



# Faire Avancer le Catalogue des Collections d'Histoire Naturelle du Monde

Donald Hobern, Alex Asase, Quentin Groom, Maofang Luo, Deborah Paul, Tim  
Robertson, Patrick Semal, Barbara Thiers, Matt Woodburn, Eliza Zschuschen

Version 2.0, 2020-04-15

# Table des matières

Colophon .....	1
Suggestion de citation .....	1
Contributeurs .....	1
Licence .....	1
URI permanent .....	1
Contrôle du document .....	1
Image de couverture .....	2
Contexte .....	3
Comment répondre à ce Note Conceptuelle .....	4
1. Utilisations pour le catalogue .....	5
1.1. Un annuaire pour soutenir la communauté des collections .....	5
1.2. Localisation de spécimens et de matériaux génétiques .....	6
1.3. Un premier pas vers la base de données des collections .....	6
1.4. Évaluation de la portée et de la valeur des collections .....	6
1.5. Augmentation de la valeur des données sur les spécimens, publications taxonomiques, etc. ....	7
1.6. Réduction de la duplication des efforts .....	7
1.7. Fondation pour des services nouveaux et enrichis .....	7
1.8. Améliorations de la citation et de la visibilité pour les collections .....	8
1.9. Prise en charge des besoins et applications nationaux et régionaux .....	8
2. Informations dans le catalogue .....	9
2.1. Portée du catalogue et définition d'une « collection » .....	9
2.2. Identifiants pour les collections .....	10
2.3. Hiérarchie des structures de collections et sous-collections .....	10
2.4. Description d'une collection .....	11
2.5. Liens de données plus étendus .....	11
2.6. Services d'information liés aux collections .....	12
3. Technologie pour le catalogue .....	13
3.1. Chemins et outils pour la publication des enregistrements de collection .....	13
3.2. Catalogues communautaires .....	14
3.3. Catalogue intégré .....	14
3.4. Systèmes de gestion de collection .....	15
3.5. Interfaces, APIs et modules client .....	15
4. Gouvernance du catalogue .....	15
4.1. Propriété des informations pour chaque collection .....	16
4.2. Communautés de pratique .....	16
4.3. Infrastructures techniques .....	16
4.4. Dispositions de gouvernance .....	17

4.5. Incitations pour les contributeurs.....	17
4.6. Financement et durabilité .....	17

# Colophon

## Suggestion de citation

Hobern D, Asase A, Groom Q, Luo M, Paul D, Robertson T, Semal P, Thiers B, Woodburn M & Zschuschen E (2020) Preprogressant le Catalogue des Collections d'Histoire Naturelle du Monde. v2.0. Copenhague: Secrétariat du GBIF. <https://doi.org/10.35035/p93g-te47>.

## Contributeurs

- **Donald Hobern** [<https://orcid.org/0000-0001-6492-4016>], Catalogue of Life | International Barcode of Life
- **Alex Asase** [<https://orcid.org/0000-0003-0116-3445>], Université du Ghana | GBIF Ghana
- **Quentin Groom** [<https://orcid.org/0000-0002-0596-5376>], Jardin Botanique Meise
- Maofang Luo, Académie Chinoise des Sciences
- **Deborah Paul** [<https://orcid.org/0000-0003-2639-7520>], iDigBio | TDWG CD Interest Group
- **Tim Robertson** [<https://orcid.org/0000-0001-6215-3617>], Secrétariat du GBIF
- **Patrick Semal** [<https://orcid.org/0000-0002-4048-7728>], Institut Royal des Sciences Naturelles de Belgique | CETAF
- **Barbara Thiers** [<https://orcid.org/0000-0002-8613-7133>], Jardin botanique de New York | Index Herbariorum
- **Matt Woodburn** [<https://orcid.org/0000-0001-6496-1423>], Musée d'histoire naturelle, Londres | TDWG CD Interest Group
- Eliza Zschuschen, Herbarium National du Suriname

Les contributeurs supplémentaires seront crédités ici dans les versions suivantes.

## Licence

Le document *Faire Avancer le Catalogue des Collections d'Histoire Naturelle du Monde* est licencié sous [Creative Commons Attribution 4.0 Unported License](https://creativecommons.org/licenses/by/4.0) [<https://creativecommons.org/licenses/by/4.0>].

## URI permanent

<https://doi.org/10.35035/p93g-te47>

## Contrôle du document

v2.0, mars 2020

Mis à jour de la v1.0, publié le 25 février 2020: Hobern D, Asase A, Groom Q, Paul D, Robertson T, Semal P, Thiers B & Woodburn M (2020) Faire Avancer le Catalogue des Collections d'Histoire

Les changements principaux par rapport à la v1.0 sont ci-dessous :

- Clarification du cadre de la consultation et définition d'une "collection d'histoire naturelle"
- Sujets et questions supplémentaires :
  - Améliorations de la citation et de la visibilité pour les collections - Q8
  - Prise en charge des besoins et applications nationaux et régionaux - Q9
  - Identifiants pour les collections - Q11
  - Hiérarchie des structures de collection et des sous-collections - Q12
- Modifications majeures au sujet et aux questions :
  - Cadre pour le catalogue et définition de la "colleciton" - Q10
- Modifications mineures suivant les webinaires préparatoires et des contributions de Ana Casino et Luc Willemse

## Image de couverture

Maryland sematophyllum mousse (*Sematophyllum marylandicum*), collecté par W.R. Buck dans Ferncliff Natural Area, Ohio State Park, Pennsylvania, États-Unis. Photo 2018 New York Botanical Garden via [The New York Botanical Garden Herbarium \(NY\)](#) [<https://www.gbif.org/occurrence/1929304566>], licencié sous [CC BY 4.0](http://creativecommons.org/licenses/by/4.0/) [<http://creativecommons.org/licenses/by/4.0/>].

