entre temps arrive OpenObs et, justement, par hasard, je regarde en juillet 2022 une donnée que je connais bien : mon Gomphe à pattes jaunes de Vallabrègues du 23/06/2014...

Première constatation : cette observation dûment enregistrée en 2014 et transmise à qui de droit dans la chaîne de circulation des données est arrivée à l'INPN le 28/02/2022 ! Soit huit ans pour que, dans le cadre de la politique publique, une information passe d'un ordinateur de Montpellier à un autre à Paris. Ce délai illustre simplement la - grande - difficulté de mise en oeuvre de cette politique... Et certainement pas pour des raisons de blocages techniques ! Mais laissons cela de côté : l'important est d'avancer, le temps perdu est... perdu.

Seconde constatation, en 2022 : l'observation est présentée au centre d'une maille de 10 kilomètres de côté ! Mon exuvie de libellule se retrouve en plein milieu d'une vigne ou d'une garrigue... Alors même que l'observation originale est précise...

De plus, le nom de la structure (Gard Nature) a disparu...

Là encore, les choses évoluent, tout doucement, et un an après, l'observation est mise à jour, dûment repositionnée et amendée... Bon. Tant mieux.

# A propos de l'observation originelle - validation

Tout travail à partir d'informations collectées nécessite une vérification de l'information : c'est le processus de validation. En matière naturaliste, en quoi cela consiste-t-il ? Tout simplement à vérifier la cohérence entre l'information apportée et la connaissance disponible au moment de son traitement.

Quelques exemples pour bien comprendre.

Une observation - réelle ! - de Proserpine *Zerynthia rumina* au coeur de l'été en Bretagne doit attirer l'attention : on va, par principe, émettre un doute sur sa validité. Pourquoi ?

1. Parce que cette espèce méditerranéenne n'est pas attendue en Bretagne, bien éloignée de son aire géographique connue.
2. Parce que ce papillon vole au printemps, en un cycle annuel unique (univoltin), et qu'une mention en-dehors de la période connue interpelle.

**Le valideur est un observateur auto-proclamé valideur : il n'y pas de règles, pas d'élections, pas de nomination particulière...** Généralement, le responsable d'un projet thématique joue ce rôle, en tant que spécialiste de la chose. Dans des structurations collectives, on peut avoir plusieurs personnes connues pour leurs compétences dans un domaine qui acceptent de passer un peu - beaucoup - de temps à relire les observations. On peut même proposer, comme nous l'avons fait pour l'ONEM et qui se fait sur iNaturalist, que la validation soit réellement collective : tout un chacun a droit d'exprimer son accord ou ses doutes, ses remarques et des arguments, à propos d'une observation. C'est d'autant plus aisé que l'on dispose d'une preuve : photo ou enregistrement audio (sous réserve que l'un ou l'autre soient de qualité suffisantes, ou montrent les critères de détermination connus et partagés).

Il est important, théoriquement et dans la pratique, que le valideur soit un tiers, c'est-à-dire ne soit pas la même personne que l'observateur. Car un regard extérieur est toujours bienvenu !

Que fait le valideur, en cas de doute ?

1. Il vérifie si l'observation dont il dispose est bien conforme à l'originale. Soit l'observation a été enregistrée pour la première fois dans la base qu'il consulte ; soit elle a été importée depuis une autre base (ce que nous faisons avec les observations de l'INPN, importées dans l'Observatoire du Patrimoine Naturel du Gard). En effet, il est vite fait d'avoir commis un impair lors du traitement des données pour l'import... Un rapide coup d'oeil sur la donnée originale, grâce à son *identifiant unique*, permet d'exclure ce biais.

Si tant est que l'on puisse accéder à l'observation à partir de l'identifiant unique : ce qui n'est pas le cas pour les données issues de l'INPN !

2. Une fois établi qu'il dispose bien de la bonne information, le valideur va consulter l'observateur, afin de lui demander plus amples renseignements, en particulier l'existence de photo ou d'enregistrement, une description orale ou écrite de l'observation.

