



Date : 08/08/2008

La bibliothèque du patrimoine de la biodiversité : partage de la littérature de la biodiversité avec le monde

Dr. Nancy E. Gwinn
Director, Smithsonian Institution Libraries

Constance Rinaldo
Librarian, Ernst Mayr Library and Archives
Museum of Comparative Zoology
Harvard University

Traduit en français par :
Dramane Sanon
(Comité national de
Politique Economique, Burkina Faso)

Meeting: 109. Science and Technology Libraries
Interprétation simultanée : Anglais, Arabe, Chinois, Français, Allemand, Russe et Espagnol

WORLD LIBRARY AND INFORMATION CONGRESS: 74TH IFLA GENERAL CONFERENCE AND COUNCIL

10-14 August 2008, Québec, Canada

<http://www.ifla.org/IV/ifla74/index.htm>

Résumé

Les dix grandes bibliothèques de musée de l'histoire naturelle, les bibliothèques de la botanique, et les institutions du Royaume-Uni et des États-Unis en 2005 se joignent pour développer une stratégie et un plan opérationnel afin de numériser la littérature publiée sur la biodiversité, conservée dans les collections respectives et de faire en sorte qu'elle soit disponible en libre accès et pour une utilisation responsable en tant que partie globale de la « biodiversité dite commune ». Le quartier général de la fondation des bibliothèques de Smithsonian Institution, le patrimoine de la biodiversité (BHL) a été reconnu comme une des quatre pierres angulaires de l'encyclopédie de la vie, un effort sans précédent pour tout le document de 1,8 millions de noms d'espèces d'animaux, de plantes, et d'autres formes de vie sur la terre. La majeure partie du financement du projet est venu de la fondation JOHN D. et CATHERINE T. MacArthur, de la fondation Alfred P. Sloan, de la fondation Gordon et Betty Moore et de la fondation Richard Lounsbury. Cet exposé fournira une vue d'ensemble de la BHL et de ses influences potentielles sur la recherche sur la biodiversité, décrira le portail de BHL et ses services de recherches innovantes, et fournira l'étude du développement d'un de ses membres : la Zoologie comparative du musée de l'Université Harvard.

Introduction

Selon un rapport du gouvernement des États-Unis, la biodiversité est un terme inventé par le scientifique internationalement reconnu E.O Wilson, comme une forme raccourcie de la « diversité biologique » en référence à la variété et au changement parmi les organismes et les complexes écologiques dans lesquels ils se produisent¹. Les sciences fondamentales pour identifier la diversité biologique sont connues comme la taxonomie, la théorie et la pratique de description, de nomination et de classification de toutes les plantes, animaux, et micro-organismes du monde, et aussi systématiques, en classant les êtres vivants au sein de groupes en se basant sur leurs évolutions d'origine. Des recherches dans ces disciplines sont le plus souvent menées dans les musées d'histoire naturelle et dans les jardins botaniques des pays développés, lesquels possèdent de riches collections exhaustives de la flore et de la faune, aussi bien que de grandes bibliothèques de la littérature qui les documentent. Les taxonomistes ont à peine effleuré les espèces connues (Godfrey 2007 ; Wheeler 2008). On estime que des 5 à 30 millions d'espèces sur la planète, seuls 1,8 million ont été identifiées et décrits.

En 1992, 150 leaders de gouvernement ont signé la Convention sur la Diversité Biologique, qui reconnaît le rôle crucial de la taxonomie dans la promotion d'un développement durable. Pour plusieurs années, les taxonomistes ont reconnu le besoin d'accélérer leur travail, avant l'éclatement des populations, les catastrophes environnementales et l'économie de développement qui réduisent la richesse des espèces existantes. Leurs travaux ont été bien mis en valeur au delà de l'acte d'identification ; cela a étendu l'utilisation et l'incidence économique pour une large portée des applications dans l'agriculture, la conservation de la biodiversité, la gestion des zones protégées, le contrôle des espèces envahissantes, la sylviculture, la reproduction des plantes, le contrôle des maladies et le commerce des produits naturels, incluant des produits pharmaceutiques (Wheeler 2008). Cependant il y a de sévères obstacles aux progrès, que l'on appelle sur le terrain « entraves taxonomiques ». Les taxonomistes effectuent en grande partie leur travail parmi les collections de spécimen et de littérature dans les nations industrialisées. Il y a peu de taxonomistes pour collectionner et travailler sur les millions d'espèces non décrites mais riches en biodiversité dans les pays pauvres économiquement². Et pour tous les taxonomistes, l'accès à la littérature appropriée peut être un processus long et coûteux (Godfrey 2007 ; Minelli 2003).

Plus que n'importe qu'elle autre science, le domaine de la biologie systématique est complètement dépendant de la littérature historique des descriptions publiées des espèces ; une publication imprimée détermine la légalité de la nomination et du crédit pour les nouvelles découvertes. Alessandro Minelli écrit : « Selon la pratique en vigueur, n'importe quel travail monographique sérieux au sujet d'un taxon donné devrait inclure un examen consciencieux de toute la littérature précédente traitant au moins une espèce appartenant à ce taxon, indépendamment qu'il ait été édité l'année dernière ou vers la fin du XVIIIe siècle en anglais, en espagnol, en russe, ou en latin. Si ces travaux incluent la description (puis) de nouveaux taxa, ou ont un impact sur le son nomenclature, ils ne peuvent pas être ignore Si ces travaux inclus la description de nouveaux taxons, ou ont une influence sur leur nomenclature, ils ne peuvent pas être ignorés³. »

Minelli décrit les écrits taxonomiques comme des documents « légaux » aussi bien que les documents scientifiques, parce qu'ils décrivent les nouvelles espèces selon les règles de codes internationaux distincts⁴. Les codes existent pour faire en sorte que tous les taxonomistes adhèrent aux principes de

¹ Le congrès des États-Unis. Service de technologie d'évaluation.1987. Des technologies de la diversité biologique, ATA-F-330 (Washington, DC : service d'impression Gouvernement des Etats Unis) : 313

². Initiative de la taxonomie globale, Convention sur la diversité biologique, [http // www.cbd.int/gti/ problem.shtml](http://www.cbd.int/gti/problem.shtml)

³ Minelli, A. 2003 « le statu de la littérature taxonomique » *Trend in Ecology and Evolution* 18(2) :75.