# Contexte

Ce document explore les besoins et les opportunités liés à l'information et aux services numériques associés aux « collections d'histoire naturelle ». Ce terme est utilisé ici pour se référer en particulier aux collections institutionnelles qui détiennent des matériaux biologiques conservés (spécimens, tissus, extraits d'ADN, etc.). Plusieurs cas d'utilisation importants concernent spécifiquement le rôle de ces collections dans le soutien de la taxonomie et d'autres domaines de la recherche biologique. Cependant, un bon nombre des exigences abordées ici sont communes à d'autres collections de sciences naturelles, en particulier les collections géoscientifiques, les collections vivantes et les collections privées. Nous espérons que ce document et les discussions prévues répondront également aux besoins de cette communauté dans son ensemble.

L'information à propos des collections d'histoire naturelle aide à cartographier le paysage complexe des ressources de recherche et aide les chercheurs à localiser et contacter les détenteurs de spécimens. Les notices de collecte contribuent au développement d'une **graphique de la connaissance de la biodiversité complètement interconnecté** [<https://doi.org/10.3897/rio.2.8767>], illustrant l'existence et l'importance des musées et des herbiers et fournissant le contexte des données disponibles sur les spécimens. Ces jeux de données peuvent également ouvrir de nouvelles avenues pour une utilisation nouvelle de ces collections et pour accélérer leur disponibilité en ligne.

Ce document explore des idées pour améliorer une collaboration mondiale en vue de construire, de maintenir et d'utiliser un **catalogue complet des collections d'histoire naturelle du monde**. Chaque idée est présentée comme un sujet distinct avec un ensemble de questions pour guider la discussion au cours de la consultation en ligne, *Progressant le Catalogue des Collections d'Histoire Naturelle du Monde* [<https://www.gbif.org/fr/news/6TvOkvpPlxRm5vHxIjYN5/>].

Au cours des dernières décennies, le domaine de l'informatique de la biodiversité s'est développé pour inclure des chercheurs et des informaticiens du monde entier, qui collaborent pour rassembler les connaissances sur les espèces et les écosystèmes de la planète sous une forme facilement utilisable.

L'informatique de la biodiversité a principalement porté sur les espèces et autres taxons (y compris leurs noms, leurs caractères de diagnostic et leurs caractéristiques), spécimens d'histoire naturelle (y compris les données sur leur collection sur le terrain, leurs mesures, images, séquences, etc.), et les observations sur le terrain (y compris les données sur les incidents, la distribution et l'abondance, les activités de suivi, la science citoyenne, la génomique et de nombreuses autres sources). Ces éléments contribuent ensemble à relever deux défis fondamentaux de la biologie: caractériser l'ensemble des espèces avec lesquelles nous partageons la planète, et la compréhension de la distribution, de la co-occurrence, des interactions et de la dynamique de ces espèces dans l'espace et dans le temps.

La communauté d'informatique de la biodiversité a également prêté attention à d'autres catégories d'information qui soutiennent ces éléments primaires, en particulier par le biais de la numérisation de la vaste littérature sur la taxonomie et la biodiversité et de travailler à l'élaboration d'un catalogue complet des collections d'histoire naturelle du monde, y compris des musées, des herbariums et une gamme de collections spécialisées.

Ces collections sont le dépôt de matériaux provenant de siècles d'investissements internationaux

pour collecter, documenter, étudier et décrire les espèces. Les spécimens et autres matériaux contenus dans ces collections ancrent notre compréhension de l'évolution et de la diversité contemporaine. Ils constituent le pont entre les connaissances historiques et les efforts continus pour décrire la vie sur Terre. Plusieurs de ces dépôts sont irremplaçables ou donnent par ailleurs un aperçu irrécupérable des distributions et des écologies passées. Ces informations sont également importantes pour la modélisation des futurs environnementaux. L'information sur les collections elles-mêmes est un outil important pour y accéder, les enrichir et les utiliser.

Plusieurs cas d'utilisation établis pour des informations normalisées de collecte se rapportent principalement à des collections biologiques préservées. Cependant, nous espérons que la consultation explorera également deux autres contextes étroitement liés : 1) les collections géologiques (souvent détenues et gérées par les mêmes institutions que les collections biologiques) et 2) les collections vivantes (se chevauchant de façon significative avec les matières et les utilisations de recherche des collections biologiques préservées). Nous saluons les contributions qui abordent ce champ d'application plus large.

## Comment répondre à ce Note Conceptuelle

Lisez les sections ci-dessous et contribuez à l'élaboration d'une feuille de route pour la création du catalogue.

