



# Optimisation de la curation de la base de données Biologics

*Correction et mise à jour taxinomiques semi-automatiques*



 **Éric Pierre & Marguerite Chartois** 

# Présentation générale de la base de données Arthropodes gérée par Biologics

Biologics est un **logiciel de gestion de base de données** payant (société belge BioAware)



- acquis en 2010 dans le cadre du projet QBOL piloté par Jean-Yves Rasplus et Astrid Cruaud
- création de la base de données Arthropodes du CBGP

Objectifs :

**1** centraliser et diffuser les données  
taxinomiques, biologiques et  
écologiques

liées aux projets de caractérisation des  
arthropodes d'intérêt agronomique



**2** gestion et valorisation des spécimens  
des collections du CBGP



# Présentation générale de la base de données Arthropodes gérée par Biologics

**Contenu de la base :** essentiellement des insectes et arachnides dont

- ravageurs des plantes cultivées, forêts et denrées stockées
- espèces envahissantes et de quarantaine pour l'Europe
- espèces prédatrices et parasitoïdes de ravageurs
- espèces d'intérêt patrimonial

## **Arthemis DB@se : interface web de Biologics**

focalisée sur les **arthropodes d'importance agronomique**  
*à destination des professionnels & scientifiques*

outil de diagnostic moléculaire (Blast : séquences COI)  
& fiches informatives (nomenclature, biologie, écologie, photos, etc.)



<https://arthemisdb.supagro.inrae.fr/>

**Certaines informations de Biologics sont partagées et utilisées par d'autres unités :**

PVBMT (Cirad / Université la Réunion)  
CRB EP-Coll (Inrae Antibes)  
CIRM-CFBP (Inrae, Beaucozéz)



# Importance de la curation des données taxinomiques dans Biologics

## → Diffusion d'informations correctes

**utiliser une nomenclature valide pour assigner correctement les séquences et rattacher toutes les données utiles à la caractérisation d'une espèce**

Corriger les coquilles d'écriture générées lors d'ajout d'information manuel ou automatique dans la base de données

## → Enrichir et actualiser Biologics

**en s'appuyant sur les publications et autres bases de données de référence**

Mise à jour des taxons de la base tenant compte des évolutions de la systématique et de la taxinomie  
(nouveaux taxons, synonymies, nouvelles combinaisons génériques, etc.)

# Importance de la curation des données taxinomiques dans Biologics

## → Diffusion d'informations correctes

**utiliser une nomenclature valide pour assigner correctement les séquences et rattacher toutes les données utiles à la caractérisation d'une espèce**

Corriger les coquilles d'écriture générées lors d'ajout d'information manuel ou automatique dans la base de données

## → Enrichir et actualiser Biologics

**en s'appuyant sur les publications et autres bases de données de référence**

Mise à jour des taxons de la base tenant compte des évolutions de la systématique et de la taxinomie (nouveaux taxons, synonymies, nouvelles combinaisons génériques, etc.)