⁴ Ce sont des codes internationaux de la nomenclature zoologique, des codes internationaux de la nomenclature botanique et des codes internationaux de la nomenclature des bactéries.

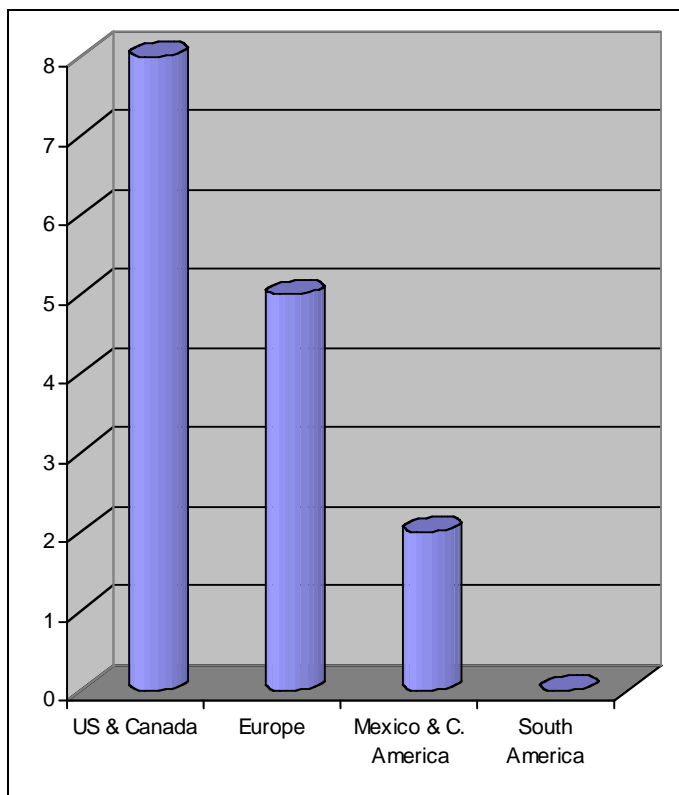
⁵ The *biologia Centrali Americana* (BCA) est un travail fondamental pour l'étude e néotropical de la flore et la faune. Il inclut presque toute chose connu concernant la diversité biologique de Mexico et le Centre de l'Amérique au moment de la publication (1879- 1915). Le travail original consite à 58

priorité, ce qui résout les problèmes provoqués par l'utilisation des homonymes ou des synonymes en les nommant. En conséquence, les taxonomistes doivent consulter toute la littérature pertinente depuis Linnaeus de manière à assurer une base solide pour leur travail.

La technologie et Internet ont finalement fourni un moyen de dissoudre l'obstacle taxonomique, au moins en partie, à travers le balayage électronique autant de la littérature et des spécimens de collections pour être partagés par la suite avec toute la communauté scientifique (Godfray 2007). RENDU ICI Il y a plus de 5,4 millions de volumes sur la biodiversité datant de 1469, 800,000 monographies et 40,000 titres de périodiques. Cinquante pour cent furent publiés avant 1923 et sont du domaine public au États-Unis. Voilà un travail important.

La planification commence

En 2003, la fondation de *la Smithsonian Institution* a accueilli une conférence des biologistes afin de discuter de ce qui était nécessaire pour améliorer efficacement la recherche biologique. Le plus grand obstacle identifié par les experts, fut l'accès à la littérature historique. Nous, dans les musées d'histoire naturelle et les bibliothèques des jardins botaniques, savons depuis des années, que les chercheurs qui ont voyagé pour consulter nos collections, passent souvent plus de leur temps devant des machines à photocopier pour rassembler le plus possible de documents avant de retourner chez eux.



La technologie numérique et Internet offre une solution ; si la littérature était numérisée pour effectuer des recherches sur la toile, les chercheurs pourraient y avoir accès de partout dans le monde. Suite à la conférence, la Smithsonian Institution a fourni le financement à ses propres bibliothèques afin de numériser la *Biologia Centrali-Américana*, 63 volumes publiés de 1879 à 1915, ce qui demeure le

volumes biologiques contenant 1284 plates illustrant 18,587 sujets. Un total de 49,392 espèces décrites, 19,263 pour la première fois. Beaucoup d'illustrations et de descriptions sont les seules qui existent de la région de biota.

travail original sur la flore et la faune du Centre-Amérique⁵. Le financement a aussi appuyé le travail initial pour le développement d'un système de codage automatique des noms scientifiques des espèces trouvées afin d'en améliorer l'accessibilité aux taxonomistes.

En 2005, au Musée d'Histoire naturelle de Londres, la fondation d'Alfred P. Sloan a financé un symposium international sous le titre : *Bibliothèque et laboratoire : le mariage de la recherche, des données et la littérature taxonomique*. Les quatre-vingt biologistes, bibliothécaires et spécialistes en informatique qui y ont assisté, ont identifié de nouveau le manque d'accès à la littérature publiée de la biodiversité comme un des principaux obstacles pour des recherches efficaces et productives (Moritz 2005).

En mai 2005, les représentants de plusieurs grandes bibliothèques de la botanique et de l'histoire naturelle se sont rencontrés au Musée National Smithsonian de l'Histoire Naturelle à Washington D.C. grâce au financement fourni par la fondation Smithsonian, l'objectif de la rencontre était de développer une stratégie et un plan efficace pour la numérisation de la littérature publiée en biodiversité et conservée dans leurs collections respectives de sorte qu'elle soit disponible en accès libre et pour une utilisation responsable comme une partie de la globalité de la « biodiversité commune » (Moritz 2002). Deux ans plus tard, les directeurs des bibliothèques du Musée de l'Histoire Naturelle de l'Amérique, des bibliothèques botaniques de l'Université de Harvard, de la bibliothèque du Musée de la Zoologie comparée de l'Université Ernst Mayr de Harvard, du Jardin botanique de Missouri, du Musée National de Londres, du Jardin botanique de New York, du Jardin Royal botanique, Kew, du Musée Field de l'Histoire Naturelle de Chicago, du laboratoire biologique de Marine, de la bibliothèque océanique de la fondation de Woods Hole et des bibliothèques de la fondation de la Smithsonian Institution se sont mis d'accord sur un protocole relatif à la création d'une Bibliothèque sur l'héritage de la biodiversité⁶. Les bibliothèques de la Smithsonian Institution ont accepté de recevoir le secrétariat de la BHL et ont fourni le premier directeur du programme, Thomas Garnett.

