



HAL
open science

BioloMICS : Un outil pour gérer les données associées aux collections des Centres de Ressources Biologiques

Emmanuelle Artige, Cécile Grondin, Jonathan Mineau, Michel-Yves Mistou

► **To cite this version:**

Emmanuelle Artige, Cécile Grondin, Jonathan Mineau, Michel-Yves Mistou. BioloMICS : Un outil pour gérer les données associées aux collections des Centres de Ressources Biologiques. NOV'AE, 2022. hal-03859206

HAL Id: hal-03859206

<https://hal.science/hal-03859206>

Submitted on 16 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BioloMICS : Un outil pour gérer les données associées aux collections des Centres de Ressources Biologiques

Emmanuelle ARTIGE¹

Cécile GRONDIN²

Jonathan MINEAU³

Michel-Yves MISTOU⁴

CORRESPONDANCE

michel-yves.mistou@inrae.fr

emmanuelle.artige@inrae.fr

RÉSUMÉ

Le logiciel commercial Biologics permet de créer une base de données relationnelle et de déployer un site web associé. Il a été choisi par plusieurs CRB de RARe pour la gestion de leurs collections. Nous faisons ici une revue extensive des fonctionnalités du logiciel, des avantages de la solution et de ses limites.

MOTS-CLÉS

Biologics, logiciel, collection, base de données, site web.

1 CBGP, INRAE, CIRAD, Institut Agro, IRD, Univ Montpellier, Montpellier, France.

2 CIRM-Levures, SPO, Univ Montpellier, INRAE, Institut Agro, Montpellier, France.

3 CIRM, SPO, Univ Montpellier, INRAE, Institut Agro, Montpellier, France.

4 CIRM, Université Paris-Saclay, INRAE, UMR 1404 MalAGE, Jouy-en-Josas, France.

BioloMICS: a tool for generating data associated with the collections of the Biological Resource Centres

Emmanuelle ARTIGE¹

Cécile GRONDIN²

Jonathan MINEAU³

Michel-Yves MISTOU⁴

CORRESPONDENCE

michel-yves.mistou@inrae.fr

emmanuelle.artige@inrae.fr

ABSTRACT

The commercial software BioloMICS allows creating a relational database and deploying an associated website. It was chosen by several BRCs of the RARe to manage their collections. Here, we provide an extensive review of the software's functionalities, the advantages of the solution and its limits.

KEYWORDS

BioloMICS, software, collection, database, website.

1 CBGP, INRAE, CIRAD, Institut Agro, IRD, Univ Montpellier, Montpellier, France.

2 CIRM-Levures, SPO, Univ Montpellier, INRAE, Institut Agro, Montpellier, France.

3 CIRM, SPO, Univ Montpellier, INRAE, Institut Agro, Montpellier, France.

4 CIRM, Université Paris-Saclay, INRAE, UMR 1404 MalAGE, Jouy-en-Josas, France.

Introduction

Les centres de ressource biologiques (CRB) doivent fournir à leurs utilisateurs des échantillons identifiés selon des standards qui évoluent avec les développements technologiques, scientifiques et réglementaires. La multiplication des informations associées aux ressources biologiques (métadonnées, données moléculaires et phénotypique, conformité réglementaire, sécurité biologique, images, etc.), leur mise à jour et leur visibilité imposent aux CRB à la fois de mettre en place des outils dédiés pour la gestion de leurs catalogues et de développer de nouvelles compétences. Pour les utilisateurs, il s'agit d'avoir un accès simple aux informations associées aux ressources pour sélectionner celles qui correspondent à leurs besoins.

L'investissement financier et humain pour le développement et la maintenance d'un logiciel ou d'un Système d'information (SI) a décidé beaucoup de Centre de ressources Biologiques (CRB) à fonctionner avec des solutions « clefs en main ». Certaines de ces solutions sont soit développées au niveau des instituts publics, soit issues d'approches collaboratives et communautaires ; elles peuvent donc être libres de droit et d'utilisation. D'autres offres sont issues du secteur privé et souvent commerciales. (Liste non exhaustive en annexe)

Une réflexion a donc été engagée dans les CRB pour identifier les outils informatiques qui permettront d'afficher, d'explorer et d'exploiter de manière optimale ces informations, et de renforcer l'attractivité des collections. Parmi ces solutions : BioloMICS, objet du présent article.

Au niveau européen, BioloMICS a été choisi comme système d'information pour l'infrastructure de recherche MIRRI (MIcrobial Ressources Research Infrastructure) qui fédère 14 CRB microbiens de 10 pays de l'UE (dont la France, représentée par INRAE-CIRM et l'Institut Pasteur). Actuellement près de 200 000 souches sont enregistrées et accessibles (500 000 à terme - https://catalog.mirri.org/page/Strains_catalog_query).

Dans le contexte de RARE, cette solution a aussi été choisie depuis 2020 par le CRB Egg Parasitoids Collection (CRB EP-Coll) [<https://www6.inrae.fr/crb-eggparasitoids-coll/>], et depuis 2010 par le CBGP (UMR Centre de Biologie pour la Gestion des Populations) pour gérer la Continental Arthropod Collection [<https://doi.org/10.15454/D6XAKL>].

