

Bilan et propositions sur la relation collections physiques - collections numériques

Line Le Gall

► **To cite this version:**

Line Le Gall. Bilan et propositions sur la relation collections physiques - collections numériques. [Rapport Technique] Muséum national d'Histoire Naturelle (Paris, France). 2018, 20 p. mnhn-02517162

HAL Id: mnhn-02517162

<https://hal-mnhn.archives-ouvertes.fr/mnhn-02517162>

Submitted on 24 Mar 2020

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.



MUSÉUM
NATIONAL D'HISTOIRE NATURELLE

Bilan et propositions sur la relation collections physiques - collections numériques

**Remise du rapport à Michel Guiraud, Direction générale déléguée
aux collections du MNHN**

Par Line Le Gall, Maître de conférences du MNHN

Avec le soutien de Dario De Franceschi

Jeudi 31 mai 2018

Lettre de mission confiée le 15 juin 2017

Chers collègues,

La numérisation de masse permet aujourd'hui de créer de véritables doublons virtuels de la collection physique. Associés à ces collections virtuelles, des outils d'aide au travail de recherche sur les images permettent à l'ensemble de la communauté scientifique de documenter les collections. La façon dont cette documentation à partir des collections virtuelles est intégrée dans les bases de collections physiques est un défi majeur pour les années à venir.

Dans le cadre d'une réflexion de fond sur les collections du MNHN, j'ai souhaité vous confier une mission d'étude sur les relations entre collections physiques et collections virtuelles et les orientations à donner en matière de gestion de l'information.

Vous me ferez parvenir pour le 31 mai 2018 le résultat de votre analyse et de vos propositions qui porteront notamment sur les aspects suivants:

- les modes actuels de gestion des collections physiques et des collections virtuelles,
- la segmentation éventuelle des utilisateurs des deux types de collection,
- l'intérêt ou la nécessité de gérer indépendamment les deux types de collection, et dans ce cas la pertinence de créer des outils pour assurer un lien,
- les moyens, conceptuels et techniques, d'intégrer les apports mutuels dans les collections de manière à assurer la convergence entre les collections si celle-ci est nécessaire,
- le paysage actuel au niveau européen et international, et les initiatives qui pourraient être prises au niveau du CETAF pour harmoniser les pratiques,
- une prospective à très long terme sur ces relations entre physique et virtuel.

Pour cette mission vous serez amenés à organiser des réunions au niveau européen et vous me communiquerez le budget qui vous paraît nécessaire pour la réalisation de cette mission.

En vous remerciant par avance de votre contribution, je vous prie d'agréer mes sentiments les meilleurs.

Michel Guiraud
Directeur général délégué aux collections

Table des matières

Lettre de mission confiée le 15 juin 2017	2
Synthèse	4
Introduction	6
Eléments de contexte scientifique	7
Les modes actuels de gestion des collections physiques et des collections numériques	10
i) Les bases de données pour la gestion collections physiques.....	10
ii) Les bases de données pour les collections numériques	10
iii) L'interopérabilité des collections « physiques » et « numériques »	12
La segmentation éventuelle des utilisateurs des deux types de collection	13
L'intérêt ou la nécessité de gérer indépendamment les deux types de collection, et dans ce cas la pertinence de créer des outils pour assurer un lien	14
i) Un changement pour les utilisateurs et gestionnaires des collections physiques.....	14
ii) Un changement pour les utilisateurs des collections numériques.....	Erreur ! Signet non défini.
Les moyens, conceptuels et techniques, d'intégrer les apports mutuels dans les collections	16
Le paysage actuel au niveau européen et international, et les initiatives qui pourraient être prises au niveau du CETAF pour harmoniser les pratiques	17
Une prospective à très long terme sur ces relations entre collections physiques et numériques	18
Remerciements.....	18
Bibliographie.....	19
ANNEXE :.....	20

Synthèse

Ce rapport sur la relation collection physique- collection numérique répond à une commande du Directeur Général Délégué aux collections par lettre de mission en date du 15 juin 2017. Les objectifs de ce rapport sont d'établir un état des lieux raisonné de la gestion des collections physiques et des collections numériques et de faire des propositions pour une cohabitation synergique de ces collections tant dans le périmètre de l'établissement que dans un contexte européen et international.

Pour mener à bien cette analyse, une dizaine d'entretiens a été conduit entre septembre 2017 et mai 2018. L'exploitation des entretiens aboutit à un état des lieux contrasté avec le constat général qu'il existe au Muséum un savoir faire séculaire des dispositifs dédiés à la gestion des collections physiques, mais que malgré de vastes chantiers de numérisation des collections, le développement d'outils pour l'exploration et l'exploitation de ces collections numériques a été initié au Muséum mais demeure un enjeu majeur pour tirer pleinement profit de cette numérisation massive.

Le rapport résume les principaux enseignements tirés de l'analyse :
En particulier sur les collections physiques les détails complets des fonds ne sont que très partiellement informatisés et parfois même partiellement connus. Il semble donc impératif de maintenir un effort constant d'informatisation des collections. De plus, les collections d'objets physiques étant au cœur de l'établissement, il faut se méfier de ne pas tomber dans le travers du tout numérique et continuer à recoller et valoriser le fond patrimonial. Un système unifié pour accéder à l'information sur des spécimens partageant des relations biotiques mais conservés dans les différentes collections fait défaut et reste à inventer pour sortir d'une logique de silo et créer davantage de continuité entre les différents ensembles de collections. Un tel système devrait aussi favoriser les services d'accès à distance.

Le Muséum assure le stockage des données des collections numériques en interne. Le stockage de ces données numériques est extrêmement coûteux tant en infrastructures qu'en consommation énergétique ce qui questionne sur la pérennité de leur conservation dans des contextes budgétaires toujours tendus pour l'établissement. Le Muséum diffuse directement les données de collections numériques via le portail en ligne sur le site Web « Science » du MNHN. Malheureusement, les collections numériques du Muséum sont moins bien référencées dans les moteurs de recherche générique tel que Google- et via les portails internationaux de partage de données que celles d'autres institutions. Un meilleur référencement du fond devrait être réalisé pour qu'il soit davantage moissonné par les grands moteurs de recherche. En outre de cette diffusion directe, tout ou partie des données sont aussi diffusées via d'autres sites institutionnels dont le Muséum est partenaire (e.g. recolnat et le GBIF). Une mutualisation des efforts de diffusions et une stratégie commune pour augmenter la visibilité des données est une piste à envisager pour une meilleure accessibilité des données.

Dans un souci de réduire la duplication des efforts entre les institutions d'Histoire naturelle, une mutualisation des outils à mettre en place permettrait une économie d'échelle mais aussi davantage d'harmonisation des pratiques. C'est dans cet esprit que

Le programme European Distributed System of Scientific Collections (DiSSCo) a été conçu. Si ce projet est financé, la participation du Muséum est une priorité. DiSSCo est une nouvelle initiative paneuropéenne d'infrastructure de recherche de 21 pays européens visant à positionner les collections européennes de sciences naturelles au centre de l'excellence scientifique axée sur les données et de l'innovation dans la recherche environnementale, le changement climatique, la sécurité alimentaire, la santé et la bioéconomie. Le consortium DiSSCo prévoit de déposer une demande de financement à l'appel à projet « socioeconomic and cultural transformations in the context of the fourth industrial revolution » dont la date butoir est le 14 mars 2019.