Un autre développement se préparait en même temps, l'Encyclopédie de la vie⁷. C'est un ambitieux et même un audacieux projet global de collaboration pour documenter avec autorité les 1,8 million d'espèces d'animaux connus, de plantes et autres formes de vie et de créer des pages web sur Internet pour chacun. Chaque page est seulement un point d'entrée, approprié à un public général mais avec des pages dirigées vers des données spécialisées pour les chercheurs. Cette encyclopédie comprend la Bibliothèque héritage de la biodiversité comme une de ses quatre pierres angulaires, afin d'apporter beaucoup de littérature historique relativement aux espèces pertinentes sur les pages web. C'était critique pour le projet, parce à travers l'Encyclopédie de la vie, la BHL recevait une première augmentation de 3 millions de dollars de la part de la fondation de John D. et Catherine T. Mac Arthur. Des financements additionnels sont venus des fondations Alfred P. Sloan, Gordon et Betty Moore, Richard Lounsbury et d'institutions de la BHL en tant que membre individuel.

POURQUOI FAIRE CELA MAINTENANT ?

Les dix membres des bibliothèques de la BHL ont au delà deux millions de volumes sur la littérature de la biodiversité réunis pendant 200 ans pour appuyer les scientifiques et les étudiants à travers le monde. Ces bibliothèques conservent définitivement ensemble une substantielle partie des connaissances de la diversité biologique publiées dans le monde. Pendant qu'il y a une multitude de projets de numérisation dans la majorité des bibliothèques de recherche ici et à l'étranger, aucun n'est

⁶ À propos de la Bibliothèque de l'héritage de la biodiversité (Biodiversity Heritage Library) <http://www.biodiversitylibrary.org/About.aspx>

⁷ (Encyclopédie de la Vie) Encyclopedia of Life <http://www.eol.org>

centré sur la discipline spécifique des institutions partenaires de la BHL et il y a ainsi un risque d'échapper les éléments significatifs de la légalité de la biodiversité. Une grande partie cette littérature est hautement spécialisée et pas souvent reproduite même dans les collections de recherche d'une université. Peu d'entre elles autorisent le libre accès aux publications numérisées. Avec ses stratégies de recherches innovatrices, la BHL a l'intention d'être « un guichet unique » pour ceux qui ont besoin de consulter la littérature sur la biodiversité.

Les coûts de la numérisation ont baissé considérablement en mode de haute production, le projet des Archives d'Internet a un coût de base de .10 cents la page. La littérature de la biodiversité est un domaine scientifique souple, bien défini et bénéficie d'une grande longévité- la littérature taxonomique courante est souvent basée sur des textes et des exemplaires qui remontent à plus de 100 ans. (Godfray 2007 ; Minelli 2003). De plus, la BHL soutient la *Global Biodiversity Information Facility* (GBIF) et d'autres initiatives internationales sur la biodiversité (Speers et Edwards 2008)⁸. Les bénéfices sont clairs : des taxonomistes et d'autres scientifiques auront accès à la littérature de la biodiversité- globalement, les scientifiques et les citoyens, dans ce monde en développement, auront finalement et facilement accès à la littérature historique. Ce qui permet de faire avancer nettement les objectifs de la Convention internationale sur la diversité biologique⁹.

DÉBUTER

Les membres de la BHL ont choisi les Archives Internet (Internet Archive) afin de fournir les services de numérisation et d'archiver les résultats des documents numérisés. Les Archives Internet encouragent d'une part la mission de la BHL de fournir la littérature en accès libre et d'autre part d'avoir la capacité de faire une grande quantité de numérisation à un coût raisonnable. Les Archives Internet ont commencé la numérisation pour les partenaires de la BHL au début de 2007. Les membres de la BHL travaillent avec trois des seize stations des centres de numérisation des Archives Internet (Boston, Washington DC, et New York), aussi bien qu'avec les plus petits centres de la fondation de la Smithsonian Institution, l'Université de l'Illinois, d'Urbana-Champaign et le Musée d'Histoire Naturelle de Londres. Cependant, les partenaires font beaucoup plus que simplement numériser les volumes.

Le personnel du jardin Botanique de Missouri a établi le portail la bibliothèque de l'Héritage de la biodiversité comme une recherche innovatrice de l'environnement avec l'objectif ambitieux de transformer la nature de la recherche scientifique, aussi bien qu'accélérer énormément la recherche dans les sciences de la vie et la conservation¹⁰. Le portail donne accès aux images numérisées (JPEG2000, PDF, et JPEG) et au texte de la littérature en Optical Character Recognition (OCR) mais cela n'est que le début. Il utilise aussi une manière intelligente de vérifier les termes reconnus par la taxonomie afin de surmonter les problèmes de noms communs versus les noms scientifiques et le changement de noms à travers le temps. Ce système de référence scientifique pour les études de la littérature scientifique offre un modèle qui diffuse, accroît les approches scientifiques et utilise l'objet de la littérature de l'histoire naturelle.

QU'EST-CE L'INTELLIGENCE TAXONOMIQUE ?

Les scientifiques utilisent les noms scientifiques pour trouver des informations sur les organismes. Un organisme peut avoir eu beaucoup de noms scientifiques à travers le temps ou de multiples noms communs qui dépendent du langage ou de la région. De plus, un nom peut référer à de multiple organisme. Ainsi il peut être difficile de récupérer l'information à propos d'un organisme même si le

⁸ Global Biodiversity Information Facility, <http://www.gbif.org>.

⁹ Convention on Biological Diversity, <http://www.cbd.int/>

¹⁰ Biodiversity Heritage Library, <http://www.biodiversitylibrary.org>

nom du courant scientifique est connu. Ce problème fut abordé par un projet international appelé uBio pour l'indexeur Universal Biological Indexer and Organizer. uBio est composé du serveur Taxonomic Name Server (TNS qui agit comme thésaurus ; NameBank, un répertoire de plus de 10,7 millions de noms biologiques enregistrés et d'identificateurs qui relient ces noms ensemble ; et la banque de classification, qui emmagasine de multiple classifications et de concepts taxonomiques¹¹. La BHL utilise « TaxonFinder », un algorithme d'intelligence taxonomique développé par les collaborateurs de uBio, pour comparer les textes de OCR avec ceux de NameBank et identifier les noms scientifiques semblables. Une fois pleinement intégré, un chercheur pourra ensuite explorer la collection de la BHL en utilisant n'importe quelle forme de nom d'organisme (Leary et al., 2008). Cela permettra aux utilisateurs des pages web de l'Encyclopedia of Life d'en retirer la littérature reliée aux espèces.