Nous présentons ici le retour d'expérience sur l'utilisation de BioloMICS par la Collection de souches au CIRM-Levures et la Collection d'Arthropodes continentaux du CBGP.

Présentation de BioloMICS

BioloMICS est une suite d'outils dont les deux principaux sont BioloMICS.net et BioloMICS.web. C'est un système de gestion de données biologiques utilisé par de nombreux acteurs dans les domaines de la biotechnologie, de l'industrie pharmaceutique, de l'alimentation, des services de santé publique, des institutions académiques, universités, collections et musées.

BioloMICS est produit par la société BioAware dont le siège est basé en Belgique. Les interlocuteurs sont francophones et anglophones. C'est un « software as a service » (SaaS) ou logiciel en tant que service, c'est à dire qu'il est installé sur des serveurs distants plutôt que sur une machine utilisateur. L'accès au logiciel en version desktop est basé sur le protocole RDP (Remote Desktop Protocol). La licence est annuelle et pour une seule instance, mais la connexion peut être multi-utilisateurs. Selon le type de licence choisie, le client aura accès ou non au module de génération de sites web (BioloMICS.Web)⁴. Les licences annuelles achetées sont flottantes, c'est à dire qu'elles peuvent être utilisées de n'importe quel ordinateur, mais une licence ne peut être utilisée par deux utilisateurs concomitamment.

Module BioloMICS.net

L'interface utilisateur de BioloMICS.net est « user-friendly » et inspirée de l'environnement Microsoft. Cette interface

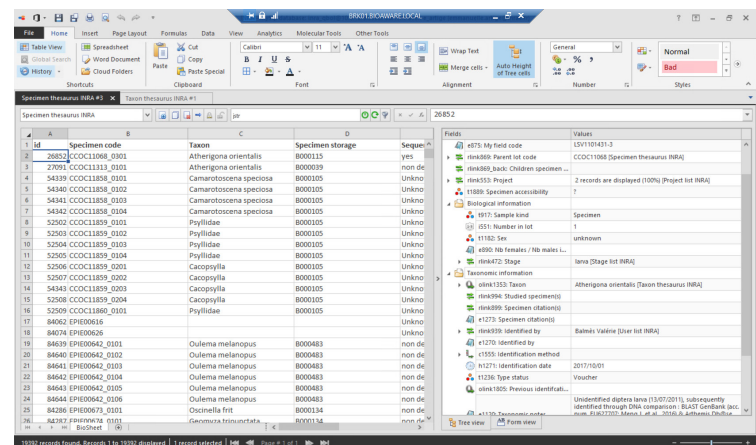


Figure 1. Capture d'écran de l'interface de BioloMICS.net.

Credit : Collection d'Arthropodes Continentaux - CBGP

permet ainsi à toute personne, informaticienne ou non, de s'approprier le logiciel et de construire sa base de données. L'outil est ergonomique, flexible et modulable, et

4 BioAware SA NV. BioloMICS Software pricing. <https://www.bio-aware.com/page/Pricing> [consulté le 11/08/2021]

peut être utilisé à différentes finalités (stockage, gestion, analyse) pour les ressources biologiques, les expériences de recherche ou les données biologiques de tous types (morphologie, physiologie, biochimie, chimie, chromatographie, électrophorèse, moléculaire, taxonomie, bibliographie, géographie, écologie, gestion du cycle de vie de l'échantillon,...).

Différents outils sont mis à disposition de l'utilisateur pour gérer et analyser les données de sa base, par exemple :

- Des outils moléculaires : enregistrement de modifications de séquences brutes, alignement par paire ou multiple de séquences, récupération automatique des profils de poids moléculaire des échantillons en utilisant l'analyse par électrophorèse sur gel.
- Pour une gestion de laboratoire encore plus intégrée, un module de système de gestion de l'information du laboratoire (LIMS) est aussi disponible ainsi qu'un outil de gestion de stockage des échantillons dans les congélateurs et autres dispositifs.
- Un outil d'analyse spectrale pour l'analyse Maldi-Tof.
- Un outil de visualisation par structuration hiérarchique des données permettant de déplacer les enregistrements en drag and drop (« cliquer et déplacer ») qui peut s'avérer utile lors de la gestion des révisions taxonomiques.
- Un outil de génération de graphiques comme sous Excel.
- Un outil de génération de distribution géographique sur carte des échantillons ayant des données GPS (style Google Earth).
- La création de tableaux de bord (« dashboard ») pour créer des affichages multiformats des données extraites (cartes, tableaux, camemberts, histogrammes, etc.). Ces tableaux de bord peuvent être publiés sur les sites web.

- L'automatisation de tâches répétitives via un module de scripts en C# et VB.

Module BioloMICS.web

BioAware enrichit son offre avec BioloMICS.web qui permet le déploiement d'une interface web à partir de la base de données. Cette fonctionnalité permet de sélectionner et de rendre disponible en ligne les ressources souhaitées d'une manière sécurisée et à jour.