**Référentiel taxinomique pour la faune et la flore de France métropolitaine et Outre-mer**

<https://taxref.mnhn.fr/taxref-web/accueil>



version v17 publiée le 10/01/2024

**Focus sur les insectes de France métropolitaine et d'Outre mer**

*Mise à jour très fastidieuse manuellement (via tableaux Excel)*



**Développement d'un pipeline efficace avec R afin d'intégrer TaxRef de façon semi-automatique**

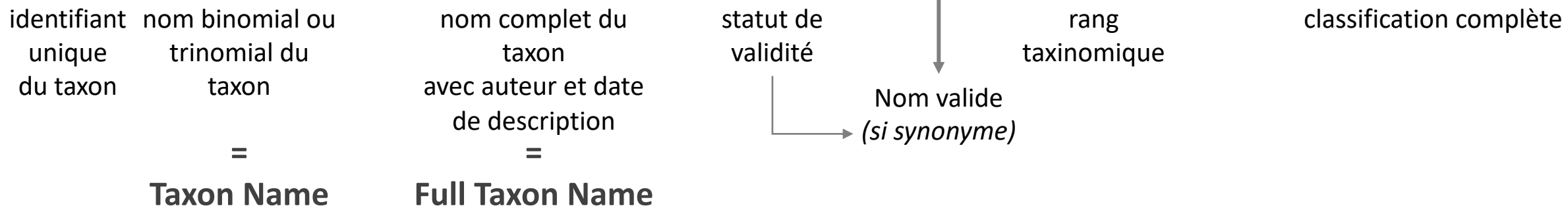


**Travail débuté au printemps 2023**

# Visualisation des champs concernés par la curation

ID	Taxon name	Complete taxon name	Taxon validity	Senior synonym.Taxon name	Taxonomic rank	Classification
334939	Mantoididae	Mantoididae Chopard, 1949	nomen valid	-	Family	Animalia, Eumeta:
332564	Parametriotidae	Parametriotidae Capuše, 1971	junior synonym	Parametriotinae	Family	Animalia, Eumeta:
336589	Phylliidae	Phylliidae Brunner von Wattenwyl, 1893	nomen valid	-	Family	Animalia, Eumeta:
375355	Hastius	Hastius Schmitz, 1946	nomen valid	-	Hyporder	Animalia, Eumeta:
376211	Archiborborus setosus	Archiborborus setosus Duda, 1921	nomen valid	-	species	Animalia, Eumeta:
376212	Archiborborus simplicimanus	Archiborborus simplicimanus Richards, 1931	nomen valid	-	species	Animalia, Eumeta:
376213	Archiborborus submaculatus	Archiborborus submaculatus Duda, 1921	nomen valid	-	species	Animalia, Eumeta:
375884	Ceratitis hamata	Ceratitis (Pardalaspis) hamata De Meyer, 1996	nomen valid	-	species	Animalia, Eumeta:
375885	Ceratitis mlimaensis	Ceratitis (Ceratalaspis) mlimaensis De Meyer, 1998	nomen valid	-	species	Animalia, Eumeta:
112399	Labops sahlbergii	Labops sahlbergii (Fallén, 1829)	nomen valid	-	species	Animalia, Eumeta:
375926	Lophogaster seyrigi	Lophogaster seyrigi Granger, 1949	junior synonym	Ivondrovia seyrigi	species	Animalia, Eumeta:
376137	Percyella guanche	Percyella guanche Bastin, Burckhardt & Ouvrard, 2023	nomen valid	-	species	Animalia, Eumeta:
112362	Perilitus marci	Perilitus marci Haeselbarth, 1999	nomen valid	-	species	Animalia, Eumeta:
375916	Schilskyana effusa	Schilskyana effusa (Majer, 1989)	nomen valid	-	species	Animalia, Eumeta:
375898	Ceratitis (Hoplolophomyia)	Ceratitis (Hoplolophomyia) Bezzi, 1926	nomen valid	-	Subgenus	Animalia, Eumeta:
375261	Zophosis (Protodactylus)	Zophosis (Protodactylus) Koch, 1952	nomen valid	-	Subgenus	Animalia, Eumeta:
372873	Monocrepidius posticus var. sticturus	Monocrepidius posticus var. sticturus Candèze, 1859	nomen valid	-	Variety	Animalia, Eumeta:
374702	Nepticula atricollis var. prunivora	Nepticula atricollis var. prunivora Skala, 1941	nomen nudum	Ectoedemia atricollis	Variety	Animalia, Eumeta:
373713	Nepticula lonicerarum var. lentinensis	Nepticula lonicerarum var. lentinensis Skala, 1935	junior synonym	Stigmella lonicerarum	Variety	Animalia, Eumeta:
373792	Nepticula plagicolella var. malicola	Nepticula plagicolella var. malicola Skala, 1939	nomen nudum	Stigmella sorbi	Variety	Animalia, Eumeta:

