

TOWARDS A ROADMAP FOR ADVANCING THE CATALOGUE OF
THE WORLD'S NATURAL HISTORY COLLECTIONS
DONALD HOBERN, SARAH VINCENT, LAURENCE LIVERMORE,
TIM ROBERTSON, JOSEPH T.MILLER, AND QUENTIN GROOM

Grant Agreement Number | 823827

Acronym | SYNTHESYS PLUS

Call | H2020-INFRAIA-2018-2020

Start date | 01/02/2019

Duration | 48 months

Work Package | 5: NA5

Work Package Lead | H Goodson, GBIF

Delivery date | 19.04.2022



SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Towards a Roadmap for Advancing the Catalogue of the World's Natural History Collections

Authors: Donald Hobern¹, Sarah Vincent², Laurence Livermore², Tim Robertson¹, Joseph T. Miller¹, and Quentin Groom³

¹ Global Biodiversity Information Facility, Copenhagen, Denmark

² The Natural History Museum, London, United Kingdom

³ Meise Botanic Garden, Meise, Belgium

1. Introduction

The creation of a collections catalogue is central to the shared vision and goals of a large number of institutions, projects, initiatives, and other stakeholders within the natural history and wider science collections landscape. However, the number and diversity of interested parties brings with it key challenges around unification of approach, interoperability of already developed and widely used systems, and the differing requirements of a wide range of user groups.

Information about these 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, showcasing the existence and importance of museums and herbaria and supplying context to available data on specimens [Page, 2016]. These records also potentially open new avenues for fresh use of these collections and for accelerating their full availability online.

There is currently no definitive estimate of the number of specimens held by collections globally, with estimates ranging between 1.2 and 2.1 billion [Ariño, 2010]. A catalogue would go some way to narrowing estimates, which in turn would provide an opportunity to gauge the economic value of collections and collection-based services [Hobern et al. 2020]. Generating value estimates could provide leverage for further funding opportunities by providing a qualitative measure of importance. Although some collection types do not easily lend themselves to valuation, alternative metrics could also be considered and utilised to showcase societal as well as economic importance.

1.1 Community consultation process

The basis for this report is the ideas paper [Hobern et al. 2020] and the subsequent community consultation on the topic [Advancing the Catalogue of the World's Natural History Collections](#), held virtually by the *alliance for biodiversity knowledge* in March and April 2020. The consultation and resulting roadmap form part of Task 5.1 under SYNTHESYS+ (Synthesys of Systematic Resources), “*Developing implementation roadmaps for priority infrastructure areas as part of cooperative RI for biodiversity*” [Smith et al, 2019]. The ideas paper was first issued for review on the 25th February 2020 and was subject to initial review and revision by task participants. Stakeholders were gathered through existing project and community networks¹, and the online consultation format endeavoured to facilitate wider participation by minimising the need for travel and allowing contributions to be collected over an extended period of time, removing time zones, travel costs and other commitments as a barrier. All materials were uploaded to the GBIF Community Forum site on 6 April 2020, and discussion threads were opened to the public from 17 – 29 April. Stakeholders were asked to consider questions relating to the 26 topic areas outlined in the ideas paper, grouped under four key categories. To remove language barriers, dedicated threads with translated

¹ GBIF website and associated communication channels (social media, mailing lists to all node managers, newsletter etc), the Alliance For Biodiversity mailing lists, SYNTHESYS+ mailing list and TDWG communication channels.

summaries were available in Spanish, French and Chinese. In addition to the ideas paper, presentations were contributed by organisations and interest groups to give participants a clear idea of the current collections information landscape, and demonstrate current tools and activities that may inform or form part of the development of a future catalogue. The timeline of the consultation process is shown in here ([ref](#)).

This paper will use the outcomes of this consultation to identify common themes, priorities, areas of consensus, and areas of dispute. These will be used to propose a vision for how a global collections catalogue may be developed, covering use cases, information, maintenance, resourcing and sustainability.

1.2 Articulating the need

The ideas paper outlines a range of potential use cases based on those collected by the [TDWG](#) Collection Description Interest Group², as well as work done by [ICEDIG](#) in preparation for [DiSSCo](#) [Hobern et al. 2020 | van Egmond et al, 2019]. These collectively illustrate the potentially extensive value and the benefits that could be offered by a global collections catalogue, whilst also highlighting the difficulties in adequately scoping the catalogue to fit the needs of a large and varied user community.

Four broad headings are described by the ideas paper:

1. Uses for the catalogue
2. Information in the catalogue
3. Technology for the catalogue
4. Governance of the catalogue

2. Current landscape

The sections in this landscape overview are based on the contributed materials for the community consultation supplemented with additional research to give an overview of the key platforms and databases, collections management systems, data standards and other community activity. The aim has been to provide background information for readers but not comprehensively cover the current landscape.

2.1 Platforms and databases

A number of existing catalogues for institution, collection and specimen-level information are already in use or development, driven by several community-driven initiatives and projects. There are other broader sources of information that could be integrated or used in a future platform. To prevent record duplication and minimise the level of resource required to create collection catalogue records, the scope, controlled vocabularies and preferred identification schema of the most relevant systems should be investigated and incorporated during development of the collection catalogue data architecture.

² https://github.com/tdwg/cd/tree/master/reference/use_cases

Atlas of Living Australia Natural History Collections - the ALA [Natural History Collections](#) page (formerly known informally as the “Collectory”) is an example of a national information resource on natural history collections. ALA has a high calibre informatics and software development team and receives strong institutional support and engagement on the national level. Collectory records do not currently use a standard vocabulary and the repository is struggling to de-duplicate collection-level records contributed for different views of the same collection [Atlas of Living Australia, 2020, Belbin et al., 2021].

CETAF Collections Registry/CETAF passports - the Consortium of European Taxonomic Facilities (CETAF) provide a central source for information about its 63 European member organisations. ‘CETAF passports’ are contributed as a condition of membership and include high-level categorisation of collections including non-mandatory collection size metrics. CETAF is currently building on the functionality of CETAF passports with the development of the CETAF Collections Registry and has proposed assigning unique institutional acronyms to each member, which may cause some overlap/conflict with existing UIDs [Semal et. al, 2019].