BioloMICS.web permet la diffusion des données contenues dans BioloMICS.net avec une interface en WYSIWYG (What You See Is What You Get). Quelques notions en langage HTML pour la création des pages d'information textuelle peuvent tout de même être utiles. L'outil autorise (via l'application) le filtrage de données, l'extraction et le formatage d'une partie des données spécifiques à un projet, un groupe taxonomique ou tout autre filtre désiré par l'utilisateur. Il génère alors automatiquement une structure de site web dans laquelle des outils de recherche permettent d'accéder aux informations filtrées de la base de données.

Plusieurs sites web peuvent être générés à partir d'une même base de données.

Sécurité, stockage et hébergement des données

Pour les aspects sécurité des données, BioAware sous-traite des tests de pénétrations à EASI, une société experte en audit sécurité des infrastructures informatiques. Tous les serveurs sont protégés par la solution SentinelOne. Les serveurs sont sauvegardés intégralement tous les jours et les copies conservées une semaine. Tout le système est dupliqué pour garantir la disponibilité. Une version de sauvegarde est activable en 2 heures maximum en cas

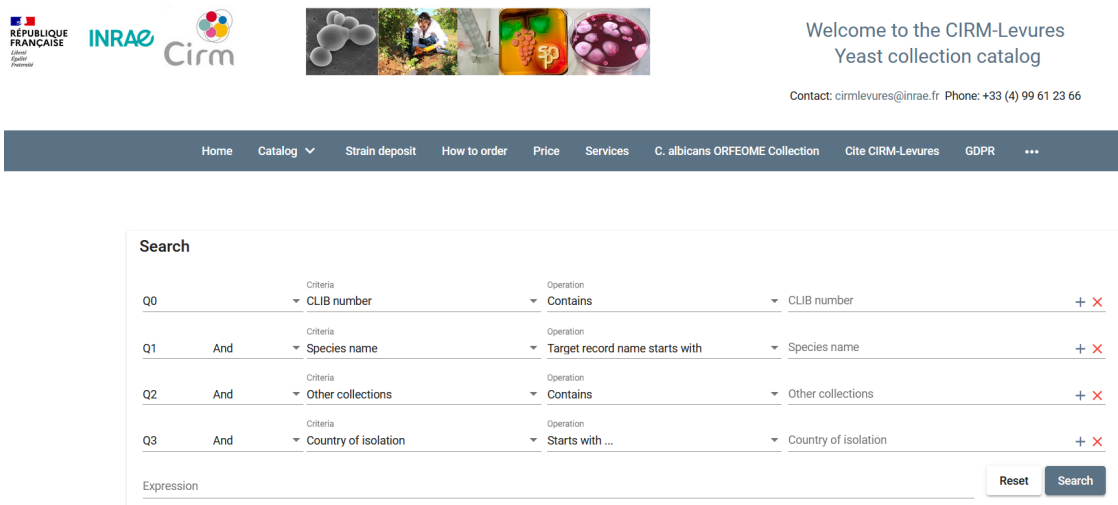


Figure 2. Capture d'écran de l'interface de BioloMICS.web (<https://cirm-levures.bio-aware.com/>)

d'effondrement du site principal (Site Recovery option) et également disponible via un autre datacenter. Une archive de chaque database client est créée journalièrement et sauvegardée sous format compressée (backup MySQL), pendant une période de 30 jours, sur les serveurs de BioAware. Cette archive est également déposée quotidiennement dans un service DropBox et mise à la disposition du client (les 5 versions les plus récentes sont gardées). Si le partenariat s'arrête, la société s'engage à restituer l'intégralité des données sous format ouvert (e.g. json/csv/Excel/MySQL/MongoDB) ou bien dans le format demandé pour exportation dans une autre base de données (avec schéma de l'architecture.)

Concernant l'hébergement des serveurs et donc des données, BioAware utilise les solutions Microsoft-Azure. La zone géographique de résidence est située en Europe (West Europe – Netherlands).

Pour les aspects d'administration, le système enregistre tous les changements dans un historique consultable. Toutes les actions, y compris celles entraînant des modifications de la structure de la base de données peuvent être annulées par les administrateurs ou l'opérateur. La restauration de la base à une date précise est possible en utilisant les sauvegardes disponibles. Le client peut demander à tout moment, ou à une fréquence souhaitée, une sauvegarde de la base de données.

Gestion des droits utilisateurs

L'accès par les utilisateurs à la base de données ou aux interfaces est géré par l'administrateur via un outil de gestion des utilisateurs. Cet outil permet de créer des groupes et d'affecter les droits (écriture, lecture, suppression) aux ressources sélectionnées (champ, table, type de données, champs et enregistrements).

Données à caractère personnel

BioAware garantit le respect du règlement général sur la protection des données (RGPD) concernant ses propres données clients. Pour ce qui est de la nature des données stockées dans le système par le client, cela n'engage que ce dernier, néanmoins le système propose une interface de gestion par le biais de deux rôles administrateurs (« Controller » et « Processor ».) Les fonctionnalités sont détaillées sur le site BioAware (<https://www.bio-aware.com/page/GDPR>).

