



***11e rencontres CollEx-Persée-FID : Evaluer la transition, consolider les résultats, renforcer la coopération**
« Regards croisés sur de nouveaux systèmes documentaires en appui à la recherche » (28-29 juin 2021)*

DATAPOC 2.0

Cécile Callou - Maître de conférences

UMS BBEES Bases de données sur la Biodiversité, Ecologie, Environnement et sociétés (MNHN- CNRS InEE)

Gildas Illien – Conservateur général/Directeur

Direction générale déléguée aux collections, MNHN

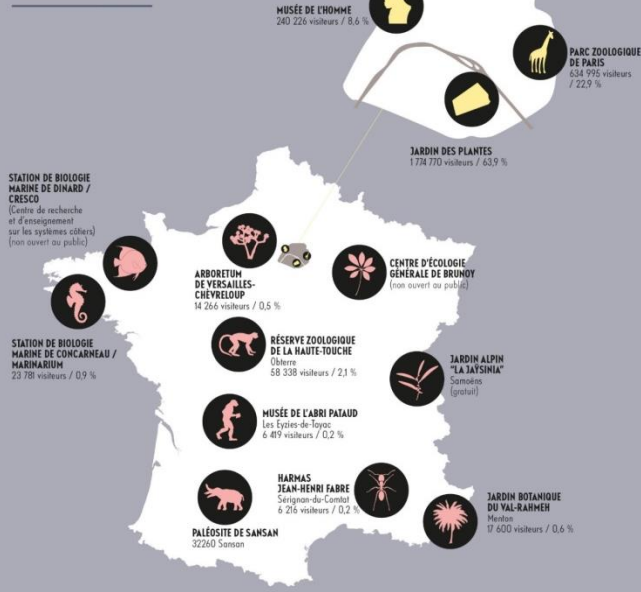
Avec la collaboration de Chloé Besombes

Direction des Bibliothèques, MNHN

Le Muséum national d'Histoire naturelle



SITES PARISIENS



En quelques chiffres

400 ans d'histoire

13 sites en France, à Paris et en région

Plus de 2.600 personnes dont 570 chercheurs

3 départements scientifiques Adaptations du vivant, Homme et environnement, Origines et évolution

66,8 millions de spécimens

5 missions fondamentales Collections, Recherche, Expertise, Diffusion, Enseignement & Formations



IIE RENCONTRES COLLEX-PERSÉE-FID
NOUVEAUX SERVICES, NOUVEAUX USAGES : QUELS RETOURS
D'EXPÉRIENCE ?
28-29 JUIN 2021

Les enjeux sous-jacents au projet datapoc

Au commencement de l'histoire du projet, des intentions stratégiques qui recoupent celles de ColLEX-Persée et les déclinent à l'échelle d'un établissement :

- **En 2016, constat d'une bibliothèque patrimoniale relativement isolée :**
 - Fréquentation et consultation de la documentation imprimée par les chercheurs en fort déclin du fait du « boom » de la documentation électronique
 - Participation limitée de la bibliothèque à la gouvernance et aux projets de développement transverses au MNHN
 - Une bibliothèque mieux insérée dans les réseaux documentaires nationaux (CADIST, Sudoc, Calames, BnF...) que dans ceux de son propre établissement
- **Repositionner la fonction documentaire et les bibliothèques dans l'écosystème scientifique du Muséum :**
 - Une nouvelle organisation qui installe un service dédié aux chercheurs avec de nouvelles propositions (science ouverte, gestion de données, archivage...)
 - Des échanges plus nourris et fréquents avec les UMR, les UMS, les départements et les instances scientifiques
 - Une reformulation de la politique documentaire et de services positionnant davantage la bibliothèque sur le champs des données et en appui de projets
- **ColLEX-Persée est apparu comme un levier national pour accompagner localement ces nouvelles orientations**
 - Participation active de politiques et de chercheurs du MNHN à tous les nouveaux de la gouvernance de ColLEX (CM, CS, COMOP, co-portage des projets)
 - Par l'apport de ressources mais aussi d'un nouveau dispositif, ColLEX-Persée a encouragé la structuration de nouveaux rapports entre collections et recherche

= La petite histoire de datapoc illustre concrètement ces évolutions.



