Advancing the Catalogue of the World's Natural History Collections

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 of Contents

Colophon

Suggested citation

Hobern D, Asase A, Groom Q, Luo M, Paul D, Robertson T, Semal P, Thiers B, Woodburn M & Zschuschen E (2020) Advancing the Catalogue of the World's Natural History Collections. v2.0. Copenhagen: GBIF Secretariat. https://doi.org/10.35035/p93g-te47.

Contributors

- 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], University of Ghana | GBIF Ghana
- Quentin Groom [https://orcid.org/0000-0002-0596-5376], Meise Botanic Garden
- Maofang Luo, Chinese Academy of Sciences
- Deborah Paul [https://orcid.org/0000-0003-2639-7520], iDigBio | TDWG CD Interest Group
- Tim Robertson [https://orcid.org/0000-0001-6215-3617], GBIF Secretariat
- Patrick Semal [https://orcid.org/0000-0002-4048-7728], Royal Belgian Institute of Natural Sciences | CETAF
- Barbara Thiers [https://orcid.org/0000-0002-8613-7133], New York Botanical Garden | Index Herbariorum
- Matt Woodburn [https://orcid.org/0000-0001-6496-1423], Natural History Museum, London | TDWG CD Interest Group
- Eliza Zschuschen, Suriname National Herbarium

Additional contributors to subsequent versions will be credited here.

Licence

Persistent URI

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

Document control

v2.0, March 2020

Updated from v1.0, published on 25 February 2020: Hobern D, Asase A, Groom Q, Paul D, Robertson T, Semal P, Thiers B & Woodburn M (2020) Advancing the Catalogue of the World's Natural History

Collections. v1.0. Copenhagen: GBIF Secretariat. https://doi.org/10.15468/doc-wnsx-ep77

The main changes from v1.0 are as follows:

- Clarification in scope of consultation and definition of "natural history collection"
- Additional topics and questions:
  - Improvements to citation and visibility for collections - Q8
  - Support for national and regional needs and applications - Q9
  - Identifiers for collections - Q11
  - Hierarchical collection structures and subcollections - Q12
- Major edits to topic and questions:
  - Scope for the catalogue and definition of “collection” - Q10
- Minor edits following preparatory webinars and input from Ana Casino and Luc Willemse

Cover image

Maryland sematophyllum moss (*Sematophyllum marylandicum*), collected by W.R. Buck in Ferncliff Natural Area, Ohiopyle State Park, Pennsylvania, United States. Photo 2018 New York Botanical Garden via The New York Botanical Garden Herbarium (NY) [https://www.gbif.org/occurrence/1929304566], licensed under CC BY 4.0 [http://creativecommons.org/licenses/by/4.0/].

Background

This paper explores needs and opportunities around digital information and services associated with "natural history collections". This term is used here to refer particularly to institutional collections that hold preserved biological materials (specimens, tissues, DNA extracts, etc.). Several important use cases relate specifically to the role of these collections to support taxonomy and other fields of biological research. However, many of the requirements discussed here are common to other natural science collections, especially geoscience collections, living collections and privately-owned collections. We hope that this paper and the planned discussions will also address the needs of this wider community.

Information about natural history collections helps to map the complex landscape of research resources and assists researchers in locating and contacting the holders of specimens. Collection records contribute to the development of a fully interlinked biodiversity knowledge graph [https://doi.org/10.3897/rio.2.e8767], showcasing the existence and importance of museums and herbaria and supplying context to available data on specimens. These records also potentially open new avenues for fresh use of these collections and for accelerating their full availability online.

This document explores ideas for improved global collaboration to build, maintain and use a comprehensive **catalogue of the world's natural history collections**. Each idea is presented as a separate topic with a set of questions to guide discussion within the online consultation, *Advancing the Catalogue of the World's Natural History Collections* [https://www.gbif.org/news/6TvOkvpPlxRm5vHxljYNN5/].