Notez que cette action est délicate à plus d'un titre :

- il faut pouvoir joindre l'observateur... Ce qui n'a pas toujours été prévu dans le parcours de la donnée... Ça prend donc du temps de recherche, puis du temps de contact et d'échanges ;
- l'observateur peut se montrer chagriné que l'on doute de sa bonne foi ! C'est humain ! Et le valideur devra en plus faire preuve de beaucoup de diplomatie et de délicatesse pour donner un tournant agréable et constructif aux échanges...

Finalement, le processus de validation a plusieurs intérêts :

- confirmer ou modifier une information (généralement l'espèce, parfois la localisation, la date) ;
- permettre à l'observateur qui se serait trompé de progresser dans sa compétence !

Ainsi cette Proserpine bretonne était bien une Proserpine... Mais pas *Zerynthia rumina* ! C'était *Proserpinus proserpina*, le Sphinx de l'épilobe, aussi nommé... Proserpine...

L'outil informatique offre aussi l'opportunité d'automatiser le processus de validation : ainsi, dans l'INPN, une espèce dans les bonnes dates et dans un département où elle est déjà connue sera validée automatiquement.

Cette validation automatisée est intéressante, mais pas forcément juste : en particulier, des papillons associés à des habitats spécifiques, par exemple *Boloria selene*, sont bien présents dans le Gard mais fortement liés à des habitats d'altitude humides et/ou acides tourbeux. Une observation en pleine garrigue attire forcément l'attention du valideur humain, mais pas du modèle national.

Autrement dit : pour notre part, lorsque nous importons des observations, elles seront traitées comme si elles n'avaient pas encore été validées. C'est aussi une forme de responsabilité que doit exercer l'utilisateur.

### **Veille et validation collectives...**

Ces deux idées ont été longuement débattues au sein de l'ONEM, lors de sa création il y a 20 ans...

Le constat était le suivant : en général, la validation est confiée à un groupe de personnes, restreint, avec des compétences reconnues. Parfois on sait qui valide , parfois on ne sait pas trop.

Le valideur, comme l'observateur, est un humain qui ne dispose pas d'autres compétences que ses connaissances propres et la connaissance partagée, au moment de la validation, par la communauté naturaliste. Chacun, observateur et valideur, peut se tromper dans l'appréciation de l'observation.

L'observateur au premier chef : c'est lui qui a réellement expérimenté une rencontre avec une plante ou un animal, lui qui s'est référé à ses connaissances propres et aux connaissances disponibles du moment pour attribuer un nom, *en bonne foi*, à son sujet d'observation.

J'insiste sur cette notion de bonne foi, car certaines observations manifestement erronées vont parfois choquer le valideur : acceptons que chacun ait le droit d'observer la nature, avec des connaissances moindres que le spécialiste ! Et dans ce cas, soyons indulgents - si l'on penche du côté des spécialistes - lorsque l'on trouve un nom de papillon des tropiques pour une simple Piéride du chou. Bien entendu que c'est faux : invitons l'observateur à se documenter plus efficacement.

La seconde difficulté est la compétence de rigueur dans l'observation. Un néophyte ne sait pas que le forme de telle tache est discriminante pour séparer trois espèces de piérides ; il a vu un papillon blanc, dans un bouquin ou sur Internet, et considère que la ressemblance est suffisante pour attribuer le même nom. Encore une fois : le processus de validation peut être une occasion de progression.

Le valideur aussi peut se tromper, car lui aussi voit sa connaissance et ses compétences progresser dans le temps, en fonction de ses propres expériences et des nouvelles lectures. Un même valideur ne portera peut-être pas le même regard sur un *Pyrgus* ou une zygène en 2010 ou en 2025... Cette évolution de la connaissance collective est encore plus criante sur d'autres groupes comme les syrphes ou les abeilles, par exemple.

Pourquoi ce développement ? Pour aboutir à cette idée que, au-delà d'une organisation de validation interne à chaque projet, il semble particulièrement intéressant d'envisager une validation collective, rendue d'autant plus facile par les outils numériques.