Fig.2 Une page de la bibliothèque de la biodiversité du patrimoine montrant la liste des noms scientifiques (en bas et à gauche, au coin le plus bas, extrait du texte OCR)

Après que les Archives Internet aient numérisé les volumes, le portail de la BHL intègre les metadonnées MARXML et une basse résolution des dossiers en JPEG. Une haute résolution des dossiers en JPEG2000 est récupérée à la volée des Archives Internet lorsque demandé par les utilisateurs et décodée au portail et être visionné grâce un navigateur web. Le texte OCR est aussi envoyé à la volée à uBio pour l'extraction du nom grâce à l'intelligence taxonomique et affiché en temps réel avec les pages images. Le portail de la BHL livre l'image et les textes des dossiers, affiche les metadonnées au complet ou en partie, applique globalement des Globally Unique Identifiers (GUIDs) pour lier aux autres services taxonomiques. Les URL persistants permettent de se relier aux enregistrements bibliographiques, volume et aux niveaux des pages dans la BHL. La direction technique de la BHL et celle du Jardin Botanique de Missouri sont constamment en train d'améliorer le portail de présentation des résultats et d'ajouter des caractères tels que les géocodages des sujets de la Bibliothèque du Congrès qui utilise le programme de Google Maps Application Interface révélant les références des bibliographies précédentes des espèces non indexées (Freeland, et al. 2008) .

¹¹ Sur le projet uBio : <http://www.ubio.org>

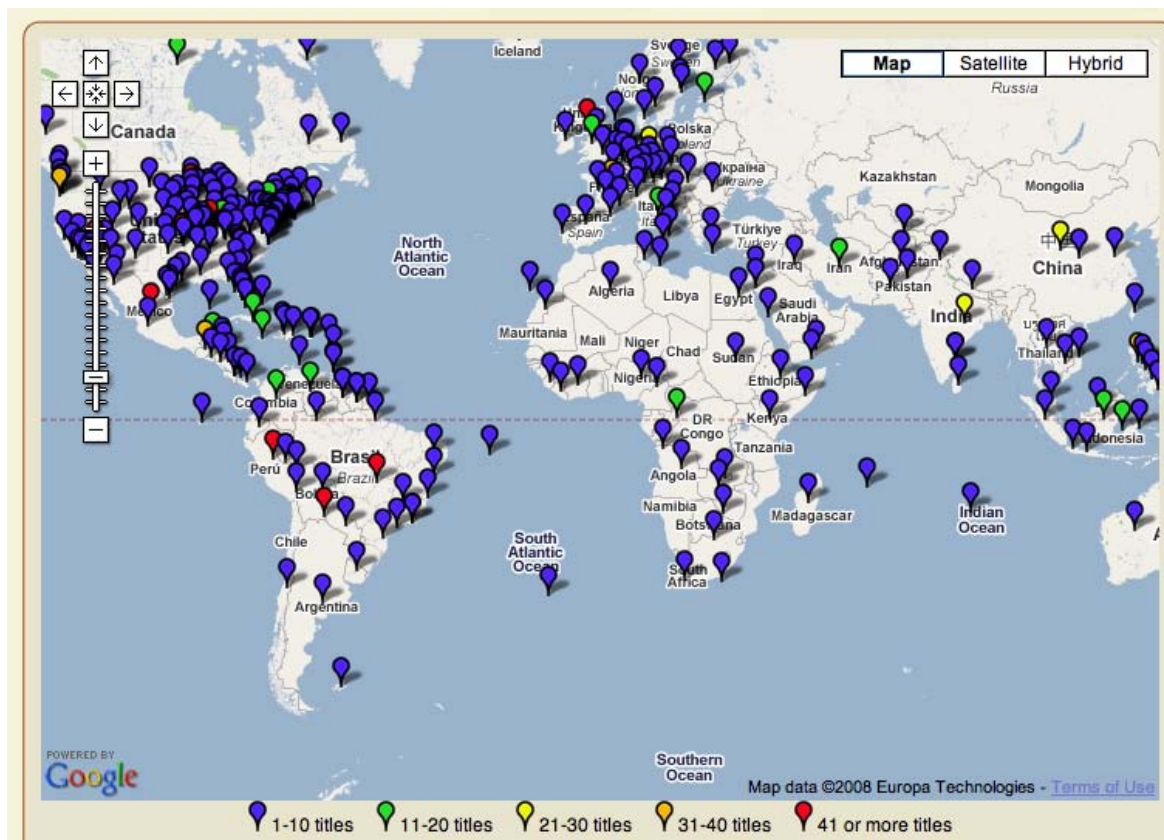


Figure 3: Les vedettes-matières de la Bibliothèque du Congrès des volumes dans la BHL sont géocodées placées sur les cartes e utilisant Google Maps API

MAIS QU'EN EST-IL DU DROIT D'AUTEUR ?

Le projet de la BHL essaye de maintenir généralement bas le risque de violation des droits d'auteur de la littérature du domaine public d'abord, recherchant des autorisations pour la numérisation, des accords alternatifs de négociation et aller de l'avant quand aucune de ces alternatives ne s'applique. La BHL a un droit de regard sur le modèle de droit d'auteur. La direction du programme de la BHL a ouvert des négociations avec une variété d'éditeurs, allant des petits, aux associations savantes et aux grandes organisations commerciales. En avril 2008, la BHL a obtenu l'autorisation de numériser les publications entières de quarante-neuf titres de musées et de petites associations d'éditeurs. La BHL numérise les demandes de publications entières jusqu'aux plus récentes publications, par des autorisations négociées. Plusieurs d'entre elles sont sur le portail de la BHL et ne coûtent rien aux associations. Les dossiers peuvent être réutilisés par les associations pour leurs propres objectifs¹². La BHL sera responsable à long terme du matériel numérisé. Certains agrégateurs de contenu et des éditeurs commerciaux ont exprimé de l'intérêt pour des accords alternatifs, comme fournir des métadonnées et des dossiers OCR pour les indexer en utilisant les outils d'intelligence taxonomiques¹³. Des discussions sont en cours avec les Archives Zoologiques (*Zoological Record*) et une collaboration avec la BHL est possible.