Over the last few decades, the field of biodiversity informatics has developed to include researchers and informaticians from all over the world, collaborating to bring together knowledge of the world's species and ecosystems in a readily usable form.

The focus of biodiversity informatics has largely been on species and other taxa (including their names, diagnostic characters and traits), natural history specimens (including information on their collection in the field, their measurements, images, sequences, etc.), and field observations (including information on occurrence, distribution and abundance surveys, monitoring activities, citizen science, genomics and many other sources). These elements together help to address two fundamental challenges in biology: characterising the set of species with which we share the planet, and understanding the changing distribution, co-occurrence, interactions, and dynamics of these species in space and time.

The biodiversity informatics community has also given attention to other categories of information that support these primary elements, especially through efforts to digitise the vast literature on taxonomy and biodiversity and work to develop a comprehensive catalogue of the world's natural history collections, including museums, herbaria and a range of specialised collections.

These collections are the repository for materials from centuries of international investment to collect, document, study and describe species. Specimens and other materials held in these collections anchor our understanding of evolution and contemporary diversity. They provide the bridge between historical knowledge and continuing efforts to describe life on Earth. Many of their holdings are truly irreplaceable or give otherwise irrecoverable insights into past distributions and ecology. Such insights are important also for modelling environmental futures. Information on the

collections themselves is an important tool for accessing, enriching and using them.

Many established use cases for standardised collection information relate primarily to preserved biological collections. This paper treats these collections as its core focus. However, we hope that the consultation will also explore two other closely related contexts: 1) geological collections (often held and managed by the same institutions as biological collections) and 2) living collections (overlapping significantly with the subject matter and research uses of preserved biological collections). We welcome inputs that address this wider scope.

How to respond to this Ideas Paper

Read the sections below and contribute to developing a roadmap for collaborative activity to build the catalogue.

We welcome contributions as follows:

- Do you represent a **stakeholder, project, database, tool, standard or community** that addresses some aspect of the topics outlined here, or do you have ideas for novel approaches to mobilise or use information on collections?
    - Please contact Donald Hobern [mailto:dhobern@gbif.org] by 3 April 2020 to contribute significant ideas or examples that will add value to the online discussions.
    - We welcome short documents or slide presentations that can be shared on the consultation website.
    - If presentations are unlikely to be clear without further explanation, please modify the slides or consider supplying the presentation as a pre-recorded video with audio commentary.
    - Please keep all materials brief and focused, so that a reader or viewer could assimilate the ideas within fifteen minutes or (ideally) less.
- Would you like to **review the preparatory webinars** outlining the scope and plan for the consultation?
    - Recordings are available online on YouTube [https://youtube.com/playlist?list=PLy6tIKN_kHB8CxNdY_x1jmmuZx4UDZ6NB] and Vimeo [https://vimeo.com/showcase/6859611].
- Would you like to **contribute to the online discussions** for the consultation?
    - Please register to join the consultation community on the GBIF Discourse site.
    - Discussions will take place between 17 and 29 April 2020.
    - We will keep you informed as more information is added to the site and ensure that you receive regular updates during the consultation
- Will you be able to expand the relevance of the consultation by **translating short summary updates** into languages other than English?
    - We expect to circulate regular short summaries (a few paragraphs every day or two) to all participants during the main consultation period to keep the discussion focused, summarise agreement, and highlight new ideas and questions.
    - We welcome assistance in translating these into languages that will make it easier for all participants to follow the discussions and know how to contribute.

- Please contact Donald Hobern [mailto:dhobern@gbif.org] if you are interested in helping.

1. Uses for the catalogue

The TDWG Collection Description Interest Group [https://github.com/tdwg/cd/tree/master/reference/use_cases] has collected use cases for natural history collection information from several major stakeholders. In addition, major European projects, including ICEDIG [https://www.icedig.eu/], are preparing for the development of the DiSSCo [https://www.dissco.eu/] infrastructure by documenting use cases for collections.