Retours d'expérience des utilisateurs

Collection de souches au CIRM-Levures

Notre choix d'utiliser le logiciel BioloMICS pour notre base

de données regroupant les données associées aux souches de la collection du CIRM-Levures (2 600 souches) s'est fait, en 2016, après le départ de notre informaticien. Ce dernier, avait développé une base de données « maison » dont aucun autre informaticien de l'unité ne connaissait l'architecture, ce qui rendait difficile sa maintenance ; les soucis se sont donc accumulés jusqu'à rendre cette base non-incrémentable, en 2016. C'est ainsi que l'équipe du CIRM-Levures a décidé de passer par un prestataire pour créer une nouvelle base de données.

BioloMICS nous est apparu comme étant la meilleure solution ; en effet, il permet de créer une base donnée sur mesure et flexible, plusieurs formats de fichiers sont acceptés et il y a un historique permettant une traçabilité de tous les changements effectués et une gestion facile des droits d'accès aux différentes données, ce qui est primordial pour un CRB. Notre choix s'est porté sur BioloMICS aussi car le personnel du CIRM-Levures n'ayant pas les compétences requises, il fallait un logiciel utilisable par des non-informaticiens et de l'aide pour créer la nouvelle architecture de la base. D'autres points ont été importants dans choix. BioloMICS est le support choisi pour la version informatique de « The Yeast », la référence bibliographique pour la taxonomie des levures dans laquelle certaines données du CIRM-Levures seront intégrées. Et BioloMICS a aussi été choisi pour la base de données de la Microbial Resource Research Infrastructure (MIRRI) qui est une Infrastructure paneuropéenne pour la préservation, la recherche systématique, la fourniture et la valorisation des ressources microbiennes et de la biodiversité. Ainsi, avoir nos données dans BioloMICS facilite grandement le transfert vers ces deux bases de données.

Après une formation en plusieurs étapes pour faciliter notre apprentissage, l'équipe de BioAware nous a aidés en nous fournissant l'architecture de la base de données de « The Yeast ». Nous avons été accompagnés dans le large choix des types de champs que nous voulions créer et dans le transfert des données extraites de notre ancienne base dans la nouvelle. Au CIRM-Levures, deux groupes d'utilisateurs ont été créés : ceux ayant les droits de modifier la base de données et ceux qui ont juste un droit de consultation des données. Cela permet de limiter les erreurs possibles. Nous avons choisi de rentrer directement nos données dans la base sans passer par des fichiers d'importation.

Une fois l'ensemble des données transférées, nous avons pu passer au chantier du site internet associé à notre base de données. Ce dernier servant de catalogue en ligne permet aux clients de la collection de consulter les données associées aux souches que nous avons choisi de rendre

visibles (<https://cirm-levures.bio-aware.com/>). Pour cette nouvelle étape, nous avons aussi été accompagnés.

Dès que les étapes précédentes ont été terminées, nous avons pu facilement créer de nouveaux champs pour augmenter les données associées à nos souches consultables sur notre catalogue en ligne ou pour lier aux souches tous les documents les concernant pour une meilleure traçabilité ou pour s'adapter à de nouvelles réglementations comme celles liées au Protocole de Nagoya. En changeant certains types de champs, nous avons aussi rendu notre catalogue plus pratique, fonctionnel et agréable visuellement (intégration d'une carte du lieu de collecte, liste des souches d'une même espèce, d'un même pays, lien vers les publications, les génomes disponibles, documents PDF téléchargeables concernant le Protocole de Nagoya, ...). En effet, la flexibilité de BioloMICS pour la création de tables et de champs permet de monter en compétence à son rythme et de faire évoluer selon nos souhaits la structure de notre base de données et les types de champs.

À la création de notre base de données, nous avons commencé avec une grande majorité de champs texte, faciles à remplir, mais qui entraînent plus facilement des erreurs et doublons dus à des fautes de frappe ou d'orthographe. Nous sommes donc actuellement en train de les remplacer pour limiter les erreurs et pouvoir modifier rapidement une information utilisée plusieurs fois dans la base de données. Nous avons également établi des liens avec d'autres bases de données comme MycoBank, qui est pour les champignons la base de données taxonomiques de référence. Ce lien permet à nos enregistrements taxonomiques d'être automatiquement mis à jour lorsque la taxonomie des espèces est modifiée dans MycoBank. Nous avons aussi créé une table Bibliographie qui permet de relier les articles scientifiques à nos souches avec un lien vers la base de données PubMed ou vers d'autres journaux. Chaque lien créé vers une autre base de données est aussi affiché sur notre catalogue en ligne pour un accès rapide aux données taxonomiques et aux articles scientifiques pour nos clients. Nous avons aussi commencé la création d'un tableau de bord qui sera utilisé pour avoir une vision globale de notre collection. Ses sorties graphiques permettront d'illustrer notre site internet et la plaquette destinée à nos clients.