Le rapport insiste sur les enjeux d'une science ouverte avec un libre accès aux données source (Open data) et la nécessité d'un renforcement de la conservation des données et de leur partage. Toutefois, l'ouverture des données impose une politique claire et transparente sur la qualité de la donnée. Les critères de la chartre de qualité devraient être davantage visible pour les utilisateurs des collections numériques. L'ouverture des données pose aussi la question de la propriété institutionnelle et intellectuelle des données. Les expéditions pour enrichir les collections physiques et les campagnes de numérisation pour enrichir les collections numériques sont largement subventionnées par des financements publics et du mécénat qui tous souhaitent être visibles. La traçabilité des financements et la propriété institutionnelle des collections tant physique que numérique semblent donc un enjeu majeur qui mériterait le développement d'outils pour en assurer la traçabilité.

L'étude des interactions entre les collections physiques et les collections numériques devrait émerger comme un nouvel objet de recherche ainsi que l'exigence de repenser les interactions entre les données et les objets. Cette réflexion pourrait avoir lieu dans le cadre plus large des réflexions qui ont actuellement lieu dans l'établissement sur la nécessité d'un cadrage et d'une impulsion pour l'ensemble de l'établissement autour de la question de la révolution et des innovations numériques.

Introduction

Le Muséum national d'Histoire naturelle est un établissement public à caractère scientifique, culturel et professionnel constitué sous la forme d'un grand établissement au sens de l'article L. 717-1 du code de l'éducation.

Selon l' Article 3 (Modifié par DÉCRET n°2014-1107 du 1er octobre 2014 - art. 2) :

« Dans le domaine des sciences naturelles et humaines, le Muséum contribue à la production, au développement et au partage des connaissances sur la diversité géologique et biologique de la Terre, sur la diversité des cultures et des sociétés et sur l'histoire de la planète.

A cette fin, il a pour mission de développer en synergie la recherche fondamentale et appliquée, l'expertise, la valorisation, l'enrichissement, la conservation et la mise à disposition des collections et des données, la formation dont l'enseignement, l'action éducative et la diffusion de la culture scientifique et technique à l'intention de tous les publics. »

Les collections scientifiques du Muséum comprennent certaines des plus grandes collections de leur genre dans le monde. Avec plus de 70 millions d'items d'une remarquable diversité (animaux vivants ou naturalisés, plantes vivantes ou en herbiers, ADN et tissus, graines, fossiles, minéraux, roches, météorites, objets ethnographiques, objets et documents scientifiques) elles constituent le cœur du travail des scientifiques de l'établissement mais aussi celui des scientifiques des autres institutions de recherche nationales et internationales. En effet, les collections d'Histoire naturelle fournissent une source essentielle pour le développement des preuves scientifiques et sont une ressource importante pour la recherche scientifique, l'éducation et la gestion des ressources (Berendsohn et al. 2017). Elles représentent les enregistrements de notre passé et de notre présent et constituent ainsi un investissement d'avenir ; elles constituent un échantillonnage spatio-temporel important qui peut être mis à profit pour relever les défis environnementaux auxquels l'humanité est confrontée. Les bases de données du Muséum couvrant les noms de taxons, les classifications et les informations basées sur les taxons, constituent une ressource précieuse qui nécessite au même titre que les collections physiques une conservation soignée et un développement minutieux afin de pouvoir être partagées et valorisées. La plupart des grandes institutions d'Histoire naturelle ont engagé ou sont en train d'engager une réflexion sur l'accessibilité de ces données (Willis et al. 2018). Pour l'avenir, il est important que le Muséum, qui a fait figure de pionnier avec la numérisation de l'herbier, poursuive ses engagements dans la numérisation et contribue au développement de moyens d'améliorer l'accès aux collections, à la fois physiquement et numériquement, pour en permettre une utilisation et un partage maximal mais également une pérennisation physique et numérique de la donnée et du spécimen.

Problématique

La numérisation massive des collections permet d'obtenir une photographie d'une collection physique à un temps donné. Toutefois, les collections du Muséum continuent d'être enrichies et s'accroissent notamment grâce aux expéditions organisées par l'établissement et celles auxquelles les chercheurs de l'établissement participent. De plus, les spécimens collections physiques évoluent au cours du temps (possible dégradation ou rénovation) et au grès de leurs études (avec éventuellement des

prélèvements destructifs, des montages pour observation, des annotations etc...). Les spécimens des collections numériques peuvent aussi faire l'objet d'annotations. A moyen et long terme, il s'établit un décalage entre une collection numérique effectuée à un temps t et la collection physique.

Objectif

L'objectif de ce rapport est d'établir un bilan sur la gestion des collections physiques et des collections numériques afin d'émettre des propositions pour une gestion plus synergique de ces collections.

Ce rapport ne prétend pas être exhaustif pour plusieurs raisons :

Pour des contraintes de temps, d'abord puisque la commande initiale comprenait une date butoir de livraison fin mai 2018.

Pour des raisons relatives à une nécessaire restriction du périmètre d'investigation : les relations entre collections physiques et collections numériques ont en effet un contour bien défini lorsque l'on s'intéresse uniquement à la gestion des données mais elles concernent aussi des domaines adjacents telle que l'utilisation qui est faite des données dans le cadre d'activités telles que la recherche, l'expertise, l'enseignement et la diffusion des connaissances.

Au regard de l'objet même d'étude : l'exploration et l'exploitation des données numériques est un champ relativement récent, en plein essor et dont la structuration est en train de se faire, car elle est étroitement dépendante des progrès technologiques. La technologie numérique permet d'ores et déjà l'accès aux données sources. Toutefois, des outils performants sont en cours de développement et devraient permettre à terme de mobiliser les données de collections dans le cadre de requêtes qui ne seraient plus forcément centrées sur des objets de collections naturelles, mais davantage sur des concepts émanant des sciences de l'évolution et de l'écologie grâce à des logiciels appropriés qui permettrait l'accès et le partage de la donnée.

Enfin la dimension internationale n'a pas pu être pleinement explorée faute de disponibilités aux dates auxquelles ont eu lieu les principales réunions. Il semble cependant crucial que le Muséum participe activement aux initiatives européennes qui sont en train de se mettre en place et dont l'objectif est d'établir des pratiques communes et de mutualiser les moyens pour une gestion standardisées et synergiques des collections numériques.

Éléments de contexte scientifique

Mobiliser les données de biodiversité dans un contexte de 6ème extinction.