sommaire

1 | **Démarche
data.museum**

2 | **Projet
Datapoc.mnhn.fr**

Appel à projet Collex-
Persée

3 | **Projet
Datapoc.mnhn.fr**

Tests utilisateurs

4 | **Projet
Datapoc.mnhn.fr**

Analyse technique

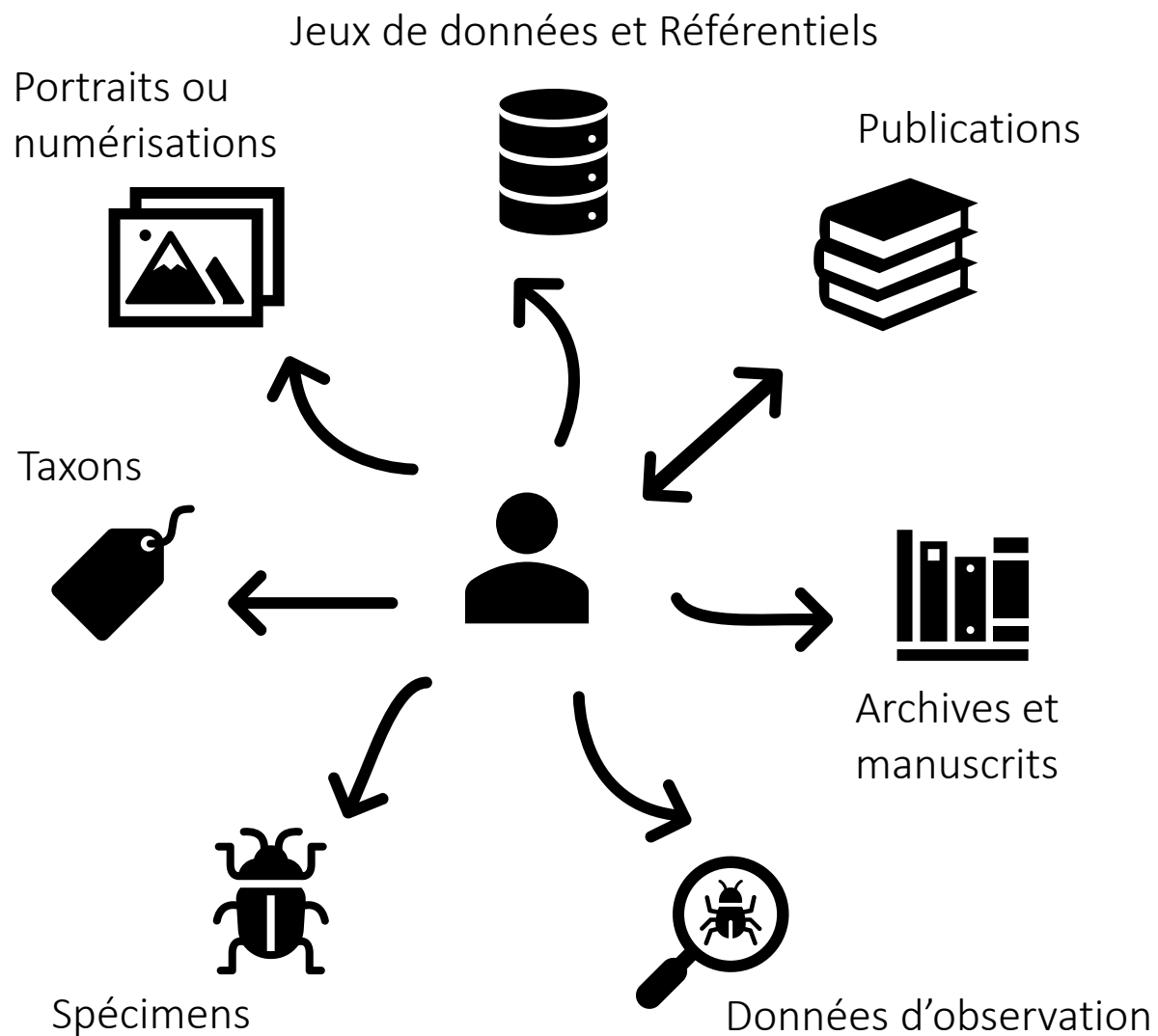
5 | **Projet Datapoc 2.0**



1

Démarche data.museum





Donnée pivot :
nom de personne

Donnée récurrente

Présente dans tous les systèmes d'information

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



Collex-Persée

Réseau Projets Ressources Groupes A propos Actualités | Agenda

Accueil > Projets >

À propos du projet

Type de projet : Services

Porteur du projet : Muséum national d'histoire naturelle

Partenaires : Muséum national d'histoire naturelle ; Direction générale déléguée aux collections ; CHRS-MNHN (UMS 3448) ; Bases de données Biodiversité ; Ecologie Environnements Sociétés ; CHRS-MNHN (UMS 2006) Patrimoine ; Patrimoine naturel ; Publications scientifiques du Muséum national d'histoire naturelle ; Agence Bibliographique de l'Enseignement Supérieur - BSCOLNAT ; Bibliothèque nationale de France ; Consortium of European Taxonomic Facilities (CETAF)