Collection d'Arthropodes Continentaux au CGBP

La Collection d'Arthropodes Continentaux (INRAE-CIRAD-Institut Agro-IRD), hébergée au CGBP, est constituée et utilisée en interne par des entomologistes comme outil

de diagnostic morphologique et génétique et comme témoin écologique et de la biodiversité. Au recrutement de la responsable technique du plateau « Collections » en 2010, les données étaient dispersées dans différents ordinateurs (sous Excel, FileMaker, Access) sans homogénéisation des données. Un de ses premiers objectifs était donc de trouver un système de gestion pour centraliser et structurer les informations sur les spécimens.

Intégrée en 2011 au programme européen FP7 Qbol sur les Organismes de quarantaine, la responsable a pu tester et prendre en main BioloMICS pour gérer les données taxonomiques, biologiques et génétiques des organismes de quarantaine européenne dont le CGBP était un des principaux acteurs en liaison avec l'Unité de recherche de Zoologie Forestières d'Orléans et le NVWA⁵ des Pays Bas pour les Arthropodes. Depuis, cette base a été déployée pour gérer l'intégralité de la Collection d'Arthropodes Continentaux du CGBP car elle répondait à différentes attentes : i) gérer une importante collection inerte d'arthropodes (environ un million de spécimens), ii) gérer d'importantes données taxonomiques régulièrement révisées (environ 60 000 espèces appartenant à une dizaine d'ordres), iii) intégrer des séquences génétiques barcodes COI, ITS, CythB... et proposer un outil de blast, iv) rendre accessible les données à l'extérieur, v) travailler sur une structure facilement modulable et ergonomique, vi) bénéficier d'une équipe de développement du logiciel professionnelle, à l'écoute et disponible (temps de réponse entre 1 minute et 24 heures).

Nous fonctionnons à ce jour avec 2 licences Ultimate. Une des deux licences est partagée entre les différents utilisateurs via un système de réservation interne. La deuxième licence est réservée à la super-administratrice pour déboguer les problèmes et répondre rapidement à une demande d'amélioration. Le CRB prend à sa charge chaque année une licence, la deuxième étant liée à des crédits qu'il nous faut trouver tous les ans (appel d'offre, projets, etc.).

La structure de la base Arthropodes a largement évolué depuis sa mise en place en 2011, et s'articule autour de quatre tables principales : Specimen, Taxonomie, Stockage et Sequences (Figure 3). Des tables, champs sont ajoutés, supprimés, modifiés pour plus de pragmatisme et en fonction des besoins ; ces manipulations faites par la super-administratrice se révèlent être simples.

Depuis 2016, nous mettons à dispositions d'autres unités de recherche utilisant BioloMICS nos tables Botanic et Taxon Arthropods. Ainsi, ces tables, développées et mises

