Integrated summary from 17 to 30 April 2020

dhobern #1 May 1, 2020, 5:16am

This is an integrated summary of all topic discussion from **all daily summaries** for the Collections Catalogue consultation.

- **1.1. Directory to support the collections community (USE)**
  - **R**: The catalogue helps establish collections as a global science infrastructure.
  - **R**: The catalogue will bring value by making acronyms for institutions and collections searchable. This is even more important with digital information.
  - **R**: The catalogue is a good place to start in finding ways to leverage collective resources and funds.
  - **R**: The catalogue is important for building a sense of community and opportunities for new collaborations
  - **R**: Index Herbariorum shows how important it is to understand and meet the needs of collection managers.
  - **R**: Can we use ROR/Grid as the underlying catalogue of institutions?
  - **R**: ROR currently contains less detail on institutions than Grid does, but ROR is apparently addressing this.
  - **R**: Taxonomists would benefit greatly from a comprehensive catalogue.
  - **R**: The catalogue will reduce the need for expensive travel by making it easier to contact collections. It will also be valuable for the wider scientific community, environmental authorities and education.
  - **R**: Data must be kept up to date.
  - **R**: Natural science collections are a more unified community than collections as a whole.
  - **R**: What is the "collections community"? Do collection managers have the same goals as administrators, taxonomists, etc.?
  - **R**: The catalogue will assist institutional managers with planning, funding and showing importance.
- **1.2. Locating specimens and genetic materials (USE)**
  - **R**: Information collected on >300 North American arthropod collections, expanding to around 750 worldwide. Most information was collected from institutional websites. A standard information form would be a great help.
  - **R**: It will be valuable to know the taxonomic scope of each collection. In the Colombian registry (RNC), only higher taxa are listed. Estimates of holdings may help with collection management.
  - **R**: Do some countries have legal restrictions on what information can be shared about collections and their specimens?
  - **R**: UK experience is that information on collections is heavily used.
  - **R**: The catalogue can play an important role in simplifying linkages between specimen data and e.g. GenBank records.
  - **R**: Different disciplines (different taxonomic groups) will need to be able to capture different sets of minimal information.
  - **R**: The catalogue would be valuable for finding (and borrowing) materials - the main challenge is keeping it sufficiently current.
- **1.3. First step towards databasing collections (USE)**
  - **R**: Step-by-step guidance is essential to enable collections staff to play their part.
  - **R**: Issues with current content in GBIF/GRSciColl, including confusion between institutions and collections.
  - **R**: Success is most likely if the catalogue is well supported by the community and there are good tools and tutorials.
  - **R**: The catalogue may indirectly lead to digitisation by raising external pressure for data to be available.
  - **R**: Can we use the four GBIF dataset classes to help collections easily evolve from no presence online to offering rich digital access?
  - **R**: Institutions need more guidance on how to share the data they have on their collections. Clear roadmaps and strategies will help.
  - **R**: What information do paleontological collections typically already hold regarding their materials?
- **1.4. Assessing the scale and value of collections (USE)**
  - **R**: Metrics may be useful to government agencies. Metrics may not need to be fully standardised. The term "value" may be problematic if administrators see some collections as less valuable.
  - **R**: It is important to consider broader societal value as well as pure economic value.
  - **R**: Regular valuation is a standard process in Australia. Calculating indicative value is possible, but care is needed in how this is presented.
  - **R**: Can we learn lessons from the One World Collection initiative?
  - **R**: The One World Collection metadata would be a great test dataset.
- **1.5. Increased value for data on specimens, taxonomic publications, etc. (USE)**
  - **R**: Examples of possible value from making linkages via the ORCIDs for people associated with collections.
  - **R**: It will be useful if the catalogue includes preservation methods. Collections may be scoped in ways that are scientifically meaningful for other purposes.
- **1.6. Reducing duplication of effort (USE)**
  - **R**: See how much work was involved in creating a catalogue of fish collections.
  - **R**: There are opportunities to make it easier for other sectors to build a catalogue like the fish collection catalogue.
  - **R**: GGBN would save resources by not having to maintain a separate catalogue. The catalogue would save time for collections staff. It may be difficult to resolve records for the same collection from different sources.
- **1.7. Foundation for new and enriched services (USE)**
  - **R**: Identify possible fundable servers for potential users outside the immediate collections community (geologists, wine industry, etc.).
  - **R**: Natural history collections also have a role as part of cultural heritage. Links to genomic information from INSDC would be useful. Environmental and regulatory agencies need information on collections. The catalogue should link to the institution's own web resources. Attempting too many functions may impact core services.
- **1.8. Improvements to citation and visibility for collections (USE)**
  - **R**: Links to examples of Pensoft semantic enrichment of publications, including resolving identifiers for collections and specimens.
  - **R**: Plazi includes resolution of collection codes from historical publications to GRSciColl. Need mechanism to add other historical collection codes found in literature.
  - **R**: The botanical code does not rigorously enforce correct use of herbarium codes.
  - **R**: Examples from cataloguing fish collections of multiple (often invalid) codes used to refer to collections.
  - **R**: Organising all identifiers used (even incorrect ones) is a useful step. Greater visibility will help biobanks. Examples of issues with citation of tissue samples. Making the catalogue a standard tool (like GenBank) will enhance visibility.
  - **R**: Some collection communities have good practices around how collections are referenced and cited, but this varies by sector.

