



BIBLIOTHÈQUES
DU MUSÉUM



{ BnF



1^{ÈRES} RENCONTRES DU RÉSEAU HORTI'DOC

LE PROJET DATAPOC, OU COMMENT FAIRE DU LIEN

UN EXEMPLE DE PROJET INNOVANT

Chloé Besombes

sommaire

1 | **Démarche data.museum**
2016 : Constat de départ

2 | **Lancement du projet**
Prestataire et développement

3 | **Phase de tests**
Sollicitation des collègues

4 | **Projet Datapoc 2.0**
La suite

5 | **Au-delà des données**
Coopération inter-établissement

6 | **Des questions ?**
N'hésitez pas !

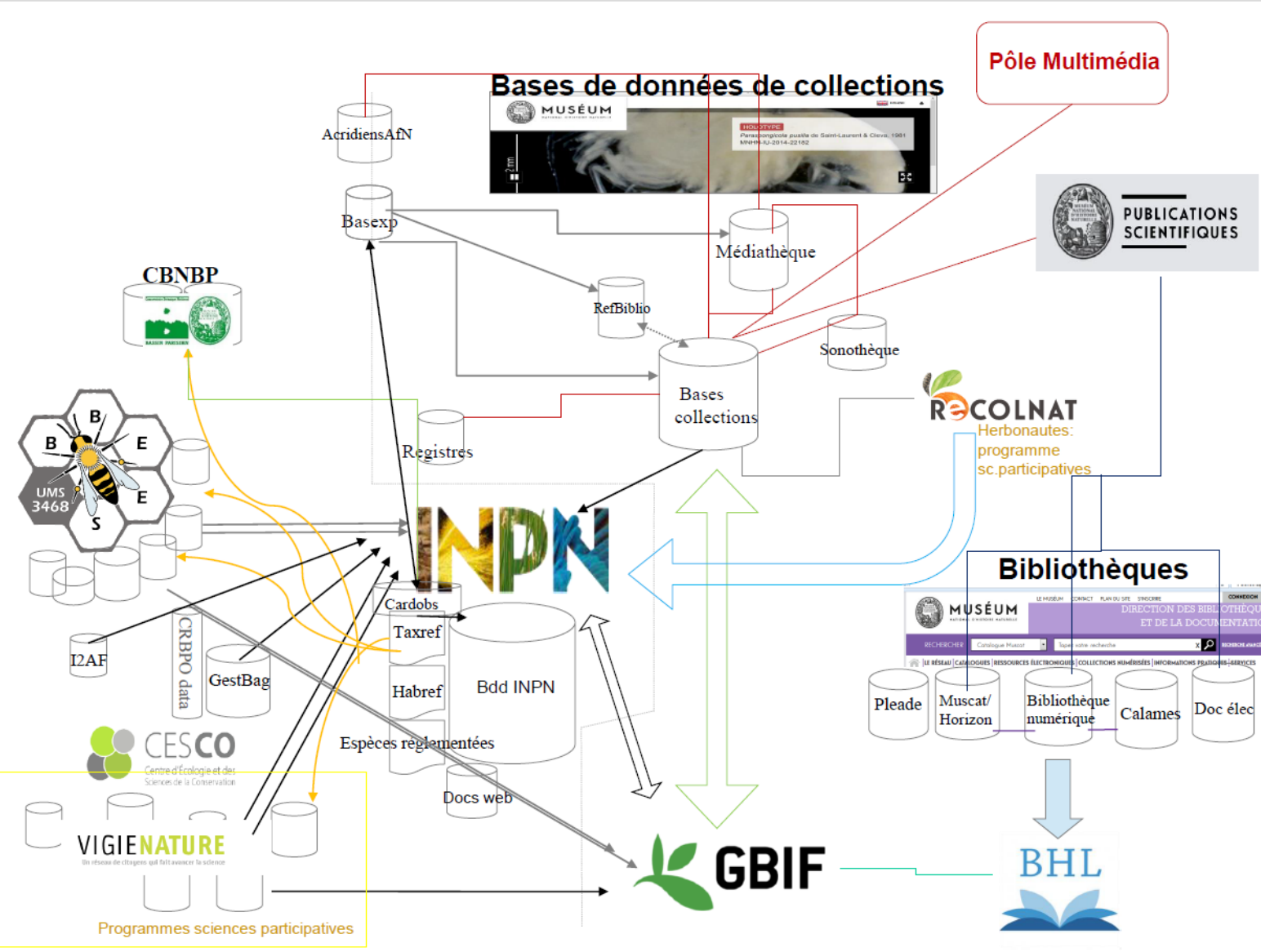


1

Démarche data.museum 2016

Constat de départ





Silos étanches

Silos étanches mais ouverts sur des outils nationaux ou internationaux

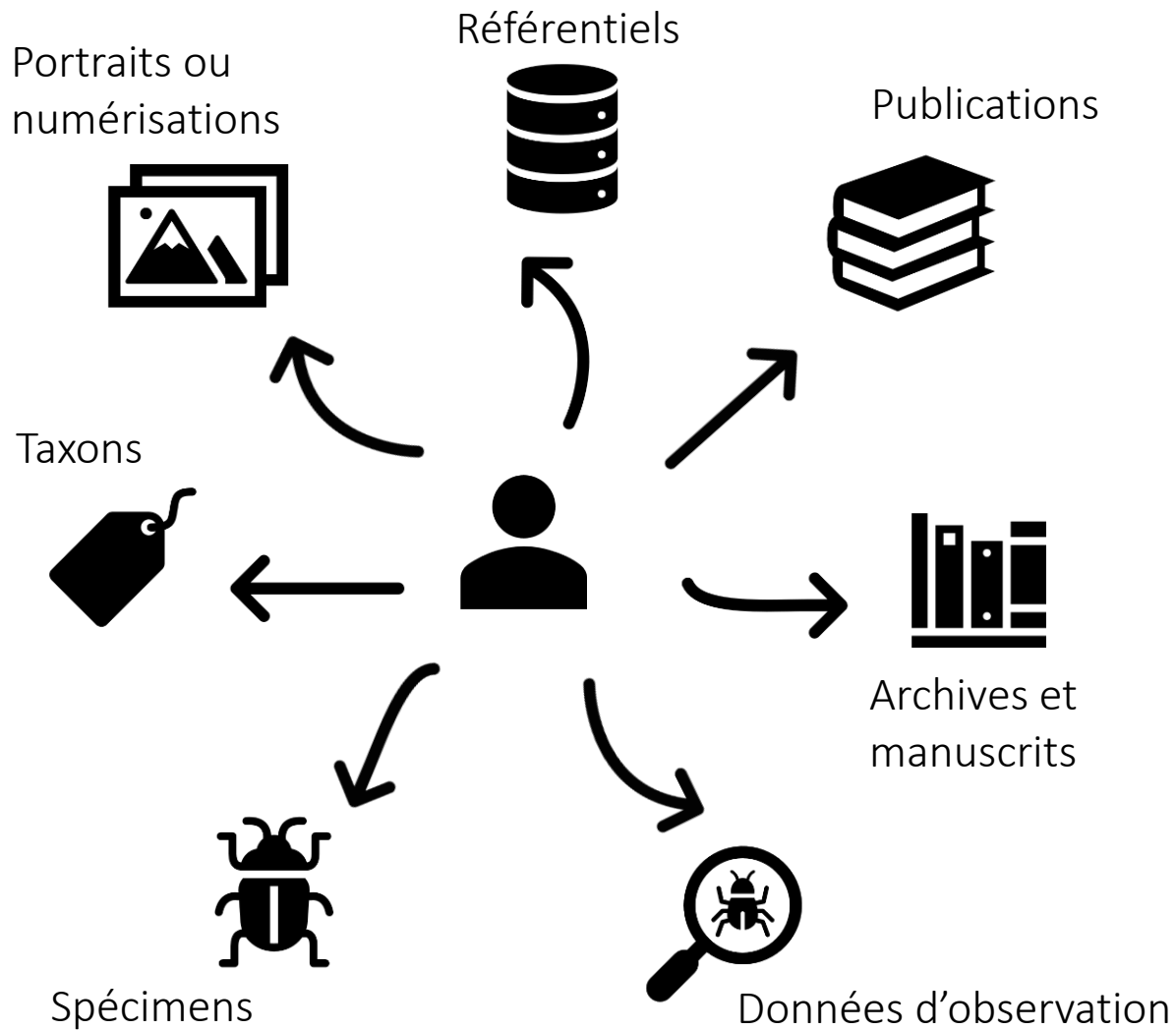
Données communes

Noms scientifiques des espèces, noms de personnes, localisations géographiques et dates

Pratiques hétérogènes

Pratiques de production et de normalisation hétérogènes entre les silos





Donnée pivot : nom de personne

Donnée stable

Présente dans tous les systèmes d'information

Stable par rapport aux autres types de données structurées dans les référentiels

Traitement hétérogène

Identifiants pérennes du côté des publications scientifiques et des bibliothèques par opposition aux bases de données naturalistes



2

Projet Datapoc.mnhn.fr Appel à projet Collex-Persée Prestataire et développement