Part Collex-Persée : 70 000 euros

Calendrier : Janvier 2019 - Janvier 2021

Projet terminé

Datapoc.mnhn.fr

Construction d'un référentiel "personnes" destiné à lier des données de collections et des jeux de données issus de la recherche en sciences naturelles



Jardin des plantes - Ecole de botanique - Allium schubertii | © MNHN - Agnès Jazouars

Contacts

Gildas Illien
Directeur des bibliothèques et de la documentation, directeur général délégué aux collections adjoint, Muséum national d'histoire naturelle. [Contact](#)

Chloé Besombes
Chef du projet datapoc.mnhn.fr, adjointe au chef du service Collecte, traitement et flux, responsable de l'équipe traitement et exposition des données, Muséum national d'histoire naturelle. [Contact](#)

[Voir plus de contacts \(1\)](#)

Partenaires



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

Appel à projet Collex-Persée

Lauréat de l'appel à projet 2018-2019

Responsable : Gildas Illien
Chef de projet : Chloé Besombes (MNHN)
Chargée de mission : Cindy Lim

Lancement du marché

Choix du prestataire : société Logilab

Logilab

Développement

Lancement du marché de Décembre 2018 à Décembre 2019

<https://www.collexpersée.eu/projet/datapoc-mhn-fr/>



IIE RENCONTRES COLLEX-PERSÉE-FID
NOUVEAUX SERVICES, NOUVEAUX USAGES : QUELS RETOURS
D'EXPÉRIENCE ?
28-29 JUIN 2021

- **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 représentation dans les différents 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.



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
Problématique des dynasties : les « Jussieu », « Cuvier », « Saint-Hilaire »

Heuristiques en plus des transformations syntaxiques

- Prise en compte de l'âge de la personne au moment de l'événement
- Dates de naissances et de mort (lorsqu'elles sont disponibles) et la date de l'événement : éviter de déterminer un spécimen avant la date de naissance
- Identifier des co-auteurs fréquents, des lieux de collecte ou des spécialités (discipline)