- R: Citation is likely to be increased if the catalogue makes links with other information tools such as ROR records.
- R: We need good citation practices to be built into the whole research lifecycle, including how we apply the FAIR principles.
- R: Users of collections need help and clarity about how to cite materials in different situations.
- **1.9. Support for national and regional needs and applications (USE)**
- R: Much of what is special about each collection relates to the significance of individual specimens.
- R: See the DiSSCo user stories as examples of the need to assess value.
- R: Help collections understand how their strengths complement one another.
- R: The RNC registry in Colombia plays an important role in permitting processes for collecting specimens.
- R: National or regional catalogues are important for demonstrating the role, value, strengths, weaknesses and uses of collections. (See also Topic 1.4)
- R Aggregating information at regional levels makes it easier to present the value of collections - comments are welcome on the design sketch offered by GBIF.
- R: Institutions don't just need metrics on size and digitisation. They also need to be able to show what is special, important or valuable about their collection.
- R: It should be possible to scope and present the catalogue to serve the perspectives of different communities.
- **2.1. Scope for the catalogue and definition of "collection" (INFORMATION)**
- R: Xylaria should certainly be included in the collection catalogue.
- R: There may be more sensitivity and legal restrictions around anthropological and paleontological collections.
- R: Examples of uses of xylaria.
- R: Natural history collection == A collection whose constituent parts (a) are derived from participant entities of natural processes and (b) have been collected to study properties of such processes.
- R: Include important historical collections that no longer exist.
- R: Organisation and support of collections varies greatly between institutions. Discussion of "research collections"/"private collections". Researchers need to know 1) what materials exist, 2) where materials are held, and 3) how it is possible to access materials. Biobanks should be included.
- R: Discussion of detailed example from INBio of how collections may be redefined over time.
- R: A good starting-point is to get a list of collections from each institution.
- R: Institutions may have "collections" of archival materials, field notebooks, multimedia, etc. - it is important to determine whether these are in scope as independent collections. (Topic 2.5 considers these as materials that at least need to be linked to a collection.)
- R: Paleontological collections have many common requirements and should be considered.
- R: The **10 Recommendations from DiSSCo** presentation includes DiSSCo's perspectives on this topic.
- R: Examples from Museum für Naturkunde Berlin of collections of additional materials (correspondence, notebooks, etc.) - it is important to be able to show the linkages between these different collections.
- **2.2. Identifiers for collections (INFORMATION)**
- R: All collections staff and other practitioners need to know and recognise their identifiers so that they can take ownership for using them.
- R: Need to consider how to resolve historical collection identifiers that represent part of a current collection.
- R: Handle historical collections, including their place now if included in a modern collection.
- R: It is unlikely that everyone will ever agree on a single definition of "collection" or "institution" so we must make it possible for them to decide how to identify their own collections.
- R: We need to be flexible about what institutions can identify as a collection.
- R: The catalogue needs to be flexible about what is considered a collection. Maybe not all collections need human-readable identifiers.
- R: Our efforts to help institutions identify their collections will help other communities such as libraries.
- R: The **10 Recommendations from DiSSCo** presentation includes DiSSCo's perspectives on this topic.
- **2.3. Hierarchical collection structures and subcollections (INFORMATION)**
- R: Hierarchies may not fit institutions that have multiple overlapping collections e.g. in different colleges. Hierarchical information may help with collection management. Different institutions may categorise collections differently.
- R: Orphaned collections that another institution adopts could be handled as items in a hierarchy.
- R: Users need pathways to find information for an institution or for a collection or for a dataset. These are different requirements.
- R: Institutions, collections and datasets should be treated as distinct elements, not a hierarchy.
- R: Discussion of relationships between institutions, a hierarchy of collections and datasets.
- **2.4. Description of a collection (INFORMATION)**
- R: Different disciplines or sectors should have the ability to agree different profiles of information to include and display.
- R: Different disciplines or sectors need different profiles instead of just more options.
- R: Suggestions for mandatory and desirable data elements in a collection record.
- R: Consensus metadata standards may not be well structured for efficient use.
- **2.5. Wider data linkages (INFORMATION)**
- R: Collections could make more use of field images. Types, taxa included, collectors, publications and staff should all be linked. Field notes should be mentioned.
- R: What can be done to help institutions in data-poor regions such as south-west Asia to participate?
- R: Need to identify benefits from participation.
- R: Linkages include references to collections intaxonomic treatments and citation of materials examined.
- R: We should consider what content is required so that the catalogue can become the source of the information needed/used by Wikipedia.
- R: Linking to aggregated information on specimens, collectors, etc. will get complicated and may not be part of the role of the catalogue of collections.
- **2.6. Information services relating to collections (INFORMATION)**
- R: All the suggested services would be valuable and help collections to work together better.
- **3.1. Pathways and tools for publishing collection records (TECHNOLOGY)**
- R: 1) Success will depend on offering good software for managing information, 2) Wikidata is likely to be a good way to share information, but not to manage it.
- R: Institutions need pathways for updates to be published automatically.
- R: Examples from Colombia and Argentina. The Argentinian SNDB registry holds less information than many of the collection systems.
- **3.2. Community catalogues (TECHNOLOGY)**
- R: Information on **bug-collections.org**, an initiative to build a catalogue of arthropod collections.
- R: GGBN is keen to participate and will retain its own identity.
- R: Can collections be given mechanisms that allow them to keep Index Herbariorum entries current and automatically update other public views of the catalogue.
- R: Access to edit records should be role-based rather than tied to a system or user.
- R: Will data standards be the same in the community catalogues and the integrated catalogue? Answer - for the most part, yes.
- **3.3. Integrated catalogue (TECHNOLOGY)**
- R: Supporting a range of different editing/curation tools and paths is critical.