1.1. A directory to support the collections community

Collections staff and taxonomists collaborate as a truly global community. Valuable specimens are distributed between institutions in all parts of the world. Researchers visit these collections or borrow specimens as part of their work. Index Herbariorum (IH) is the directory of information on the world's herbaria (addresses, contacts, specialties, size, etc.). It is a well-managed resource and highly regarded as a tool by the botanical community. No full equivalent exists globally for other natural history collections, although national/regional infrastructures such as the ALA collections pages, the iDigBio US Collections List, and the CETAF profiles serve similar roles. GBIF has recently integrated the Global Registry of Scientific Collections (GRSciColl) into its registry as a framework that can be extended with richer information curated by collections communities.

Q1. Would the collections community benefit from a comprehensive directory of all natural history collections? Who would make use of such a directory? (The focus here is on the catalogue as a directory of known institutions and information required to contact them.)

1.2. Locating specimens and genetic materials

Taxonomic studies and other research projects normally depend on researchers (or their contacts) knowing which institutions hold relevant specimens or other materials. This is complicated by the history of expeditions and collecting activities. Specimens have been scattered across all continents. Only a small proportion of these specimens have been databased in forms that can be accessed through GBIF or other portals. A catalogue providing at least summary information on taxonomic and geographic scope for each collection could assist researchers in locating relevant materials.

Q2. Would summary information on every collection's materials be a useful tool? Who would use this information? What is the minimum level of information (and what is ideal) to support these users?

1.3. A first step towards databasing collections

The information needed to build the catalogue of collections closely matches the metadata required to publish a specimen dataset to GBIF and other portals. A record that describes a collection could be

treated as a minimal first step, perhaps leading through processes such as Join The Dots and onwards to comprehensive digitisation. A comprehensive catalogue of such records could guide efforts to prioritise further digitisation, by highlighting collections with holdings of particular relevance or by assisting the development of collaborative digitisation networks like the ADBC Thematic Collection Networks.

Q3. Can publishing a collection record to a catalogue assist collections in moving towards full digitisation? What incentives or support do collections need to make this a worthwhile step?

1.4. Assessing the scale and value of collections

Estimates of the number of specimens held by collections run into billions, but no definitive number exists. A catalogue could help to narrow these estimates and to assess the economic value of these irreplaceable holdings. This information may help to justify the scale of effort and funding needed to digitise collections and make their data accessible for universal, reliable and persistent use.

Q4. Would more accurate estimates of the scale and value of collections be useful? How might these be used and by whom?

1.5. Increased value for data on specimens, taxonomic publications, etc.

Accurate information on any collection can be used as a reference or as linked data associated with specimen records and other data objects. Users of specimen records need contextual information about the collection that holds the specimen, for example to communicate with collection managers about individual specimens, to offer corrections to specimen data, or simply to determine whether the collection is likely to hold quantities of similar specimens. It may be inefficient to embed all of this information within the specimen record. Holding a single authoritative copy assists with keeping the collection information current. The collection record may also contain information on taxonomic or geographic scope or other aspects that can resolve potential ambiguities within a specimen record. Links to current collection records will also enhance taxonomic publications referencing their materials. This is particularly important because catalogue numbers and other specimen identifiers used in publications may not link to digitised information on the specimens. Linking to the collection simplifies future access and may enable digital links to be inferred in future.

Q5. How could a comprehensive collections catalogue contribute to improvements to other categories of biodiversity data? What requirements would these improvements place on the catalogue?

1.6. Reducing duplication of effort

Although no complete catalogue of collections exists, the need for such information leads to such

data repeatedly being published in different formats for different portals, project documentation, metadata for other data, etc. This duplication results in confusion as outdated information remains on the web. Mechanisms that always link to a single continuously updated version (and a version history) would address these issues.

Q6. Can we identify savings in time and costs that would arise from a well-managed shared catalogue of collections?