Nous saluons les contributions de la manière suivante :

- Représentez-vous une **partie prenante, un projet, une base de données, un outil, une norme ou une communauté** qui aborde certains aspects des sujets décrits ici, ou avez-vous des idées de nouvelles approches pour mobiliser ou utiliser les données sur les collections?
  - Veuillez contacter **Donald Hobern** [<mailto:dhobern@gbif.org>] avant le 3 avril 2020 pour apporter des idées ou des exemples significatifs qui ajouteront de la valeur aux discussions en ligne.
  - Nous accueillons les documents courts ou les diapositives qui peuvent être partagés sur le site Web de consultation.
  - Si les présentations ne sont pas claires sans plus d'explications, veuillez modifier les diapositives ou envisager de fournir la présentation comme une vidéo pré-enregistrée avec un commentaire audio.
  - Veillez à ce que tous les documents soient concis, afin qu'un lecteur ou un spectateur puisse assimiler les idées dans les quinze minutes ou (idéalement) moins.
- Aimerez-vous **consulter les webinaires préparatoires** qui décrivent la portée et planifier la consultation ?
  - Les enregistrements sont disponibles en ligne sur **YouTube** [[https://youtube.com/playlist?list=PLY6tIKN\\_kHB8CxNdY\\_x1jmmuZx4UDZ6NB](https://youtube.com/playlist?list=PLY6tIKN_kHB8CxNdY_x1jmmuZx4UDZ6NB)] et **Vimeo** [<https://vimeo.com/showcase/6859611>].
- Vous aimeriez **contribuer aux discussions en ligne** pour la consultation?
  - Veuillez vous inscrire pour rejoindre la communauté de consultation sur le site du GBIF Discours.
  - Les discussions se déroulera entre le 17 et le 29 avril 2020.
  - Nous vous tiendrons informé quand plus d'informations sont ajoutées au site et nous vous

assurerons de recevoir des mises à jour régulières lors de la consultation

- Serez-vous en mesure d'élargir la pertinence de la consultation en **traduisant des mises à jour sommaires courtes** dans des langues autres que l'anglais?
  - Nous prévoyons de faire circuler régulièrement de courts résumés (quelques paragraphes chaque jour ou deux) à tous les participants durant la période de consultation principale afin de garder la discussion centrée, résumer l'accord et mettre en évidence de nouvelles idées et questions.
  - Nous sommes heureux de recevoir toute assistance pour leurs traductions dans des langues qui permettront à tous les participants de suivre plus facilement les discussions et de savoir comment contribuer.
  - Veuillez contacter **Donald Hobern** [mailto:dhobern@gbif.org] si vous êtes intéressé à aider.

## 1. Utilisations pour le catalogue

Le **TDWG Collection Description Interest Group** [https://github.com/tdwg/cd/tree/master/reference/use\_cases] a rassemblé des cas d'utilisation pour des données de collection d'histoire naturelle de plusieurs parties prenantes. En outre, les grands projets européens, y compris **ICEDIG** [https://www.icedig.eu/], se préparent au développement de **DiSSCo** [https://www.dissco.eu/] infrastructure en documentant les cas d'utilisation des collections.

### 1.1. Un annuaire pour soutenir la communauté des collections

Le personnel des collections et les taxonomistes collaborent en tant que communauté véritablement mondiale. De précieux spécimens sont distribués entre institutions dans toutes les parties du monde. Les chercheurs visitent ces collections ou empruntent des spécimens dans le cadre de leurs travaux. L'Index Herbariorum (IH) est le répertoire des informations sur l'herbarie du monde (adresses, contacts, spécialités, dimensions, etc.). Il s'agit d'une ressource bien gérée et hautement considérée comme un outil par la communauté botanique. Il n'existe pas d'équivalent complet pour les autres collections d'histoire naturelle, bien que les infrastructures nationales/régionales telles que les pages des collections de l'ALA, la liste iDigBio US Collections et les profils CETAF remplissent des rôles similaires. Le GBIF a récemment intégré le Registre Mondial des Collections Scientifiques (GRSciColl) dans son registre en tant que cadre pouvant être étendu avec des données plus riches organisées par les communautés de collections.

**Q1.** La communauté des collections bénéficierait-elle d'un répertoire complet de toutes les collections d'histoire naturelle ? Qui pourrait utiliser un tel annuaire ? (L'accent est mis sur le catalogue en tant que répertoire des institutions connues et des informations nécessaires pour les contacter.)

## 1.2. Localisation de spécimens et de matériaux génétiques

Les études taxinomiques et les autres projets de recherche dépendent généralement des chercheurs (ou de leurs contacts) qui savent quelles institutions possèdent des spécimens pertinents ou d'autres matériaux. Ce qui est compliqué par l'histoire des expéditions et des activités de collecte. Les spécimens ont été dispersés sur tous les continents. Seule une faible proportion de ces spécimens ont été placés sous forme de formulaires accessibles via le GBIF ou d'autres portails. Un catalogue fournissant au moins des renseignements sommaires sur la portée taxonomique et géographique de chaque collection pourrait aider les chercheurs à localiser les matériaux pertinents.

**Q2.** Les informations sommaires sur les matériaux de chaque collection seraient-elles un outil utile ? Qui utiliserait ces informations ? Quel est le niveau minimum d'informations (et quel est l'idéal) pour soutenir ces utilisateurs ?

## 1.3. Un premier pas vers la base de données des collections

Les informations nécessaires pour construire le catalogue des collections correspondent de près aux métadonnées requises pour publier un jeu de données de spécimens sur le GBIF et d'autres portails. Un enregistrement qui décrit une collection peut être considéré comme une première étape minimale, peut-être par le biais de processus tels que 'Join The Dots' et continuer à la numérisation complète. Un catalogue complet de ces enregistrements pourrait guider les efforts pour donner la priorité à la numérisation, en mettant en évidence les collections avec des fonds d'une importance particulière ou en aidant le développement de réseaux de numérisation collaborative tels que les réseaux de collections thématiques ADBC.

