

# Numériser et promouvoir les collections d'histoire naturelle

**MARC PIGNAL**

pignal@mnhn.fr

**ÉVA PÉREZ**

eperez@mnhn.fr

Muséum national d'histoire naturelle

*Marc Pignal est botaniste. Il a été responsable de l'herbier national à Paris. Il a étudié sur le terrain la flore de Guyane et du Brésil, mais aussi des Comores et de Nouvelle-Calédonie. Il est actuellement chargé de mission au MNHN, dirige des programmes liés à la numérisation des collections, comme le projet franco-brésilien REFLORA ou l'infrastructure biosanté e-ReColNat.*

*Eva Pérez est ingénieur botaniste. Après des travaux de terrain en Argentine, puis en Guyane française, elle a participé à l'opération de numérisation de l'herbier national de Paris comme expert en techniques d'herbiers. Elle travaille actuellement sur le montage de l'infrastructure en biosanté e-ReColNat.*

Les collections naturalistes ont de commun avec les fonds documentaires le devoir de conservation des objets et de leur diffusion auprès des publics. Conserver et diffuser sont l'avvers et le revers d'une même médaille.

Mais le matériel d'histoire naturelle se révèle en général plus divers en taille et en composants, plus fragile et parfois beaucoup plus ancien puisqu'il couvre des périodes beaucoup plus longues (particulièrement si l'on songe à la paléontologie ou à la géologie). Contrairement aux livres, les objets naturalistes sont des données primaires comme pourraient l'être des fonds d'archives, et leur compréhension requiert une certaine connaissance technique et biologique.

## Des collections indépendantes des modes et des objectifs à court terme

### Conserver n'est pas seulement étudier

La politique de conservation des objets d'histoire naturelle suit une dynamique entièrement différente de celle de son étude. À ce sujet, un métier, indépendant de la recherche, est décrit au sein même de la branche d'activité professionnelle (BAP) du ministère de l'Enseignement supérieur et de la Recherche (MESR), à savoir le BAPF : Information, Documentation, Culture, Communication, Édition, TICE. Celui-ci, complémentaire du monde de la recherche ne peut cepen-

dant lui être inféodé. La conservation s'occupe de la pérennité des collections et doit être capable de juger si l'atteinte à un objet pour des besoins de recherche est pertinente ou non.

Par ailleurs, l'accès des objets à tous les chercheurs, dans le respect de la conservation des objets, est indispensable. Des conflits d'intérêt peuvent apparaître lorsqu'une équipe de recherche est responsable administrative d'une collection patrimoniale.

## Des objets conservés pour de nouvelles questions scientifiques

L'évolution des idées et les grandes questions liées à la biologie forment un nouveau contexte dans lequel les objets participent à la connaissance de l'environnement, jusqu'à la modélisation du changement global, mais aussi à l'histoire des sciences et des usages. Le public s'élargit donc à mesure que l'on s'éloigne de la systématique.

## Un procédé industriel pour plus de trois siècles de conservation

Les techniques de virtualisation des collections ouvrent celles-ci à des publics nouveaux et des utilisations nouvelles. Elles peuvent aussi inquiéter les gestionnaires actuels. Ces technologies ouvrent la perspective d'une virtualisation du travail sur des objets naturalistes. Elles participent à la production, la mise en valeur et la diffusion de connaissances scientifiques.

En prenant exemple sur la numérisation des herbiers à grande échelle, voici un court descriptif des atouts et des limites de ces approches nouvelles.

### L'exemple des herbiers

Les planches d'herbiers témoignent de l'existence d'une plante à un endroit et à un moment donné. Plus de trois siècles de prospections ont permis de constituer en France

des collections de plus de 20 millions d'exemplaires.