Livrables

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

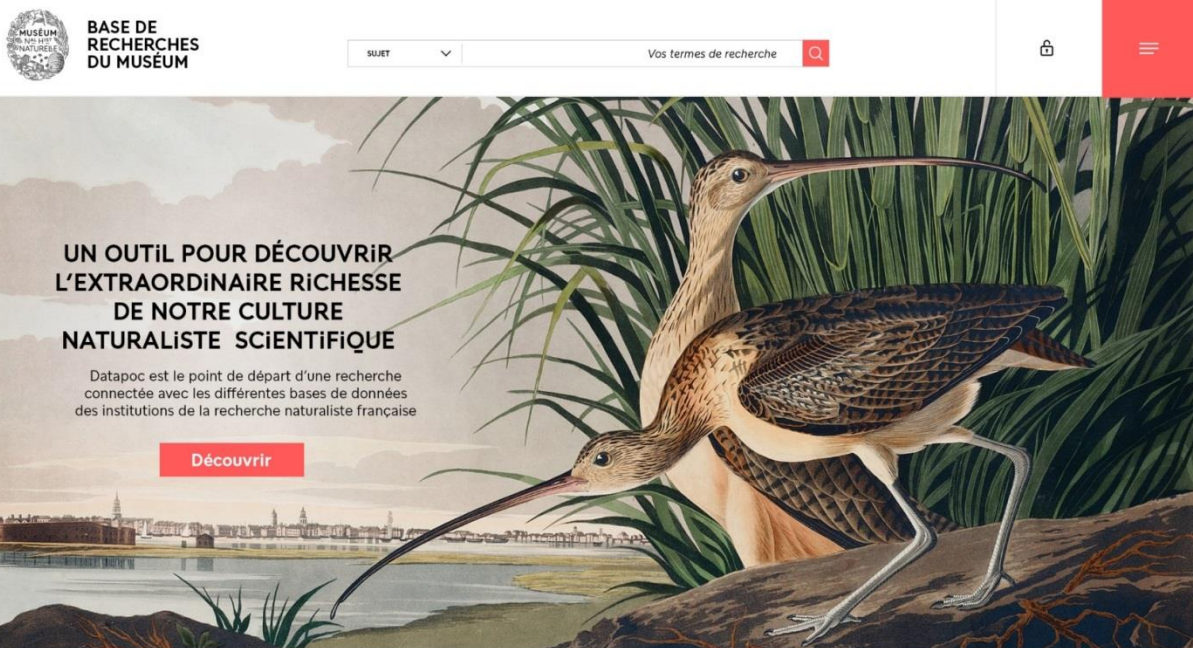
Interface selon les standards du web

L3 : Rapport sur les mécanismes

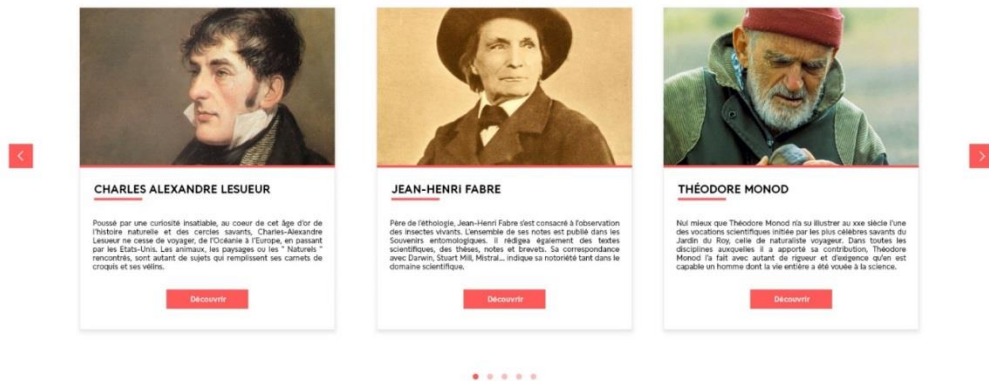
Calcul des heuristiques, scores d'alignements, itération des mécanismes

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.



Portraits de Naturalistes



Page « Personne » naturaliste

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 » taxonomique

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

<https://datapoc.mnhn.fr/>



Fabre, Jean-Henri



Détails
<div>Dates</div> <div>1823 - 1915</div>
<div>Identifiants</div> <div><div><div>idRef</div><div>[1.00]</div><div>026856328</div></div><div><div>Wikidata</div><div>[0.99]</div><div>Q383348</div></div><div><div>ISNI</div><div>[0.98]</div><div>0000 0001 2132 6263</div></div><div><div>BNF</div><div>[0.98]</div><div>11902115n</div></div><div><div>BHL</div><div>[0.98]</div><div>1675</div></div><div><div>VIAF</div><div>[0.98]</div><div>51689251</div></div><div><div>IPNI</div><div>[0.98]</div><div>18621-1</div></div><div><div>Harvard Botanists index</div><div>[0.98]</div><div>40886</div></div></div>
<div>Rôles</div> <div>Auteur de taxons, Collecteur, Déterminateur, Observateur, Auteur de documents</div>

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 éminent, un écrivain passionné par la nature 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 l'homme de lettres réunis et, à ce titre, est au 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)

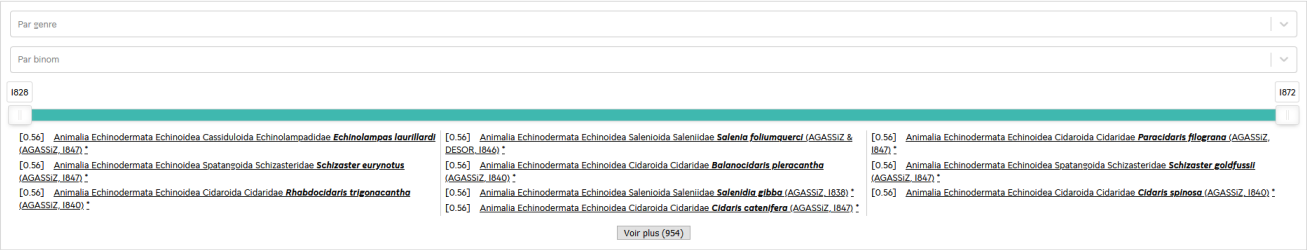
Identifiants pérennes

Sélection de plusieurs référentiels nationaux et internationaux
IdREF : identifiant pivot