5 The Netherlands Food and Consumer Product Safety Authority.

Arthropods Database structure in BioLMICS

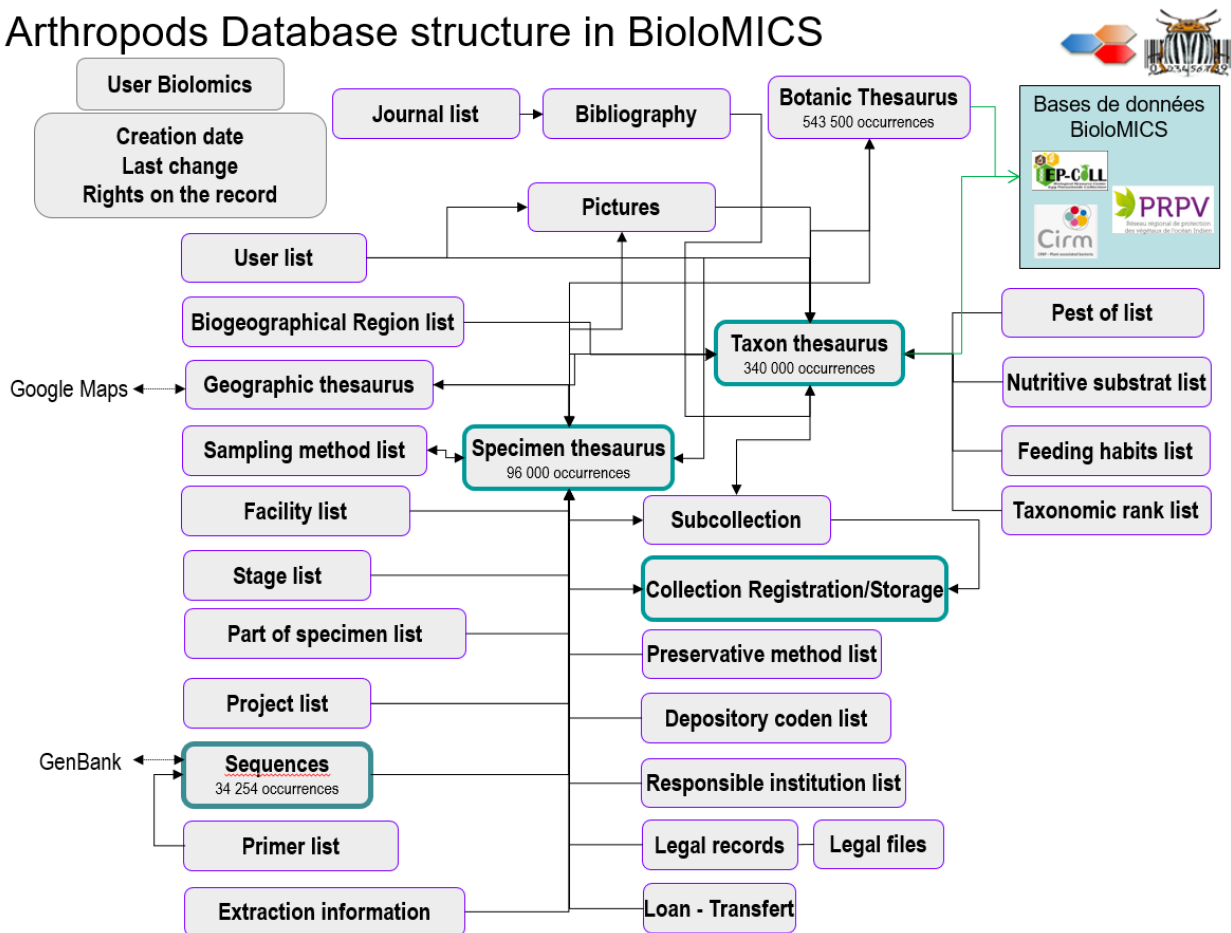


Figure 3. Structure de la base de données Arthropodes – CBGP : visualisation des tables et leurs liaisons.

à jour au CBGP, sont utilisées par la base PRPV du Pôle de la protection des plantes de La Réunion⁶, celles du CRB Ep-Coll⁷ et du CIRM-CFBP⁸ pour incrémenter leurs propres bases de données. Seuls les droits de lecture ont été pour l'instant accordés à ces unités, mais une collaboration plus poussée pourrait naître si des spécialistes de ces domaines voulaient coopérer.

L'équipe BioAware a répondu au fil des années à de nombreuses demandes d'amélioration de notre part comme la gestion en ontologie des enregistrements dont nous nous servons aujourd'hui pour gérer nos tables Taxon, Botanic, Collection Registration et Geography. Ainsi, nous pouvons par exemple par un simple glisser-déplacer d'une boîte d'échantillons sous une visualisation hiérarchique changer la localisation de tous les échantillons inclus.

Deux personnes sont référentes pour la base : une super-administratrice, chargée de la structuration et des droits, et un

administrateur, responsable des tables Taxonomie et Botanique ainsi que de l'intégration de données pour différents projets. Ces personnes sont les interlocuteurs privilégiés des utilisateurs internes et de l'équipe de BioAware en cas de problème ou de demande d'amélioration. Une dizaine de personnes (chercheurs et personnel technique) ont été formées en interne au CBGP et ont acquis différents droits en fonction de leur implication : lecture, ajout/modification/suppression des données, intégration en lot sur toutes ou certaines tables, administration de certaines tables, voire de certains champs, dans les tables. Les utilisateurs sont ainsi rattachés à des groupes dont les droits sont clairement définis.

Des « templates » (modèle de fichier) Excel standardisés, établis à destination des gestionnaires de collections et des porteurs de projets, permettent *in fine* l'importation en lots des données dans la base via un système de mapping (templates Specimen, Extraction, Sequences et Images).

6 Pôle de la protection des plantes, Saint Pierre de la Réunion. Réseau régional de protection des végétaux de l'océan Indien. <https://db.e-prpv.org/>.

7 Biological Resource Centre "Egg Parasitoids Collection - Ep-Coll" (<https://doi.org/10.15454/AY4LMT>). INRAE-Sophia-Antipolis.

8 CIRM-CFBP : Collection Française de Bactéries associées aux Plantes, INRAE, Angers-Beaucouzé.

BioLoMICS permettant de gérer facilement et finement les droits de diffusion, les porteurs de projets peuvent choisir de mettre leurs données sous embargo pour qu'elles soient en accès restreint jusqu'à la fin du projet de recherche. La sauvegarde des données de la base archivée pendant deux mois sur les serveurs de BioAware est doublée d'un back-up sql tous les 6 mois et d'une sauvegarde annuelle des données des tables en CSV sur les serveurs du CBGP.

Le module BioLoMICS.WEB nous permet de générer des sites internet ciblés ; ainsi si Arthemis DB@se⁹ rend accessible les données d'une grande diversité d'espèces d'arthropodes, notamment les ravageurs et les auxiliaires des cultures, PhylAphid DB@se¹⁰ est dédié aux données Pucerons. Sur ces sites, nous mettons à disposition du public des informations non confidentielles : données passeport de nos ressources, séquences validées d'ADN produites. Même si nous partageons les séquences barcodes validées sur Genbank, il est important pour le CBGP de gérer et de mettre à disposition un outil d'identification moléculaire spécialisé sur les arthropodes terrestres ne regroupant que des séquences validées par des experts en entomologie selon des procédures définies¹¹ afin d'éviter l'écueil de séquences incorrectement assignées à un groupe taxonomique, ou alors des séquences rattachées à un niveau taxonomique grossier. Ainsi, nous proposons à l'utilisateur d'utiliser l'outil de « BLAST » lui permettant de soumettre une séquence pour la comparer aux séquences de notre base et/ou de Genbank, d'afficher des arbres phylogénétiques sous différentes méthodes de reconstruction (UPGMA, WPGMA, Single linkage, Complete linkage, Neighbor joining, Ward, UPGMC, WPGMC) et d'avoir accès aux données du spécimen et de la séquence.