Techniquement, les herbiers sont des objets commodes : deux dimensions, un format plus ou moins standardisé. C'est la raison pour laquelle des collections géantes ont vu le jour. L'herbier national français conserve environ 9 millions de spécimens. Ils peuvent s'assimiler en plus à des livres diffusés à peu d'exemplaires grâce au principe des doubles déposés dans plusieurs institutions. Les techniques

d'indexation s'inspirent d'ailleurs de celles pratiquées dans les bibliothèques.

### Des différences entre herbiers et bibliothèques

La plupart des collections d'herbiers disposent rarement d'inventaires. Elles se sont constituées selon le principe des poupées russes, les fonds s'intègrent à d'autres collections pour aboutir *in fine* dans les grands instituts. L'un des défis est de savoir exactement quels spécimens constituent les fonds d'herbiers.

Les herbiers demeurent plus fragiles que les livres à la manipulation et aux attaques des insectes.

Pour toutes ces raisons, la numérisation en nombre des planches d'herbier était une approche innovante pour atteindre les objectifs d'inventaire et de diffusion. Sa réalisation peut être considérée comme un projet pilote applicable en partie à d'autres collections.

Au cours des trois années de production de la numérisation de l'herbier national de Paris, plus de 6 millions d'images ont été mises en ligne au fur et à mesure de leur réalisation. La rationalisation du travail et la mise en place de chaînes de production a produit jusqu'à 12 000 images par jour.

Le frein majeur se situait dans le tri et le classement des collections non intégrées dans la collection générale. L'opération demandait une approche plus artisanale. Les connaissances botaniques se sont révélées utiles pour comprendre et interpréter les étiquettes des spécimens qui sont à la base du classement.

### Numérisation de masse ou numérisation sélective ?

Deux philosophies s'affrontent : la numérisation d'objets choisis contre la numérisation industrielle de tous les objets d'une collection. La première est sans doute intellectuellement la plus satisfaisante ; elle est pratiquée d'ailleurs depuis des années dans énormément de collections. La pratique de sélectionner prioritaire-

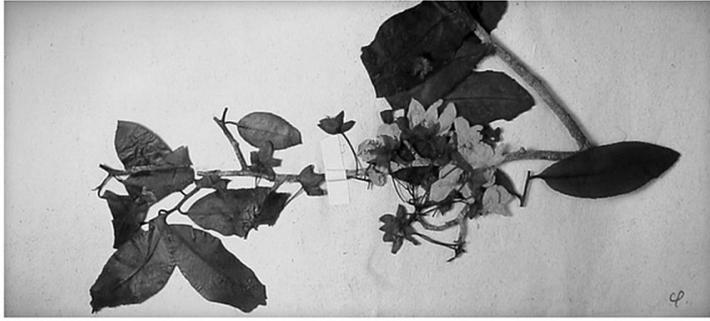


Chlorophytum de Madagascar (image d'herbier).

Les herbonautes  
L'herbier numérique collaboratif citoyen vient d'ouvrir

Qui sommes-nous ? Missions Accès rapide Se connecter

### Qui sommes nous ?



Ce site vous propose de participer à la création d'une base de données scientifique à partir des millions de photos des plantes de l'herbier de Paris. A vous d'explorer ces plantes et leurs étiquettes pour déterminer quand et où elles ont été récoltées. Les données que vous trouverez sont des éléments précieux pour mieux connaître la biodiversité et mesurer et prévoir son érosion. (en savoir plus)

### Choisissez votre première mission

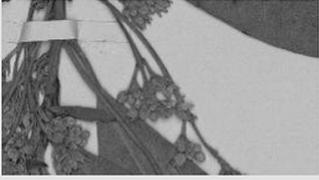


**Suivez Benjamin Balansa**  
Revivez les voyages de Benjamin Balansa

**76%**  
Objectif : 40000 contributions

Participer

Saurez-vous identifier le pays de ce spécimen ?