Peu de controverses demeurent sur le fait que la Terre est actuellement confrontée à un changement global exceptionnel d'origine anthropogénique (Barnosky et al. 2011). En conséquence de ces perturbations majeures, l'érosion de la biodiversité s'accélère à la surface de notre planète et entraîne une crise majeure qualifiée de « 6ème extinction ». Dans ce contexte, les collections d'Histoire naturelle constituent un réservoir énorme d'informations sous-exploitées pour étudier les effets du changement global sur la biodiversité. Les collections d'Histoire naturelle étant un bien public, leur préservation,

leur gestion, leur développement, leur valorisation ainsi que leur accès pour une utilisation appropriée constituent un enjeu majeur.

Un nouveau prisme d'étude de la biodiversité : émergences des notions d'holobiomes/rerelations durables entre les organismes.

Comme en témoigne la première conférence internationale sur l'Holobionte qui s'est tenue au Muséum du 17 au 21 avril, l'appréhension de la diversité dans sa dimension mosaïque est une thématique scientifique en plein essor. L'essentiel des collections d'Histoire naturelle a été mis en place avant l'émergence de ce concept ; néanmoins beaucoup d'entre elles hébergent encore au moins une fraction d'organismes en interaction et peuvent ainsi constituer un objet d'étude d'un nouveau genre. Toutefois, les bases de données ne permettent pas actuellement de documenter suffisamment les interactions entre les organismes, et les collections numériques de documents de format variés pourraient constituer un moyen de faire émerger ces notions d'interactions sur la base des collections initialement mises en place en se centrant majoritairement autour du concept de l'espèce.



Un spécimen abrite toute une communauté, le microbiote, qui est au moins partiellement présent dans les collections mais les bases de données ne permettent pas de stocker ce type d'information.

Les collections numériques une alternative pour l'étude de la biodiversité dans le contexte de durcissement de l'accès aux ressources génétiques lié aux législations sur le partage des avantages découlant de leur utilisation (Accès et Partage des Avantages, APA) et à l'interdiction du prêt des spécimens types décidée par le Président du Muséum en date du 30 mai 2017

Suite aux accords de Nagoya, il se met en place dans chaque pays un contexte législatif où les ressources génétiques font l'objet de réglementations toujours plus contraignantes. Les collections numériques peuvent être vues comme une alternative d'accès aux données de ces ressources génétiques en s'affranchissant des démarches déclaratives. Toutefois il faut souligner que la notion de « ressource biologique » fait l'objet d'un débat dans la communauté. En particulier, la potentielle inclusion des « Digital Sequence Information » dans le périmètre des APA (Convention on Biological Diversity 2018) laisse craindre un frein d'envergure pour la recherche sur la biodiversité (Royal Society of Biology 2017). Si pour l'heure les documents numériques tels que des images 2D, images 3D, documents audio, documents vidéo ne sont pas dans le périmètre

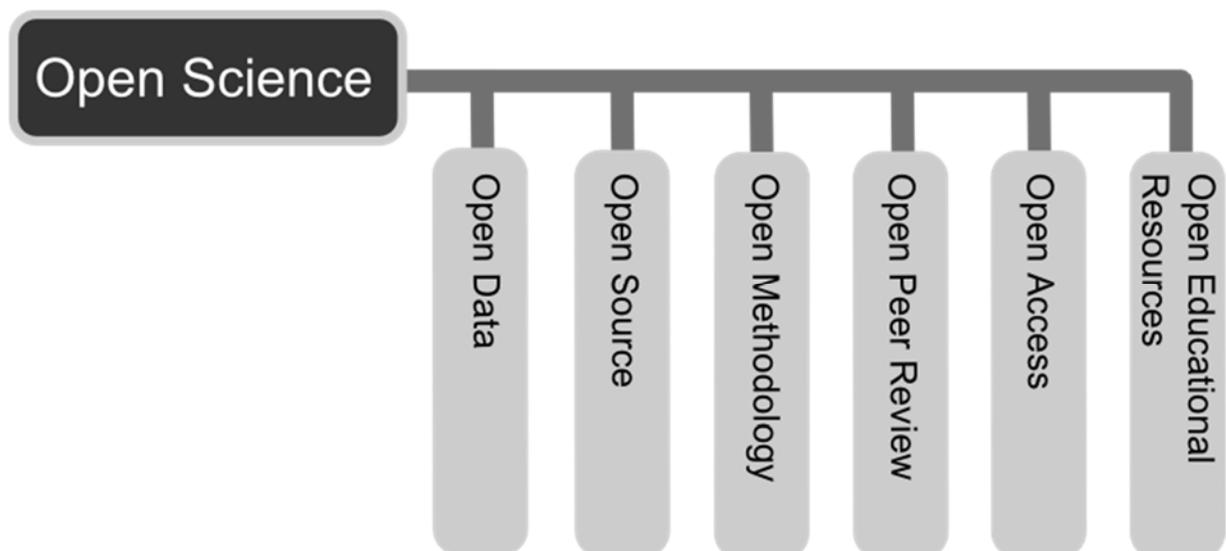
des APA ; il faut rester vigilants face à la montée de lobbying. De plus, en dépit de leur fort intérêt, les collections numériques ne peuvent remplacer les collections physiques dans de nombreux aspects des activités de recherche, notamment celles liées à la systématique où l'accès à de nombreux caractères nécessitent des investigations invasives ou non (e.g. analyses isotopiques, analyses chimiques des composés organiques). Les collections numériques ne pourront donc pas pallier entièrement l'arrêt du prêt des types (Staff Paris Herbarium 2018).

Vers une pratique d'Open science par défaut

La science ouverte est un mouvement dans lequel la science et les données sont perçues comme un bien commun, rendu accessible, notamment grâce aux outils du web, tant aux pairs qu'à tous publics intéressés. Cette transparence dans la pratique scientifique est largement promue par les instances européennes et relayées par les instances de la recherche françaises, notamment l'HCERES.

Le cadre législatif français organisant l'ouverture des données publiques a été profondément remanié par la « Loi Gratuité », par la « Loi pour une République numérique », puis par la codification de toutes ces dispositions au sein du Code des relations entre le public et l'administration (CRPA). L'ouverture des données publiques est devenue un principe par défaut, plusieurs régimes d'ouverture sont réunifiés et la loi étend la politique d'open data à un nombre accru d'acteurs publics : collectivités locales, services publics industriels et commerciaux...

La mise en place de collections numériques participe pleinement à cette dynamique de science ouverte. Cependant, l'ouverture des données nécessite la mise en place de procédure de validation et une réflexion pour établir une politique claire et transparente sur les propriétés institutionnelles/intellectuelles des données, un engagement sur le niveau de qualité des données.



Les six principes de l'Open Science ou Science ouverte (source Wikipedia)

Les collections d'Histoire naturelles : une preuve scientifique

De récentes affaires de suspicions de fraudes scientifiques ont agitées la communauté française des sciences biologiques et remis au centre la question des preuves

scientifiques et de leur conservation. Dans ce contexte, les collections d'Histoire naturelle ont été conçues pour accueillir des « spécimens référents » notamment pour la pratique de la systématique. En effet, les codes de nomenclatures prévoient le dépôt dans une institution d'un spécimen type porte-nom pour chaque nouveau taxon de rang spécifique ou infra. Les collections sont donc un socle indispensable à l'avancée de la connaissance. Les collections physiques permettent de conserver ces spécimens, et grâce aux collections numériques il est maintenant possible de les donner à voir pour une pratique vertueuse d'une science transparente, partagée et donc ouverte.