Partie description

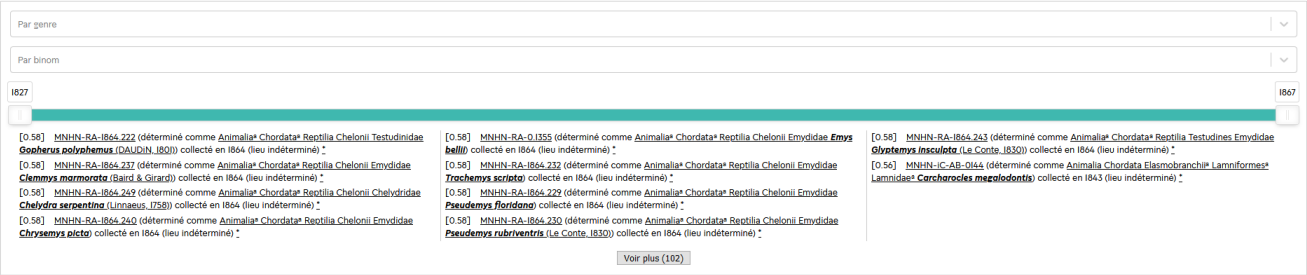
Biographie extraite de Wikipédia





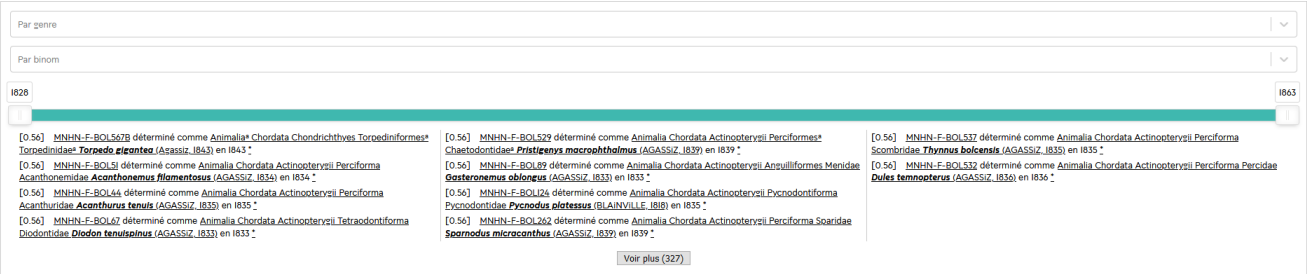
Visualisations graphiques [Sankey](#) [Sunburst](#)

Collecteur



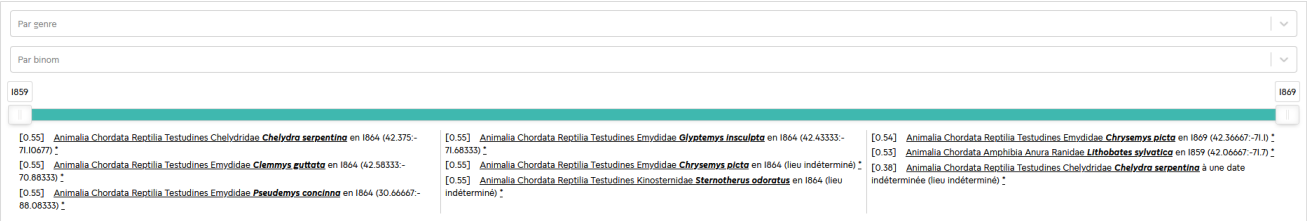
Visualisations graphiques [Sankey](#) [Sunburst](#)

Déterminateur



Visualisations graphiques [Sankey](#) [Sunburst](#)

Observateur



Page Personne

Exemple de la page de Louis Agassiz (1807-1830)