**Q3** La publication d'un dossier de collection dans un catalogue peut-elle aider les collections à se diriger vers la numérisation complète ? Quels sont les incitants ou le soutien dont les collections ont besoin pour en faire une étape digne d'intérêt ?

## 1.4. Évaluation de la portée et de la valeur des collections

Les estimations du nombre de spécimens détenus par les collections s'élèvent à des milliards, mais aucun nombre définitif n'existe. Un catalogue pourrait aider à réduire ces estimations et à évaluer la valeur économique de ces dépôts irremplaçables. Ces informations peuvent aider à justifier l'ampleur des efforts et du financement nécessaires pour la numérisation des collections et rendre leurs données accessibles pour une utilisation universelle, fiable et persistante.

**Q4.** Des estimations plus précises de la portée et de la valeur des collections seraient-elles utiles ? Comment pourraient-elles être utilisées et par qui ?

## 1.5. Augmentation de la valeur des données sur les spécimens, publications taxonomiques, etc.

Des informations précises dans toute les collection peuvent être utilisées comme référence ou comme données liées associées à des enregistrements de spécimens et à d'autres objets de données. Les utilisateurs d'enregistrements de spécimens ont besoin d'informations contextuelles sur la collection qui contient le spécimen, par exemple pour communiquer avec les gestionnaires de collection à propos des spécimens individuels, pour apporter des corrections aux données des spécimens ou simplement pour déterminer si la collection est susceptible de contenir des quantités de spécimens similaires. Il peut être inefficace d'intégrer toutes ces informations dans l'enregistrements du spécimen. La détention d'une seule copie faisant autorité aide à maintenir l'information recueillie à jour. Les enregistrements de collection peuvent également contenir des renseignements sur la portée taxonomique ou géographique ou d'autres aspects qui peuvent résoudre les ambiguïtés potentielles dans un enregistrements de spécimen. Les liens vers les enregistrements de collection actuels amélioreront également les publications taxonomique faisant référence à leurs matériaux. Ceci est particulièrement important parce que les numéros de catalogue et autres identifiants de spécimens utilisés dans les publications peuvent ne pas être liés à des informations numérisées sur les spécimens. La connexion à la collection simplifie l'accès futur et peut permettre à l'avenir de déduire des liens numériques.

**Q5** Comment un catalogue de collections complet pourrait-il contribuer à améliorer des autres catégories de données sur la biodiversité ? Quelles conditions ces améliorations mettraient-elles sur le catalogue ?

## 1.6. Réduction de la duplication des efforts

Bien qu'aucun catalogue complet de collections n'existe, le besoin de telles informations conduit à ce que de telles données soient publiées de manière répétée dans différents formats pour différents portails, documentation du projet, métadonnées pour d'autres données, etc. Cette duplication est source de confusion, car des informations obsolètes restent sur le Web. Les mécanismes qui se lient toujours à une seule version continuellement mise à jour (et à une histoire de versions) résoudront ces problèmes.

**Q6.** Pouvons-nous identifier les économies de temps et de coûts qui résulteraient d'un catalogue de collections partagé bien géré ?

## 1.7. Fondation pour des services nouveaux et enrichis

Un annuaire complet pourrait servir de base à de nouveaux outils qui renforcent les efforts

taxonomiques et la coopération entre tous les détenteurs de collection. Un exemple pourrait être le développement de systèmes de prêts distribués ou de la numérisation à la demande, tel que prévu pour le système de prêts et de visites européens de DiSSCo (ELViS). Un catalogue pourrait également servir de vitrine aux institutions pour mettre en valeur leurs propriétés et leurs caractéristiques uniques, comme dans le concept visuel partagé par le GBIF pour les pages de collection. Le suivi et les rapports du GBIF sur l'utilisation des données sur la biodiversité dans les publications de recherche pourraient alimenter de nouveaux services fournissant des paramètres standards et aider les collections à mesurer et à rendre compte de leur impact.

**Q7** Quels autres services pourraient être développés sur les fondations d'un catalogue de collections ? Celles-ci attireraient-elles des investissements pour financer le développement et soutenir la maintenance du catalogue?

## 1.8. Améliorations de la citation et de la visibilité pour les collections

La valeur de la recherche est principalement mesurée en termes de visibilité et d'impact de la littérature publiée. Les collections d'histoire naturelle sont mal reconnues par de telles mesures et leur importance est presque cachée dans le contexte des outils de recherche fondateurs. Les utilisateurs des collections sont régulièrement invités à citer les spécimens examinés et à référencer la collection. Cependant, la citation est souvent **manquante, incomplète ou ambiguës** [<https://fistfulofcinctans.wordpress.com/2016/06/23/how-and-why-to-cite-museum-specimens-in-research/>]. Des infrastructures de recherche telles que **OpenAIRE** [<https://explore.openaire.eu/search/find>] en Europe cartographie de plus en plus non seulement les liens entre les chercheurs et les publications, mais aussi les ensembles de données, les projets, les fournisseurs de contenu et les organisations. Un catalogue pourrait aider à normaliser la citation des collections, rendant leur impact visible à travers de tels graphiques de connaissances. Les revues et les comités de rédaction pourraient être encouragés à exiger des identificateurs de collection standard partout où les collections sont référencées.

**Q8** Comment un catalogue complet pourrait-il promouvoir la citation et l'attribution des collections ? Que peut-on faire pour encourager une large utilisation standardisée des identifiants du catalogue ?