- R: Colombian law gives responsibility to the RNC catalogue to the Alexander von Humboldt Institute.
- **3.4. Collection management systems (TECHNOLOGY)**
- R: Collection management systems can play a valuable role as a source of statistics on digitisation progress.
- R: GBIF is already integrating digitisation progress information from Index Herbariorum into GRSciColl.
- R: Work is under way on a MIDS (minimum information about a digital specimen) standard.
- R: Details on the Minimum Information about a Digital Specimen (MIDS) standard.
- R: A “software summit” could be an opportunity to bring together developers of collection management systems and other software together to align plans.
- R: EML documents combined with TDWG CD elements may be a workable model for data exchange.
- R: Additional informaton on the Minimum Information for a Digital Specimen standard.
- R: Direct linkages with collection management systems could be an effective option.
- R: Symbiota supports data import via Excel, CSV or from IPT servers. In Costa Rica, collection systems have been expected to provide annual estimates of holdings.
- R: Discussion about how collection management systems could become key tools for managing collection records.
- R: For many collections, the “collection management system” is Excel. This needs to be covered.
- **3.5. Interfaces, APIs and client modules (TECHNOLOGY)**
- R: Content of collection records should be interpreted and validated as far as possible so it can be used as data.
- R: OpenRefine may be a valuable tool for reconciling data from different sources. It may help to link information about the same collection even when identifiers are not fully standardised.
- R: Design mock-ups need to show more than just how the catalogue can be viewed by humans.
- R: Technical aspects associated with GBIF implementing new interfaces to the catalogue.
- **4.1. Ownership of information for each collection (GOVERNANCE)**
- R: Collection administrators are still not as engaged as is necessary. In Argentina, access to collection information has sometimes been restricted. Indigenous labels and worldviews should be included where relevant.
- R: Good presentation of information will encourage institutions to take active control of their collection records.
- R: We should consider what will help institutions prioritise keeping information current.
- R: Personal development is needed to ensure new staff understand the importance of collection information and how to maintain it. This may require assigned roles in some institutions, but always requires recognition for the work involved.
- **4.2. Communities of practice (GOVERNANCE)**
- R: Major challenges may occur in cases where e.g. a regional catalogue and a disciplinary catalogue both include the same collection in their scope.
- R: In Australia, the ALA can act as a knowledge broker between collections and international catalogues.
- R: Publishers are an important user community.
- R: SPNHC has a key role to play in supporting communities of practice.
- R: Communities of collections work well when organised along taxonomic lines. Example of UK experiences through NatSCA. In considering communities of practice, it is important to focus on the people involved.
- **4.3. Technical infrastructures (GOVERNANCE)**
- R: The informatics components should focus on supporting the information needs of collections.
- **4.4. Governance arrangements (GOVERNANCE)**
- R: Discussion of relationship between GBIF/GRSciColl and community catalogues (collaboration or competition, unified or integrated). Alongside the FAIR principles, the Indigenous Data Sovereignty working group has articulate the CARE principles to support ethical use of materials from indigenous groups.
- **4.5. Incentives for contributors (GOVERNANCE)**
- R: Sometimes there may be benefits in publishing information on behalf of collections, but they must have confidence that they are in control. Incentives may include advertising and recognition for the collection, tools that simplify work inside the collection, more credibility when seeking funding. Collections would benefit if funding is available for visits by taxonomists and for software licences.
- **4.6. Funding and sustainability (GOVERNANCE)**
- R: Resources are not evenly distributed globally - we should consider ways to enable altruism and partnerships to support less-resourced collections.
- R: GBIF and CETAF have key roles to play as part of the funding and sustainabilily model, but securing long-term funding will be hard.
- R: Need to be able to identify meaningful metrics to justify long-term support. Government agencies, major world collections and GBIF may play important roles in securing funding.