Partie naturaliste

- Auteurs de taxon
- Collecteurs
- Déterminateurs
- Observateurs

En lien avec les spécimens de la famille

IIE RENCONTRES COLLEX-PERSÉE-FID
NOUVEAUX SERVICES, NOUVEAUX USAGES : QUELS RETOURS
D'EXPÉRIENCE ?
28-29 JUIN 2021

Éléments de bibliographie

Auteur

Par source

1833

2003

1834

1835

1836

1837

1838

1839

1840

1841

1842

1843

1844

1845

1846

1847

1848

1849

1850

1851

1852

1853

1854

1855

1856

1857

1858

1859

1860

1861

1862

1863

1864

1865

1866

1867

1868

1869

1870

1871

1872

1873

1874

1875

1876

1877

1878

1879

1880

1881

1882

1883

1884

1885

1886

1887

1888

1889

1890

1891

1892

1893

1894

1895

1896

1897

1898

1899

1900

1901

1902

1903

1904

1905

1906

1907

1908

1909

1910

1911

1912

1913

1914

1915

1916

1917

1918

1919

1920

1921

1922

1923

1924

1925

1926

1927

1928

1929

1930

1931

1932

1933

1934

1935

1936

1937

1938

1939

1940

1941

1942

1943

1944

1945

1946

1947

1948

1949

1950

1951

1952

1953

1954

1955

1956

1957

1958

1959

1960

1961

1962

1963

1964

1965

1966

1967

1968

1969

1970

1971

1972

1973

1974

1975

1976

1977

1978

1979

1980

1981

1982

1983

1984

1985

1986

1987

1988

1989

1990

1991

1992

1993

1994

1995

1996

1997

1998

1999

2000

2001

2002

2003

Voir plus (66)

Sur l'auteur

Par source

1841

1

- [\[0-99\]](#) [Un savant séducteur \(2007\) :](#)
- [\[0-99\]](#) [Commemorative notice of Louis Asassiz \(1873\) :](#)
- [\[0-99\]](#) [Louis Asassiz 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 Asassiz \(1960\) :](#)
- [\[0-99\]](#) [Un Anti-évolutionniste exemplaire \(1974\) :](#)
- [\[0-70\]](#) [Reading the shape of nature \(1990\) :](#)
- [\[0-70\]](#) [Stations of the Thayer Expedition to Brazil, 1865-1866 \(1977\) :](#)
- [\[0-70\]](#) [Catalogue des archives de Louis Asassiz \(1807-1873\) \(1973\) :](#)

Voir plus (104)

Archives et iconographie

Auteur

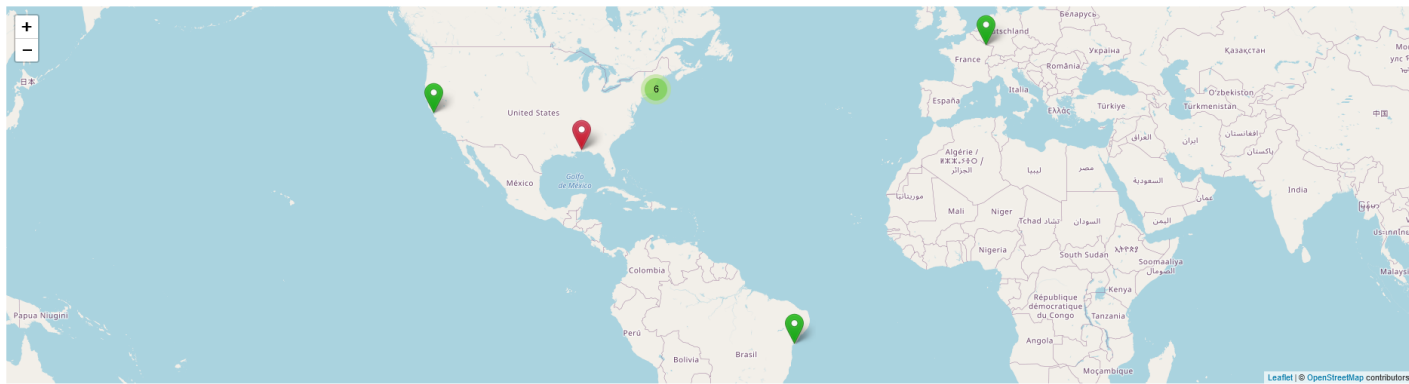
Aucun élément trouvé

Sur l'auteur

Aucun élément trouvé

Carte

En vert les collectes et en rouge les observations.