The Global Registry of Scientific Collections (GRSciColl, including GRBio) was initially developed as a global ‘clearing house’ of information for institutions and collections before being incorporated by GBIF in 2019. GRBio held information on biodiversity collections and was a subset of GRSciColl which is open to all categories of scientific collection. Although its content is currently incomplete, GRSciColl is considered a viable framework for expansion and is currently in a new phase of development. So far synchronisation has been established with Index Herbariorum (see below) and content from the iDigBio collection database has been integrated, with GRSciColl now powering the [iDigBio collection portal](#). GBIF are now actively developing the [codebase](#) with a focus on the role based authentication model necessary to enable wider contributions. The [key priorities](#) going forward are: reduction in record duplication, improved documentation, pooling of a greater resource of editors, allowing anyone to propose changes, establishing a master data management solution, and improving the UI. [GBIF Secretariat 2021]

iDigBio web portal - iDigBio is the US national resource for digitised information about vouchered natural history collections. The iDigBio specimen portal makes available millions of records from neontological and paleontological specimens curated at museums and other institutions in the US. The data held in their repository follows the Darwin core and Audubon core data standards and iDigBio has contributed upwards of 1.5k collection-level records to GRSciColl to date [iDigBio, 2021]

Index Herbariorum – is the most successful and established collections catalogue, covering the world’s botanical collections. Herbaria can provide/edit their records and updates can be provided through email or other channels. Existing tools such as the Integrated Publishing Toolkit which currently facilitate the creation of EML metadata could be developed for the delivery of collection records. It has already been integrated into GRSciColl.

Wikidata is already recognised as an identifier broker with potential to advance biodiversity knowledge graph development [Sachs et al, 2019] and is already being used by successful community initiatives like [Bionomia](#) [Shorthouse, D. 2020]. It could be used to semantically link people, taxa, places, collections, institutions and more.

The Global Research Identifier Database (GRID) and the **Research Organisation Registry (ROR)** are existing databases of globally unique persistent identifiers and associated metadata for education and research-related organisations across all disciplines. Each service holds data on more than 100,000 organisations [ROR, 2021; GRID 2021], and their identifiers are interoperable. GRID is a commercial product managed and owned by Digital Science. GRID provided the seed data for ROR, which is a community-led initiative. These databases could potentially be used as a starting point for institutional identifiers.

2.2 Collections management systems

Collections Management Systems (CMS) are used to organise, control and manage natural history collections. They are used for a range of purposes including inventory management, creation and publication of descriptive specimen and collection metadata, risk management, collection conservation and assessment, exhibition management, loans and research requests, and as stores of legal information regarding the acquisition and use of collections.

Collections management systems are likely to be one of the fundamental sources of natural history collections data but they pose a number of challenges to creating a catalogue of collections. There is a great variety of different systems in use. A survey of European collections conducted by DiSSCo [Casino et al, 2017] identified over 37 different “systems” ranging from in-house (e.g. Kotka, PlutoF), general database management systems (e.g. Microsoft Access, Filemaker), commercial (e.g. Adlib, ActiMuseo, EMu) and open source solutions (e.g. DINA, Koha, Specify). Some respondents did not use any sort of CMS and stored their collections data in spreadsheets or text documents. The number of systems in use across the world will be even greater although some are more frequently used in particular countries (e.g. Specify in North America) and for particular taxonomic groups (e.g. BRAHMS for plants).

There are no studies evaluating these various CMS as a source of broad collections metadata. CMS interoperability has been studied at a limited scale with a focus on specimen/observation data [Dillen et al, 2019] with the conclusion that we are far from being able to seamlessly import and export data between different CMS.

2.3 Data standards and interoperability

The standards discussed below are highly relevant to the collection catalogue problem space. Within the scope of this report, they should be considered opportunities to promote compatibility with collection-level data standards, avoid duplication of effort and proliferation of competing standards and as potential sources for collection-level data enhancement and standardisation.

Standard	Description	More information
Darwin Core (DwC)	Darwin Core is the most widely used standard for sharing biodiversity observation and specimen data. It	Wieczorek et al. 2012

	builds on existing metadata standards (like Dublin Core) and is supported by the majority of specimen-level data repositories and community tools/platforms.	
ABCD	The Access to Biological Collections Data (ABCD) Schema is also standard to Darwin Core for specimens. ABCD is a comprehensive, complex, structured standard for biodiversity data.	(ref) Access to Biological Collection Data task group. 2007; Fichtmueller et al. 2019
ABCDEFG	ABCDEFG (Access to Biological Collection Databases Extended for Geosciences) is an extension to ABCD developed to support palaeontological, mineralogical and geological digitized collection data.	(ref) Petersen et al. 2018
TDWG Attribution project	A collaboration between TDWG and the Research Data Alliance to enhance existing and create new standards for giving attribution for the maintenance, curation, and digitization of physical and digital objects with a special emphasis on biodiversity collections.	(ref) Thessen et al. 2019
AudubonCore	Audubon Core (AC) is a set of vocabularies designed to represent metadata for biodiversity multimedia resources/collections. The vocabularies address such concerns as management of media and collections, descriptions of content, taxonomic, geographic, and temporal coverage, and appropriate ways to retrieve, attribute and reproduce them.	(ref) Morris et al. 2013
Natural Collections Descriptions (NCDs)	The NCD standard arose from an earlier TDWG attempt to define a collection-level data standard. NCDs are actively used by several platforms outlined in 2.1., but subsequent development efforts stalled and as a result this standard has not been more widely taken up. The TDWG CD model (see below) is acknowledged as the natural successor/continuation of the NCD standard.	(ref) Natural Collections Descriptions interest group 2008
TDWG Collection Descriptions (CDs)	Building on earlier work in the NCD standard, the TDWG Collection Descriptions (CD) data standard will define a set of classes and properties that can be used to represent groups of collection objects and their associated information. These incorporate common characteristics used to describe, group and break down collections, metrics for quantifying those collections, and properties such as persistent identifiers for tracking collections and managing their digital counterparts.	[Woodburn et al. 2020] (ref)

	Coupled with flexible underlying data models, the CD standard is intended to support use cases from simple, high-level collections summaries to detailed quantitative collection breakdowns and assessments.	
--	--	--

2.4 Community activity and stewardship

GRSciColl

Edits to the GRSciColl catalogue are currently made by GBIF Secretariat data managers, the iDigBio data managers, changes originating from Index Herbariorum and individuals from institutions and national nodes within the GBIF network. These edits are coordinated by the GBIF Secretariat.

3. Community Priorities

This section presents the community's priorities for a collection-level catalogue by summarising notable areas of consensus and concerns that emerged during the consultation process. We have followed the four high level categories (Use, Information, Technology, Governance) and their 25 subcategories in the community forum. There are a few instances where we have referenced comments from other subcategories if the responses were off topic or formed a more cohesive summary as part of another subcategory. Where applicable there is a link (ref) to original text in the GBIF Community Forum. Integrated summaries of all forum threads can be found [here](#).

3.1 Use

3.1.1 [Directory to support the collections community](#)

By establishing natural history collections as a global scientific infrastructure we make it easier to foster new collaborations, resource research, fund opportunities and support sustainable data infrastructure. By standardising our institutional acronyms and the collections held within them, we improve collection discoverability and citability, making it easier to demonstrate impact and importance ([ref](#)). We can make use of existing persistent identifiers (PIDs) in [GRID](#) or [ROR](#), so we are not establishing a set of new PIDs and benefit from integration and re-use [Addink 2020] ([ref](#)).