## 1.9. Prise en charge des besoins et applications nationaux et régionaux

Bien que cette consultation vise à encourager le développement d'informations standardisées pour toutes les collections à l'échelle mondiale, chaque pays ou région peut avoir des besoins ou des utilisations pour cette même information dans les applications et les services locaux. Il est important d'identifier une envergure de ces besoins et de s'assurer qu'ils sont traités dans le cadre d'une solution collaborative. Une approche inclusive incitera à travailler ensemble pour rendre l'information sur chaque catalogue aussi complète, actuelle et exacte que possible. Les exigences

sont relativement bien comprises de la part de l'Europe (par exemple DiSSCo) et des États-Unis (par exemple iDigBio), mais d'autres régions peuvent avoir des besoins subtilement ou sensiblement différents.

**Q9** Quels sont les besoins ou utilisations possibles au niveau national et régional ? Les portails nationaux ou les réseaux spécialisés nécessitent-ils des informations qui ne sont pas couvertes par les normes actuelles de métadonnées de collection? Existe-t-il des infrastructures de recherche régionales ou des sites Web publics importants qui incluent (ou devraient inclure) des informations sur les collections locales? Y a-t-il des utilisations régionales importantes qui ne soient pas adressées ailleurs dans ce document ?

## 2. Informations dans le catalogue

Nous devons développer une vision partagée du contenu du catalogue et de la manière dont il sera lié à d'autres produits d'information.

### 2.1. Portée du catalogue et définition d'une « collection »

La portée du catalogue doit être définie. L'usage principal envisagé est la liste et la description des collections contenant des spécimens biologiques conservés, dénommés ici "collections d'histoire naturelle". La consultation se concentrera sur le développement d'une solution solide et efficace pour les collections d'histoire naturelle, mais il sera précieux d'explorer les besoins et les opportunités d'autres types de collections de sciences naturelles. Certaines d'entre elles peuvent s'intégrer facilement dans le champ d'application du catalogue. Dans d'autres cas, le travail sur le catalogue pourra être bénéfique à ces autres communautés. Notez en particulier que: 1) de nombreuses institutions détiennent des collections biologiques et géologiques et peuvent les gérer comme une entité unifiée ; 2) DiSSCo inclut les collections géologiques dans son champ d'application et d'autres réseaux tels que iDigBio incluent à tout le moins des collections paléontologiques ; 3) GRSciColl a été créé un registre de toutes les collections scientifiques; et 4) la norme 'Collection Description' du TDWG est extensible à différents types de collections.

**Q10.** Quelle est la définition pour nos objectifs (critères minimaux et suffisants) d'une collection d'histoire naturelle? Comment les collections sont-elles liées et diffèrent des 1) institutions, 2) jeux de données et 3) événements de collecte (e. g. expéditions) ? Les catégories suivantes doivent-elles être incluses ? Sinon, y a-t-il des liens importants ou des opportunités dont on devrait encore tenir compte?

- Collections géologiques et paléontologiques
- Collections anthropologiques
- Collections ethnobotaniques
- Collections de bois (xylaria)
- Banques de tissus, dépôts d'ADN et collections de plaquettes
- Collections vivantes (collections microbiennes, zoos, aquarias, jardins botaniques, banques de semences)
- Collections personnelles

## 2.2. Identifiants pour les collections

La plupart des collections sont déjà identifiées par un ou plusieurs codes de collection et peuvent avoir déjà des identifiants web (URLs, DOIs, etc.) dans une ou plusieurs bases de données. Le catalogue pourrait réutiliser l'un ou l'autre de ces identifiants ou aider à supporter un schéma normalisé. GRSciColl est là pour aider à la normalisation des codes de collection et des identifiants lisibles par machines, mais plusieurs autres efforts sont également en place. Les identifiants uniques pour chaque collection seront importants afin de maximiser les liens croisés entre information et référence standardisée, mais d'autres identifiants existants devraient idéalement conduire à la même information et être reconnus comme des synonymes de ces identifiants préférés.

**Q11.** Quels schémas d'identification (IH collection codes, GRSciColl URIs, etc.) existent déjà et devraient être maintenus sous une forme ou une autre ? Ces systèmes suivent-ils une définition cohérente de collection d'histoire naturelle ? Quelles sont les caractéristiques des identifiants qui sont importantes pour l'utilisation par les machines et les humains? Y a-t-il des avantages à choisir un système d'identification particulier (e.g. [DOIs](https://www.doi.org/) [https://www.doi.org/] ou [ROR](https://ror.org/) [https://ror.org/]) ? Que peut-on faire pour promouvoir l'utilisation de ces identifiants préférés ?

## 2.3. Hiérarchie des structures de collections et sous-collections

Au sein de l'IH, chaque herbarium correspond généralement à une institution avec son propre code de collection, son adresse postale, etc. Au sein de la zoologie, les musées sont souvent structurés en collections de groupes taxonomiques différents, voire hiérarchisées. Les spécimens récoltés lors des expéditions célèbres ou par des chercheurs importants peuvent avoir leur propre identité et apparaître comme des collections spéciales. Par conséquent, les conservateurs et les chercheurs peuvent vouloir désigner des ensembles de spécimens différents (potentiellement superposés)

comme des collections séparées avec leurs propres noms, identifiants et descriptions.

**Q12.** Le catalogue doit-il supporter les relations hiérarchiques entre collections (et enregistrements de collection)? Dans l'affirmative, comment les relations parent-enfant pourraient fonctionner et pouvons-nous inférer les informations du parent à l'enfant ou vice versa?