Participer

### Activité

 **brindesoleil** a identifié les botanistes pour le spécimen *Persea americana* (MNHN/P/P02004326) dans la mission **Fruits exotiques I**  
Il y a 33 minutes

 **fouf** a identifié le pays pour le spécimen *Justicia calycina* (MNHN/P/P04367062) dans la mission **Juste le pays (épisode 2)**  
Il y a 8 minutes

 **brindesoleil** a identifié la région pour le spécimen *Jacaratia corumbensis* (MNHN/P/P06586450) dans la mission **Fruits exotiques I**  
Il y a 44 minutes

Page d'accueil du site *Les herbonautes*.

ment les types nomenclaturaux des collections en est un exemple. Excellente sur le principe, plus facile à financer dans des petits projets, elle se révèle coûteuse à grande échelle et surtout lente à l'excès. Peu de données sont mises en ligne et bien des publics en demandent plus.

Il s'avère que la seconde option est moins coûteuse et plus enrichissante que la première. En effet, d'une part, en coût complet, la sélection peut être plus onéreuse que la numérisation d'un fonds entier. D'autre part, cette mise à disposition élargie se révèle plus riche et plus à même de satisfaire un public toujours plus varié.

Cependant, la numérisation industrielle est-elle applicable à toutes les collections ? La diversité des collections en trois dimensions impose de mettre au point d'autres méthodes. Des méthodes variées en raison des multiples contraintes. Les insectes – petits, fragiles, nombreux – ne se traiteront pas de la même façon que des fossiles de grands mammifères, ou bien des roches. Les équipes américaines qui travaillent sur ce sujet explorent en particulier l'optimisation des tâches et la mise en place de *workflows* efficaces.

Le renseignement des objets est une autre affaire ; leur virtualisation ouvre la porte à des outils innovants tels que les sciences participatives, ou plus classiquement, les procédures d'OCR des documents imprimés comme les étiquettes d'herbiers.

## Sciences participatives

Cette appellation regroupe un ensemble d'outils disparates qui font appel au volontariat et à la culture individuelle pour participer à la gestion de l'information scientifique. Le terme anglo-saxon de *Citizen Science*, science citoyenne, exprime assez bien le fait que chacun peut contribuer avec ses compétences. Cela passe par le renseignement d'objets, l'indexation, la vérification de documents numériques... Dans le domaine des sciences naturelles, la science participative intervient aussi lors de campagnes d'observation de plantes ou d'animaux vivants (oiseaux, papillons etc.).

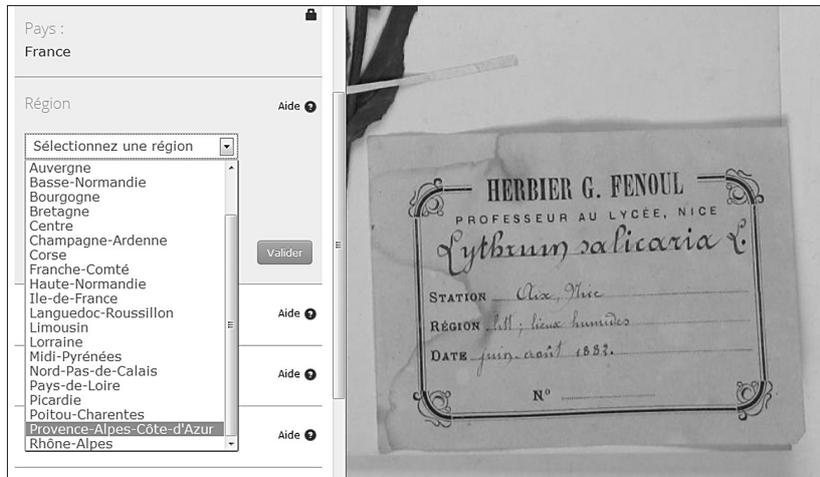
La participation d'un public sensibilisé offre bien des avantages. Outre la publicité qu'offre l'ouverture sur le web, elle augmente aussi les regards

croisés sur les objets, et par voie de conséquence l'expertise qu'on leur porte. Cette capacité qui prépare le travail de validation a souvent été négligée : dans le cas du site *Les herbonautes*<sup>1</sup>, qui permet la participation des internautes à l'informatisation des herbiers, la démarche de faire renseigner en aveugle plusieurs participants autorise une pré-validation efficace.