ustacea, Hexapoda, Insecta, Dictyoptera, Mantodea  
 , Lepidoptera, Glossata, Heteroneura, Gelechioidea  
 matodea, Verophasmatodea, Areolatae, Phyllioidea  
 alcidoidea, Chalcididae, Haltichellinae, Haltichellini  
 idea, Sphaeroceridae, Copromyzinae, Archiborborus  
 idea, Sphaeroceridae, Copromyzinae, Archiborborus  
 idea, Sphaeroceridae, Copromyzinae, Archiborborus  
 Jacini, Ceratitidina, Ceratitis, Ceratitis (Pardalaspis)  
 acini, Ceratitidina, Ceratitis, Ceratitis (Ceratalaspis)  
 a, Miroidea, Miridae, Orthotylinae, Halticini, Labops  
 uryctinae, Holcobraconini, Ivondroviina, Lophogaster  
 era, Sternorrhyncha, Psylloidea, Triozidae, Percyella  
 onioidea, Braconidae, Euphorinae, Perilitini, Perilitus  
 Aelyridae, Dasytinae, Chaetomachnini, Schilskyana  
 . Tephritidae, Dacinae, Dacini, Ceratitidina, Ceratitis  
 lea, Tenebrionidae, Pimeliinae, Zophosini, Zophosis  
 i, Elateridae, Agrypninae, Oophorini, Monocrepidius  
 loidea, Nepticulidae, Nepticula, Nepticula atricollis  
 lea, Nepticulidae, Nepticula, Nepticula lonicerarum  
 dea, Nepticulidae, Nepticula, Nepticula plagicolella



# Démarche de curation des données taxinomiques de Biologics



Biologics

TaxRef

TaxRef

## Sélection des Insectes de France et d'Outre Mer

318 203 lignes



Table Taxon

v.17



Importation dans

R

**Mise à l'écart des taxons pour lesquels :**

**Problèmes taxinomiques non résolus :**

- absence de descripteur (ex : *Lysandra albicans x caelestissima*)
- date de publication inconnue (ex : *Aphis dragocephalus*)