- **Définition d'un corpus**

Du jardin au Muséum : en 516 biographies, Philippe Jaussaud & Edouard-Raoul, Publications du Muséum national d'Histoire naturelle, 2004.

- **Complétion et ajustement : 500 noms (XVII^e-XXI^e)**


En fonction du taux de représentation dans les différentes disciplines. Noms de naturalistes historiques et contemporains.

- **Travail d'alignement en collaboration avec l'Abes**

Chantier qualité mené dans la Direction des bibliothèques et de la documentation : taux de couverture de 85% avec le référentiel national IdRef.



Appel à projet CollEx-Persée

- Lauréat de l'appel à projet 2018-2019
- Choix du prestataire: société 
- Développements: Décembre 2018 - Décembre 2019

Livrables du projet

- L1 : Jeu de données structurées à partir des différents silos de données sélectionnés
Données naturalistes (science.mnhn.fr), imprimés (publications scientifiques et catalogue Muscat des bibliothèques), données d'observation (INPN)
- L2 : Site web et API
Interface selon les standards du web
- L3 : Rapport sur les mécanismes
Calcul des heuristiques, scores d'alignements, itération des mécanismes
- L4 : Préconisations pour le passage à l'échelle
Estimation de la fiabilité des heuristiques et des coûts humains de validation. Mesure de l'impact sur les différents silos.

Poc : proof of concept /
preuve de concept