## Exemple des herbonautes

Ce néologisme désigne les internautes qui vont participer à la saisie des métadonnées liées aux spécimens d'herbiers. Ainsi, le site qui leur est consacré propose donc au grand public de contribuer à l'enrichissement d'une base de données scientifique à partir des millions de photos des plantes de l'herbier national situé au Muséum national d'histoire naturelle de Paris, fonds qui sera enrichi par les photos produites durant le montage de l'infrastructure bio-santé

1. Muséum national d'histoire naturelle, *Les herbonautes* : <http://lesherbonautes.mnhn.fr>



Interface permettant à l'herbonaute de renseigner les métadonnées liées à une image. Ici, renseignement de la région où a eu lieu la collecte.

e-ReColNat<sup>2</sup>, présentée ci-dessous. Avec l'aide des internautes, le site a l'ambition de faire d'une banque d'images une base de données accessible à tous.

Le site propose des «missions». Chacune d'entre elles comporte entre 250 et 2 500 images de spécimens qui répondent à une question scientifique. L'internaute choisit au hasard une image qu'il devra identifier et renseigner. Il explore alors l'image de la plante et son étiquette et détermine où et par quel(s) botaniste(s) elle a été récoltée.

Afin de valider les données, chacune de ces images sera proposée plusieurs fois sur le site. En cas de contradiction, un forum permet aux participants de discuter et de choisir la donnée la plus pertinente. À l'issue de la mission, un compte rendu est diffusé sur le site et les informations sont incorporées à la base de données Sonnerat (base du réseau des herbiers de France), et va alimenter celles de l'Inventaire national du patrimoine naturel<sup>3</sup> ainsi que les grandes bases internationales sur la biodiversité (GBIF).

2. Lire le descriptif du projet : [http://media.enseignementsup-recherche.gouv.fr/file/Infrastructures\\_2/98/7/E-RECOLNAT\\_206987.pdf](http://media.enseignementsup-recherche.gouv.fr/file/Infrastructures_2/98/7/E-RECOLNAT_206987.pdf)

3. Muséum national d'histoire naturelle [ed.], *Inventaire national du patrimoine naturel*, 2003-2013 : <http://inpn.mnhn.fr>

L'herbonaute peut être un passionné de sciences naturelles, un amateur de géographie, un féru de paléographie ou juste un accro des «enquêtes» qui va consacrer un peu de son temps pour aider à l'informatisation des herbiers. Il n'est pas nécessaire d'être compétent en botanique. Il suffit juste d'être un peu curieux...

Les données reportées par les amateurs sont des éléments précieux pour améliorer la connaissance de la biodiversité et de sa dynamique : cela peut notamment aider à mesurer l'érosion de la diversité végétale, ce qui est important dans le contexte de crise d'extinction actuelle.

### Gestion de la validation des données

Les métadonnées qui renseignent les objets ne font que préparer le travail d'indexation des objets. Une validation est indispensable, bien que celle-ci puisse être nuancée.

À ce propos, il faut différencier la mise en ligne des données et leur validation. Mais force est de constater qu'il vaut mieux mettre en ligne 200 000 enregistrements dont on sait comprendre les erreurs, plutôt que 2 000 complètement validés. L'essentiel est d'informer clairement les utilisateurs du degré de validation des éléments présentés.

Un exemple d'utilisation vient à l'appui de cette démonstration. Lors

de la recherche d'une localité inconnue à l'occasion d'une informatisation, il est courant d'utiliser un moteur de recherche et de découvrir des documents «OCRisés» remplis d'erreurs de reconnaissance, mais contenant les informations suffisantes pour valider l'existence d'un lieu ou d'une personne.