There are many collections that are mostly invisible due to the predominant specimen-based approach to digitisation and its higher cost compared to collections-level data when sharing digital collections ([ref](#)). We recognise that understanding and serving the needs of different users will be important and that keeping the collections data up-to-date will be a challenge ([ref](#)).

3.1.2 [Locating specimens and genetic materials](#)

A catalogue of summary information about the items in collections would be a highly useful resource, if the summary information was relevant, reliable and could be kept up-to-date.

Previous initiatives relating to the creation and aggregation of collection-level catalogue records have led to increased use of and interest in items in the collection ([ref](#)). Summary collection information acts as a 'signpost' for end-users to help them narrow down where items of interest may be held in the world's collections and lays the groundwork for further investigation and communication with collection managers. Collection-level records would also facilitate development of key networks and linkages between specimen data and existing, related data silos such as International Nucleotide Sequence Database Collaboration (INSDC) databases ([ref](#)), providing additional points of discovery and entry for underserved or non-traditional users.

The minimum level of information that would be needed for useful collection records is likely to vary across disciplines, user groups and geopolitical contexts ([ref](#)). There was a general consensus that mandatory fields should include taxonomy, holding institution and collection scale metrics ([ref](#)). These could be augmented by non-mandatory fields to allow additional data to be shared where available ([ref](#)).

3.1.3 [First step towards databasing collections](#)

We need to give guidance and support to the community, especially collections staff. This includes providing good tools and tutorials to curate, update and disambiguate collections data ([ref](#)). The community will need region-specific roadmaps and strategies as levels of support and incentivisation vary ([ref](#)). Current emphasis on publishing specimen records lessens potential data sharing of less well-resourced collections that are effectively excluded. The GBIF dataset classes offer a hierarchy of complexity and can serve as a stepwise path towards a goal of specimen digitization of a collection.

Publishing a metadata-only dataset could be sufficient to advertise a collection and information about its holdings. The collection would become Findable, even if not digitally Accessible, Interoperable or Reusable.

If a collection is then in a position to add a checklist dataset summarising species held - this was quite a common category of web page 15 years ago - the collection could be listed in simple ways on GBIF species pages, again further raising its profile for wider access and use. This adds some Interoperability. Databasing as DwC specimen data then takes things forward and allows for full FAIRness.

3.1.4 [Assessing the scale and value of collections](#)

Estimates of collection size are already widely held and used by collection-holding institutions, but these metrics are decentralised and typically provide little information on the assessment methodology used.

High-level estimates of collection size would be useful to external stakeholders such as government agencies ([ref](#)). Collection size estimates can be used to represent the 'value' of

collections on the national and global scale and would be invaluable in helping the community to 'build funding cases, show current (often national) capacity, and highlight gaps' ([ref](#)) [Leggatt, 2019 | Council of Australasian Museum Directors, 2018].

To be useful, such estimates would need to be either comparable/developed under a shared methodology (e.g., the One World Collection project) ([ref](#)), or contain sufficient methodological information to allow users to assess the suitability of the record for inclusion in comparison or aggregation operations made against the catalogue data and/or exclude certain methodologies from these operations ([ref](#)). The former approach would make the catalogue easier to use, but the latter would facilitate data collection and re-use of existing information.

Standardised methodologies for valuing collections based on scale and scope are already in active use ([ref](#)), but there is risk attached to following a single dogma in this respect: the value of collections will ultimately depend on the requirements of those seeking them ([ref](#)).

3.1.5 [Increased value for data on specimens, taxonomic publications, etc.](#)

We recognise better linkage of collections metadata, such as major collectors, with other external identifiers and authorities like [ORCID](#), [Wikidata](#), and [VIAF](#) will improve discoverability both inside and outside our community ([ref](#)). If we could combine collector information with taxa present in collections (e.g. at a checklist as opposed to a specimen/occurrence level) we would have a better understanding of what makes a collection unique ([ref](#)). Sharing more detailed preservation methods including ambient environmental conditions, including for non-molecular collections, is important for collections users ([ref](#)).

3.1.6 [Reducing duplication of effort](#)

A large amount of information about collections is already available on institutional websites, but this requires effort to pull together and maintain. The provision of a template or other pro-forma data collection mechanism to let collection managers update summary data quickly and easily would be helpful ([ref](#)). Some institutions already record curatorial assessments for their collections; mutual support of these assessments in a world collections catalogue and capturing additional information required at the same time would be beneficial ([ref](#)).

Providing reusable collections data and standardised institution and collection names would reduce the overhead on other specialised collection catalogues like the [Global Genome Biodiversity Network](#) (GGBN). They currently maintain their own general collections registry and could instead focus more time and community effort on collections biobank metadata ([ref](#)).

There are recent discipline-based examples of assessing states of collections [Cobb et al, 2019 | Sierwald et al 2018] including and creating a U.S. fish collection catalogue using a programmatic approach but this is still relatively high effort and misses significant collections

[Singer, Love & Page 2018] ([ref](#)). This programmatic approach could be applied to other disciplines but would need community support to generalise the code ([ref](#)).

3.1.7 [Foundation for new and enriched services](#)

A collections catalogue would make us more findable and accessible to new audiences and users ([ref](#)). Without an awareness of what resources are available in our collections or clear channels to contact the collection managers, potential funders will overlook our holdings ([ref](#)). We should be making our biodiversity information more available to environmental managers, policy makers and other government agencies ([ref](#)). Our collections have a role as part of cultural heritage within the wider arts and humanities research community ([ref](#)). Some of the person linkages required are described in section 3.1.5.

Cooperation with other initiatives like the International Nucleotide Sequence Database Collaboration (INSDC) is crucial in linking sequences that lack references to collections and corresponding voucher specimens and samples. Building tools to help researchers submit better metadata is important ([ref](#)).

When considering new and enriched services we should also be mindful of focus, delivery and utility. While new downstream use is important we should consider focusing narrowly on what queries the catalogue can support best in the short to medium term and that correspond to a sufficiently important audience (e.g. large, high impact, well-resourced, etc) ([ref](#)). We can look at other adjacent sectors for analogue data infrastructures and what makes their core services successful [Leonelli, 2013].

3.1.8 [Improvements to citation and visibility for collections](#)

Research value is primarily measured in terms of visibility and impacts from published literature. To be recognised by such measures, the citation and attribution of natural history collections needs to be agreed and standardised across the community and made visible and useful to stakeholder groups such as publishers, funding bodies and data aggregators. ([ref](#))

Understanding the community's existing practices and data quality issues in this area is key to successfully developing the collection catalogue so that citation of collection-level records is sustainable, measurable and more fit-for-purpose than current practices ([ref](#)). Outcomes from this analysis, such as comprehensive lookup tables of identifiers used for particular collections or institutions (even if not unique within the wider sector) ([ref](#)), could improve discoverability of collections from the end-user perspective and feed back into current initiatives around unlocking the historic scholarly record ([ref](#)) and aid in the discovery and embedding of linkages between related outputs ([ref](#)).