: people of collection
/ les gens des collections



Mécanismes

Unifier les noms, même lorsque les présentations sont différentes

« Lamarck », « J.B. Lamarck », « Jean-Baptiste Lamarck », « Lamarck, chevalier de Monet », « Lamarck, J.-B. », « JB Lamarck », etc.

Désambigüiser les homonymes et pouvoir les rapprocher s'il s'agit de la même personne
Les dynasties comme par les exemples les « Jussieu »

Heuristiques en plus des transformations syntaxiques

- Prise en compte de l'âge de la personne au moment de l'événement : éviter de prendre en compte une collecte de spécimen à l'âge de deux ans
- Dates de naissances et de mort (lorsqu'elle existe) et la date de l'événement : éviter de déterminer un spécimen avant la date de naissance
- Identifier des co-auteurs fréquents ou de la géographie ou de la discipline



Organisation du site

Page « Personne » naturalistes

Zone de description biographique avec identifiants pérennes et des rôles (auteur, collecteur, déterminateur, observateur, auteur/sujet de publications, d'archives ou iconographie et cartographie)

Page « Famille » famille taxonomique

Auteurs, collecteurs, déterminateurs et observateurs en lien avec les spécimens de cette famille. Sujet des publications, archives ou iconographies et cartographie



Page Personne

(Exemple de la page de Louis Agassiz)