## 2.4. Description d'une collection

Le **TDWG Collection Descriptions (CD) Interest Group** [<https://www.tdwg.org/community/cd/>] est en train de développer le **standard CD pour les descriptions de collection** [<https://github.com/tdwg/cd>] (dérivé du standard TDWG Natural Collections Description (NCD)). Les réseaux existants et les systèmes institutionnels utilisent une variété de formats différents ou de variantes de normes de métadonnées pour leurs enregistrements de collection, par lequel l'interopérabilité entre ces ressources (et de ce fait l'agrégation des données) est limitée. Pour surmonter cette barrière, il faut de la clarté autour de facteurs tels que les normes et les vocabulaires préférés, les champs obligatoires et la compatibilité entre les informations sous différents formats.

**Q13.** Quelles informations descriptives devraient être considérées comme obligatoires ou souhaitables pour chaque collection ? Est-ce que le TDWG CD couvre tout ce qui est nécessaire? Sinon, quelles améliorations seraient nécessaires ? Combien de ces informations doivent être normalisées pour un traitement automatique (plutôt que juste pour les lecteurs humains)?

## 2.5. Liens de données plus étendus

Les informations du catalogue de collection peuvent être liées à un large éventail d'autres informations de biodiversité (spécimens, séquences, jeux de données, images, publications, etc.) pour supporter l'accès et l'exploration de l'information.

**Q14.** Quelles informations devraient être liées aux enregistrements de collection ? Nous devrions nous concentrer sur l'établissement de liens qui justifieront réellement les coûts de création et de maintien de ceux-ci. Ceux qui suivent semblent être de bons candidats, mais d'autres sont possibles. Dans chaque cas, nous devrions déterminer si le lien doit être bidirectionnel :

- Spécimens détenus par une collection
- Spécimens types détenus par une collection
- Espèces/taxons représentés dans une collection (avec/sans nombre de spécimens)
- Séquences, images et autres préparations de la collection (mais celles-ci sont sans doute mieux traitées comme informations sur les spécimens plutôt que sur la collection)
- Jeux de données (listes d'espèces, occurrences, événements d'échantillonnage) associés à la collection
- Expéditions de collecte réalisées par ou contribuant à la collection (modélisées comme des événements d'échantillonnage ?)
- Collecteurs associés à une collection
- Publications basées sur le matériel de la collection
- Chercheurs/staff associés à la collection
- Carnets de notes de terrain

## 2.6. Services d'information liés aux collections

La valeur principale du catalogue de collection peut apparaître dans les services d'information qui peuvent être offerts autour de l'information gérée. La prise en considération de ces services peut aider à clarifier les exigences en matière de contenu.

**Q15.** Que voulons-nous faire du catalogue, au-delà d'avoir des données ouvertes propres et complètes sur chaque collection? Les services potentiels suivants sont susceptibles d'être candidats, mais d'autres sont possibles. Dans chaque cas, le service dépendrait-il d'un partenariat avec d'autres référentiels numériques (par exemple **BHL** [<https://www.biodiversitylibrary.org/>], **GBIF** [<https://www.gbif.org/fr/>], **CoL** [<http://www.catalogueoflife.org/>])?

- Évaluer la croissance, l'étendue et la valeur des collections mondiales
- Alimenter un **tableau de bord de numérisation des collections** [<https://zenodo.org/record/2621055#.Xn1lqIgzabi>] afin de surveiller et mettre en évidence les progrès
- Découvrir l'emplacement des matériaux biologiques ou la présence probable de matériaux biologiques pour n'importe quel taxon
- Développer des services de découverte afin d'accéder à l'information sur les spécimens types ou communiquer avec la collection appropriée lorsque le spécimen n'est pas numérisé
- Identifier les sections de collections qui doivent être numérisées pour répondre à des questions spécifiques
- Analyser les écarts entre des données de spécimens publiés et le catalogue de collection afin d'orienter la numérisation qui comblera au mieux les lacunes taxonomiques, géographiques ou autres.
- Découvrir les avoirs qui rendent une collection particulière unique, et en augmente ainsi la valeur
- Développer et financer des programmes de numérisation collaborative axés sur la compréhension globale des avoirs du réseau
- Développer des systèmes interinstitutionnels de prêts et d'ateliers taxinomiques
- Développer des modèles de citation pour les collections et suivre leur impact
- Effectuer une évaluation des risques sur la santé ou la stabilité d'une collection

## 3. Technologie pour le catalogue

Un large éventail d'outils différents sont déjà utilisés dans le cadre des métadonnées de collection, la gestion partielle de catalogues tels que IH, GRSciColl, le Registre des collections du CETAF et les pages de collections nationales. Celles-ci varient en termes de capacités techniques et de pérennité. Certaines sont bien soutenues par les communautés existantes et pourraient faire partie d'une solution interconnectée. Le but de cette consultation est de déterminer quels composants, suffisamment mûrs et stables, peuvent contribuer à une telle solution et de déterminer quels autres composants devraient encore être développés.