1.7. Foundation for new and enriched services

A comprehensive directory could serve as a foundation for new tools that enhance taxonomic efforts and cooperation between all collection holders. One example might be the development of distributed loans systems or on-demand digitisation, as planned for the DiSSCo European Loans and Visits System (ELViS). A catalogue could also serve as a showcase for institutions to highlight their holdings and unique features, as in the visual concept shared by GBIF for collection pages. GBIF tracking and reporting on the use of biodiversity data in research publications could feed into new services that provide standard metrics and help collections to measure and report their impact.

Q7. What other services could be developed on the foundations of a collection catalogue? Would these attract investment to fund the development and support the maintenance of the catalogue?

1.8. Improvements to citation and visibility for collections

Research value is primarily measured in terms of visibility and impacts from published literature. Natural history collections are poorly recognised by such measures and their importance as foundational research tools is almost hidden. Users of collections are regularly urged to cite specimens examined and reference the collection. However, citation is often lacking, incomplete or ambiguous [https://fistfulofcinctans.wordpress.com/2016/06/23/how-and-why-to-cite-museum-specimens-in-research/]. Research infrastructures such as OpenAIRE [https://explore.openaire.eu/search/find] in Europe increasingly map not only linkages between researchers and publications but also datasets, projects, content providers and organisations. A catalogue could help to standardise citation of collections, making their impact visible through such knowledge graphs. Journals and editorial boards could be encouraged to require standard collection identifiers wherever collections are referenced.

Q8. How might a comprehensive catalogue promote citation and attribution for collections? What can be done to encourage wide standardised use of identifiers from the catalogue?

1.9. Support for national and regional needs and applications

Although this consultation aims to encourage the development of standardised information for all collections globally, each country or region may have needs or uses for this same information to in local applications and services. It is important to identify a range of these needs and to make sure they are addressed as part of a collaborative solution. An inclusive approach will bring incentives to work together to make information on each catalogue as complete, current and accurate as possible. Requirements are relatively well understood from Europe (e.g. DiSSCo) and the United States (e.g. iDigBio), but other regions may have subtly or significantly different needs.

Q9. What national and regional needs or possible uses should be considered? Do national portals or specialist networks require information not currently addressed by data standards for collection metadata? Are there significant regional research infrastructures or public websites that include (or should include) information on local collections? Are there regionally important uses that are not addressed elsewhere in this document?

2. Information in the catalogue

We need to develop a shared vision for the content that the catalogue should hold and how it interlinks with other information products.

2.1. Scope for the catalogue and definition of “collection”

The scope for the catalogue needs to be defined. The core use case under consideration is the listing and description of collections holding preserved biological specimens, referred to here as "natural history collections". The consultation will focus on developing a solution that is robust and effective for natural history collections, but it is valuable to explore needs and opportunities around other types of natural science collection. Some of these may fit readily within the scope of the catalogue. In other cases, work on the catalogue may offer benefits to these other communities. Note in particular: 1) many institutions hold both biological and geological collections and may manage these as a unified whole; 2) DiSSCo includes geological collections within its scope, and other networks such as iDigBio include at least paleontological collections; 3) GRSciColl was established to hold records on any scientific collection; and 4) the TDWG Collection Description standard is extensible for different collection types.

Q10. What is the definition for our purposes (minimal and sufficient criteria) of a natural history collection? How do collections relate to and differ from 1) institutions, 2) datasets and 3) collecting events (e.g. expeditions)? Should the following categories be included? Otherwise, are there important linkages or opportunities that should still be considered?

- Geological and paleontological collections
- Anthropological collections
- Ethnobotanical collections
- Wood collections (xylaria)
- Tissue banks, DNA repositories and slide collections
- Living collections (microbial collections, zoos, aquaria, botanic gardens, seed banks)
- Personal collections

2.2. Identifiers for collections

Most collections are already identified by one or more collection codes and may have existing web identifiers (URLs, DOIs, etc.) in one or more databases. The catalogue may reuse one or other of these identifiers or may help to support a standardised scheme. GRSciColl exists to assist with standardisation of collection codes and machine-readable identifiers, but several other efforts are also in place. Unique identifiers for each collection will be important to maximise cross-linkage of information and standardise citation, but other existing identifiers should ideally resolve to the same information and be recognised as synonyms for the preferred identifiers.