Description

Portraits, dates de naissance et mort, identifiants pérennes et biographie

Partie naturaliste

Auteur de taxon, collecteur, déterminateur, observateur

Partie bibliographique

Auteur ou sujet des publications, archives et iconographie. Cartographie



Page Personne

(Exemple de la page de Jean-Henri Fabre)



Fabre, Jean-Henri

Détails	
Dates	
1823 - 1915	
Identifiants	
idRef [1.00]	026856328
Wikidata [0.99]	Q313348
iSNI [0.98]	0000 0001 2132 6243
BNF [0.98]	11902115n
BHL [0.98]	1675
ViAF [0.98]	51689251
iPNI [0.98]	18621-1
Harvard Botanists index [0.98]	40896
Rôles	
Auteur de taxons, Collecteur, Déterminateur, Observateur, Auteur de documents	

Identifiants pérennes

Sélection de plusieurs référentiels nationaux et internationaux

Description

Jean-Henri Casimir Fabre, né le 21 décembre 1823 à Saint-Léons (Aveyron), mort le 11 octobre 1915 à Sérignan-du-Comtat (Vaucluse), est un homme de sciences, un humaniste, un naturaliste, un entomologiste et un poète français de langues d'oc (et à ce titre félibre) et française, lauréat de l'Académie française et d'un nombre élevé de prix.

Il peut être considéré comme l'un des précurseurs de l'éthologie, science du comportement animal, et de l'écophysiologie.

Ses découvertes sont tenues en haute estime en Russie, aux États-Unis, en Chine, en Corée et surtout au Japon où Jean-Henri Fabre est considéré comme le modèle accompli de l'homme de sciences et de programme des enseignements de l'école primaire. Il est aussi mondialement connu pour ses *Souvenirs entomologiques*, qui ont été traduits en quinze langues.

« Un grand savant qui pense en philosophe, voit en artiste, sent et s'exprime en poète », c'est ainsi que Jean Rostand qualifie Jean-Henri Fabre.

Via [Wikipédia](#) (CC BY-SA 3.0)

Partie description

Biographie extraite de Wikipédia

Éléments de bibliographie

Auteur

Par source

1833 2003

[0.99] The classification of insects from embriyological data (1849) ;
 [0.99] Essay on classification (1962) ;
 [0.99] An essay on classification (1859) ;
 [0.99] Nomenclatoris zoologici index universalis (1848) ;
 [0.99] Catalogue raisonné des familles, des genres et des espèces de la classe des Echinodermes (1846) ;
 [0.99] Monographie des poissons fossiles du vieux grès rouge ou système Dévonien (old red sandstone) des îles Britanniques et de Russie ;
 [0.99] Nomenclator zoologicus, continens nomina systematica generum animalium tam viventium quam fossilium ;
 [0.99] Recherches sur les poissons fossiles (1834) ;
 [0.99] Bibliographia zoologica et geologica ;
 [0.99] Histoire naturelle des poissons d'eau douce de l'Europe Centrale (1842) ;

Voir plus (66)

Sur l'auteur

Par source

1841 2007

[0.99] Un savant séducteur (2007) ;
 [0.99] Commemorative notice of Louis Agassiz (1873) ;
 [0.99] Louis Agassiz et son séjour à Neuchâtel de 1832 à 1846 (1907) ;
 [0.99] Un naturaliste du dix-neuvième siècle (1875) ;
 [0.99] Nature and the American mind (1974) ;
 [0.99] Louis Agassiz (1960) ;
 [0.99] Un Anti-évolutoniste exemplaire (1974) ;
 [0.70] Readings the shape of nature (1990) ;
 [0.70] Stations of the Thayer Expedition to Brazil, 1865-1866 (1977) ;
 [0.70] Catalogue des archives de Louis Agassiz (1807-1873) (1973) ;

Voir plus (104)