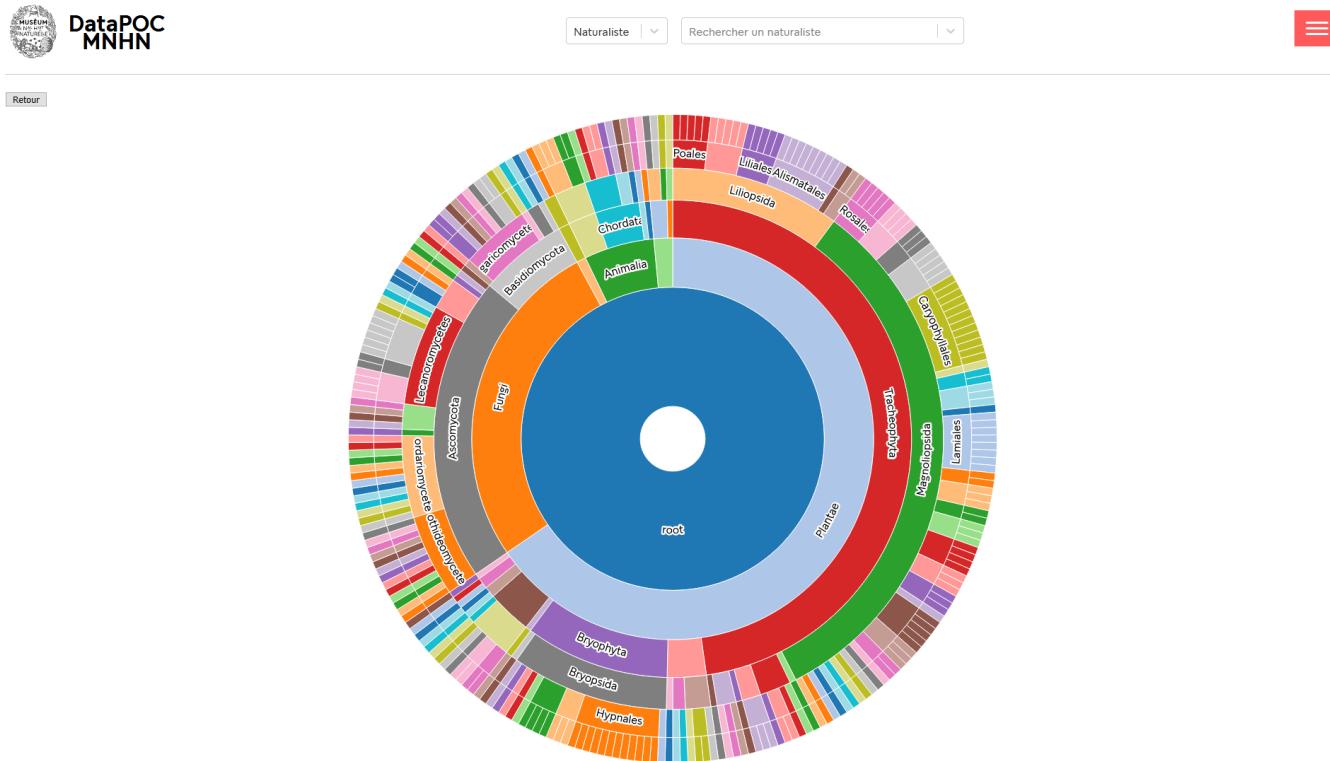


Page Personne

Exemple de la page de Louis Agassiz (1807-1830)

Partie bibliographie

- Éléments de bibliographie en tant qu'auteur ou sujet :
 - Publications scientifiques
 - Catalogue des bibliothèques
 - HAL (Hyper Articles en Ligne)
 - BHL (Biodiversity Heritage Library)
 - Références bibliographiques (science.mnhn.fr)
- Archives et iconographie en tant qu'auteur ou sujet : collections iconographiques numérisées des bibliothèques.
- Cartographie des événements :
 - En vert, les collectes
 - En rouge, les observations



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'événement (ici publication)


Éléments de bibliographie


Auteur


Par source


1855


 [0.99] [Bestioles](#) (2017) *


 [0.99] [Souvenirs entomologiques](#) *


 [0.99] [La vie des araignées](#) (1949) *

 [0.99] [Une ascension au mont Ventoux](#) (2013) *

 [0.99] [Les inventeurs et leurs inventions](#) (1986) *

 [0.99] [Les merveilles de l'instinct chez les insectes](#) (1989) *

 [0.99] [Le scorpion languedocien](#) (2002) *

 [0.99] [Les insectes](#) (2013) *

[Voir plus \(284\)](#)

Sur l'auteur

Par source

1886

II

Meta données

Poids de la forme de titre : 1

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

- [bib] [0.99]** *J.-H. F. [1] [1] "Q026856328", 1.0)" => "Q026856328" (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]** *J.-H. F. [1] [0.2] "[1823-01-01; 1915-01-01]" => "1966" (Score d'évènement (peut être posthume))*
- [bib] [0.99]** *J.-H. F. [1] [0.05] "1823-01-01" => "1823" (Score pour la date de naissance)*
- [bib] [0.99]** *J.-H. F. [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'événement (ici publication)





DataPOC
MNHN

Taxonomie

Rechercher

Page Famille

Exemple de la page des *Erebidae*



Erebidae

Détails

Règne

Animalia^a

Division

Arthropoda

Classe

insecta

Ordre

Lepidoptera

Famille

Erebidae

Numérisations

Photographies et autres représentations du spécimen

Phylogénie

Avec les principaux rangs

Partie naturaliste

Auteurs de taxon, Collecteurs, Déterminateurs, Observateurs

Partie bibliographie

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

Carte : collectes et observations

Crédit : Akyurek M. (CC BY)



IIE RENCONTRES COLLEX-PERSÉE-FID

NOUVEAUX SERVICES, NOUVEAUX USAGES : QUELS RETOURS

D'EXPÉRIENCE ?