Les modes actuels de gestion des collections physiques et des collections numériques

i) Les bases de données pour la gestion et la valorisation des collections physiques

Les bases de données pour la gestion des collections physiques ont été développées sous ORACLE et leur consultation est possible via l'interface JACIM qui est le fruit d'un travail et d'échanges réguliers entre les informaticiens et les gestionnaires des collections ainsi que les chercheurs. Cet outil est en perpétuelle évolution pour répondre aux besoins et aux attentes des gestionnaires de collections et des chercheurs. Les procédures de sauvegarde des données sont mises en place et réalisées par la DSI (Direction des services informatiques du Muséum).

L'un des principaux enjeux des bases de données de collections réside dans le fait que les objets en collection sont de natures extrêmement variées et que les pratiques qui en découlent sont donc diverses. Il a donc été mis en place des ensembles qui correspondent approximativement à de grands groupes taxonomiques au sein desquels il existe une certaine harmonisation des pratiques et des logiques de conservation. Toutefois, même à l'intérieur d'un seul ensemble, se pose la question de ce qu'est l'objet biologique en collection. Il peut s'agir d'un fragment d'individu, d'un individu entier, d'une communauté.

Cette diversité d'échelle biologique des objets est difficile à prendre en compte dans les bases de données dont la clef d'entrée principale est un objet quel que soit son échelle biologique. Il est difficile de décrire complètement un objet de collection s'il comporte plusieurs taxa comme par une dalle à empreintes de feuilles fossilisées ou une planche d'herbier de mousses.

Par ailleurs, il apparaît que le degré d'informatisation des collections est très variable en fonction des fonds.

Il semble primordial que la base de données soit adaptée à la gestion des objets pour les collections physiques (et pas forcément par une approche taxon centrée). Poursuivre les efforts d'informatisation de la donnée est une priorité absolue.

ii) Les bases de données pour les collections numériques

Depuis une quinzaine d'années, de nombreuses initiatives de numérisation à haut débit ont vues le jour tant dans les domaines scientifiques (Tulig et al. 2012) que dans d'autres

domaines telle que la numérisation des archives tant publiques que privées. Ce flot massif de données numériques offre de nombreuses possibilités jusque lors inaccessibles. Cependant, le développement d'outils pour l'exploration et l'exploitation de ces volumes importants de données demeure un enjeu majeur pour tirer pleinement profit de cette numérisation massive.

Les objets numérisés sont de formats variés. Parmi eux on peut citer :

- images 2D
- images 3D
- documents audio
- documents vidéo

Parmi la diversité des ensembles de collections du Muséum, tous ces formats de documents sont représentés. Toutefois, les techniques de numérisation d'image 2D étant les plus simples sur le plan technique, les objets bi dimensionnels tels que les herbiers ont été très largement numérisés (Le Bras et al. 2017). Cette numérisation permet un partage de la donnée qui est disponible sous licence Creative Commons CC0 ce qui est conforme avec les pratiques d'Open Science. Toutefois, au delà de l'ensemble de botanique qui constituait un chantier de numérisation relativement simple et très similaire dans sa mise en œuvre a une numérisation d'un fond d'archives, l'essentiel des autres collections du Muséum est tridimensionnel ce qui pose des problèmes techniques tant pour l'acquisition que le stockage de la donnée. En outre les caractères accessibles par une numérisation surfacique ne sont pas toujours les plus utilisés pour la recherche scientifique. Une réflexion sur la plus-value pour la communauté scientifique de la numérisation d'une collection devrait avoir lieu et conduire à une priorisation des chantiers.

Aujourd'hui, il n'existe pas à proprement parler de bases de données indépendantes pour les collections numériques ; les documents numériques sont liés aux spécimens de la collection physique (voir schéma). Chaque document numérique bénéficie de l'attribution d'un identifiant pérenne accessible par une url unique.

Exemple :

<https://science.mnhn.fr/institution/mnhn/collection/f/item/r09452?listIndex=1&listCount=120>

L'utilisation de ces identifiants uniques est à poursuivre. Il faudrait aussi veiller à préserver autant que possible les informations liées aux modes de numérisation (date, projet de numérisation, méthode de validation, etc). Il est probable que dans le futur un même objet puisse être numérisé à plusieurs reprises avec éventuellement des technologies différentes. L'ensemble des documents numériques devrait rester disponible et chacun avoir un identifiant unique.

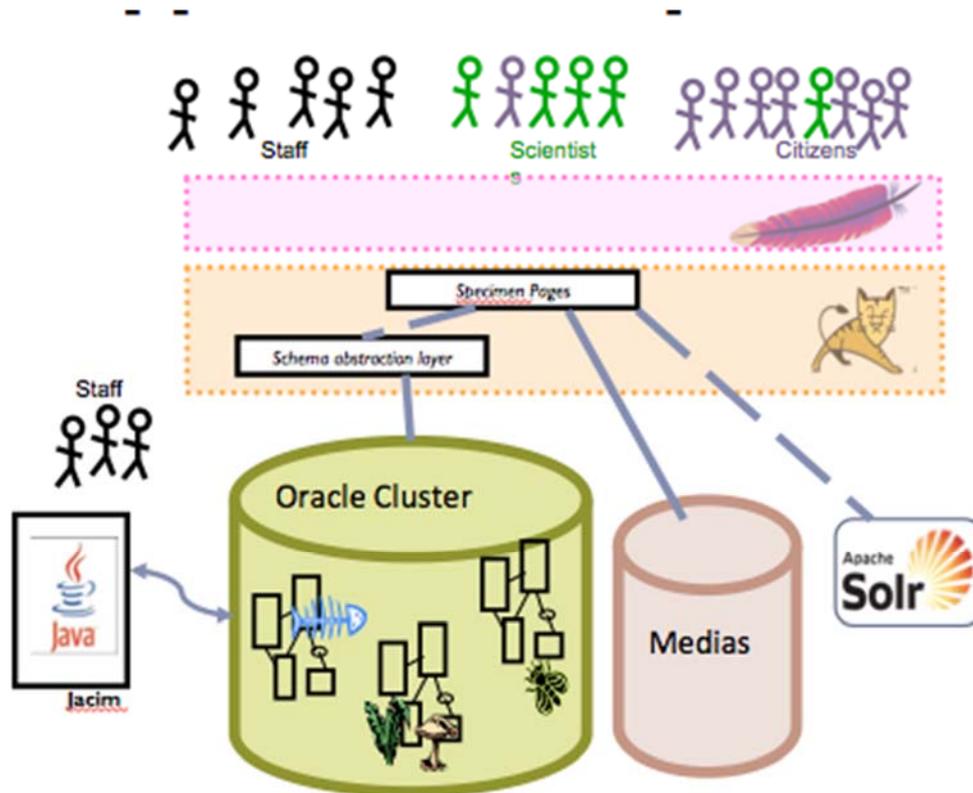


Schéma conceptuel de l'interface « science » permettant l'accès aux collections numériques et aux informations qui leurs sont liées (Source : aimablement fourni par Simon Chagnoux)