Q11. What identifier schemes (IH collection codes, GRSciColl URIs, etc.) already exist and need to be maintained in some form? Do these schemes follow a consistent definition of a natural history collection? What characteristics of identifiers are important for use by machines and humans? Are there benefits in selecting any particular identifier scheme (e.g. DOIs [https://www.doi.org/] or ROR [https://ror.org/] identifiers)? What can be done to promote use of the preferred identifiers?

2.3. Hierarchical collection structures and subcollections

Within IH, each herbarium record usually corresponds to an institution with its own unique collection code, street address, etc. Within zoology, museums are often structured as a set of collections with differing and possibly hierarchical taxonomic scope. Specimens collected on famous expeditions or by significant researchers may have their own identity and appear as special collections. As a result, curators and researchers may wish to refer to different (potentially overlapping) sets of specimens as separate collections with their own names, identifiers and descriptions.

Q12. Should the catalogue support hierarchical relationships between collections (and collection records)? If so, how do parent-child relationships work, and do we infer information from parent to child or vice versa?

2.4. Description of a collection

The TDWG Collection Descriptions (CD) Interest Group [https://www.tdwg.org/community/cd/] is currently developing the CD standard for collection descriptions [https://github.com/tdwg/cd] (evolving from the earlier TDWG Natural Collections Description (NCD) standard). Existing networks and institutional schemes use a variety of different formats or variants of metadata standards for their collection records, as a result of which interoperability between these resources (and hence data aggregation) is limited. To overcome this barrier, clarity is needed around factors such as preferred standards and vocabularies, mandatory fields and compatibility between information in different formats.

Q13. What descriptive information should be considered mandatory or desirable for each Collection? Does the TDWG CD work supply everything needed? Otherwise, what enhancements are necessary? How much of this information needs to be normalised for machine processing (rather than just for human readers)?

2.5. Wider data linkages

Information in the collection catalogue may be linked to a wide range of other biodiversity information (specimens, sequences, datasets, images, publications, etc.) to support information access and exploration.

Q14. What information should be linked to collection records? We should focus on making linkages that will actually justify the costs of creating and maintaining them. The following are likely to be candidates, but others are possible. In each case, we should determine whether the linkage needs to be bidirectional:

- Specimens held by a collection
- Type specimens held by a collection
- Species/taxa represented in a collection (with/without specimen counts)
- Sequences, images and other preparations from the collection (but these may be better treated as information about specimens rather than about the collection)
- Datasets (checklists, occurrences, sampling events) associated with the collection
- Collecting expeditions carried out by or contributing to the collection (modeled as sampling events?)
- Collectors associated with a collection
- Publications based on materials from the collection
- Researchers/staff associated with the collection
- Field notebooks

2.6. Information services relating to collections

The main value from the collection catalogue may appear in the information services that can be offered around the information managed. Considering these services may help to clarify the content requirements.

Q15. What do we want to do with the catalogue, beyond having clean and comprehensive linked open data about each collection? The following potential services are likely to be candidates, but others are possible. In each case, would the service depend on a partnership with other digital repositories (e.g. BHL [https://www.biodiversitylibrary.org/], GBIF [https://www.gbif.org/], CoL [http://www.catalogueoflife.org/])?

- Assess the growth, scale and value of the world's collections
- Provide collection digitisation dashboard [https://zenodo.org/record/2621055#.Xn1lqIgzabi] to monitor and highlight progress
- Discover the location of biological materials or the likely presence of biological materials for any taxon
- Develop discovery services for accessing information on type specimens or communicating with the relevant collection where the specimen is not digitised
- Identify sections of collections that should be digitised to answer specific questions
- Match gap analysis of published specimen data against the collection catalogue to prioritise digitisation for filling taxonomic, geographic, or other gaps.
- Discover holdings that make a particular collection unique, and therefore of even higher value
- Develop and fund collaborative digitisation programmes focused on understanding of the holdings of the network as a whole
- Develop cross-institutional loan systems and taxonomic workbenches
- Develop citation models for collections and track their impact
- Perform risk assessment of the health or stability of a collection

3. Technology for the catalogue

A wide range of different tools are already in use for authoring collection metadata, curating partial catalogues such as IH, GRSciColl CETAF Collections Registry and national collections pages. These vary in their technical capabilities and sustainability. Some are well supported by existing communities and could form part of an interconnected solution. A goal for this consultation is to identify which components are mature and stable and can contribute to such a solution and to identify what other components may need to be developed.