Advancing the Catalogue of the World's Natural History Collections

---

dhobern #2 May 1, 2020, 5:17am

This is an integrated summary of all discussion points on documents, presentations and the process from **all daily summaries**.

- **Comments on this virtual consultation process**
- R: Discussion around the levels of activity in this consultation. In future consider more visibility via social media, use “likes” to get more signs of participation.
- R: It is important to engage with less well resourced collections. Discussion of possible reasons why some stakeholder may have chosen not to participate.
- **Document: 10 recommendations from DiSSCo**
- R: We may need to build a hierarchically organised catalogue of all scientific collections with separate interfaces and branding for important subsets.
- R: Ethnobotanical and zooarchaeological collections are two more examples of collections which may need to be accommodated from multiple perspectives.
- R: Zoos, aquaria and culture collections are sufficiently different that they would need separate support. It would be best to focus on preserved biological collections first.
- R: Multiple collection abbreviations may need to be associated with each collection identifier.
- **Document: GBIF Services and Support for the Collections Catalogue**
- R: Many ideas in response to the GBIF paper on possible web interfaces for collection information.
- R: Collection catalogue interfaces need to show good provenance information on the source of content.
- **Presentation: ALA Collectory**

- **R** Has the ALA solved issues recognising the same occurrence record from different sources?
- **Presentation: CETAF and DiSSCo Collections Registry**
  - **R**: Discussion of the range of codes associated with RBINS/ISRcNB/KBIN.
  - **R**: Example of the range of codes associated with RBINS/ISRcNB/KBIN and how codes are presented by the CETAF Profiles.
- **Presentation: TDWG Collection Descriptions Data Standard Task Group**
  - **R**: The CD standard should include an element to show which collection management software is used by the collection.
  - **R**: Discussion of possible benefits in exposing information on the collection management system in the collection description record.
  - **R**: Registration, or automated discovery, of endpoints associated with collection management systems could be a very flexible approach.
- **Presentation: The Specifications of Earth Science Collections (CETAF)**
  - **R**: Are the vocabularies that this presentation discusses already stable and ready for wider use?
  - **R**: Standards and vocabularies are still under development for earth science and paleobiology materials. DiSSCo Prepare is working on these.
  - **R**: Status of vocabularies for earth science collection categories.