Archives et iconographie

Auteur

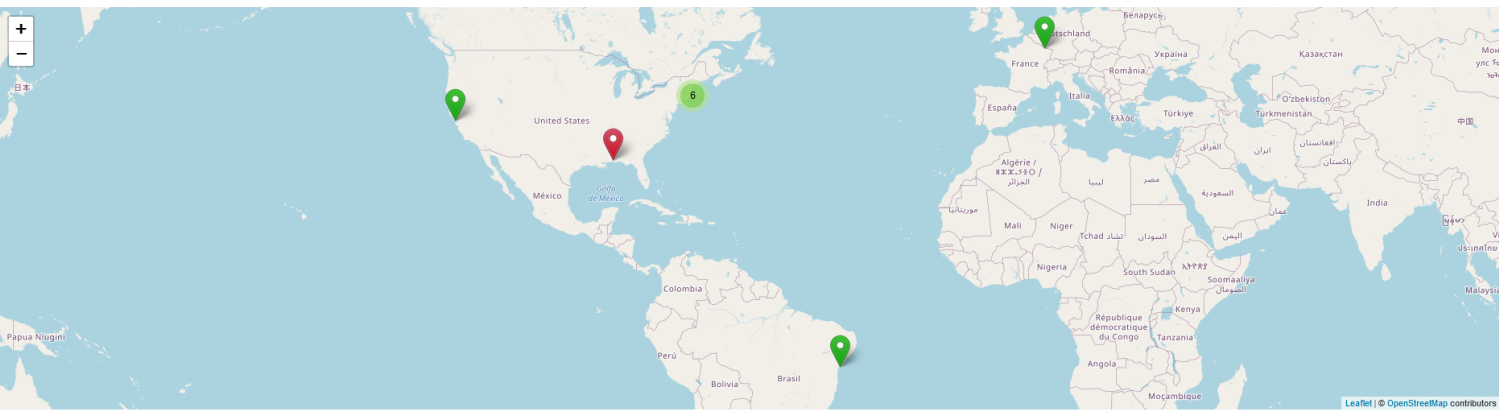
Aucun élément trouvé

Sur l'auteur

Aucun élément trouvé

Carte

En vert les collectes et en rouge les observations.



Page Personne

(Exemple de la page de Louis Agassiz)

Affichage des résultats d'alignement

❖ Partie naturaliste

- Auteurs de taxon
 - Collecteurs
 - Déterminateurs
 - Observateurs
- En lien avec les spécimens de la famille

❖ Partie bibliographique

- Éléments de bibliographie en tant qu'auteur ou sujet: publications scientifiques, catalogue des bibliothèques ,HAL (Hyper Articles en Ligne),BHL (Biodiversity Heritage Library
- Archives et iconographie en tant qu'auteur ou sujet : collections iconographiques numérisées des bibliothèques.
- Cartographie des événements

Éléments de bibliographie

Auteur

Par source

1855

- bib [0.99] [Bestioles \(2017\) *](#)
- bib [0.99] [Souvenirs entomologiques *](#)
- bib [0.99] [La vie des araignées \(1949\) *](#)
- bib [0.99] [Une ascension au mont Ventoux \(2013\) *](#)
- bib [0.99] [Les inventeurs et leurs inventions \(1986\) *](#)
- bib [0.99] [Les merveilles de l'instinct chez les insectes \(1989\) *](#)
- bib [0.99] [Le scorpion languedocien \(2002\) *](#)
- bib [0.99] [Les insectes \(2013\) *](#)

Voir plus (284)

Sur l'auteur

Par source

1886

Meta données

Poids de la forme de titre : 1

légende : [score] [poids] "référence" "cible" (description)

- bib [0.99] J.-H. F. [1] [1] "(026856328, I.0)" => "026856328" (Score pour l'identifiant idRef)
- bib [0.99] J.-H. F. [1] [0.4] "fabre jean henri" => "fabre jean henri" (Score entre les noms)
- bib [0.99] Henri F. [1] [0.2] "[1823-01-01, 1915-01-01]" => "1966" (Score d'évènement (peut être posthume))
- bib [0.99] Jean-H. [1] [0.05] "1823-01-01" => "1823" (Score pour la date de naissance)
- bib [0.99] Jean-H. [1] [0.05] "1915-01-01" => "1915" (Score pour la date de décès)
- bib [0.99] [La vita di J. H. Fabre *](#)
- bib [0.99] [The life of Jean Henri Fabre \(1921\) *](#)
- bib [0.99] [Fabre, poet of science \(1913\) *](#)
- bib [0.99] [Le génie de J.-H. Fabre \(1924\) *](#)