La possibilité que donne BioLoMICS de générer des sites WEB multiples, ciblés sur un projet ou un groupe taxonomique à partir d'une base de données unique, est un outil important de communication et de diffusion de l'information vers la communauté scientifique. Il nous permet de diffuser en temps réel les données de la base mais nécessite, là-aussi, un investissement en temps pour apprivoiser le module de création de site web.

Google Analytics fourni par BioAware nous permet de mesurer l'audience des sites web générés.

Synthèse des retours utilisateurs

En conclusion, pour nous, le système d'information dédié aux ressources biologiques BioLoMICS offre des services de qualité, car il est en constante évolution (de nombreuses options n'ont pas encore été testées faute de temps et d'un nombre de licence insuffisant). Ce système nous permet de gérer sereinement le cycle de vie de nos échantillons, de les référencer et de les valoriser grâce au module de création de sites web. Ce système ergonomique et facile d'utilisation pour des non-informaticiens nécessite toutefois l'implication de référents, en tant qu'administrateurs de la base et premiers interlocuteurs des utilisateurs, et de la société BioAware ; sans eux, la base de données ne pourrait être maintenue tant structurellement que pour l'implémentation des données. Un atout de ce produit réside aussi dans la réactivité du personnel de la société BioAware.

Intérêts de la solution

- Une gestion de la structuration de la base de données en WYSIWYG (What You See Is What You Get) complète et facile : possibilité d'ajout, de modification, de mise à jour ou de suppression de tables, champs, enregistrements, groupe d'utilisateurs et d'utilisateurs.
- Une gestion des droits des utilisateurs simple et à la carte : chaque utilisateur reçoit des droits de lecture, d'écriture et de suppression sur les tables, les champs et les enregistrements.
- Un historique des modifications permet à l'utilisateur ou à l'administrateur d'annuler des modifications en un clic.
- Un SAV disponible, compétent et un accompagnement pas à pas appréciable, notamment pour le développement des bases de données et des sites web.
- Une formation gratuite en français ou en anglais.
- Une aide en ligne en anglais régulièrement mise à jour.
- Des requêtes avancées avec des opérateurs booléens et logiques peuvent être effectuées, ce qui permet des recherches fines.
- La possibilité de créer ses propres vues sur des tables données. Cela permet de n'afficher que certains champs pour une meilleure visualisation de l'information.
- Une exportation et une importation sous des formats communs et standardisés : texte délimité, html, XML, MS-Excel ou MS-Word, Fasta, etc.

9 INRAE-CBGP, 2021, «Arthemis database», <https://doi.org/10.15454/TBGRIB>. [online database accessed 2021-08-09] : <https://arthemisdb.supagro.inra.fr/>.

10 Coeur d'acier et al., 2014. DNA barcoding and the Associated PhylAphidB@se Website for the identification of European Aphids (Insecta: Hemiptera: Aphididae). Plos One, 9 (6) to use the database in a publication. DOI:10.1371/journal.pone.0097620.

11 European and Mediterranean Plant Protection Organization [OEPP/EPPO], PM 7/129 (2) DNA barcoding as an identification tool for a number of regulated pests 2021, Bulletin 51, 100-143, ISSN 0250-8052. DOI: 10.1111/epp.12724.

- L'importation de grandes quantités de données en une seule fois est possible.
- Les liens vers d'autres bases de données BioloMICS : une table peut être partagée par plusieurs comptes clients différents.
- Participation au développement de BioloMICS : les souhaits des clients sont pris en compte dans les futurs développements du système si le même souhait est remonté par plusieurs clients.

Points de vigilance

- Le coût annuel de la licence représente une somme importante pour certains laboratoires.
- Les requêtes ne peuvent se faire qu'entre tables liées directement. De ce fait, la structuration de la base relationnelle doit prendre en compte cette limite. De plus, il est complexe d'exporter directement le schéma d'une base de données relationnelle multi-tables dans un autre système.
- Un investissement d'une ou de plusieurs personne(s) du laboratoire est indispensable pour pouvoir concevoir, maintenir et développer les bases de données. Ces personnes ressources sont les interlocuteurs privilégiés avec BioAware.
- La sensibilisation et la formation des utilisateurs doivent être intégrées comme bonnes pratiques.
- La production et l'utilisation d'outils standardisés (e.g. fichier Excel formaté) est indispensable pour l'intégration de données venant de multiples sources.
- BioAware est une PME fondée et animée par un scientifique et l'évolution de la société à moyen terme est indéterminée.
- L'évolution des conditions d'accès sécurisés aux serveurs qui vont limiter la flexibilité d'utilisation par plusieurs utilisateurs au sein d'un CRB.