### 3.1. Chemins et outils pour la publication des enregistrements de collection

Les informations existantes sur les collections sont éditées et maintenues de différentes manières. IH permet aux herbiers de fournir ou de modifier leurs enregistrements et offre un support aux

herbiers leurs mises à jour par e-mail ou par d'autres canaux. D'autres communautés, comme les portails nationaux, offrent d'autres moyens aux collections pour fournir ou mettre à jour leurs informations. Plusieurs outils aident les éditeurs de données à créer des métadonnées EML pour la publication de données sur le GBIF ou ailleurs. Ceux-ci pourraient évoluer vers des enregistrements de collection dans des formats préférés. L'**Integrated Publishing Toolkit (IPT)** [<https://www.gbif.org/fr/ipt>] pourrait être amélioré pour offrir l'enregistrement de collection comme l'un des types d'enregistrements principaux qui peuvent être partagés. Cela permettrait aux collections de publier un ou plusieurs enregistrements de collection comme un petit jeu de données indépendant ou de réseaux de collections pour gérer et publier un jeu de données comprenant de nombreux enregistrements de collection. Wikidata pourrait également servir d'outil ou de plate-forme pour éditer des informations du catalogue et le rendre largement accessible et réutilisable.

**Q16.** Quels outils, bases de données et sites Web existants peuvent aider à mobiliser et à maintenir les informations de collection ? Est-il possible d'identifier des outils ou des chemins supplémentaires qui devraient être développés ou supportés ?

## 3.2. Catalogues communautaires

IH est le catalogue le mieux établi et dessert une grande communauté de collections, mais beaucoup d'autres communautés sont importantes, comme les efforts ciblés au niveau régional ou national, tels que les profils institutionnels du CETAF, les portails web d'**iDigBio** [<https://www.idigbio.org/portal/collections>] et d'ALA, et l'**Initiative pour une seule collection mondiale** [<https://biss.pensoft.net/article/38772/>], et les efforts alignés thématiques, comme le **Répertoire mondial des collections culturelles** [<http://www.wfcc.info/ccinfo/index.php/home/content>] et le portail du **Réseau mondial de la biodiversité du génome** [[http://www.ggbn.org/ggbn\\_portal/members/index](http://www.ggbn.org/ggbn_portal/members/index)]. Un catalogue global complet devrait s'assurer que les besoins de ces différentes communautés sont satisfaits et soutenir leur fonctionnement continu et leur indépendance partout où elles sont utiles aux collections. Comprendre ces exigences est essentiel à la planification de la mise en œuvre technique et à la gouvernance du catalogue.

**Q17.** Quels catalogues répondent déjà aux besoins de certaines communautés de collections? Comment un catalogue intégré pourrait-il soutenir ces communautés ? Quelles communautés ont besoin d'une identité et/ou d'une plateforme propre? Quelle est la meilleure façon d'inclure ces communautés dans une solution interconnectée? Est-il opportun que le contenu soit créé et amélioré par un public plus large (e. g. via Wikidata)?

## 3.3. Catalogue intégré

Le GBIF a pour mission de fournir un soutien à l'échelle mondiale aux solutions informatiques de la biodiversité et a élargi son Registre afin d'héberger les données historiquement maintenues dans GRSciColl. Le contenu de GRSciColl est incomplet et doit être considéré comme un cadre à l'expansion vers des métadonnées de collection plus riches qui représentent correctement les besoins et les intérêts des collections. Le GBIF peut servir de contexte à l'intégration et à la déduplication des informations de collection à partir de différentes sources et à l'interconnexion de ces informations avec d'autres données sur la biodiversité. Le GBIF a besoin de conseils sur la

meilleure façon de répondre aux besoins et à la gestion de l'identité des collections et de leurs communautés, au fur et à mesure qu'il développe de tels services.

**Q18.** Y a-t-il des problèmes si GBIF fournit l'hébergement et le support du catalogue mondial au travers de son Registre ? De quoi faut-il s'assurer pour que cela réponde aux besoins des collections et soit complètement adopté et endossé par la communauté des collections? Quels défis faut-il relever pour minimiser tant la duplication du contenu que les efforts au travers d'un catalogue intégré?

### 3.4. Systèmes de gestion de collection

La plupart des collections d'histoire naturelle conservent des données sur leurs spécimens dans un système de gestion de collections (CMS) tel que **Specify** [<https://www.sustain.specifysoftware.org/about/>], **Symbiota** [<http://symbiota.org/docs/>], **EMu** [<https://emu.axiell.com/>], **DarWIN** [<https://biss.pensoft.net/article/39054/>] ou **BRAHMS** [<https://dps007.plants.ox.ac.uk/bol/>]. Certains de ces outils pourraient s'enrichir d'une interface directe avec le catalogue de collection, fournissant ainsi des métadonnées et des statistiques constamment à jour.

**Q19.** Quelles sont les exigences présentes ou futures pour les interfaces directes entre les plates-formes CMS et le catalogue de collections ? Y a-t-il des opportunités spéciales qui devraient être envisagées ? Les plateformes CMS pourraient-elles devenir une source de métadonnées pour les collections institutionnelles dans un catalogue mondial ?

### 3.5. Interfaces, APIs et modules client

La valeur d'une ressource commune partagée peut être maximisée en s'assurant que les interfaces et les API répondent aux besoins de tous les groupes clés de parties prenantes, en ce compris la résolution des problèmes liés à la diffusion de contenu dans toute la mesure possible dans plusieurs langages. Certains besoins peuvent être résolus en proposant des composants clients réutilisables qui peuvent être intégrés dans d'autres applications.

**Q20.** Quelles interfaces et API sont nécessaires pour maximiser l'accès au catalogue de collections? Comment le catalogue peut-il au mieux prendre en charge les diverses communautés d'utilisateurs, en ce y compris les locuteurs de langues différentes ?