Previous initiatives around standardising citation and attribution have stalled due to lack of uptake ([ref](#)); a critical mass of adopters is required before stakeholders outside of the core community (e.g., publishers and aggregators/content banks) will change their working practices to incorporate a particular standard. Additional barriers to user uptake include a lack of guidance around citation/attribution practices for both collection users and collection-

holding institutions (ref) and uncertainty around proper citing procedure for collection data from aggregators and other secondary sources (ref). It may be too difficult to get authors to consistently use a standard abbreviation. It might be easier to simply link multiple abbreviations to a single, stable PID (ref).

Engagement may be encouraged via links with other data repositories, especially those with established infrastructure and dataflows related to the identification and resolution of research citations. ROR, GBIF and Wikipedia, for example, already integrate with Datacite and Crossref (ref), both of which provide impact metrics that would incentivize both contribution to the catalogue and adherence to related standard citation practices (ref).

3.1.9 [Support for national and regional needs and applications](#)

One of the biggest issues we face is demonstrating the role and value of collections (value is covered in more detail in section 3.1.4). This is often a national challenge because this is where the funding lies, but on occasion becomes a continental or global challenge (ref). A more integrated model of the natural world, founded on observations and collections, would provide evidence to see where we are deficient in data, and to identify which organisations might coordinate to fill these gaps at a national or regional level (ref).

Uniqueness of collections can help focus prioritisation for digitisation (and other activities) at a national and regional level. It can act as a starting point for understanding how to effectively collaborate and pool resources (ref).

In other sections it was noted that some countries have minimal online catalogues, or resources in different languages that could make them less internationally discoverable (ref). National legislation can play an important role in incentivising data sharing and coordinating national activities (e.g. the *Registro nacional de Colecciones in Colombia*) (ref). This may be an example other countries or research councils could adopt.

3.1 Recommendations

- A collection catalogue should mandate a minimum number of standard fields such as: taxonomy, holding institution and collection scale metrics which could be augmented with additional fields where available.
- Strong guidance and support materials must be available to the community to support the catalogue.
- Continued methodological standardization that maintains flexibility needs to be continued.
- Collection records should maintain linkages with other external identifiers and authorities.
- Collections catalogue should be built so that it can be a national resource.

3.2 Information

3.2.1 [Scope for the catalogue and definition of “collection”](#)

The definition of a natural history collection is broad and perhaps delimited more by the goals and uses of the collection than its contents. At its core a natural history collection represents biological and geological diversity on earth, but it does not exclude objects such as extraterrestrial geology, nor anthropological artifacts. Even living collections, whether in an active or dormant state, can be included. Furthermore, the collection objects themselves are not necessarily items of biological and geological diversity, but include associated materials, such as field notebooks, photographs and ethnobotanical objects. Collections can be eclectic or have a specific focus and *raison d'être*, such as a xylaria.

Yet, there is even an array of different goals for a collection. Some are purely taxonomic, but there are others that focus on education, history, material science etc. Even the regulations, management and ethical considerations do not unify the group. Living collections, human remains and objects of cultural significance have specific requirements that must be considered specifically. One cannot even state that a collection should exist, as collections that have been destroyed or been divided up still need to be identifiable even if their original contents are unknown.

In some cases collections are defined at the institutional level, as is true for most herbaria listed in Index Herbariorum. However, there are many other collections where the material is divided by curatorial practices, perhaps taxonomically or by the collections origins. Ultimately it may be easier to define a collection in the Catalogue retrospectively by the need for them to be listed within the Catalogue and by the advantages they gain from being listed. Collections often map to organisational structure and to curatorial approaches rather than adhering to consistent definitions. This conflation of institutional structure with institutional collection(s) is too frequent to have occurred by chance; it seems reasonable to assume that operational concerns and priorities (e.g., naming/defining a collection to reflect acquisition or provenance events) play a key role in shaping the community conception of a ‘collection’.

Differentiating a “natural history” collection is important but we need to consider and link to collections that are often treated as adjunct collections (archives, field notebooks, registers, photographic collections) and born digital collections (e.g. sound records, camera trap images) which may be discrete collections in their own right (see [Section 3.2.5](#)).

There was broad consensus that the scope should be broad and inclusive, including all natural science collections that are useful for natural science, natural history or natural heritage. This includes xylaria, paleontological collections and anthropological collections. Some of these collections will have sensitivity and legal restrictions that need managing when sharing their descriptions.

3.2.2 [Identifiers for collections](#)

Multiple collection identification schemes exist and are actively used, often in parallel - a situation which reflects the flexible definition of a collection as discussed earlier ([ref](#)). A

number of identifier schemes are provided by/derived from data platforms and services: GRSciColl, ROR, ALA Collectory and the GBIF Registry ([ref](#)). Identifiers for an organisation or unit within an organisation have also been widely adopted as a shorthand to refer to the collections they hold, even if the original entity no longer exists in an operational sense ([ref](#)).

It may be the case that only these more traditionally conceived-of collections (e.g. a specific herbarium) need a human-readable identifier due to their historical use in previous and current registries ([ref](#)). We need to avoid conflating the purpose of and requirements for human and machine-readable identifiers: machine-readable identifiers need to be globally unique, persistent and resolvable. They should provide unambiguous identification of a collection -- even if the contents or environment of the collection changes over time -- and facilitate wider data linkages. Human identifiers need to be succinct, descriptive, memorable and, if not unique and persistent, flagged clearly enough to enable software systems to distinguish and accommodate this ([ref](#)).

One approach to selecting a particular identification scheme or range of schemes would be to prioritise those that correlate most closely with the class(es) of collection definition schemes used by the catalogue ([ref](#)). It would also be prudent to prioritise identification schemes on their technical capacity, accessibility, underlying infrastructure and accompanying data services ([ref](#)).

Usage of preferred identifiers could be promoted by the development of resources and activities focused on community engagement and increasing the wider awareness of the benefits and availability of the selected schemes ([ref](#)).

3.2.3 [Hierarchical collection structures and subcollections](#)

Hierarchical relationship structures would be useful for collections which have changed ownership or location over the course of their lifespan. For example, a subcollection record could be linked to a 'parent' collection record to reflect provenance and facilitate discovery ([ref](#)). Hierarchies are less suitable for use in scenarios where a single collection object falls under the scope of several different collections ([ref](#)). This scenario is common and, unless carefully handled, could lead to double-counting and inflation of collection size metrics.

A non-hierarchical system could be a suitable grouping mechanism for more traditional parent/child relationship use-cases if relevant conceptual entities (e.g. institution, collection, dataset) were defined, standardized and incorporated during its design ([ref](#)). The development of distinct classes of collection-level records and relationships in this fashion would also be key to safely representing non-hierarchical relationships between collections while enabling aggregators and other platforms to logically constrain the operations that can be applied to different classes of catalogue record ([ref](#)). The nature and scope of each class of collection record needs to be communicable to end-users to allow for different search strategies based on their data requirements ([ref](#)).