LA GOUVERNANCE DE LA BHL

Le patrimoine de la bibliothèque de la biodiversité n'est pas incorporé et ainsi n'est pas une entité légale. À travers la signature des accords des Memoranda, chaque Institution membre s'est engagée à

¹² Can I Use Your Images, <http://www.biodiversitylibrary.org/Copyright.aspx>.

¹³ Tools, <http://biodiversitylibrary.org/Tools.aspx>

collaborer afin de construire la bibliothèque du patrimoine de la biodiversité et de rechercher du financement additionnel pour l'appuyer. Les directeurs des bibliothèques membres formant le conseil de direction se rencontrent au moins une fois par année pour constater les progrès, discuter des récentes publications et élire les responsables. Un Comité exécutif se rencontre de façon hebdomadaire en téléconférence et est composé du Directeur du programme de la BHL, du Président du comité institutionnel, du vice-président, du secrétaire et du directeur technique. Comme les membres ont mis en application le processus de numérisation, des problèmes pratiques et des obstacles sont apparus. L'étude du cas de la bibliothèque de Ernst Mayr au Musée de la Zoologie Comparative illustre certains de ceux-ci.

L'ETUDE DE LA BIBLIOTHEQUE D'ERNST MAYR

Pendant que les coûts de la numérisation de la page étaient bas au départ, ils se sont élevés au dessus des objectifs prévus. La première étape était d'acheter les outils d'analyse de la collection d'OCLC et d'ajouter les archives de toutes les bibliothèques membres¹⁴. La fondation Smithsonian a payé la subvention de la première année. Cela a pris plus d'une année pour avoir cet outil et de le mettre à jour, principalement parce que toutes les bibliothèques n'étaient pas membres initialement de OCLC, et, dans le cas de l'Université de Harvard, les archives des bibliothèques de la botanique et de la zoologie ne pouvaient pas être séparées des autres bibliothèques de Harvard. L'outil d'analyse des collections d'OCLC a permis une bonne vision des forces de la collection de l'institution et on a autorisé le groupe à évaluer le nombre des pages du domaine public qui pouvait être numérisée plus ou immédiatement et d'identifier des sujets avec des titres uniques dans les bibliothèques membres. Un nombre de titres hors du projet de numérisation de la BHL ont doublé ceux de la BHL. Afin de minimiser la duplication, le personnel de la bibliothèque révise d'autres projets ; si le matériel est en libre accès et peut facilement être intégré, dans certains cas, les membres de la BHL peuvent modifier les plans de numérisation.

Avec la contribution du comité de direction d'EOL, les membres de la BHL ont choisi les périodiques comme la première priorité, parce que les périodiques sont essentiels pour les scientifiques qui constituent le premier public pour la parution initiale de la BHL. Le Musée de l'Histoire Naturelle de Londres a développé un outil afin d'éviter de numériser les doubles parmi les institutions membres. Une base de données fut produite incluant les catalogues d'archives en ligne de tous les périodiques des institutions membres avec une boîte de contrôle pour le tri, c'est ainsi qu'est né la liste des périodiques. Comme les titres sont identifiés pour la numérisation, une institution pourra accéder à cette base de données afin d'être sûre qu'aucune autre personne n'a soumissionné sur eux. S'il n'y a pas d'offre, ou si les offres couvrent seulement une partie de périodiques, l'institution est libre de réclamer le titre, ou au moins les parties de titres qui n'ont pas été souscrits, pour la numérisation. La bibliothèque de MBL/WHOI a construit un outil de doubles de monographie qui montre que des monographies ont été numérisées; il n'y a pas encore de processus fixé pour les monographies.

Les membres de la BHL ont mis rapidement un « wiki » en projet afin de faciliter les communications parmi les bibliothèques très éloignées. Le « wiki » a été utilisé pour maintenir les minutes des téléconférences et des rencontres, des présentations et des procédures et pour coordonner les questions techniques, la qualité des discussions et les travaux de collection. Cela est devenu évident en mai 2008 pour les membres directeurs de gérer au jour le jour la planification et le processus, le rythme de travail nécessitant de meilleures communications. La direction avait besoin de comparer ses expériences plus directement afin d'assurer un rythme du travail efficace et éviter « de réinventer la roue ». Sous la conduite de Suzanne Pilsk des bibliothèques de la Smithsonian Institution, la direction participe maintenant à deux ou plus des téléconférences mensuelles. Ces conférences sont devenues essentielles pour développer, gérer efficacement et coordonner le rythme du travail pour le projet de la BHL. La direction des bibliothèques membres a découvert que chacun

¹⁴ WorldCat Collection Analysis, www.oclc.org/collectionanalysis/default.htm

était en train d'expérimenter de la frustration avec une variété de problèmes de rythmes de travail et était surpris par les changements dans la quantité de travail à réaliser.

James Hanken, le directeur du Musée de la Zoologie comparée a dégagé du financement pour les cinq prochaines années afin de numériser les collections de la bibliothèque Ernst Mayr pour la bibliothèque du patrimoine de la biodiversité. Les fonds touchent la numérisation et les coûts de démarrage du processus. Avec ce financement, la bibliothèque a engagé deux étudiants du Simmons College Library afin de préparer les collections pour la numérisation. Le premier rôle des étudiants est de placer un code à barres, extraire, contrôler le matériel de bibliothèque pour une numérisation adéquate et placer un identificateur. Dans la bibliothèque d'Ernst Mayr, les services de la direction technique utilisent les rapports générés par le système intégré de bibliothèque (ILS) pour créer « les picklists » d'articles identifiés pour la numérisation. Les listes sélectionnés sont des listes d'articles disponibles à numériser, basées sur les critères de la date et selon l'ordre des rayons. Les « les picklists » rendent le travail plus facile étudiants et aux autres directions pour aller sur les rayons et sortir exactement les bonnes pièces. « Les picklists » sont éditées selon le contrôle des partenaires de la BHL, utilisant les outils décrits plus haut pour minimiser la numérisation des doubles. Puis la direction des services techniques vérifie les pièces extraites en se basant sur les listes, produit la liste finale pour la livraison au centre de numérisation, vérifie les livres au départ et de nouveau au retour en regard de la première liste d'emballage et surveille le travail des étudiants. Une base de donnée fourni un inventaire de toutes les pièces numérisées pour la bibliothèque de Ernst Mayr. Les articles rejetés soit par la bibliothèque soit par le centre de numérisation, sont renvoyés via le système des bibliothèques intégrées. Le bibliothécaire des collections spéciales fournit les listes de sujets spécifiques (par exemple, herpétologie, ichtyologie) dans les collections spéciales de monographies et extrait les articles appropriés pour le code à barres et le départ pour le centre denumérisation. Le superviseur du tirage extrait les pièces et les remet ensuite dans les rayons au retour.