## 4. Gouvernance du catalogue

Les standards et les outils ne sont qu'une partie de la solution. Pour que le catalogue réussisse et fournisse de la valeur, il doit être accepté par les parties prenantes qu'il représente. en particulier les institutions détentrices de collections et les communautés qui soutiennent les collections. Il est important d'identifier les parties prenantes qui ont besoin de s'approprier chaque aspect de la collection et de comprendre comment elles peuvent être activées, habilités et avoir les ressources

afin d'assumer ces responsabilités. Des mécanismes sont également nécessaires pour faire face à des situations dans lesquelles des besoins ou des intérêts peuvent entrer en conflit.

## 4.1. Propriété des informations pour chaque collection

L'hypothèse de base est que chaque institution devrait avoir la responsabilité et le contrôle de l'information sur ses collections. Cependant, il peut être approprié de déléguer partiellement ou entièrement cette responsabilité, à des communautés thématiques, régionales ou nationales qui ont des curateurs de données capables d'assurer la qualité et la normalisation des enregistrements de collections. Dans certains contextes, où les institutions ont, pour quelque raison que ce soit, n'ont pas fourni des informations faisant autorité ou n'ont pas les ressources pour le faire, il peut y avoir des raisons d'autoriser ou d'encourager une plus large base d'utilisateurs à contribuer et à améliorer les enregistrements de collections. Dans tous les cas, un historique de l'information est nécessaire afin que les utilisateurs puissent comprendre et répondre aux changements apportés par d'autres.

**Q21.** Comment gérer la propriété et le contrôle d'accès des enregistrements de collections ? Comment les éditeurs appropriés devraient-ils être reconnus et validés ? Y a-t-il des situations où une intervention automatique ou humaine sera nécessaire pour résoudre les désaccords ou les différences ?

## 4.2. Communautés de pratique

Les communautés telles que IH, CETAF, ALA, iDigBio, etc., jouent un rôle important de soutien des collections et de promotion des pratiques basées sur les normes. Dans de nombreux cas, ces communautés ont un haut niveau de compréhension et participent étroitement au développement de solutions informatiques sur la biodiversité. Leurs rôles et leurs droits doivent être bien définis et pris en compte dans toute solution intégrée.

**Q22.** Quels sont les besoins de ces communautés pour pouvoir mener à bien leur travail et soutenir leurs collections? Comment une approche intégrée peut-elle améliorer leurs aides ? Quels risques doivent être pris en compte ?

## 4.3. Infrastructures techniques

Les Infrastructures d'information sur la biodiversité telles que GBIF, DiSSCo, iDigBio et d'autres plates-formes nationales et régionales sont généralement financées dans le cadre élargi d'objectifs de science ouverte pour les infrastructures de recherche. Leur participation peut constituer un pont important entre les besoins des communautés de collection, de financement et d'expertise en solutions informatiques. Les rôles et les responsabilités doivent toutefois être bien définis afin d'assurer que les besoins des chercheurs et des communautés d'utilisateurs restent centraux. Il est important de définir clairement comment ces infrastructures techniques peuvent participer au mieux à la solution globale, notamment en démontrant les bénéfices nécessaires pour obtenir un financement durable pour un catalogue intégré et pour toutes les parties qui le constituent.

**Q23.** Quelles infrastructures techniques doivent être engagées dans le cadre de la solution? Comment leurs rôles et leurs besoins seront-ils au mieux équilibrés avec ceux des collections et de leurs communautés ?

## 4.4. Dispositions de gouvernance

Une solution complexe, basée sur les communs, dépendra pour sa réussite à long terme d'un modèle de gouvernance qui garantira à toutes les parties la confiance que leurs intérêts soient bien servis et protégés. Le modèle devrait trouver le juste équilibre entre l'assurance d'une collaboration saine et la minimisation des frais généraux en termes de réunions, de rapports, etc.

**Q24.** Existe-t-il des modèles appropriés qui peuvent être adoptés ou étendus afin de soutenir la gouvernance du catalogue? Peut-elle être gérée dans le contexte d'une organisation ou d'une institution existante ?

## 4.5. Incitations pour les contributeurs

L'effort pour que chaque institution s'inscrire et de gérer ses propres enregistrements de collections sera sans doute relativement faible. Toutefois, la stabilité du système dépendra des efforts continus, de ces institutions ou d'autres parties, pour corriger les erreurs et les informations périmées. Il devrait y avoir des avantages clairs ou des incitants qui encourageraient les parties prenantes à contribuer à cet effort. Un objectif clé devrait être de s'assurer que le catalogue contribue utilement au travail des gestionnaires de collection et des taxonomistes. La reconnaissance des contributions devrait aussi être précieuse.

**Q25.** Quelles sont les incitants pour les différents contributeurs à maintenir les informations dans le catalogue ? Comment les maximiser ?

## 4.6. Financement et durabilité

Les besoins en financement dépendront entre autre de l'approche adoptée pour construire le catalogue. Les coûts seront plus élevés si le soutien central nécessaire au maintien du contenu est plus important. Même si le contenu est largement géré gratuitement par la communauté internationale, soutenir une telle infrastructure fiable nécessite des efforts et des investissements à long terme (voir par exemple le **modèle CoreTrustSeal** [<https://www.coretrustseal.org/>] pour les dépôts de confiance).

**Q26.** Comment financer les aspects techniques et de gouvernance? Un financement externe est-il envisageable? Quels autres modèles peuvent être réalisables (contributions provenant des collections, inclusion dans la mission financée du GBIF ou tout autre hôte)?