3.2.4 [Description of a collection](#)

The usage of the [TDWG CD standard for collection descriptions](#) for the collection catalogue is broadly supported by the community: the only additional field suggested during the consultation was one containing information on how to cite the collection ([ref](#)).

The TDWG CD model centres on a small number of mandatory fields and a larger range of optional fields. This approach allows different classes of collection description records to be described using dimensions most appropriate for the discipline, while sufficiently controlling core field data quality and integrity to allow some level of class interoperability ([ref](#)). The flexibility to describe different collections using optional, discipline-specific fields is widely seen as essential to successful uptake and use of a collection-level data standard and accompanying discovery systems and catalogues ([ref](#)).

Controlled vocabularies should be identified or developed for as many fields as is feasible ([ref](#)). Fields most urgently in need of a controlled vocabulary could be identified via analysis of existing specimen-level records containing equivalent DwC fields ([ref](#)).

Any consensus/community level collection data standard should not be considered complete until it has undergone adoption or testing in institutional data workflows and projects to ensure that it is fit-for-purpose ([ref](#)). Real-life testing and early adoption of the standard for a small set of use-cases/collection description classes would facilitate the identification and subsequent development of those fields most suited for machine access ([ref](#)).

3.2.5 [Wider data linkages](#)

Selecting a limited number of fields for use as linkages/identifiers to external services will enable discovery by non-traditional users, e.g. visitors to a wikipedia page following a citation link to the collection catalogue ([ref](#)). It will also improve the usability of the collection catalogue by allowing users to easily navigate to external, authoritative sources of information on topics associated with the specified collection ([ref](#)).

Fields selected for use in this manner need to be carefully evaluated and prioritised: creating and maintaining linkages between data silos is a non-trivial undertaking and the benefits to contributors, system providers and external data sources must be clearly defined ([ref](#)). There is general consensus that the following core fields should be explored: collector, species/taxa, specimen-level information, notable and/or primary collectors and associated publications ([ref](#)). Linkages should be bidirectional wherever feasible, taking into account each external data source's sustainability and technical capacity in areas such as link resolution, identifier integrity and reporting ([ref](#)).

Fieldwork notes and images, type specimens, and taxonomic treatments were also mentioned as possible candidates for linkage ([ref](#)), but these fields may be more appropriately and usefully associated with specimen-level records ([ref](#)). External linkages with sources that provide usage and impact metrics could be valuable mechanisms for boosting engagement. Without support and clearly defined benefits for catalogue

contributors, this may lag in existing areas of poor data-density such as south-west Asia ([ref](#)).

3.2.6 [Information services relating to collections](#)

All of the information services proposed (listed below) were recognised as adding value to the catalogue. Partnerships with existing digital repositories (e.g., CoL, GBIF, BHL) to deliver shared or complementary services would be beneficial for encouraging both development progress and collaboration within the existing ecosystem of research infrastructure services, tools and platforms ([ref](#)).

3.2 Recommendations

- A collection catalogue would be broad and inclusive to be used across many disciplines that maintain collections.
- Collection identifiers initiation must be accompanied by community engagement.
- Controlled vocabularies should be identified or developed for [TDWG CD standard for collection descriptions](#).
- Core fields should be used for linking to external data.

3.3 Technology

3.3.1. [Pathways and tools for publishing collection records](#)

Good software and infrastructure will be critical to building a global collections catalogue - its creation and maintenance is likely to be one of the more significant costs ([ref](#)). The proposed approach would use a single master record for each collection and utilise existing publishing mechanisms to keep them up-to-date ([ref](#)). We could use Wikidata as a broker between other identifier systems being mindful that it is not an authoritative source ([ref](#)). Wikidata could also allow other members of the community to enhance data about our collections and would make the collections data more discoverable.

There are national platforms that could be integrated with a global collections catalogue (Colombia's *Registro Nacional de Colecciones* and Argentina's *Sistema Nacional de Datos Biológicos*) but update frequency and data richness compared to the individual contributing collections should be evaluated ([ref](#), [ref](#)).

3.3.2. [Community catalogue](#)

There are several community catalogues that are established, well-used and will want to retain their own identity. These catalogues (like Index Herbariorum or GGBN) could act as the primary copy which are then synchronised to GRSciColl ([ref](#)). In some cases institutes themselves will maintain their own information on local systems, or get support for publishing this data at a national level (e.g. iDigBio or Atlas of Living Australia) ([ref](#)). This will require careful consideration of how to model and manage role-based access permissions for editing collections information and nominating which source(s) should be used as the

primary copy. The data standards used across the community catalogues and the global catalogue will normally be the same, but where there are differences they will be mapped to ensure they are still discoverable ([ref](#)).

There are other community initiatives that are building discipline specific catalogues (like bug-collections.org) – there should be discussions between GBIF and these communities to understand how they can contribute to or use GRSciColl functionality ([ref](#)).

3.3.3. [Integrated catalogue](#)

A successful integrated catalogue needs tools to easily customise, create, and update collections records. This will be a combination of manual and automated approaches, including tools to support the community resolve and map informal collection identifiers ([ref](#), [ref](#)).

3.3.4. [Collection management systems](#)

While CMS hold the potential to be efficient data sources for a collection catalogue, they should not be a barrier to participation: a significant proportion of organisations manage their collections data solely through spreadsheet tools ([ref](#)). The GBIF IPT goes some way to reducing similar participation barriers at the specimen level ([ref](#)), but still requires a degree of infrastructure and technical resources that should not be assumed to be widely accessible ([ref](#)). The GBIF IPT does not facilitate round tripping of data. Collection catalogue records are likely to be simpler to create and fewer in number than specimen-level records, so provision of a simple web-form could be a suitable mechanism by which CMS-free contributors can add data to the catalogue ([ref](#)).

For organisations where the CMS plays a central role in all aspects of the collection data lifecycle, the ability to manage collection-level records in the same system would have significant benefits. Inclusion of collection record management functionality would reduce double-entry of data, enable links between specimen and collection records, simplify high-level reporting, enable better tracking of digitisation progress, promote consistency between common fields and potentially drive workflows around automated enhancement of specimen level records ([ref](#)).

CMS systems could automate the creation and updating of collection-level records: both descriptive and quantitative collection metadata could be produced by aggregating specimen-level records over a limited set of dimensions ([ref](#)). Specify and Symbiota both already hold some capacity for interoperability with IPT and EML: a similar approach incorporating fields from the TDWG CD standard may be a suitable mechanism for data exchange between CMS and the collection catalogue ([ref](#)).