Fig4 Une des stations de numérisation des Archives Internet à la Bibliothèque publique de Boston.

Au moins 2, 5 FTE (incluant le temps de Constance Rinaldo comme secrétaire de la BHL) sont consacrés au projet. De plus, les systèmes du bureau de l'Université de Harvard ont travaillé avec les Archives Internet afin d'assurer des liaisons faciles avec la norme Z 39.50, d'aider à la planification et d'ajouter les URLs au HOLLIS et avec le catalogue de l'Université de Harvard en ligne en retour du matériel numérisé. Le lien avec la Z39.50 permet à IA d'extraire des métadonnées du catalogue de l'Université de Harvard afin de rendre populaire les archives de IA. L'autorisation d'utiliser la Z39.50 gratuitement aux entités qui ne sont pas de Harvard se fait uniquement sur la base du cas par cas ; obtention cette permission a pris plus de six mois. En fin de compte, c'est important pour les utilisateurs de Harvard qu'ils aient un accès immédiat aux collections de Harvard disponibles électroniquement. Ainsi l'étape finale est de fournir l'accès aux matériels numérisés à travers le portail de Harvard. Cela signifie que les URL accentués des objets numérisés et générés de Harvard doivent être disponible dans HOLLIS. À ce moment-là, les URL produits par le portail de la BHL doivent être entrés à la main, mais le bureau d'information des systèmes l'Université de Harvard est en train de travailler avec la bibliothèque d'Ernst Mayr afin de développer une solution automatisée.

Pour la bibliothèque d'Ernst Mayr, le bureau général du comité de l'Université de Harvard a déterminé que des classes de textes qui peuvent être numérisées avec un minimum de risques inclus l'Université de Harvard, les publications du Musée de la Zoologie comparative (MCZ), le matériel du droit d'auteur américain d'avant 1923 et celui non américain d'avant 1908. De plus, là où les permissions ont été accordées gratuitement, des périodiques d'autres sociétés et de musées sociétés peuvent être numérisés.

Développer un contact a été un long processus (deux ans) dans les deux sens entre Harvard et les avocats des Archives Internet. Bien que la numérisation opère depuis décembre 2007, le contrat vient juste d'être accordé. La bibliothèque d'Ernst Mayr a loué une compagnie de bibliobus pour faire la navette des livres entre la bibliothèque d'Ernst Mayr à Cambridge, MA et le centre de numérisation de la bibliothèque du Consortium de Boston dans la bibliothèque publique de Boston. L'accès à la bibliothèque d'Ernst Mayr est en train d'être remis en question par les transporteurs à cause du circuit de circulation modifié par la nouvelle construction et des difficultés de circuler autour des vieux bâtiments.

Plusieurs des publications anciennes, particulièrement celles de MCZ ont des encarts ou sont plus grandes que la taille moyenne. Initialement, les Archives Internet n'avaient pas de solution pour les encarts et le personnel avait retiré et identifié les volumes avec des encarts, et ceux-ci étaient pas numérisés. D'autres publications qui causent un rejet ou des questionnements pour la numérisation comprennent des marges étroites, la taille et des fragiles ou de pauvres reliures. Le personnel de l'IA au centre de numérisation de Boston ont été très flexible et est volontaire pour numériser des pièces inhabituelles et de fournir un test de simple numérisation pour que la meilleure décision puisse être prise pour le matériel de la bibliothèque Ernst Mayr. Une solution d'encart est en train d'être testée à Boston et bien que ce soit encore pour une dimension limitée (18'' par 24''), c'est déjà une amélioration. Les encarts numérisés par IA varient du supérieur à l'original à aussi bon que l'original mais la visualisation est adéquate. Pour mieux voir les encarts, l'image JPEG2000 doit être affichée. Le PDF et d'autres vues d'encarts ne sont pas adéquates dans plusieurs cas. Peu de livres sont rejetés maintenant qu'il y a une solution pour les encarts mais la numérisation des encarts augmente le temps nécessaire à numériser une pièce ainsi que le coût. La bibliothèque Ernst Mayr envoie actuellement approximativement 200 volumes chaque semaine à l'établissement de numérisation de l'IA à la bibliothèque publique de Boston.