Les collections numériques présentent l'avantage d'être consultable par quiconque ayant à disposition une interface connectée à Internet ce qui devrait être davantage valorisé en permettant un retour des utilisateurs qui pourraient contribuer à signaler des ressources, à relever des inexactitudes etc.

iii) L'interopérabilité des collections « physiques » et « numériques »

La numérisation des collections d'Histoire naturelle appelle de nouvelles opportunités d'analyse des données et par conséquent de nouveaux besoins en accessibilité de ces données. Poursuivre l'effort d'informatisation des spécimens des collections paraît une priorité. En effet, il existe de très fortes disparités d'informatisation des collections, certaines étant très lacunaire. Il y a donc un risque, avec la généralisation des consultations via le site « Science », que des collections majeures mais peu renseignées tombent dans l'oubli puisque les requêtes qui concerneront des champs non renseignés, ne permettront pas de faire ressortir ces spécimens puisque l'utilisateur ne verra que ce qui est visible.

Il semble donc indispensable de communiquer l'état actuel de l'informatisation des divers ensembles de collections pour que les utilisateurs puissent évaluer la pertinence des réponses qu'il obtient au regard du degré d'informatisation des collections.

La segmentation éventuelle des utilisateurs des deux types de collection

Si les rapports d'activités des différentes Unités de Gestion des Collections permettent de suivre l'évolution des activités sur les collections d'objets en recensant les visites et les prêts via l'interface web colhelper, l'absence d'outils performants de mesure de l'audience des sites Internet tant pour les bases de données Muséum que pour les BD agrégatives (i.e. GBIF, RECOLNAT, EUROPEANA, etc) obère une analyse des usages des collections numériques.

A minima, il faudrait pouvoir disposer de statistiques sur les usages numériques et de l'Internet comprenant notamment :

Le nombre de visiteurs uniques,

Le nombre de pages vues

Le nombre de pays ayant consulté au moins une page

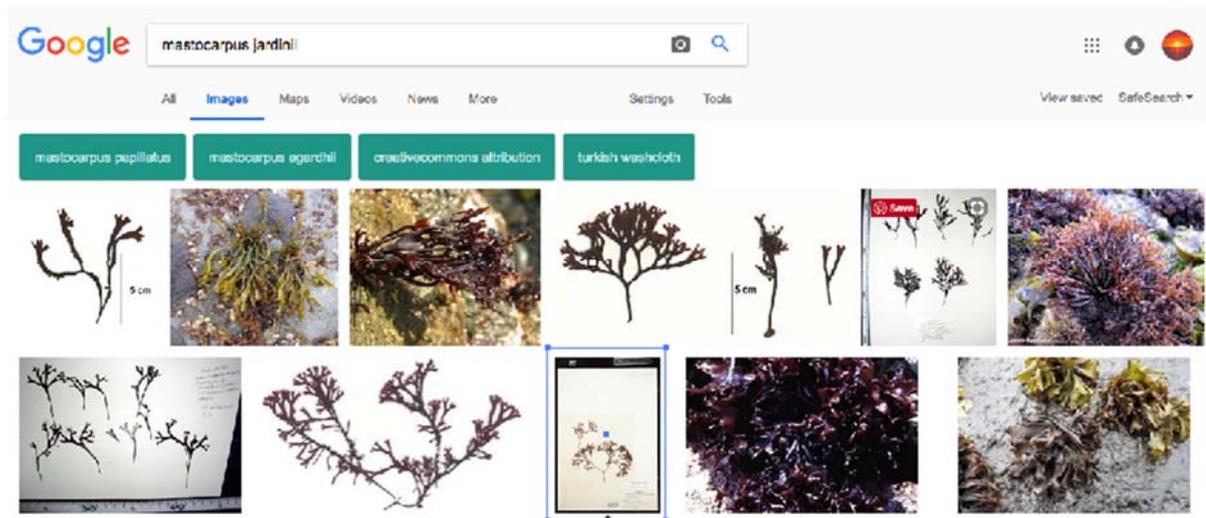
Un regard sur quelles sont les collections numériques les plus visitées permettrait de mieux percevoir la diversité des communautés d'utilisateurs. Cette analyse permettrait de disposer d'une donnée supplémentaire lors des arbitrages des priorités des numérisations à venir.

Il faudrait augmenter la visibilité des collections numériques en optimisant leurs référencement pour améliorer leur accessibilité et favoriser leur consultation par un public plus varié.

Aujourd'hui l'interface web « science » est essentiellement conçue pour une consultation à travers le formulaire :

<https://science.mnhn.fr/institution/mnhn/search>

Ce formulaire ne permet des requêtes que très formatées (essentiellement un nom de taxon) bien adaptées à la communauté scientifique (en particulier aux taxonomistes et systématiciens) et à celle des amateurs éclairés. Cependant ce formulaire est peu ergonomique pour un public large y compris les enseignants et les professionnels de la diffusion des connaissances qui s'interrogent davantage sur des concepts que sur des taxons. Actuellement, il est plus que probable que les utilisateurs des collections numériques soient principalement des scientifiques (professionnels ou non). Parmi les développements facilement envisageables pour élargir le public de la collection numérique, il faudrait favoriser le repérage par les moteurs de recherche des images des collections numériques afin que l'utilisation du formulaire ne soit plus l'unique entrée de cette collection. Par exemple une recherche dans un moteur tel que Google image sur un nom de taxon d'algues permet de visualiser des collections américaines (e.g. University of California, Berkeley, University of North Carolina...) parmi les premiers référencement de Google.



Résultat d'une recherche sur Google image à partir d'un nom de taxon d'algue. Des Herbiers américains sont parmi les premières références (l'image encadrée en bleue est un scan d'Herbier de l'University of New Carolina)

L'intérêt ou la nécessité de gérer indépendamment les deux types de collections, et dans ce cas la pertinence de créer des outils pour assurer un lien

i) Un changement pour les utilisateurs des collections numériques

Jusque récemment l'accès aux collections d'Histoire naturelle était l'apanage quasiment exclusif des scientifiques. La numérisation des collections permet une accessibilité de ces données à une vaste communauté de scientifiques et de citoyens. Ce changement ouvre la voie à une participation active de cette nouvelle communauté à la conservation de la donnée. Les données liées à la collection physique pourraient alors être enrichies par des données issues de la collection numérique. Les programmes de sciences participatives tels que les Herbonautes (<http://lesherbonautes.mnhn.fr/>) ont été un vrai succès. Cependant ils fonctionnent sur des projets clairement délimités mis en place par des chercheurs du Muséum. Les projets de sciences participatives se conçoivent dorénavant en considérant les citoyens comme des acteurs à part entière du projet (co-construction de projet). Pour davantage les impliquer dans la démarche sans les contraindre aux projets émanant des chercheurs, il faudrait leur laisser la possibilité d'intervenir sur toutes les données pour lesquelles ils souhaitent s'investir.