Voir plus (86)

Page Personne

(Exemple de la page de Jean-Henri Fabre)

Fonctionnalités

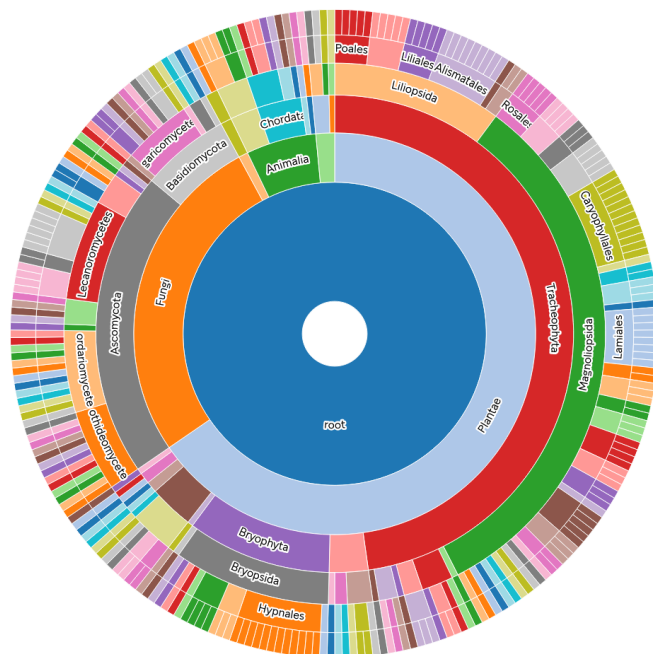
- Vignettes des spécimens
- Filtres
- Représentations graphiques (Sunburst et Sankey)
- Scores d'alignement

Scores d'alignement

Métadonnées. Calculs des scores d'alignement

Comparaison de données : identifiants pérennes, dates de naissance et mort, date de l'évènement (ici publication)





Page Personne

(Exemple de la page de Jean-Henri Fabre)

Représentation « Sunburst »

Représentation en fonction des collectes



Page Famille

(Famille taxonomique : exemple de la page des *Rosaceae*)

Numérisations

Photographies et autres représentations du spécimen

Phylogénie

Avec les principaux rangs



Rosaceae

Détails

Règne
Plantae

Division
Magnoliophyta

Classe
Magnoliopsida^a

Ordre
Rosales^a

Famille
Rosaceae

(CC BY)



Auteurs de taxon

Par naturaliste

Par genre

Par binom

1862

1873

[0.56] Saporta, Gaston de a créé le taxon *Plantae* Maenoliohytae Maenolioopsida* Rosales* Rosaceae *Cotoneaster assimiflora* (SAPORTA, 1873) ;

[0.36] Bastard, Toussaint a créé le taxon *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Rosa sylvia* (BASTARD) ;

[0.36] Benoist, Raymond a créé le taxon *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Polylois subintera* (BENOIST) ;

[0.36] Bois, Désiré a créé le taxon *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Cotoneaster adpressus* (BOIS) ;

[0.36] Boreau, Alexandre a créé le taxon *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Rosa lewisii* (BOREAU) ;

[0.36] Bosc, Louis-Augustin-Guillaume a créé le taxon *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Crotaeus celsianus* (BOSC) ;

[0.36] Bouvet, Georges a créé le taxon *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Rubus semiflexuosus* (SUDRE & BOUVET) ;

[0.36] Bronenart, Adolphe a créé le taxon *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Prunus venenata* (BRONEN) ;

[0.36] Bureau, Édouard a créé le taxon *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Neillia thibetica* (BUREAU & FANCH) ;