**Informations incomplètes associées aux échantillons du Plateau Collection**

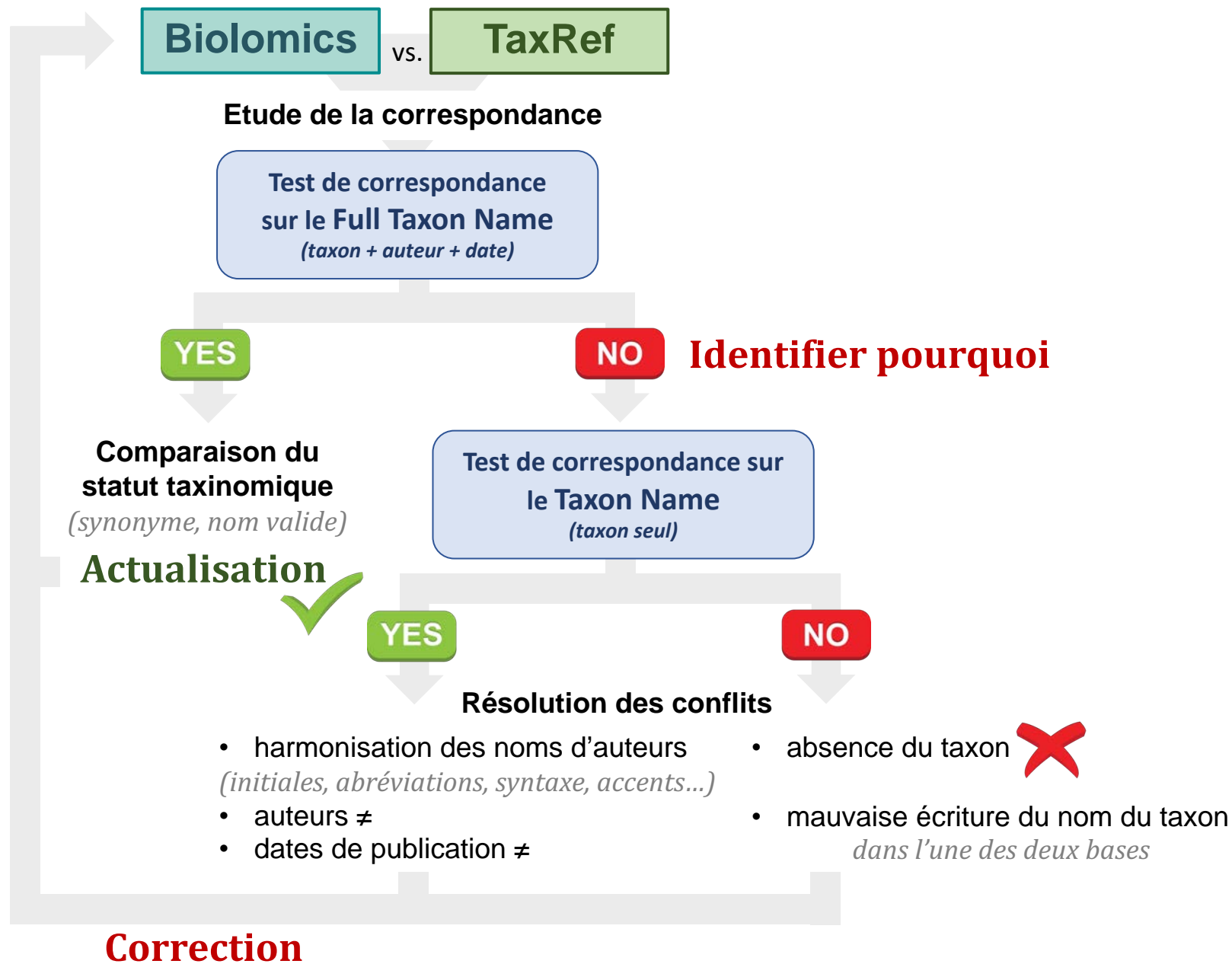
- identification en cours (ex : *Amasa* sp.)
- identification incomplète (ex : *Gryllotalpa cf. debilis* Gerstaecker, 1869)
- description en cours de nouvelles espèces (ex : *Cohicaleyrodes* n. sp. 1)

Correction des coquilles

- Recherche d'enregistrements doublons
- Harmonisation des expressions (ex : *abréviations, virgule entre auteur et date*)
- Correction des coquilles (ex : *espaces manquants, parenthèses et autres caractères surnuméraires*)
- Vérification de la nature des informations dans les champs concernés (ex : *présence de l'auteur-date dans Full Taxon Name*)
- Vérification de la cohérence des informations entre les champs (ex : *Taxon Name cohérent avec la classification*)

Harmonisation des champs pour  
comparaison des deux bases de données

# Démarche d'actualisation des données taxinomiques de Biologics





## → Terminer la correction globale de la base

- Corriger les dernières coquilles
- Correction / Harmonisation des noms d'auteurs

## → Comparer Biologics à la dernière version de TaxRef

- Mettre à jour les statuts des taxons
- Ajouter à Biologics les taxons de TaxRef qui en sont absents

## → Communiquer à TaxRef les éventuelles coquilles ou erreurs que nous avons corrigé afin qu'ils mettent à jour leur base de données

### Objectif final

Intégrer chaque année  
de façon semi-automatique dans  
Biologics les mises à jour de TaxRef

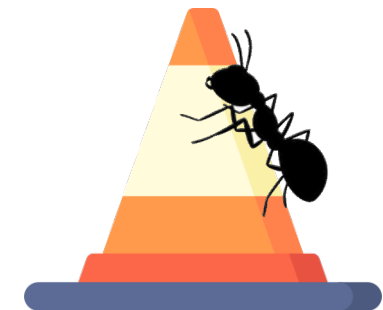
~ **8000**

*Taxons corrigés en 2023 !*

**27%**

*Correspondance actuelle  
Biologics – TaxRef sur le  
Full Taxon Name*

*avant harmonisation des auteurs*



**WORK in  
PROGRESS**



**Merci de votre attention !**