3.1. Pathways and tools for publishing collection records

Existing information on collections is edited and maintained in different ways. IH allows herbaria to provide or edit their records and offers support for herbaria to provide updates via email or other channels. Other communities such as national portals have other pathways for collections to provide or update information. Several tools help data publishers to create EML metadata for publishing data to GBIF and elsewhere. These could evolve to deliver collection records in preferred formats. The Integrated Publishing Toolkit (IPT) [https://www.gbif.org/ipt] could be enhanced to offer collection

records as one of the core record types that can be shared. This would allow collections either to publish one or more collection records as a small standalone dataset or collection networks to manage and publish a dataset comprising many collection records. Wikidata could also serve as a tool or platform for editing catalogue information and making it widely accessible and reusable.

Q16. Which existing tools, databases and websites can help to mobilise and maintain collection records? Is it possible to identify additional tools or pathways that need to be developed or supported?

3.2. Community catalogues

IH is the best established catalogue servicing a large community of collections, but many other communities are important, including regionally or nationally focused efforts, such as CETAF's institutional profiles, the web portals of iDigBio [https://www.idigbio.org/portal/collections] and the ALA, and the One World Collection initiative [https://biss.pensoft.net/article/38772/], and thematically aligned efforts, such as the World Directory of Culture Collections [http://www.wfcc.info/ccinfo/index.php/home/content] and the Global Genome Biodiversity Network portal [http://www.ggbn.org/ggbn_portal/members/index]. A comprehensive global catalogue should ensure that the needs of these different communities are met and support their continued operation and independence wherever is valued by collections. Understanding these requirements is essential in planning the technical implementation and governance of the catalogue.

Q17. What catalogues already address the needs of some communities of collections? How can an integrated catalogue support these communities? Which communities require a separately branded identity and/or platform? What is the best way to include these communities as part of an interconnected solution? Is there a role for content to be created and improved by a wider audience (e.g. through Wikidata)?

3.3. Integrated catalogue

GBIF has the mission to provide global-scale support for biodiversity informatics solutions and has expanded its Registry to host the data historically maintained as GRSciColl. GRSciColl content is incomplete and is best seen as a framework for expansion with richer collection metadata that properly represents the needs and interests of collections. GBIF can serve as the context for integration and deduplication of collection information from different sources and for interlinking this information for other biodiversity data. GBIF requires guidance on the best way to support the needs and branding of collections and their communities as it develops such services.

Q18. Are there issues with GBIF providing hosting and support for the global catalogue through its Registry? What is required to ensure that this meets the needs of collections and is fully adopted and owned by the collections community? What challenges need to be addressed to minimise duplication of content and effort within an integrated catalogue?

3.4. Collection management systems

Most natural history collections maintain data on their specimens in a collection management system (CMS) such as Specify [https://www.sustain.specifysoftware.org/about/], Symbiota [http://symbiota.org/docs/], EMu [https://emu.axiell.com/], DarWIN [https://biss.pensoft.net/article/39054/] or BRAHMS [https://dps007.plants.ox.ac.uk/bol/]. Some of these tools could develop to interface directly with the collection catalogue, providing up-to-date metadata and metrics.

Q19. What present or future requirements are there for interfaces directly between CMS platforms and the collection catalogue? Are there special opportunities that should be considered? Could CMS platforms become a source of metadata for institutional collections within a global catalogue?