[0.36] Cambessis-Jacques a créé le taxon *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Potentilla doublonneana* (CAMBESSIS) ;

Voir plus (1093)

Collecteurs

Par naturaliste

Par genre

Par binom

1700

2017

[0.56] Balansa, Benjamin a collecté MNHN-P-P03728016 (déterminé comme *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Rubus moluccanus* (L.) en 1868 (-21.31752;165.4692)) ;

[0.56] Hallé, Nicolas a collecté MNHN-P-P00647031 (déterminé comme *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Rubus aimatus* (GUST.) en 1964 (lieu indéterminé)) ;

[0.56] Jovet-Ast, Suzanne a collecté MNHN-P-P0292486 (déterminé comme *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Malus acerba* (MÉRAT) en 1961 (Forêt de Coye (Oise)) ;

[0.56] Baynal-Rouges, Aline a collecté MNHN-P-P03302612 (déterminé comme *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Malus pumila* (MILL.) en 1965 (Arboretum Segrezianum, St Sulpice de Favières (S.O.)) ;

[0.56] Coste, Hippolyte a collecté MNHN-P-P00933152 (déterminé comme *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Potentilla grandiflora* (L.) en 1905 (lieu indéterminé)) ;

[0.55] Baynal, Jean a collecté MNHN-P-P04176672 (déterminé comme *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Potentilla norvegica* (L.) en 1954 (lieu indéterminé)) ;

[0.55] Delavay, Jean-Marie a collecté MNHN-P-P0214361 (déterminé comme *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Photinia lasiocarpa* (FRANCH. C.K. SCHNEID.) en 1885 (lieu indéterminé)) ;

[0.53] Jovet, Paul a collecté MNHN-P-P00503303 (déterminé comme *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Prunus mahaleb* (L.) en 1926 (lieu indéterminé)) ;

[0.53] Weddell, Huzues Alzerman a collecté MNHN-P-P03340763 (déterminé comme *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Rubus urticifolius* (POIR.) en 1844 (lieu indéterminé)) ;

[0.53] Armonin, Gérard Guy a collecté MNHN-P-P00684078 (déterminé comme *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Achanea arvensis* (L.) en 1972 (lieu indéterminé)) ;

Voir plus (1508)

Déterminateurs

Par naturaliste

Par genre

Par binom

1802

2008

[0.56] Jovet, Paul a déterminé MNHN-P-P00503293 comme *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Potentilla tomentella* (CRANTZ NECK.) en 1926 ;

[0.56] Guillaumin, André a déterminé MNHN-P-P02143233 comme *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Eriobotrya sequinii* (H. LÉV. CARDOT ex GUILLAUMIN) en 1924 ;

[0.52] Prévost, Jean a déterminé MNHN-P-P00220291 comme *Plantae* Tracheophyta* Maenolioopsida* Rosales* Rosaceae *Prunus myrtifolia* (L. Urb.) en 2000 ;

Page Famille

(Famille taxonomique : exemple de la page des *Rosaceae*)

Partie naturaliste

- Auteurs de taxon
- Collecteurs
- Déterminateurs
- Observateurs

En lien avec les spécimens de la famille

Partie bibliographique

- Éléments de bibliographie ayant pour sujet la famille en question : publications scientifiques, catalogue de la Direction des bibliothèques et de la documentation, HAL (Hyper Articles en Ligne) et BHL (Biodiversity Heritage Library)
- Archives et iconographie : collections iconographiques numérisées de la DBD.
- Carte : en vert les collectes, en rouge les observations



3

Projet Datapoc.mnhn.fr Phase de tests



3 PHASES DE TESTS

Retour des tests

Ergonomie de la plateforme à améliorer des points restent à éclaircir

L'apport des scores d'alignement, contextualiser les identifiants pérennes, nuances entre les différentes parties naturalistes peu abordables de l'extérieur (auteur, déterminateur, collecteur et observateur)

Mieux faire apparaître les liens entre les pages

Les pages de science.mnhn.fr apparaissent comme une continuité de la plateforme datapoc.mnhn.fr. Rapidement un sentiment de se perdre au milieu des pages sans pouvoir revenir au résultat précédent.