En effet, revenant à l'INPN, chacun peut, en consultant le site, en téléchargeant des données naturalistes, se rendre compte de l'existence d'une information douteuse. En général, cette information est alors écartée. Mais pourquoi ne pas intégrer dans le processus de validation le fait qu'un utilisateur lambda, peut-être non connu et reconnu pour ses compétences, ait la possibilité de participer à cette validation ? Comment ? En ayant la possibilité de contacter aisément l'observateur, ou tout au moins le gestionnaire de la base de donnée, pour l'informer de ses doutes, les argumenter.

En posant ces questions, on en vient à une autre notion tout à fait importante et particulière aux données naturalistes : la teneur de l'information est susceptible d'être modifiée dans le temps. Exemple : modification de l'espèce parce que, des années après, en visant nos photos, nous nous rendons compte que nous nous étions trompés, que notre Petite Tortue était en fait une Grande Tortue.

Alors on modifie notre donnée originale, dans la base de donnée de départ.

Et il faut que cette modification puisse être répercutée tout au long des chaînes de transmission et de voyage des données naturalistes ; quand je change de Tortue dans l'Observatoire du Patrimoine Naturel du Gard, il faut que cette modification se fasse aussi dans les bases de données du CEN Occitanie, du SINP Occitanie, de l'INPN, du GBIF (système mondial)...

Pour faciliter validation et suivi de la donnée, l'informatique nous a fait un cadeau formidable : l'identifiant unique !

# Identifiant unique et paternité de la donnée

Ces deux notions sont importantes et intimement associées, la première purement informatique, la seconde nettement anthropique...

Prenons comme exemple une observation cible, citée deux fois dans OpenObs, sous les adresses <https://openobs.mnhn.fr/openobs-hub/occurrences/580b04ed-d098-48a8-b2ad-9f2c5cdfb114> et <https://openobs.mnhn.fr/openobs-hub/occurrences/f9a2c7fa-4213-427f-9565-0a74f630a72f>.

Comment savons-nous que ce sont des doublons ?

On retrouve un sujet identique (même espèce), le même nombre d'individus, les mêmes observateurs, la même date, les mêmes coordonnées géographiques et la même précision d'information géographique.

Dans le cas présent, étant l'un des auteurs de l'observation, je peux effectivement attester qu'il n'existe qu'une seule mention originale.

La question est donc simple : comment éviter les doublons ?

La réponse est tout aussi simple : en respectant l'identifiant unique !

## Comment choisir un identifiant unique ?

Commençons par rappeler l'importance de cette notion au sein d'une base de données (quelles que soient les informations contenues).

L'identifiant unique au niveau de la base de données, qui se présente comme un tableur, est un champ spécifique, numéroté par exemple 1,2, etc. ; ce numéro est généralement incrémenté automatiquement, ce qui fait que toutes les bases de données naturalistes ont potentiellement une observation avec l'identifiant 12345... Insuffisant pour assurer la différenciation entre deux données lorsque l'ensemble des observations est compilé dans une nouvelle base «rassembleuse» comme l'INPN.

Pourtant, une base de données associée à un site Internet est unique au monde... En effet, une adresse Internet (par exemple <http://naturedugard.org>), dit aussi *nom de domaine*, est unique au monde, par obligation, sans quoi le système ne fonctionnerait pas (on arriverait un coup sur deux sur une mauvaise page...). Cette idée, développée voilà une quinzaine d'années par le GBIF et Tela-Botanica, constitue une nouveauté par rapport au temps d'avant, quand Internet n'existait pas...

Aussi, la donnée <http://naturedugard.org?identifiant=12345> est unique dans la base de données associée au site web, et donc unique dans le monde entier.

Mais ! Car il y a un mais (sinon, ce ne serait pas drôle...) !

Je peux me faire pirater mon site... Je peux arrêter mon site... Je peux oublier de payer mon prestataire fournisseur de nom de domaine... Bref : il se peut tout à fait que le site web <http://naturedugard.org> change de main, avec une base de données toute neuve qui repart de 1 ! Cela peut aussi provenir de moi : un problème ou un choix de vider la base de données et de repartir de 1...

Selon ces considérations, nous pourrions donc trouver plusieurs observations ayant l'identifiant unique 12345 avec la même adresse web...