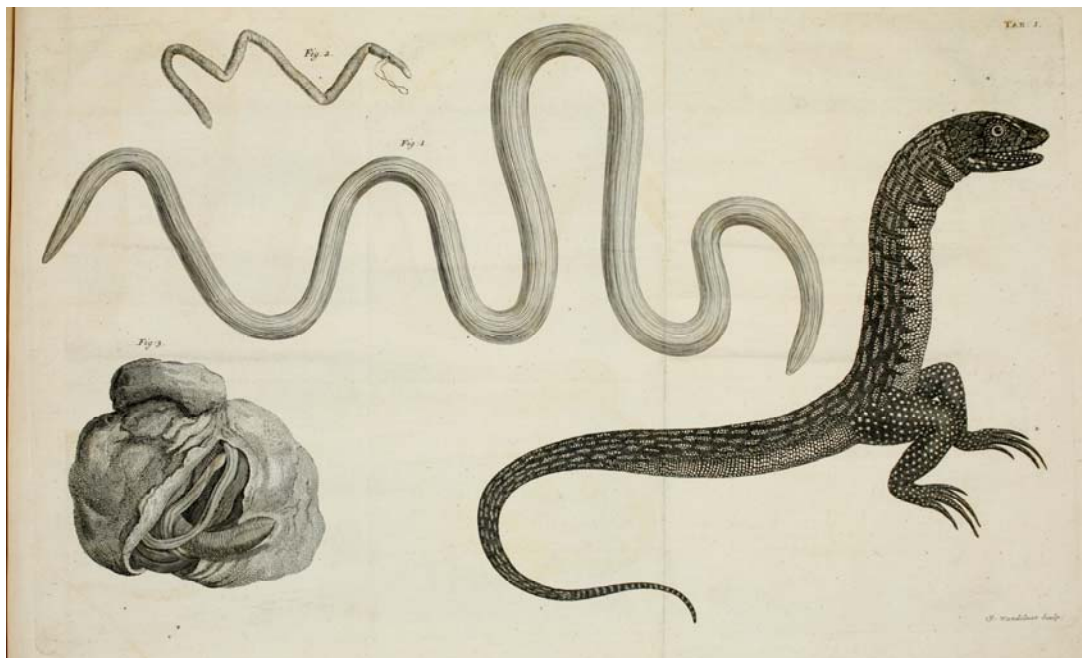


Fig5. Un simple encart de Klein, Jacob Theodor. 1755 Tentamen herpetologiae. Leidae & Gottingae : Apud Eliam Luzac, Jun. L'original de l'encart était 10'' par 15''.

LES RÉSULTATS EN COURS

Au 20 avril 2008, il y avait plus de 3,5 million de pages, comprenant 4,000 titres (plus de 9,000 volumes) accessibles à travers le portail de la BHL. Le projet a démontré que :

- Le concept de production de numérisation de collections générales est possible ;
- Ils y a de hauts niveaux de précision de l'OCR des imprimés des 19^e et 20^e siècles ;
- L'intelligence taxonomique (des noms d'espèces trouvés) à travers des millions de pages contre 11 millions de noms dans la banque de noms est hautement effectif ; et
- Séparées administrativement et géographiquement des institutions peuvent collaborer sur un projet complexe à niveau multiple et parvenir à des résultats concrets dans un domaine connaissance spécifique.

Beaucoup de littérature se trouve dans le monde civilisé, dans l'hémisphère nord, mais tel que mentionné plutôt, la plupart de la biodiversité est dans le monde en développement et dans l'hémisphère sud. Un des désirables dénouements du projet est de rapatrier la littérature de ces pays d'origine en la rendant accessible à tout individu qui se connecte à Internet. Il est évident que ça fonctionne suite au message suivant que la bibliothèque Smithsonian recevait d'un chercheur du Musée de l'histoire naturelle de Pérou :

Ma profonde gratitude pour me permettre l'accès à la version numérisée de ce très rare « Bulletin des Séances de la Société Entomologique de France ». Cela a été très important pour mes travaux sur la base de données des noms des papillons du monde de pouvoir consulter ces périodiques, conservés par très peu de bibliothèques dans le monde. Je ne peux pas insister assez sur l'importance d'avoir accès aux versions électroniques de la littérature, spécialement pour nous les chercheurs qui ne pouvons pas bénéficier de bibliothèques bien fournies....Je souhaite seulement qu'il ait beaucoup plus ressources électroniques comme celles-ci. S'il vous plaît, continuez cet excellent travail !

Une autre d'Hawaï :

Aloha. J'habite une grande île d'Hawaï à 300\$ dollars par avion de Honolulu au Musée Bishop. Même quand je peux aller au Musée (où j'étudie la peau des oiseaux Hawaïens), ils n'ont pas tous les oiseaux (moho apicalis, l'Oahu moho a disparu)...J'ai cherché ce texte durant vingt ans. Mahalo nui loa pour tous vos durs travaux. La lecture de ces pages signifient beaucoup pour moi et pour d'autres... Je ne vous remercierai jamais assez, ni insisterai sur l'importance de votre site web. Merci de mettre les articles sur le web d'une manière facile pour le repérage.

L'Encyclopédie de la Vie et d'autres projets informatiques sont des exploitations de données de la BHL pour des informations localisées dans la littérature taxonomique léguée. Un nombre de sociétés savantes ont été attirées par les modèles de droits d'auteurs et des permissions attribuées pour la numérisation du matériel courant. De plus, le projet a généré une stimulation au sein de la communauté internationale et beaucoup d'opportunités pour développer des partenariats. Il a été appuyé aussi par la traditionnelle écoute de bibliophiles et de scientifiques.

OÙ ALLONS-NOUS À PARTIR DE MAINTENANT?

Comme une étude l'a démontré, il y a des problèmes à résoudre liés à la numérisation et les types de documents que constitue la bibliothèque du patrimoine de la biodiversité. Finalement, le projet doit fournir une analyse du niveau d'article des périodiques, ce qui nécessite une certaine adaptation des indices bibliographiques actuels de la littérature de la biodiversité, autant que le développement des outils automatiques pour une analyse structurée. Ce développement peut être combiné avec la création du « catalogue unifié » pour fournir « un guichet unique » de l'accès à la littérature.

Les partenaires de la bibliothèque de la biodiversité sont actuellement centrés au niveau anglo-américain, bien que beaucoup de collections contiennent un accroissement de matériel non anglophone. L'objectif clé est de développer des partenariats globaux incluant la communauté entière. Le consortium projette éventuellement de s'élargir et d'attirer des partenaires spécialement en Europe, en Asie et dans les pays en développement. Le meilleur moyen est d'inclure plus de langues est le partenariat avec des autres pays et faire en sorte que la BHL soit utilisée en plusieurs langues. La BHL est en train travailler pour engager des partenaires européens dans les projets, tels que l'Institut Européen de distribution de technologie (EDIT)¹⁵ et les Ressources de Synthèses et Sémantiques (SYNTHESESYS)¹⁶. EDIT est composé de 27 pays européens, d'Amérique du nord et de Russie dont les objectifs sont de surmonter les obstacles taxonomiques par battissant une compétence taxonomique par l'organisation de formation et d'information. L'objectif du projet de la SYNTHESESYS est de développer une infrastructure européenne coordonnée pour les chercheurs dans les sciences naturelles. La BHL espère accroître ces projets dans un partenariat avec d'autres pays et ainsi développer une infrastructure de milieux non anglophones. Graham Higley, le président du Comité d'Institution de la BHL et leader du projet de la SYNTHESESYS et le directeur du Programme de la BHL, Tom Garnett, ont travaillé inlassablement afin de garantir que les Européens, les Africains, les Australiens, les Américains du sud et les pays d'Asie soient inclus dans les discussions concernant la littérature. Des discussions informelles ont eu lieu déjà avec l'Académie de science Chinoise, du Brésil, la bibliothèque de numérisation Européenne et du Musée fur Naturkunde der Humboldt- Universität zu Berlin. Le gouvernement Australien semble vouloir financer la numérisation comme un élément de l'atlas de la vie australienne. La BHL continuera aussi de travailler avec les éditeurs commerciaux pour une juste et équitable utilisation de leurs publications.