Plateforme très prometteuse volonté d'en voir plus

Proposer un filtre géographique, pouvoir exporter les données sous la forme d'un fichier .csv, ...



4

Projet Datapoc 2.0 2020-2022

Nouveau départ



Nécessité d'un nouveau départ

Technologie peu répandue, difficile à s'approprier dans l'écosystème SI du Muséum

Mises à jour très lourdes, imports complets nécessaires et impossible à envisager dans le cadre d'un passage à l'échelle

Site web rigide, peu modulable pour le rendre plus ergonomique

Manipulation et récupération des données à améliorer

Nécessité de développer de nouvelles fonctionnalités dans la perspective d'une réinformatisation du SI collections du Muséum



À propos du projet

Porteur du projet :
Muséum national d'histoire naturelle

Partenaires :
Agence bibliographique de l'Enseignement supérieur (Abes) -
Bibliothèque nationale de France (BnF) -
Consortium Biodiversity Heritage Library (BHL)

Part Collex-Persée :
100 000 euros

Calendrier :
Octobre 2020 - octobre 2021

Datapoc 2.0

Lier les données pour les consolider et ouvrir de nouvelles perspectives de recherche sur l'histoire de la biodiversité



Illustration Datapoc 2.0 | © MNHN/Camille Dégardin

Contacts

Cindy Lim
Chargée de mission
Datapoc 2.0 [Contact](#)

Cécile Callou
Responsable
scientifique du projet [Contact](#)

Chloé Besombes
Chef du projet
Datapoc 2.0 [Contact](#)

Partenaires



<https://www.collexpersée.eu/projet/datapoc-2-0/>



**BIBLIOTHÈQUES
DU MUSÉUM**

Appel à projet Collex-Persée

Lauréat de l'appel d'offre 2019-2020

Elaboration du nouveau projet

Rédaction du cahier des charges, mise en place d'une procédure d'achat innovant, choix du prestataire.

Développements



Décembre 2021-Décembre 2022

DATAPOC

1ÈRES RENCONTRES HORTI'DOC – 16 JUIN 2022

CHLOÉ BESOMBES

5

Au-delà des données

Un autre bilan du projet



Retour d'expérience

En 5 ans, la dynamique engagée autour du projet datapoc a contribué :

- **à sortir la bibliothèque de son isolement métier et de ses murs**
 - Reconnaissance des capacités de ses personnels à conduire des projets et à comprendre des données hors de leur « zone de confort » bibliographique
 - Montée en compétence des personnels de bibliothèque sur les champs scientifiques (taxonomie, écologie...) au cœur des missions du MNHN
- **à co-construire une vision partagée autour des données et un projet d'établissement ambitieux**
 - Constitution d'une communauté multi-acteurs motivée et en confiance associant personnels IST, chercheurs, informaticiens
 - Le travail réalisé avec datapoc (cartographie des bases de données, données de référence, ingénierie algorithmique, fonctionnalités de recherche...) a vocation à être capitalisé au sein d'un projet plus global de refonte des SI collections et recherche dans le cadre d'un « Plan Marshall » numérique qui émerge aujourd'hui au MNHN.
 - Preuve de concept : pas seulement au niveau technique, mais aussi de la capacité à travailler ensemble et autrement.
- **au niveau national, et dans le cadre du GIS COLLEX-Persée**
 - Perception différente du MNHN : pas seulement un lieu d'excellence disciplinaire mais aussi un établissement capable de s'impliquer dans la gouvernance et la modernisation des réseaux documentaires et scientifiques.



6

Questions

Avez-vous

des questions ?



MERCI !

ADRESSE DE LA PLATEFORME : [DATAPOC.MNHN.FR](https://dataoc.mnhn.fr)
POUR TOUTES REMARQUES OU SUGGESTIONS : DATAPOC@MNHN.FR
