2.5. Wider data linkages (INFORMATION)

dhobern #1 March 27, 2020, 10:26am

> This is topic 2.5. in the *Information* section of the ***Advancing the Catalogue of the World's Natural History Collections*** consultation. Use this topic to discuss the questions listed below.

Background
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.

Other materials

The following contributed materials are particularly relevant to this topic:

- **Presentation: CETAF and DiSSCo Collections Registry**
- **Document: GBIF Services and Support for the Collections Catalogue**
- **Presentation: ALA Collectory**

Questions

- 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

Document: GBIF Services and Support for the Collections Catalogue

Presentation: ALA Collectory

Presentation: CETAF and DiSSCo Collections Registry

2.1. Scope for the catalogue and definition of "collection" (INFORMATION)

Document: GBIF Services and Support for the Collections Catalogue

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

建立《全球自然历史馆藏名录》

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

建立《全球自然历史馆藏名录》

Collections Catalogue - Daily Summaries

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

建立《全球自然历史馆藏名录》

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

建立《全球自然历史馆藏名录》

Integrated summary from 17 to 30 April 2020

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

建立《全球自然历史馆藏名录》

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

Advancing the Catalogue of the World's Natural History Collections

建立《全球自然历史馆藏名录》

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

Presentation: TDWG Collection Descriptions Data Standard Task Group

---

Maofang #4 April 23, 2020, 3:38am

We got a response from @strawberry posted in Chinese:

The translation from her post:

How to motivate participations from poor data density areas, such as Southwest Asia? For example, promoting the catalogue in those areas via some acquaintances.

1 Like

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

建立《全球自然历史馆藏名录》

Integrated summary from 17 to 30 April 2020

---

dhobern #5 April 23, 2020, 11:57pm

Thanks @strawberry and @Maofang . This is an excellent question. We cannot expect anyone to participate unless they can see how they will benefit from their participation. I would like to know how we can help collection in areas like south-west Asia to get real benefits in the following ways:

- Better visibility and higher reputation for their collection nationally and internationally
- Opportunities for new partnerships to study the species in their collections
- Help in making data on their collections and specimens available across the Internet

I think we could help by trying to do the following:

- Make it easier for collections to start sharing information - make it easier to publish information (See Topic 1.3)
- Make information available in more languages and through more national portals
- Improve our understanding of what is most needed by collections in countries that are not yet involved (See Topic 1.9)

I am very interested in any ideas you have on these topics.

1 Like

Debbie #6 April 26, 2020, 6:14pm

I'm wondering what such a resource offers them in return? What (services?) are we going to offer as incentives to contribute? How will we say "membership" or "participation" has its privileges?

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

建立《全球自然历史馆藏名录》

Integrated summary from 17 to 30 April 2020

---

agosti #7 April 27, 2020, 7:18am

Material citation and taxonomic treatment are decisive because especially the treatment includes a scientists expertise as standardized statement about these speicmen from a particular collection.

Using just publications means loosing the precision of a scientists opinion by first in many ways not providing access to the content because it is closed access, secondly because the precise links is lost because the semantics the scientists adds to an article is lost.

A scienist's statement in an article is as follows: This specimen from collection A is part of taxon B. She makes this by using a materials citation of the specimen as part of a taxonomic treatment (which is part of a publication).

1 Like

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

建立《全球自然历史馆藏名录》

Integrated summary from 17 to 30 April 2020

---

qgroom #8 April 27, 2020, 7:23am

> dhobern:
>
> What information should be linked to collection records?

One of the topics that has come up in earlier conversations on the Collections Catalogue is Wikipedia.
Wikipedia is one of the first places people go to get information on all sorts of topics, including collections and their hosting institutions. Wikipedia is only a secondary source of information, therefore it relies upon primary sources of information, such as the Collection Catalogue.
What this means in practice is that the Collection Catalogue would need to provide the sort of information that Wikipedia wants. This might include more general information about the collection, its history and the people associated with it. It also means that the Catalogue should be easily and stably citable.
There may be also be information that a collection might consider important to share in a primary source, but is not strictly relevant to the Collection Catalogue. For this reason there should perhaps be sufficient scope for free-text entries in the Catalogue.

1 Like

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

建立《全球自然历史馆藏名录》

Integrated summary from 17 to 30 April 2020

---

cboelling #9 April 28, 2020, 10:40am

Most of the information suggested as candidates for inclusion is associated with individual specimens in a collection. In a catalogue of collections such information would, as I understand, need to be aggregated on the level of the containing collection.

For example, if information on the collectors of the specimens is desired, a list of collectors would need to be aggregated from the specimen level and connected to the collection record. In the case of a specialist collection which was brought into being to assemble all specimens collected by a particular collector, that list would have a single entry.

Such aggregated information would need to be provided by the data providers or, in an advanced scenario, might be scraped by the catalogue from information endpoints representing the collection contents.

I think while all of this candidate information - and the corresponding services described in **topic 2.6** - is potentially useful one should carefully consider in which cases such aggregated information corresponds to use cases of the catalogue. This might get really complex really quickly, and for some of these enquiries the information facilities offered by the collection resp. its holding institution might be better suited.

3.4. Collection management systems (TECHNOLOGY)

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

建立《全球自然历史馆藏名录》

Integrated summary from 17 to 30 April 2020

---

dhobern #10 April 28, 2020, 12:22pm

Thanls **@cboelling** . The intention was never to embed all of this additional aggregated information into the catalogue. The goal was to identify classes of information to/from which a user may reasonably expect to be able to navigate from a catalogue record. Associated infrastructures such as GBIF can provide the corresponding services but some planning is necessary to ensure that identifiers map correctly and are reused consistently.

neilcobb #11 April 28, 2020, 3:56pm

I am baffled that collections are not more proactive in supplementing specimens with field images or more specifically actively coordinating iNaturalist projects and other similar ventures. There are plenty of limitations and challenges with field images, but they are getting better. People are taking images from more aspects and using better lenses and they make the data public within in weeks (passing research-grade certification) and not 5 years that we see with speicmen records. iNat research-grade records have 627 citations to date in publications.

Collections Catalogue - Daily Summaries

Integrated summary from 17 to 30 April 2020

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

建立《全球自然历史馆藏名录》

pzermoglio #12 April 29, 2020, 6:42pm

From **@maperalta** in the **Spanish thread**

Information related to the collection records:
• Type specimens in a collection: Yes
• Species/taxa represented in a collection (with/without specimen counts): without counts, only main groups.
• Include at least main collectors associated.
• Include publications based on materials from the collection of at least the last 5 years.
• Yes to include researchers/staff associated with the collection with institutional email address.
• Only mention if there are field notes.
There should be a formalized commitment so that links are bidirectional, as these are problematic because it means to take responsibility for the content on a third party site.