¹⁵ The European Distributed Institute of Taxonomy, EDIT, is the collective answer of 27 leading European, North American and Russian institutions to a call of the European Commission, issued in 2004, for a network in "Taxonomy for biodiversity and Ecosystem Research." This project started on 1 March 2006 and will last 5 years." <http://www.e-taxonomy.eu>

¹⁶ SYNTHESESYS is "the European Union-funded Integrated Infrastructure Initiative grant. This five year project, which began in February 2004, comprises 20 European natural history museums and botanic gardens [and] aims to create an integrated European infrastructure for researchers in the natural sciences." <http://www.synthesys.info/index.htm>

Les liaisons aux autres types de données (exemple, molécule, morphologique, phénotype) sont les solutions à l'ensemble du plan. Il sera nécessaire d'obtenir une rentabilité égale des coûts et d'accélérer l'entrée des collections spéciales et les articles avec des encarts ou ceux qui sont trop grands. La reconnaissance des Caractères Optiques (OCR) doit être améliorée pour les plus vieilles publications et ceux en caractères autres que romains. Il est probable que des audiences soit étendues à travers des outils de réseautage accommodant et le contenu reproposé pour de nouvelles audiences. BHL fut développée pour les scientifiques, mais l'audience s'élargira en autant que les outils deviendront disponibles.

L'architecture fondamentale de la BHL nécessite un renforcement. La BHL se déplace de l'application environnemental d'internet .Net vers l'architecture Fedora Commons¹⁷. Fedora Commons fournit une ouverture de source architecturale qui peut générer beaucoup de formes du contenu de numérisation. Cela permettra une structure d'ouverture pour la description, la conservation et la récupération du matériel numérisé dans les différentes manières requises par les scientifiques et les bibliothécaires. Par exemple, *Biologia Centralia Americana* a été décrite comme seul titre avec 63 volumes ou comme 21 titres, chacun avec de multiple volumes. Fedora Commons peut organiser une telle variation et il a la capacité d'être une architecture persistente et durable pour la bibliothèque du patrimoine de la biodiversité.

Le partenariat de la BHL est en train de travailler avec toute la communauté taxonomique, les éditeurs, les Archives Internet et d'autres organisations, pour garantir que la littérature de la bibliothèque du patrimoine de la biodiversité soit disponible pour tous, des étudiants cherchant l'information sur les insectes dans les jardins aux scientifiques enquêtant sur les espèces en danger au Brésil. C'est une grande vision. Mais si la BHL est capable d'appuyer la recherche de la biodiversité en aidant à éliminer un fois des plus grands obstacles à l'identification des organismes du monde vivant, on pourra l'évaluer de grand succès.

¹⁷ About Fedora Commons <http://www.fedora-commons.org/about>

REFERENCES ET LIENS INTERNET

- Biodiversity Heritage Library: <http://www.biodiversitylibrary.org> (accessed April 23, 2008)
- Biodiversity Heritage Library Prospectus: <http://www.sil.si.edu/bhl/supportdocuments/BHLP-prospectus10-05.pdf> (accessed April 23, 2008)
- Convention on Biological Diversity <http://www.cbd.int/>
- Encyclopedia of Life: <http://www.eol.org> (accessed April 23, 2008)
- European Distributed Institute of Technology (EDIT) <http://www.e-taxonomy.eu/>
- Freeland, C., M. Kalfatovic, J. Paige, and M. Crozier. 2008. "Geocoding LCSH in the Biodiversity Heritage Library," *The Code4Lib Journal*, Issue 2. 11 pp.
- Global Biodiversity Information Facility (GBIF) <http://www.gbif.org/>.
- Godfray, H.C.J., B.R. Clark, I.J. Kitching, S.J. Mayo, M.J. Scoble. 2007. "The Web and the Structure of Taxonomy," *Systematic Biology*, 56 (6): 943-955.
- Leary, P. R, D. P. Remsen, C. N. Norton, D.J. Patterson and I.N. Sarkar. 2008. "uBioRSS: Tracking Taxonomic Literature Using RSS," *Bioinformatics* 23(11): 1434-1436.
- Minelli, A. 2003. "The Status of Taxonomic Literature," *Trends in Ecology and Evolution* 18(2): 75-78.
- Moritz, Thomas. 2002. "Building the Biodiversity Commons" *D-Lib Magazine* 8 (6):
- Moritz, T. 2005. "Macro-economic Case for Open Access." In Proceedings of a Conference in Association with an International Conference on the Barcode of Life, organised by Graham Higley and Chris Lyal (Natural History Museum, London), Anna Weitzman and Tom Garnett (Smithsonian Institution, Washington), and Tom Moritz (American Museum of Natural History, New York).
- OCLC Collection Analysis Tool <http://www.oclc.org/collectionanalysis/default.htm>
- Speers, L, and J. Edwards. 2008. "International Infrastructure of Enabling the New Taxonomy: the Role of the Global Biodiversity Information Facility (GBIF). In *The New Taxonomy*, The Systematics Association Special Volume Series 76 edited by Quentin D. Wheeler, Boca Raton, FL.: CRC Press, Taylor & Francis Group, pp 87-94.
- Synthesis of Systematic Resources (SYNTHESYS) <http://www.synthesys.info/index.htm>
- uBio <http://www.ubio.org/>
- U.S. Congress. Office of Technology Assessment. 1987. *Technologies To Maintain Biological Diversity*, OTA-F-330 (Washington, DC: U.S. Government Printing Office)
- Wheeler, Q. 2008. "Toward a New Taxonomy." In *The New Taxonomy*, The Systematics Association Special Volume Series 76 edited by Quentin D. Wheeler, Boca Raton, FL.: CRC Press, Taylor & Francis Group, pp. 1-18
- Wheeler, Q. "Taxonomic Shock and Awe." In *The New Taxonomy*, The Systematics Association Special Volume Series 76 edited by Quentin D. Wheeler, Boca Raton, FL.: CRC Press, Taylor & Francis Group