Ajoutons alors un troisième critère : la date de saisie. **Si l'on considère l'adresse web + l'identifiant unique à la saisie + la date de saisie, alors on obtient une information unique dans l'espace (le monde entier) et le temps (puisque le site web ne peut pas être double à un instant donné, mais seulement par succession temporelle).**

Ainsi cette adresse <http://www.naturedugard.org/obs.php?obs=174060&saisie=2014-07-26> est résolument unique. On devrait même ajouter l'heure précise de saisie pour être tout à fait dans le vrai.

Et c'est très facile à faire :

- l'identifiant est un champ de base... de base de données ;
- l'adresse web est indispensable à tout site Internet ;
- la date de saisie peut être enregistrée de façon automatisée.

Autre avantage de cet identifiant unique : il suffit de le mentionner lorsque l'on rapporte une observation pour que le lecteur, par curiosité ou questionnement, puisse facilement ouvrir la page spécifique de l'enregistrement original.

Et l'on peut partager, dupliquer, réorganiser autant de fois que l'on veut la donnée : on a toujours accès à l'original pour peu que l'on ait pris soin de conserver cette adresse...

### Et la paternité dans tout ça ?

L'observateur qui a à coeur de diffuser et partager ses observations apprécie d'être reconnu comme l'auteur des données naturalistes générées.

Si les outils modernes invitent à utiliser des pseudonymes, songeons cependant que dans le cadre de l'utilisation des données, par exemple pour l'Atlas des papillons de jour et zygènes du Gard, nous remercierons uniquement des personnes dûment identifiées, c'est-à-dire, dans nos sociétés, avec un nom et un prénom... Travailler sous pseudo présente le risque d'écarter les observations d'un travail collectif, ou un usage sans mention des auteurs. Pour notre part en tout cas, remercier Asterix30, blurpp ou Albert Einstein nous semble difficile.

Parfois l'auteur de la donnée naturaliste a souhaité être associé à une structure (une personne morale) comme une association, un bureau d'étude, un service de collectivité publique ou de l'Etat...

Il nous semble important, moralement (ça n'a aucune incidence sur la gestion informatique), que ces éléments constitutifs de la donnée soient respectés et transmis en bonne et due forme.

### Prenons un exemple...

Mon observation de Faucon crécerelle *Falco tinnunculus* (taxref : 2669) à Beaucaire (30) le 04/05/2008, par Jean-Laurent Hentz, intervenant au nom de l'association Gard Nature.

Cette observation est enregistrée, pour la première fois, dans l'OPNG sous le numéro 12345 (lien exact actuel : <http://www.naturedugard.org/obs.php?obs=12345>).

Imaginons que...

Cette observation soit transmise au CEN Occitanie : dans la base du CEN il y aura :

- un champ identifiant unique de la base du CEN, par exemple n°123456 ;
- un champ identifiant unique global (GUID) = <http://www.naturedugard.org/obs.php?obs=12345&saisie=2010-03-01> ;
- un champ observateur, avec Hentz Jean-Laurent ;
- un champ structure associée, avec Gard Nature.

Cette observation, transmise par le CEN via le SINP à l'INPN, va évoluer éventuellement de la sorte :

- un champ identifiant unique de la base de l'INPN, par exemple n°123qgsbwn653451fcjnjwtVJK ;
- un champ identifiant unique de la base de la structure ayant transmis l'observation, par exemple <http://www.naturedugard.org/obs.php?id=123456> si c'est passé par le CEN, ou <http://www.naturedugard.org/obs.php?obs=12345&saisie=2010-03-01> si ça arrive directement par l'Observatoire du Gard ;
- un champ identifiant unique global (GUID) = <http://www.naturedugard.org/obs.php?obs=12345&saisie=2010-03-01> ;
- un champ observateur, avec Hentz Jean-Laurent ;
- un champ structure associée, avec Gard Nature.

Idem si ça part au GBIF : conservation du GUID.

**Enfin, le GUID est le seul champ qui devrait être transmis et conservé de façon unique et durable, sans possibilité de modification, dans toutes les bases utilisant une même donnée.** Sauf exception : changement de nom de domaine, comme cela s'est produit pour [observation.org](http://observation.org). Mais on peut alors trouver des astuces de redirection automatisée des adresses Web.