Elements of this architecture are already operating in GRSciColl, which automatically integrates collection-level metrics and digitisation progress indicators derived from records harvested from Index Herbariorum ([ref](#)). The MIDS (minimum information about a digital specimen) metadata standard may be an appropriate digitisation progress metric, but further

thought is required on how this could be best adapted to reflect digitisation status at the collection level ([ref](#))

3.3.5. [Interfaces, APIs and client modules](#)

A “one-size-fits-all” approach rarely works when attempting to integrate data from a variety of systems. Flexibility and agility will be important when designing the interfaces and underlying APIs ([ref](#)). The users of a global collections catalogue will have varying technical capabilities and we need to ensure participation for all, so we need to support spreadsheet uploads and web form editing. In terms of APIs and harvesting data we need to take a gradual approach at connecting, partnering and building on established infrastructures wherever possible.

Interpreting and validating data will be critical when building the global collections catalogue. Lessons from [Bionomia’s](#) implementation of an [OpenRefine reconciliation endpoint](#) would be useful in designing services. Careful consideration and potentially editing the collections model in Wikidata would allow us to more easily use Wikidata in our own reconciliation efforts and share our data more effectively ([ref](#)). We should interpret and validate the content of collection records as much as possible so it can be used as data which increases its utility and value. We will need to design and understand how we display human and machine readable operational data, metadata, standard compliance, update mechanisms and provenance ([ref](#)).

3.3 Recommendations

- A single master record for each collection is required and existing publishing mechanisms should be used to keep them up-to-date
- The existing community catalogues should retain their own identity and synchronized with the global system
- Link data from existing CMS to reflect digitisation status at the collection level
- System should display both human and machine readable data

3.4 Governance

3.4.1. [Ownership of information for each collection](#)

The starting assumption for owning and managing collections information is that each institution would be responsible for its own collections. There are conditions when responsibility and access control may be delegated to a third party where local resources are limited or non-existent ([ref](#), [ref](#)). Indigenous labels and worldviews should be included in collections descriptions where possible ([ref](#)).

Even when there are local resources we will need to encourage active maintenance through mixed approaches like training and educational outreach, how we present data to users, and how we recognise and credit editors ([ref](#), [ref](#), [ref](#)). Formally incorporating the maintenance of collections information into an organisational role(s) would be ideal but this has been challenging in the past ([ref](#)).

Although it is assumed that institutions, and by implication curators, will provide and maintain collection information there is an obvious concern that they may not be engaged with this international initiative to take ownership of their information. This may require a certain amount of training, but also incentivizing. Curators are not necessarily the beneficiaries of better collection data at a global level and without sufficient incentive will not engage.

3.4.2. [Communities of practice](#)

There are existing communities where collections (or parts of collections) are already described in multiple collection catalogues due to an overlap in scope ([ref](#)). We need to avoid duplication of effort wherever possible through integration and interoperability.

There are several examples of national organisations which may act as intermediaries, or already curate national collections data (e.g. [NatSCA's FENSCORE](#), [iDigBio](#) and [Atlas of Living Australia](#)) and could champion the global catalogue at a national level using established networks and infrastructure ([ref](#), [ref](#)).

Publishers are a significant part of existing communities of practice: they are probably one of the biggest users of collection codes and could effectively promote their use and encourage linkage. They are also a source of collections data not found elsewhere, such as private collections ([ref](#)).

Further discussions are required on how best to encourage, support and engage existing communities as they will be critical in driving voluntary contributions. The overall approach to building a global catalogue will almost certainly at some level be a mosaic that federates across different communities and services, each with their own focus and strengths ([ref](#)).

3.4.3. [Technical infrastructures](#)

Limited discussion and was covered more in 3.3 Technology.

3.4.4. [Governance arrangements](#)

This discussion was merged into 3.4.2. Communities of Practice.

3.4.5. [Incentives for contributors](#)

Promoting collections by supporting branding and rolling highlights can raise awareness of collections and act as a free advertisement. We can consider developing functionality that allows contributors to generate metrics for reporting to stakeholders and supporting funding requests, directing internal curatorial efforts, understanding the value of collections and finding potential collaborators. Formally acknowledging collections work in metrics and metadata is not always considered a positive: there are concerns that it could be used to control work (through performance management?) and that public recognition was undesirable ([ref](#)).

Financial support for training courses and potentially collections improvement projects could be provided to those who maintain their collections records. Free licenses to data (collections?) management systems and technical support for those that require it.

While not an incentive, lowering the technical barriers for editors and contributors makes participation more likely ([ref](#)). A sense of ownership is important for long-term engagement and enabling contributors rather than serving contributors is more sustainable ([ref](#)).

3.4.6. [Funding and sustainability](#)

Both the governance and technical infrastructure will require funding and support - either from formal inclusion in the GBIF mission or from elsewhere. In addition to GBIF there are regional consortiums, either nationally or international (e.g CETAF), that would benefit from a collections catalogue and have a vested interest in ensuring long-term sustainability ([ref](#)). Even with their support long-term funding will be challenging. Government agencies, including research councils, and large collections are also potential sources of funding and support ([ref](#)).

While some regions will be able to contribute staff time and potentially funding, there are areas where there are economic or legal constraints that make economic contributions unfeasible. To ensure global participation and sustainability we will need to consider how we can support less-resourced regions ([ref](#)).

Metrics and performance indicators will be needed to justify long-term support with stakeholders. Sustained growth, data quality and fitness-for-use are some of the potential metrics that need monitoring ([ref](#)).

3.4 Recommendations

- Mechanisms for outreach and training are critical for success
- Share governance through existing communities of practice
- Formal acknowledging of collections work in metrics and metadata is critical but not sufficient
- Metrics and performance indicators are needed

4. Developing and implementing a roadmap

4.1 GBIF's GRSciColl Catalogue priority roadmap 2021

This roadmap builds on the previous work of GRSciColl Catalogue that connected Index Herbariorum, imported the iDigBio content and linked GBIF occurrence records to the entities in GRSciColl. The roadmap identifies six key priorities to progress.

Reduce the amount of duplicate records

The connection with Index Herbariorum and import of iDigBio enriched the catalogue, but also increased the number of duplicate entities that can't be automatically handled. This will be addressed by:

- Documenting guidelines on how a data manager can resolve duplicate issues [[REG-316](#)]. The guidelines will provide example scenarios, explain the recommended approach to defining codes and explain the implications on external systems (see master data management below).
- Develop tools that help identify potential duplicates alerting them to managers [[REG-191](#)]

Allow anyone to propose changes

The current processes are weak, and don't capture the proposed change in a structured manner.

- Develop an interface allowing anyone to propose a change to any/all fields and state whether they have authority to approve it. Changes are then to be reviewed and applied by the editorial team [[REG-CONSOLE-376](#)].

Improve documentation

- Document the technical aspects of the system focusing on the data model [[REG-317](#)], authorization rules [[REG-310](#)] and the details around master data management (see below).
- Document the guidelines for data editors including the decision process of merging entities and assigning IDs and codes [[DP-3](#)] [[REG-316](#)].