Il faudrait réfléchir à ce que l'ensemble des données disponibles via le site « Science » puisse bénéficier du retour de la communauté des usagers dans un souci de progresser sur la qualité de la donnée.

Les modalités de validation de ces informations citoyennes restent à inventer mais plusieurs pistes pourraient être explorées :

- une validation par un gestionnaire de la collection.
- une validation citoyenne (par des pairs citoyens) avec établissement d'un seuil au delà duquel la contribution est validée.
- mise en place de règles de priorisation de synchronisation dans les données nouvelles versus données existantes

Afin que la communauté des utilisateurs des collections numériques continue à croître il faudrait aussi mener une réflexion sur les pratiques de ces usagers et davantage cerner leurs besoins.

L'enjeu résidera alors à adapter les catalogues de collections à l'environnement du Web pour fournir aux utilisateurs des données pertinentes pour leurs besoins. Il est vraisemblable que la communauté des usagers soit éclectique et que les usages soient par conséquent variés. Le développement d'outils favorisant l'ergonomie et la sélection des seules données attendues serait une plus value pour améliorer le confort des usagers. Une des stratégies pourrait consister à sortir les données des silos afin de créer un réseau de données interconnectées en utilisant les règles des ontologies informatiques (Holetschek et al. 2016). L'un des principaux enjeux réside dans le fait que les utilisateurs mobilisent généralement des concepts (populations, espèces, communautés, holobiomes, écosystèmes) alors que la base de données est initialement conçue pour gérer des données de la biodiversité basées sur des objets physique. Des réflexions sur la formalisation de concepts taxonomiques a déjà été initiées (de Jong et al. 2015).

ii) Un changement pour les utilisateurs et gestionnaires des collections physiques

Historiquement, seule la consultation des collections physiques faisait foi et toutes les annotations étaient réalisées directement sur cette collection. Les gestionnaires des collections dédiaient une partie de leur temps à l'accueil des visiteurs fortement impliqués dans la réalisation des annotations. Si on promeut une pratique nouvelle dans laquelle les utilisateurs des collections numériques pourraient annoter la collection numérique, il faut envisager que la consultation de la collection physique ne soit plus l'unique référence.

Il faudrait alors concevoir un nouvel itinéraire dans lequel la consultation simultanée de la ressource physique et numérique serait favorisée afin que l'ensemble des annotations (collection numériques et physiques) puisse être considéré concomitamment.

Des outils favorisant la visualisation simultanée de plusieurs documents numériques ont d'ores et déjà fait l'objet de développement (Suhrieb et al. 2017) et ceux mis en place dans le cadre du laboratoire de RECOLNAT sont opérationnels (<https://lab.recolnat.org/>). Il faudrait toutefois voir à intégrer ces fonctionnalités à JACIM avec d'éventuelles adaptations. Enfin, un changement de pratique dans l'activité ne pourra se faire sans un accompagnement du personnel (agents des collections et chargés de conservation) qui pourrait prendre la forme d'animations scientifiques sous forme de séminaires et/ou d'ateliers. Le surcroît d'activité lie aux changements de ces pratiques impacte directement les gestionnaire des collections, toutefois, il ne faut pas que la gestion des collections numériques (accroissement, suivi et pérennisation de la donnée) se fasse au détriment de la collection physique. Il semble pertinent de souligner ici que les métiers de gestion de collections physiques et de gestion de collections numériques sont différents.

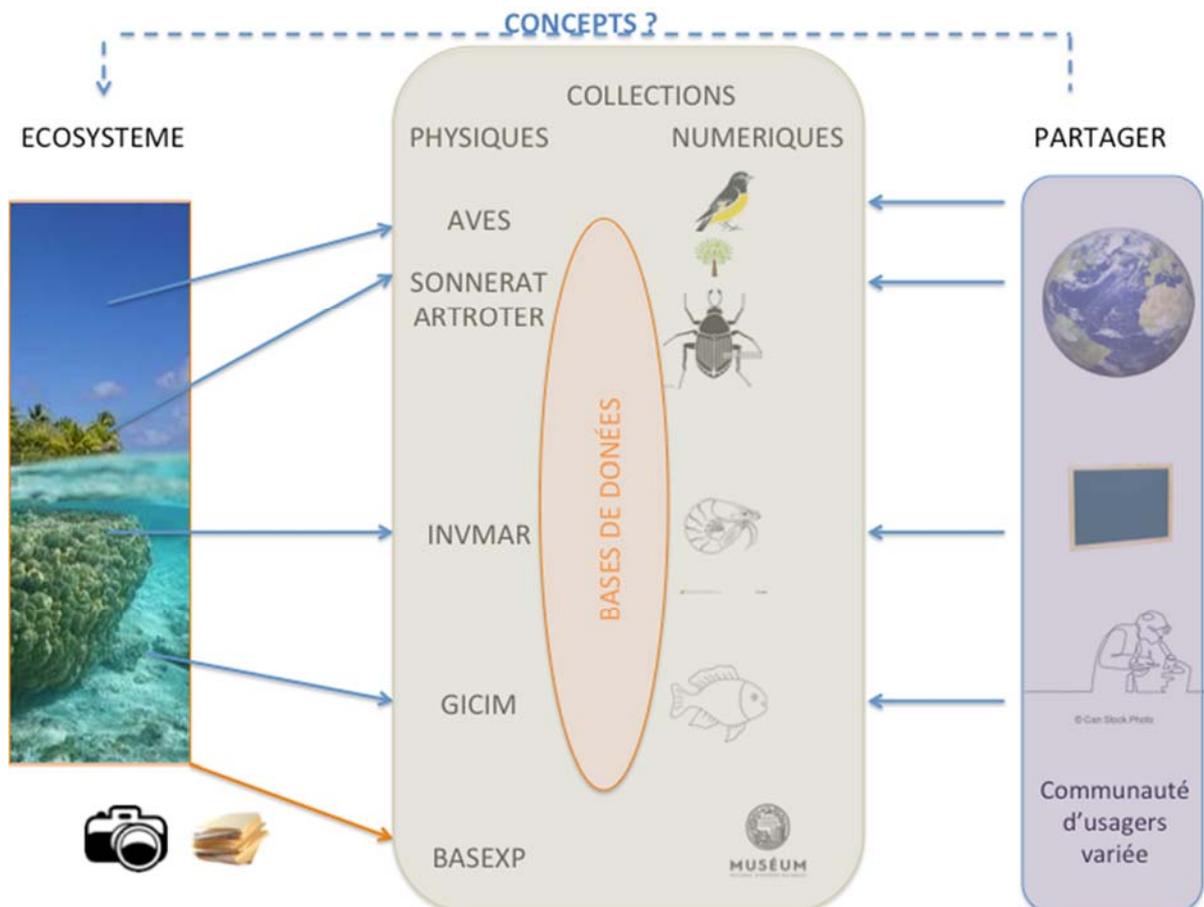


Schéma de bancarisation de la donnée. L'organisation de la donnée avec une logique de silos correspondant aux grands ensembles ne permet pas aujourd'hui d'effectuer des requêtes sur des concepts liés aux différents degrés d'organisation de la biodiversité (e.g. holobiome, communauté, écosystème)...