Conclusion

Nous avons vu que le logiciel BioloMICS est bien adapté à l'usage des CRB pour la gestion et la visibilité de leur catalogue. Néanmoins, et loin s'en faut, la majorité des collections d'intérêt agronomique ne sont pas maintenues dans un CRB, mais sont conservées dans les unités par les équipes de recherche. Les outils utilisés pour leur gestion n'offrent pas nécessairement toutes les garanties de traçabilité, de conformité réglementaire, d'intégrité des données et de pérennité qu'elles méritent. D'autre part, ces collections historiques ou de travail sont rarement exposées et pourraient être plus visibles, ce qui donnerait aux équipes qui les maintiennent un surcroît de reconnaissance et des opportunités de collaboration.

Ces quelques remarques finales, pour insister sur l'intérêt de mettre en place, à l'échelle de l'Institut, des départements et de DISC, une réflexion sur la gestion des collections qui pourrait se concrétiser par une liste d'outils de gestion des collections ayant l'agrément de l'institut et dont la mise en œuvre permettrait d'obtenir un label collection-INRAE. Nous rappellerons pour finir que la nouvelle version du règlement intérieur des unités (ns2021-55b : <https://intranet.inra.fr/NS/ns2021-55.pdf>) indique, dans son article 16 Sauvegarde des collections, qu'il est de la mission des unités, des chercheurs, des centres et du CIRM de s'assurer de la conservation du patrimoine biologique de l'Institut. ■



Cet article est publié sous la licence Creative Commons (CC BY-SA). <https://creativecommons.org/licenses/by-sa/4.0/>

Pour la citation et la reproduction de cet article, mentionner obligatoirement le titre de l'article, le nom de tous les auteurs, la mention de sa publication dans la revue « NOV'AE », la date de sa publication et son URL.

Annexe :

Liste non exhaustive d'outils et systèmes d'information

- **BioMICS** (<https://www.bio-aware.com/>) : Système d'information dédié à la gestion, l'analyse et la publication de données biologiques (description détaillée dans cet article.) Solution commerciale BioAware.
- **COLLEC-SCIENCE** (<https://www.collec-science.org/>) : Gestion des données d'échantillonnage et des stocks (mode Web). Écrit en PHP, il fonctionne avec une base de données PostgreSQL. Logiciel sous licence libre AGPL INRAE.
- **e-SIToul-Barcode** (<https://get.genotoul.fr/systeme-dinformation/#barcode>) : Système d'information dont le but est de permettre la traçabilité et la gestion d'objets par code barre (1D ou 2D). Outil développé et maintenu par INRAE depuis 2005. Cet outil fait partie de la suite logiciels du système d'information e-SIToul, développé par GeT-PlaGe et installé sur GeT-TQ et GeT-TRiX, permettant de mutualiser des ressources et les bases de données associées.
- **FREELIMS** (<https://freelims.org/>) : Système de gestion des informations de laboratoire (LIMS) basé sur le cloud pour la gestion des données de laboratoire. Solution gratuite sous conditions.
- **Gfi-eDataBiotec** (<https://www.inetum.com/fr/healthcare>) : Recensement, identifications, suivi, stockage de recherche pour les ressources des CRB et Cancéropôles. Société commerciale INETUM.
- **LABWARE** (<https://www.labware.com/>) : Systèmes de gestion de l'information de laboratoire et cahiers électroniques pour les laboratoires de Santé. Solution commerciale.
- **Livextens et TDHarmony** (<https://www.technidata-web.com/fr/>) : Informatique de laboratoire, suite de logiciels pour laboratoires d'analyses médicales, CRB et biobanques. Société commerciale Technidata.
- **MBioLIMS BioBanking** (<https://www.modul-bio.com/nos-solutions/mbiolims/>) : Gestion des échantillons et données associées. Logiciel commercial dédié aux CRB, biobanques, cohortes et aux sociétés de diagnostic. Société MODUL-BIO.
- **OLGA** (<http://crb-tropicaux.fr/>) : Application Web créée en 2012 permettant la gestion des données liées aux ressources génétiques végétales et aux collections des CRB. Solution portée par le CIRAD, INRAE et l'IRD.
- **SeedUSoon** (<https://www.cite-des-energies.fr/biam/seedusoon/>) : Logiciel pour la gestion de ressources biologiques (e.g. stocks de semences) et l'échange de lignées végétales. Licence gratuite CEA au bénéfice d'établissements de recherche et/ou d'enseignement à but non lucratif.
- **SPECIFY** (<https://www.specifysoftware.org/>) : Logiciel open-source pour le traitement des données associées aux spécimens dans les CRB. Il a été élaboré en partenariat avec des musées d'histoire naturelle.

Certains sites web proposent des listes ou des classements comparatifs des outils ou SI dédiés à la gestion des données génétiques de tous types :

- <https://www.biobanking.com/top-lims/>
- <https://www.g2.com/categories/lims>
- <https://blog.genofab.com/the-best-lims-software-in-2021>
- <https://digital.com/software-and-service-reviews/>
- <https://www.trustradius.com/laboratory-information-management>