# Identifiant unique : bis !

Nous allons insister ici sur des usages très précis de ce GUID (pour Globally Unique IDentifier), et qui nous intéressent particulièrement dans la gestion des bases de données naturalistes.

## Un outil de dédoublement...

L'INPN reçoit des données des différents SINP, des données en direct, des données du GBIF, etc. Comme chaque base, bien que requérant à peu près les mêmes champs, est unique dans sa conception et sa structure, il peut y avoir, au gré des imports, réorganisation des champs (notamment d'éventuelles concaténations) et exports, des lignes contenant la même information mais présentées différemment, que l'on ne saura distinguer à première vue.

Le respect du GUID permet tout de suite de retrouver les doublons ! Et ce, de manière automatisée et sans erreur. Ce que la concaténation de champs ne permet pas pour cause d'écritures diversifiées de certaines informations (observateurs, sources et structures notamment).

Et cela vaut pour toute base de donnée intégrative : le SINP, le GBIF ou toute autre !

## Un outil de mise à jour...

Nous avons insisté, plus avant, sur le fait que la donnée naturaliste est susceptible d'évoluer dans le temps. Or notre proposition de GUID (encore une fois, issue du GBIF et de Tela-Botanica) offre une assurance statique, justement qui n'évolue pas dans le temps.

On associera alors avec grand intérêt un second champ : date (et heure) de modification !

Exemple : j'ai envoyé mon observation avec mon GUID de Faucon crécerelle en 2010 dans la cadre de la collecte d'informations pour les ZNIEFF du Languedoc-Roussillon. Cette donnée est arrivée un jour à l'INPN. Elle existe donc :

- dans l'Observatoire du Patrimoine Naturel du Gard ;
- dans la base du CEN Occitanie ;
- dans la base du SINP de l'Occitanie ;
- dans la base de l'INPN ;
- dans la base du GBIF via export annuel de l'INPN.

Mettons que cette observation bénéficie d'une photo. Cette image n'est pas transmise (parfois une adresse d'illustration est indiquée), ni stockée ni conservée dans le cadre de l'INPN.

Imaginons qu'un observateur curieux, ou un validateur, vienne à consulter l'observation originale, voie la photo, et soit en mesure de nous dire : «Vous vous êtes trompés, les gars, c'est un Faucon crécerellette ! » Arguments à l'appui...

Dans ce cas (cette façon de procéder arrive régulièrement), l'observateur ou le gestionnaire de l'Observatoire va modifier la donnée, éventuellement en précisant dans le commentaire les arguments et l'évolution de cette donnée. Qui conserve son GUID initial ! Incluant la date de saisie initiale ! Mais toute modification, toute intervention sur la teneur ou la structure de la donnée pourra mettre à jour un champ dédié : date de modification. Généralement sous une forme 2024-04-01 12:59:23.

Lors des futurs imports de données, on pourra ainsi demander aux ordinateurs (plus performants que nous) d'importer :

- toute donnée avec un GUID inconnu ;
- toute donnée avec un GUID connu mais une date de modification ultérieure à la dernière date d'import.

Et ne pas s'occuper des observations avec un GUID connu et une date de modification antérieure à la dernière date de mise à jour des données.

# A propos des coordonnées géographiques

Dans notre usage courant des données naturalistes, nous utilisons les coordonnées géographiques les plus précises possibles, notamment pour attribuer les observations à une commune. Pourquoi ? En particulier parce que, en France, l'échelle communale joue un rôle primordial dans les politiques et les actions d'aménagement du territoire. Les élus locaux, à travers leurs démarches de cartes communales et de PLU (Plan Local d'Urbanisme), décident de l'évolution des usages des sols, en particulier en ce qui concerne l'urbanisme : nouveaux lotissements, nouveaux bâtis collectifs (écoles, halles de sports, salles communales...). De ce fait, une connaissance des enjeux naturalistes pouvant être rapportés à un territoire communal est crucial.