Grow the pool of editors

- Present the system at the global nodes meeting and openly invite node managers to assign staff
 - Identify specific tasks we would ask them to do, arranging into a TODO list so it is clear for contributors and community involvement can easily be measured.
- Review the authorization rules to ensure that editors can be granted access to work on only those areas they are responsible for [[REG-310](#)]

Define and implement the master data management solution

There are potentially multiple sources of truth for the metadata in the catalogue which needs to be resolved; a problem known as [master data management](#). For example we have information available in a dataset metadata description, an existing GRSciColl entry and an Index Herbariorum record.

- Define, implement and document the approach taken by the catalogue for handling differing views of metadata [[REG-319](#)]

- An approach could be as follows:
 - For each institution and collection entry in the catalogue, a single source of truth is identified for the key metadata (title, description etc). This may be one of:
 - An entry from Index Herbariorum, or other system that is automatically integrated through harvesting
 - Metadata for a dataset registered in GBIF (i.e. an EML file) [\[REG-305\]](#)
 - An entry made directly into the catalogue through the user interface, or pushed through the API by an application (e.g. a collection management system)
 - The core metadata is never changed in GRSciColl for externally sourced entities, and edits must be applied in the system providing the master record.
 - The entries in GRSciColl may be enriched with the following fields:
 - Additional identifiers to link to alternative views or aid discovery

Develop a richer user interface

- Implement a new user interface for the GrSciColl based on the [visual concepts](#) including:
 - Institution and collection search and detail pages
 - Integration of specimen-related occurrences (search, maps, gallery, detail, clustering)
 - Capability for any user to “suggest a change”
- Explore citation tracking based on data mediated through GBIF for GRSciColl institutions and collections. For more information, follow the discussion here [\[REG-323\]](#).
- Launch the new site considering aspects of branding/naming with a call for institutions to review their data and clear instructions on how to suggest edits.

Areas not covered in this immediate roadmap

This is a non exhaustive list of items that are not part of our immediate priorities in 2021 but that we aim to address in the longer term (possibly starting end 2021):

- Explore adding DOIs or similar to institutions and collections [\[REG-320\]](#)
- Explore synchronization with the NCBI BioCollections [\[REG-307\]](#)
- Explore integration of the TDWG Collection Descriptors [\[REG-176\]](#)
- Explore synchronization with CETAF [\[REG-322\]](#)
- Explore integration of external identifiers [\[REG-274\]](#)
- Improve the staff entities (people added as contacts to collections and institutions) [\[REG-321\]](#)

4.2 2021 GBIF Progress report on roadmap implementation

The Secretariat created a road map in 2021 for the GBIF Registry of Scientific Collections (GRSciColl), focusing on the development necessary to allow wider external contribution, and to mature the processes around editing. Editors may now be given scoped responsibility at institutional or national level. Induction webinars have been held with several nodes, resulting in a global team of 45 editors and 12 mediators actively curating content in the registry as of July 2021. Training videos are being developed.

The iDigBio collection catalogue is now powered by GRSciColl, through its open APIs. iDigBio data managers edit directly through the online editing interface.

Documentation for editors is in progress and can be expected in Q3 2021. All GRSciColl fields are associated with an English-language description available in the online forms. Capabilities for anyone to suggest a change were deployed in May. Proposed data changes are reviewed by the pool of editors and mediators before being applied.

Capability for user interface translations is set up to support multilingual content. Editors and external communities in support of the catalogue are invited to propose translations to support their work. This has already been taken up by the Society for the Preservation of Natural History Collections (SPNHC) Biodiversity Crisis Response Committee.

A service has been deployed allowing the linking of collections in GRSciColl to specimen records in GBIF. This has resulted in 134 million records being linked to GRSciColl entries. A basic data dashboard is now available for institutions and collections such as this example.

Options for a richer user interface for GRSciColl are being considered within the hosted portal framework (see Activity 1c). There is work remaining on the data model, and how to make best use of the data clustering to link related data before this can progress. It is anticipated the collections catalogue user interface will continue into 2022.

Adoption of persistent identifiers, such as Research Organization Registry (ROR) identifiers is actively being discussed within European nodes (e.g. through DiSSCo) with some nodes piloting use of ROR IDs on their entries.

Integration with the CETAF registry remains an objective, but has not yet started due to the effort required to enable external editors and focus on content issues. It is expected to extend into 2022. Piloting a profile of the TDWG Collection Descriptions to capture collection-level metadata has also been delayed and will be considered for 2023.

4.3 GBIF's GRSciColl Catalogue priority roadmap 2022

- Complete outstanding tasks to deploy an enriched GRSciColl providing search and access of collections, specimens and people
- Focus on content of GRSciColl: cleanup of existing entries and registration of new ones by promoting use and giving training and support to editors, and promoting consistent use of codes within data shared

- Seek to identify links between journal articles and collections based on the collection codes, within the framework of the EU-funded BiCIKL project
- Support user interface translations for GRSciColl
- Carried over from 2020: Explore synchronization of content with the Consortium of European Taxonomic Facilities (CETAF) Registry (under development)

Appendix

Category	Topic
Use	1.1. Directory to support the collections community
Use	1.2. Locating specimens and genetic materials 32
Use	1.3. First step towards databasing collections 38
Use	1.4. Assessing the scale and value of collections 16
Use	1.5. Increased value for data on specimens, taxonomic publications, etc. 14
Use	1.6. Reducing duplication of effort 19
Use	1.7. Foundation for new and enriched services 9
Use	1.8. Improvements to citation and visibility for collections 9
Use	1.9. Support for national and regional needs and applications 10
Information	2.1. Scope for the catalogue and definition of “collection” 21
Information	2.2. Identifiers for collections 24
Information	2.3. Hierarchical collection structures and subcollections 11
Information	2.4. Description of a collection 19
Information	2.5. Wider data linkages 11
Information	2.6. Information services relating to collections 8
Technology	3.1. Pathways and tools for publishing collection records 12
Technology	3.2. Community catalogue 11
Technology	3.3. Integrated catalogue 11
Technology	3.4. Collection management systems 24
Technology	3.5. Interfaces, APIs and client modules 8
Governance	4.1. Ownership of information for each collection 15

Governance	4.2. Communities of practice 12
Governance	4.3. Technical infrastructures 7
Governance	4.4. Governance arrangements 10
Governance	4.5. Incentives for contributors 10
Governance	4.6. Funding and sustainability 11
Language	Adelantando el Catálogo de Colecciones de Historia Natural del Mundo 29
Language	Progressant le Catalogue des Collections d'Histoire Naturelle du Monde
Language	建立《全球自然历史馆藏名录》 8
Process	Comments on this virtual consultation process 31

Table x - Discussion topics from the virtual consultation

References

Access to Biological Collection Data task group. 2007. Access to Biological Collection Data (ABCD), Version 2.06. Biodiversity Information Standards (TDWG)
<http://www.tdwg.org/standards/115>

Addink, Wouter (2020) <https://dissco.tech/2020/04/11/identifiers-for-our-institutes-grid-and-ror/>