Les moyens, conceptuels et techniques, d'intégrer les apports mutuels dans les collections

Les collections numériques devraient permettre de dépasser la logique « d'ensembles » avec laquelle sont gérées les collections physiques pour des raisons évidentes liées aux pratiques de conservation. Une base de données sur les liens entre les organismes pourrait être mise en place pour décrire les relations observées soit lors de la collecte (notamment lors des grandes expéditions qui ambitionnent des échantillonnages multitaxons) soit lors d'une observation a posteriori (comme par exemple le projet de Florent Martos d'étude des organismes de la rhizosphère sur les spécimens d'Herbier). Les relations entre les organismes devraient dorénavant pouvoir être consignées dans une base de données de JACIM en permettant de bancariser les liens entre des spécimens d'ensemble différents. En vu d'un codage efficace pour caractériser ces liens une réflexion sur les différentes modalités de relation durable entre les organismes devrait avoir lieu.

Afin que les collections d'histoire naturelle soient davantage mobilisées dans la pratique de la recherche au delà de la systématique, il faudrait envisager de mettre en place des

fonctionnalités qui permettent de gérer des données de la biodiversité en prenant davantage en compte ses différentes échelles. A titre d'exemple, au cours des 10 dernières années de nombreuses initiatives ont vu le jour pour mettre en place des bases de données de traits. Le développement d'outils permettant de faire le lien entre des bases de données de traits et celles du Muséum serait une avancée considérable pour la recherche scientifique et la valorisation des collections dans des approches d'écologie fonctionnelle.

Les collections d'Histoire naturelle sont aussi un formidable objet pour étudier l'histoire des pratiques scientifiques et les itinéraires des savoirs. Toutefois, ce témoignage historique est peu visible à travers les collections numériques. Les évolutions des classifications pourraient être davantage visible si des approches multi classificatoires étaient entreprises.

Le paysage actuel au niveau européen et international, et les initiatives qui pourraient être prises au niveau du CETAF pour harmoniser les pratiques

Dans un souci de réduire la duplication des efforts entre les institutions, la mutualisation des moyens et des pratiques est à promouvoir.

Le Muséum est signataire de La Déclaration de Bouchout

(<http://www.bouchoutdeclaration.org/>) pour la gestion ouverte des connaissances sur la biodiversité et la France s'est engagée en signant la charte européenne d'accès aux infrastructures de recherche

(https://ec.europa.eu/research/infrastructures/pdf/2016_charterforaccessto-ris.pdf)

Le Muséum est déjà fortement impliqué dans le paysage européen des infrastructures de recherche. Membre du CETAF (consortium of European Taxonomic Facilities/ 33 institutions/21 pays) il contribue activement aux réflexions et au développement de nouveaux outils mutualisés (e.g. colhelper). Au delà du cercle du CETAF, 114 institutions /21 pays participent à une nouvelle initiative paneuropéenne d'infrastructure de recherche, DiSSCo (European Distributed System of Scientific Collections). Cette initiative ambitionne de positionner les collections européennes de sciences naturelles au centre de l'excellence scientifique en axant sur les données et l'innovation dans la recherche environnementale, le changement climatique, la sécurité alimentaire, la santé et la bioéconomie. Le pilotage de ce projet est actuellement assuré par Naturalis, Leiden.

Si la candidature du programme DiSSCo est retenue, ce que nous devrions savoir officiellement en septembre 2018, la participation active du Muséum au programme DiSSCo est une priorité.

Le Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation a exprimé son soutien à cette initiative en date du 26 Juillet 2017. Le consortium DiSSCo prévoit de déposer une demande de financement à l'appel à projet « socioeconomic and cultural transformations in the context of the fourth industrial revolution » dont la date butoir est le 14 mars 2019.

Une prospective à très long terme sur ces relations entre collections physiques et numériques

La numérisation des collections d'Histoire naturelle appelle de nouvelles opportunités d'analyse des données et par conséquent de nouveaux besoins en accessibilité de ces données. Si l'utilisation des bases de données via l'interface JACIM pour les gestionnaires et d'autres interfaces pour mettre à disposition du public les contenus de la collection numérique est fonctionnelle, un continuum reste à inventer pour que les utilisateurs puissent mobiliser les données correspondant à leurs attentes... Limitation des BD qui gèrent des objets physiques ou numériques alors que la science s'interroge sur des concepts émergents de ces objets mais les dépassant... A l'instar de ce qui se développe dans le domaine des métadonnées documentaires qui mettent en place des outils de gestion de concepts bases sur Un code de catalogage conçu pour répondre aux défis du Web (modèles FRBR/FRAD)

Outre le développement d'outils informatiques, une piste à creuser pour ouvrir davantage les collections numériques serait de mettre en place un service de médiation. Comme il existe des médiateurs pour la visite de collections d'objet, un métier au cœur de la mission du Muséum, une équipe de médiation des collections numériques pourrait contribuer à mieux connaître la communauté des utilisateurs des collections numériques et leurs attentes. Cette équipe de médiateurs pourrait ainsi proposer de nouveaux développements en concertation tant avec les équipes en charge du développement des bases de données ou de la conservation des collections physiques qu'avec les scientifiques.

Le Muséum dispose d'un Centre de recherche sur la conservation pour développer des recherches sur la conservation des collections physiques.

A l'instar de ce centre, développer une cellule de recherche multidisciplinaire et multi métiers sur les collections numériques permettrait de développer des solutions pour surmonter les divers verrous qui subsistent encore pour tirer pleinement profit de la numérisation des collections.

Cette cellule pourrait avoir comme mission initiale la mise en place d'un mode de gouvernance transversal des collections numériques avec un responsable clairement identifié.

Résumé des 11 principales recommandations

1. Poursuivre les efforts d'informatisation de la donnée
2. Assurer la pérennité du stockage des données numériques
3. Avoir une politique claire et transparente sur la propriété intellectuelle et institutionnelle de la collection numérique
4. Mettre en place une chartre de qualité de la donnée et la communiquer
5. Augmenter l'accessibilité de la collection numérique
6. Tirer davantage profit de la communauté d'utilisateurs de la collection numérique pour enrichir la base de données
7. Concevoir un nouvel itinéraire de consultation simultanée de la ressource physique et numérique afin d'accéder concomitamment aux annotations sur les deux types de collection.