Des informations encore plus précises sont requises lors de discussions à propos de projets particuliers, portés par des privés ou des collectivités publiques : par exemple dans le cadre d'une enquête publique, tout citoyen est invité à s'exprimer sur la documentation fournie, qui ne concerne que quelques parcelles à l'intérieur d'une commune. Dans ce cas particulier, il convient, si l'on veut pouvoir apprécier la qualité de la prise en compte de la faune et de la flore au sein dudit projet, de pouvoir disposer, le cas échéant, d'informations que l'on puisse rapporter à cette échelle géographique très resserrée. Que l'on devrait retrouver sur le site <https://depot-legal-biodiversite.naturefrance.fr/>

C'est en ce sens que nous procédons à une première lecture du jeu de données issues de l'INPN.

31 311 observations sont ainsi données selon une précision dite XY Centroïde ; le cas échéant, ces informations devront être écartées des analyses et interprétations.

# A propos des noms d'observateurs

Insistons encore sur le fait que, dans l'usage courant, les auteurs d'un travail d'analyse de données naturalistes auront plaisir à citer et remercier les observateurs, sans qui l'information ne serait pas disponible. Ces observateurs, personnes physiques, ont un nom et un prénom, et ne sauraient se résumer à quelque pseudo, tout rigolos qu'il soient...

Un même observateur, lors de son inscription sur un site de base de données, peut choisir de le faire de différentes manières :

- Hentz Jean-Laurent ;
- HENTZ Jean-Laurent
- Jean-Laurent Hentz ;
- Jean Laurent Hentz ;
- J.-L. Hentz ;
- Jean-Laurent H. ;
- etc.

Chacune de ces écritures est unique pour l'ordinateur. Et lorsque l'on veut intégrer des données avec l'objectif que les observateurs soient dûment nommés à chacune des utilisations de leurs observations, ou proposer une liste des observateurs en ordre alphabétique, il faut bien pouvoir individualiser chaque observateur... Ce qui est simple à l'échelle d'une unique base : on va demander à chaque observateur d'être inscrit une seule fois, dans un sens précis. Mais devient difficile dès que l'on échange des données, surtout si l'on a un champ qui permet de citer plusieurs observateurs !

Autre exemple :

- Philippe Baldi, Jean-Laurent Hentz ;
- Jean-Laurent-Hentz, Philippe Baldi ;
- Baldi P. & Hentz JL ;
- etc.

## **Proposition : vers une mise en commun des noms d'observateurs ?**

L'outil informatique est bien plus performant sur des nombres. Généralement dans une base de donnée on attribue à chaque observateur un identifiant numérique. Peut-on imaginer, au moins à l'échelle nationale ou européenne (ou mondiale ?), de créer un outil d'identification unique des observateurs ?

Un tel outil faciliterait grandement le travail d'intégration des données d'une base dans une autre, participerait à la garantie du respect de la paternité.

# Conclusion

L'Inventaire National du Patrimoine Naturel et son outil de diffusion des observations OpenObs constituent des éléments phare pour l'accès à une connaissance naturaliste partagée à l'échelle de la France.

La robustesse de la structuration informatique de ces outils est éloquente.

Nous nous permettons, au travers de ce document, de critiquer, d'un point de vue technique et des usages, l'organisation de l'information mise à disposition du public via OpenObs, et suggérons des évolutions, qui nous sembleraient être des améliorations, en particulier sur les aspects suivants :

- l'organisation de la présentation des champs d'information, en mettant en avant les informations constitutives de l'observation, les métadonnées d'accompagnement puis les métadonnées de gestion et d'administration de la base de données ;
- la prise en compte d'un véritable identifiant unique dans l'espace et dans le temps, très important pour des questions de dédoublement des données, d'accès à l'information originale, de validation et de suivi de l'évolution de la donnée.

Nous attirons aussi l'attention sur l'usage des pseudos, invitant les observateurs à se référencer comme personnes physiques, portant un nom et un prénom ; dans la même idée nous rappelons la notion de paternité de la donnée. Ceci en espérant que l'INPN sera un modèle en la matière, modèle applicable par ruissellement dans l'ensemble des bases de données naturalistes...

Pour aller plus loin nous formulons l'idée et l'hypothèse d'un référencement des observateurs, à l'échelle Française pour le moins, qui viendrait faciliter les échanges de données et contribuerait au respect de la paternité des observations naturalistes.