Ariño, Arturo H. 2010. "Approaches to Estimating the Universe of Natural History Collections Data". *Biodiversity Informatics* 7 (2). <https://doi.org/10.17161/bi.v7i2.3991>

Atlas of Living Australia (2020) Atlas of Living Australia Strategy 2020-2025, Atlas of Living Australia, Publication Series No 2, Canberra, Australia, pp. 14.
<https://www.ala.org.au/app/uploads/2020/07/ALA-Strategy-2020-25-Public-June-6-2020.pdf>

Belbin L, Wallis E, Hobern D, Zenger A (2021) The Atlas of Living Australia: History, current state and future directions. *Biodiversity Data Journal* 9: e65023.
<https://doi.org/10.3897/BDJ.9.e65023>

Casino, Ana, Gödderz, Karsten, Raes, Niels, Addink, Wouter, Koureas, Dimitris, & Hutson, Anna. (2019). DiSSCo Partner Capabilities Survey 2017 [Data set]. Zenodo.
<http://doi.org/10.5281/zenodo.2653708>

Cobb NS, Gall LF, Zaspel JM, Dowdy NJ, McCabe LM, Kawahara AY. 2019. Assessment of North American arthropod collections: prospects and challenges for addressing biodiversity research. *PeerJ* 7:e8086 <https://doi.org/10.7717/peerj.8086>

Council of Australasian Museum Directors (2018) Australian framework for the valuation of public sector collections for general purpose financial reporting.
<https://camd.org.au/files/2018/11/CAMD-Collections-Valuation-Framework-1-Nov-2018.pdf>

Dillen M, Groom Q, & Hardisty A. (2019, April 30). Interoperability of Collection Management Systems. Zenodo. <http://doi.org/10.5281/zenodo.3361598>

Fichtmueller, D., Berendsohn, W. G., Droege, G., Glöckler, F., Güntsch, A., Hoffmann, J., ... & Reimeier, F. (2019). ABCD 3.0 Ready to use. Biodiversity Information Science and Standards, 3, e37214.

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>.

iDigBio (2021) iDigBio Portal. Retrieved March 2nd, 2021 from <https://www.idigbio.org/portal>

Leggatt, J. (2019) <https://www.intheblack.com/articles/2019/12/01/can-you-put-a-price-tag-on-heritage-assets> [archived on <https://web.archive.org/>]

Leonelli, S. Global data for local science: Assessing the scale of data infrastructures in biological and biomedical research. *BioSocieties* 8, 449–465 (2013).
<https://doi.org/10.1057/biosoc.2013.23> (final draft)

Morris, Robert A, Vijay Barve, Mihail Carausu, Vishwas Chavan, José Cuadra, Chris Freeland, Gregor Hagedorn, Patrick Leary, Dimitry Mozzherin, Annette Olson, Gregory Riccardi, Ivan Teage, and Greg Whitbread. 2013. "Discovery and Publishing of Primary Biodiversity Data Associated With Multimedia Resources: The Audubon Core Strategies and Approaches". *Biodiversity Informatics* 8 (2). <https://doi.org/10.17161/bi.v8i2.4117>.

Natural Collections Descriptions interest group (2008_ Natural Collections Descriptions (NCD), version 2008-08-12. Biodiversity Information Standards (TDWG)
<http://www.tdwg.org/standards/312>

Page R (2016) Towards a biodiversity knowledge graph. *Research Ideas and Outcomes* 2: e8767. <https://doi.org/10.3897/rio.2.e8767>

Petersen, M., Glöckler, F., Kiessling, W., Döring, M., Fichtmüller, D., Laphakorn, L., Baltruschat, B., and Hoffmann, J.: History and development of ABCDEFG: a data standard for geosciences, *Foss. Rec.*, 21, 47–53, <https://doi.org/10.5194/fr-21-47-2018>, 2018.

Robertson, T (2021) GRSciColl Roadmap. Retrieved: March 04, 2021.
<https://github.com/gbif/registry/blob/master/roadmap-grscicoll.md>

Sachs J, Page R, Baskauf S, Pender J, Lujan-Toro B, Macklin J, Comspon Z (2019) Training and hackathon on building biodiversity knowledge graphs. *Research Ideas and Outcomes* 5: e36152. <https://doi.org/10.3897/rio.5.e36152>

Semal P, Tilley L, Theeten F, Casino A (2019) The new CETAF Registry of collections and integration of the current CETAF passport: A collection information hub for the European Natural Science Community [PowerPoint presentation]. Available at: <https://collections.naturalsciences.be/cpb/cetaf-passport-and-collections-registry-manual>

Shorthouse, D. (2020). How it works. Retrieved April 02, 2021, from <https://bionomia.net/how-it-works>

Sierwald, P., Bieler, R., Shea, E.K., Rosenberg G. "Mobilizing Mollusks: Status Update on Mollusk Collections in the U.S.A. and Canada," American Malacological Bulletin 36(2), 177-214, (1 December 2018). <https://doi.org/10.4003/006.036.0202>

Singer RA, Love KJ, Page LM (2018) A survey of digitized data from U.S. fish collections in the iDigBio data aggregator. PLoS ONE 13(12): e0207636. <https://doi.org/10.1371/journal.pone.0207636>

Smith VS, Gorman K, Addink W, Arvanitidis C, Casino A, Dixey K, Dröge G, Groom Q, Haston EM, Hobern D, Knapp S, Koureas D, Livermore L, Seberg O (2019) SYNTHESYS+ Abridged Grant Proposal. Research Ideas and Outcomes 5: e46404. <https://doi.org/10.3897/rio.5.e46404>

Thessen, A. E., Woodburn, M., Koureas, D., Paul, D., Conlon, M., Shorthouse, D. P., & Ramdeen, S. (2019). Proper Attribution for Curation and Maintenance of Research Collections: Metadata Recommendations of the RDA/TDWWG Working Group. Data Science Journal, 18(1), 54. DOI: <http://doi.org/10.5334/dsj-2019-054>

van Egmond, E., Willemse, L., Paul, D., Woodburn, M., Casino, A., Gödderz, K., Vermeersch X., Bloothoofd, J., Raes, N. (2019, March 31). Design of a Collection Digitisation Dashboard (Version 1.0). Zenodo. <http://doi.org/10.5281/zenodo.2621055>

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLOS ONE 7(1): e29715. <https://doi.org/10.1371/journal.pone.0029715>

Woodburn M, Paul DL, Addink W, Baskauf SJ, Blum S, Chapman C, Grant S, Groom Q, Jones J, Petersen M, Raes N, Smith D, Tilley L, Trekels M, Trizna M, Ulate W, Vincent S, Walls R, Webbink K, Zermoglio P (2020) Unity in Variety: Developing a collection description standard by consensus. Biodiversity Information Science and Standards 4: e59233. <https://doi.org/10.3897/biss.4.59233>