### Communiquer sur les données diffusées

La mise à disposition de documents qui ne constituent ni des synthèses, ni des ressources croisées, demande un exercice de communication inhabituel. Les herbiers diffusés peuvent receler des données fausses, erreurs de terrain ou de détermination, ou plus simplement de mélange d'étiquettes. Il faut avertir sans détour tout utilisateur du danger que constitue l'utilisation d'une donnée primaire.

Dans le même esprit, les données obtenues par la science participative sont aussi entachées de l'erreur potentielle que constitue la copie des données.

### Les inquiétudes des gestionnaires de collections

#### À nouvel outil, nouvelles contraintes

Les gestionnaires des collections peuvent s'inquiéter à juste titre de l'accès à un public plus nombreux et d'horizons plus variés. Quelles implications cet accès aura sur les collections physiques, à commencer par la révélation de l'existence d'objets précieux?

Cette inquiétude s'accroît lorsqu'on laisse la possibilité aux internautes de compléter les informations. Cependant, bien des pare-feux existent. L'art de la gestion des accès a encore de beaux jours devant lui.

#### Faut-il brûler les collections?

Les gestionnaires de collections peuvent se rassurer : le virtuel ne remplacera en aucun cas les collections

physiques. Leur accès sera toujours indispensable pour avancer dans le domaine de la diversité biologique. Les collections virtuelles viennent en complément et permettent une ouverture inédite. Elles offrent un avantage supplémentaire en ménageant les collections physiques en termes de conservation. Elles sont un nouveau moyen de disposer d'une sauvegarde d'objets toujours fragiles au même titre que les documents d'archives. Elles permettent une préparation plus fine et précise des demandes de consultation de la collection physique, voire une estimation de la pertinence de cette dernière.

## E-ReColNat

Le programme e-ReColNat<sup>4</sup> est l'un des onze lauréats de l'appel à projets « Infrastructures nationales

---

4. E-ReColNat a comme partenaires principaux le MNHN, l'université Montpellier 2, le PRES Clermont université, l'université de Dijon, l'IRD, le Cnam, l'Inra, l'association Tela Botanica et la société Agoralogie. Par ailleurs, plus de soixante institutions ont accepté de fournir leurs données et leurs images.

en biologie et santé». À l'image d'un portail comme Joconde (portail des collections des musées de France), il se propose de rassembler sur une même plateforme, l'ensemble des images et des données des collections naturalistes. Son ambition est aussi de mettre en place une interopérabilité entre les objets et leurs utilisateurs par le moyen de « bureaux virtuels » qui permettront non seulement de consulter les collections, mais aussi d'interagir et de compléter les informations.

### Un bureau virtuel

Le travail à partir des photos d'une collection est déjà pratiqué sur de nombreux sites liés à l'histoire naturelle. L'étape suivante réside dans la possibilité pour les utilisateurs de compléter les informations, que ce soit directement dans les bases de données, ou dans un espace d'étude partagé qui peut éventuellement être rendu public.

Il sera possible de créer ainsi un véritable espace interactif dans lequel les internautes ne sont plus seulement consultants des données, mais acteurs de leur enrichissement...

## Conclusion

La numérisation des collections naturalistes est un mouvement de fond qui va, qu'on le veuille ou non, modifier complètement le rapport aux objets et la façon dont on les consulte.

Toutes les grandes institutions au niveau international ont dans leur stratégie de développement des projets de « digitalization », et nombre d'entre eux sont déjà initiés. Il y a fort à parier que ce courant soit de plus en plus rapide. D'ici cinq ans, le contexte en sera complètement modifié.

Il est normal qu'il suscite l'inquiétude des professionnels, car il va à l'encontre de bien des pratiques classiques. Cependant, seuls les praticiens des collections ont un regard complet sur leur domaine. Le débat s'enrichira toujours des échanges d'idées entre tous les acteurs de l'histoire naturelle. ●

Septembre 2013