28-29 JUIN 2021

3

Projet Datapoc.mnhn.fr

Tests utilisateurs



Phasage prévu

1.
Première vague de tests : le groupe
data.muséum

Comité de pilotage du projet

2.
Deuxième vague de tests :
Gestionnaires de bases de données

Collègues de la DGD-C et collègues du
réseau Recolnat

3.
Troisième vague de tests : Extérieurs
aux Muséums

Étudiants du Muséum et usagers des
bibliothèques du Muséum



01 / DÉMARCHE
DATA.MUSEUM

02 / PROJET
DATAPOC.MNHN.FR
AAP COLLEX-PERSÉE

03/ PROJET
DATAPOC.MNHN.FR
TESTS UTILISATEURS

Retour des tests

Ergonomie de la plateforme à améliorer

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.

Prévoir des outils pour améliorer la qualité des données

Des erreurs d'attribution (collecte, identification, observation) à corriger dans les bases-sources

Plateforme très prometteuse

Volonté forte d'aller plus loin (Proposer un filtre géographique, pouvoir exporter les données sous la forme d'un fichier .csv, ...)

Intérêt pour les services offerts : accès aux données facilité, enrichissement par des données extérieures etc.



4

Projet Datapoc.mnhn.fr

Analyse technique



Nécessité d'un nouveau projet

01 / DÉMARCHE
DATA.MUSEUM

02 / PROJET
DATAPOC.MNHN.FR
AAP COLLEX-PERSÉE

03/ PROJET
DATAPOC.MNHN.FR
TESTS UTILISATEURS

04 / PROJET
DATAPOC.MNHN.FR
ANALYSE TECHNIQUE

Technologie Open Source mais 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


Développer de nouvelles fonctionnalités dans la perspective d'un nouveau SI Muséum



5

Projet Datapoc 2.0





Réseau
Projets
Ressources
Groupes
A propos
Actualités | Agenda

Accueil
Projets

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

Appel à projet Collex-Persée

Lauréat de l'appel à projet 2019-2020

Responsable scientifique : Cécile Callou
 Chef de projet : Chloé Besombes (MNHN)
 Chargée de mission : Cindy Lim

Lancement du marché

Partenaires



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



IIE RENCONTRES COLLEX-PERSÉE-FID

NOUVEAUX SERVICES, NOUVEAUX USAGES : QUELS RETOURS
 D'EXPÉRIENCE ?

28-29 JUIN 2021

01 / DÉMARCHE
DATA.MUSEUM

02 / PROJET
DATAPOC.MNHN.FR
AAP COLLEX-PERSÉE

03 / PROJET
DATAPOC.MNHN.FR
TESTS UTILISATEURS

04 / PROJET
DATAPOC.MNHN.FR
ANALYSE TECHNIQUE

05 / PROJET DATAPOC 2.0

Objectifs

Passage à l'échelle

Développer la qualité des alignements de données à partir des alignements proposés

Développer l'interopérabilité, par la diffusion des informations au format RDF et leur exposition par l'intermédiaire de services web (API) documentés

Rendre interopérables les bases de données du Muséum

Consolider les données et optimiser le réseau des systèmes d'information du Muséum

Favoriser l'ouverture, la visibilité et le référencement des données, produites ou récoltées par les personnel du Muséum, par le biais de leur diffusion, leur valorisation et un encouragement fort à leur réutilisation (Science Ouverte).

Créer une interface publique et des services innovants permettant à tout public (chercheurs, acteurs de la recherche et du monde naturaliste, développeurs) d'explorer les données afin d'obtenir des résultats originaux, ou d'enrichir ces données (grâce au développement d'outils prédictifs pour consolider les bases par exemple).

Contribuer aux réseaux et infrastructures nationaux et européens, documentaires et naturalistes, autour de l'interopérabilité des référentiels.



Retour d'expérience : au-delà du projet

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.

= COLLEX-Persée a été un facteur de succès, accélérateur et facilitateur de coopération et d'innovation dans le contexte du MNHN.



MERCI DE VOTRE ATTENTION !

IIE RENCONTRES COLLEX-PERSÉE-FID
28-29 JUIN 2021