3.5. Interfaces, APIs and client modules

The value of a shared commons-based resource can be maximised by ensuring that interfaces and APIs support the needs of all key stakeholder groups, including addressing issues around content delivery to the fullest extent possible in multiple languages. Some needs may be addressed by offering reusable client components that can be embedded in other applications.

Q20. What interfaces and APIs are required to maximise access to the collection catalogue? How can the catalogue best support diverse user communities, including speakers of different languages?

4. Governance of the catalogue

Standards and tools are only one part of the solution. For the catalogue to succeed and provide value, it must be accepted by and deliver value to the stakeholders it represents, in particular the collection-holding institutions and the communities that support collections. It is important to identify the stakeholders that need ownership for each aspect of the collection building and to understand how they can be enabled, empowered, and resources to take on these responsibilities. Mechanisms are also needed to deal with situations in which needs or interests may come into conflict.

4.1. Ownership of information for each collection

The basic assumption is that each institution should have primary responsibility and control for information on its collections. However, it may be appropriate to delegate full or partial responsibility to thematic, regional or national communities that have data curators able to ensure the quality and standardisation of collection records. In some contexts, where institutions have for any reason not provided authoritative information, or do not have the resources to do so, there may be reason to allow or encourage a wider user base to contribute and improve collection records. In all cases, a version history is required for the information, so that users can understand and respond to changes

made by others.

Q21. How should ownership and access control for collection records be managed? How should appropriate editors be recognised and validated? Are there situations where automated or human intervention will be required to resolve disagreements or discrepancies?

4.2. Communities of practice

Communities such as IH, CETAF, ALA, iDigBio, etc. play an important role supporting collections and promoting standards-based practices. In many cases, these communities have a high level of understanding and participate closely in the development of biodiversity informatics solutions. Their roles and rights need to be well defined and supported in any integrated solution.

Q22. What do these communities require to be able to carry out their work efficiently and support their collections? How can an integrated approach enhance their offerings? What risks need to be addressed?

4.3. Technical infrastructures

Biodiversity information infrastructures such as GBIF, DiSSCo, iDigBio and other national and regional platforms are usually funded in the context of broader open science goals for research infrastructures. Their participation can provide an important bridge between the needs of the collection communities and funding and expertise for informatics solutions. Roles and responsibilities must however be well defined to ensure that the needs of researchers and user communities are central. It is important to define clearly how these technical infrastructures can best participate in the overall solution, including demonstrating the benefits required to secure sustained funding for an integrated catalogue and for all the component parts.

Q23. What technical infrastructures need to be engaged as part of the solution? How are their roles and needs best balanced with those of the collections and of their communities?

4.4. Governance arrangements

A complex, commons-based solution will depend for its long-term success on a governance model that provides confidence to all parties that their interests are served and protected. The model should find the right balance between ensuring the health of the collaboration and minimising associated overheads in terms of meetings, reporting, etc.

Q24. Are there appropriate models that can be adopted or expanded to support the governance of this catalogue? Can it be managed in the context of an existing organisation or institution?

4.5. Incentives for contributors

Relatively little effort may be required for each institution to register and manage its own collection records. However, the stability of the system will depend on continued effort from these institutions or from other parties to correct errors and outdated information. There should be clear benefits or incentives encouraging stakeholders to contribute this effort. A key goal should be to ensure that the catalogue contributes usefully for the work of collection managers and taxonomists. Acknowledgement of contributions may also be valuable.

Q25. What are the incentives for different contributors to maintain information in the catalogue? How can these be maximised?

4.6. Funding and sustainability

Funding needs will depend on other aspects of the approach adopted to build the catalogue. Costs will be higher if more central support is required to maintain the content. Even if the content is largely managed for free by the international community, sustaining a reliable infrastructure requires effort and long-term investment (see for example the CoreTrustSeal model [https://www.coretrustseal.org/] for trusted repositories).

Q26. How can the governance and technical aspects be funded? Is external funding likely? What other models may be feasible (contributions from collections, inclusion within the funded mission for GBIF or some other host)?