8. Mettre en place d'une base de donnée au sein de JACIM permettant de documenter les différentes relations biotiques interspécifiques.
Dans un second temps, une réflexion devrait avoir lieu pour développer les ontologies nécessaires pour qualifier plus subtilement ces relations (e.g. parasitismes, mutualismes, compétitions, prédatons).
9. Mise en place d'une médiation entre les usagers web des collections et les gestionnaires des données.
10. Mise en place d'une cellule de recherche multidisciplinaire sur les collections numériques
11. Promouvoir une participation active du Muséum au programme DiSSCo

Remerciements

Je tiens à remercier Michel Guiraud pour la confiance qu'il m'a accordée en me confiant cette mission. Je souhaite aussi exprimer ma gratitude à l'ensemble des participants aux entretiens pour le temps qu'ils m'ont consacré, les idées et connaissances qu'ils ont bien voulu partager et pour les échanges fructueux que nous avons pu avoir. Mes sincères remerciements s'adressent aussi à Nadia Améziane qui a bien voulu faire une relecture de ce rapport dans un délai très bref.

Bibliographie

- BARNOSKY A. D., MATZKE N., TOMIYA S., WOGAN G. O. U., SWARTZ B., QUENTAL T. B., MARSHALL C., MCGUIRE J. L., LINDSEY E. L., MAGUIRE K. C., MERSEY B. & FERRER E. A. 2011 - Has the Earth's sixth mass extinction already arrived?. *Nature* 471 (7336) : 51–57.
- BERENDSOHN W., MÜLLER A., KOHLBECKER A., GÜNTSCH A., LUTHER K. & PLITZNER P. 2017 - The CDM Applied: Handling of Names, Taxa and Concepts in a Conservation Context. *Proceedings of TDWG* 1: e20364.
- CONVENTION ON BIOLOGICAL DIVERSITY 2018 - Fact-finding and scoping study on digital sequence information on genetic resources in the context of the convention on biological diversity and the Nagoya protocol. Ad hoc technical expert group on digital sequence information on genetic resources.
- HOLETSCHEK J., BAUMANN G., KOCH G. & BERENDSOHN W. G. 2016 - Natural History in Europeana - Accessing Scientific Collection Objects via LOD. In: (Ed. by E. Garoufallou, I. Subirats Coll, A. Stellato & J. Greenberg), *Metadata and Semantics Research* pp. 223–234. Springer International Publishing, Cham.
- DE JONG Y., KOUWENBERG J., BOUMANS L., HUSSEY C., HYAM R., NICOLSON N., KIRK P., PATON A., MICHEL E., GUIRY M., BOEGH P., PEDERSEN H., ENGHOFF H., VON RAAB-STRAUBE E., GÜNTSCH A., GEOFFROY M., MÜLLER A., KOHLBECKER A., BERENDSOHN W., APPELTANS W., ARVANITIDIS C., VANHOORNE B., DECLERCK J., VANDEPITTE L., HERNANDEZ F., NASH R., COSTELLO M., OUVARD D., BEZARD-FALGAS P., BOURGOIN T., WETZEL F., GLÖCKLER F., KORB G., RING C., HAGEDORN G., HÄUSER C., AKTAÇ N., ASAN A., ARDELEAN A., BORGES P., DHORA D., KHACHATRYAN H., MALICKY M., IBRAHIMOV S., TUZIKOV A., DE WEVER A., MONCHEVA S., SPASSOV N., CHOBOT K., POPOV A., BORŠIĆ I., SFENTHOURAKIS S., KÖLJALG U., UOTILA P., OLIVIER G., DAUVIN J.-C., TARKHNISHVILI D., CHALADZE G., TUERKAY M., LEGAKIS A., PEREGOVITS L., GUDMUNDSSON G., ÓLAFSSON E., LYSAGHT L., GALIL B., RAIMONDO F., DOMINA G., STOCH F., MINELLI A., SPUNGIS V., BUDRYS E., OLENIN S., TURPEL A., WALISCH T., KRPACH V., GAMBIN M., UNGUREANU L., KARAMAN G., KLEUKERS R., STUR E., AAGAARD K., VALLAND N., MOEN T., BOGDANOWICZ W., TYKARSKI

P., WĘSŁAWSKI J., KĘDRA M., M. DE FRIAS MARTINS A., ABREU A., SILVA R., MEDVEDEV S., RYSS A., ŠIMIĆ S., MARHOLD K., STLOUKAL E., TOME D., RAMOS M., VALDÉS B., PINA F., KULLANDER S., TELENUS A., GONSETH Y., TSCHUDIN P., SERGEYEVA O., VLADYMYROV V., RIZUN V., RAPER C., LEAR D., STOEV P., PENEV L., RUBIO A., BACKELJAU T., SAARENMAA H. & ULENBERG S. 2015 - PESI - a taxonomic backbone for Europe. *Biodiversity Data Journal* 3: e5848.

LE BRAS G., PIGNAL M., JEANSON M. L., MULLER S., AUPIC C., CARRÉ B., FLAMENT G., GAUDEUL M., GONÇALVES C., INVERNÓN V. R., JABBOUR F., LERAT E., LOWRY P. P., OFFROY B., PIMPARÉ E. P., PONCY O., ROUHAN G. & HAEVERMANS T. 2017 - The French Muséum national d'histoire naturelle vascular plant herbarium collection dataset. *Scientific Data* 4: 170016.

ROYAL SOCIETY OF BIOLOGY 2017 - Response from the Royal Society of Biology to Defra's request for views and relevant information on potential implications of the use of Digital Sequence Information (DSI) on genetic resources for the three objectives of the Convention on Biological Diversity (CBD) and for the objective of the Nagoya Protocol on Access and Benefit Sharing (ABS).

STAFF PARIS HERBARIUM 2018 - Loan of Types from Paris Herbarium (P & Pc). *Taxon* 67 (2) : 467–467.

SUHRBIER L., TSCHÖPE O., BERENDSOHN W. & GÜNTSCH A. 2017 - Integrating AnnoSys into your specimen data portal. *Proceedings of TDWG* 1: e20313.

TULIG M., TARNOWSKY N., BEVANS M., KIRCHGESSNER A. & THIERS B. 2012 - Increasing the efficiency of digitization workflows for herbarium specimens. *ZooKeys* 209: 103–113.

WILLIS K. J., PATON A. J. & SMITH R. J. 2018 - Science Collections Strategy 2018–2028. Royal Botanic Gardens, Kew.

ANNEXE :

Personnes ressources auprès des quelles ont été conduite un entretien entre septembre 2018 et mai 2019.

Thierry [Bourguoin](#), Vice Chair executive committee CTAF
 Simon [Chagnoux](#), Chef de projet - Développeur Informatisation des Collections
 Frédérique [Chlous](#), Directrice du Département Hommes et environnement
 Gaël [Clément](#), Directeur du Département Origines et Evolution
 Dario [De Franceschi](#), ex Chargé d'ensemble de la Paléontologie
 Vanessa [Demanoff](#), Responsable de la Cellule Ingénierie de Projet
 Bruno [Denetière](#), Chargé de conservation
 Gildas [Illien](#), Directeur des bibliothèques et de la documentation / Adjoint au Directeur général délégué aux collections
 Florent [Martos](#), Chargé de conservation
 Serge [Muller](#), Chargé d'ensemble de la Botanique
 Marc [Pignal](#), Directeur de l'infrastructure RECOLNAT
 Germinal [Rouhan](#), Chargé de conservation
 Jian-Sheng [Sun](#), Directeur du Département Adaption du vivant