1.3. First step towards databasing collections (USE)

dhobern #1 March 27, 2020, 9:55am

> This is topic 1.3. in the *Uses* section of the ***Advancing the Catalogue of the World's Natural History Collections*** consultation. Use this topic to discuss the questions listed below.

Background
The information needed to build the catalogue of collections closely matches the metadata required to publish a specimen dataset to GBIF and other portals. A record that describes a collection could be treated as a minimal first step, perhaps leading through processes such as Join The Dots and onwards to comprehensive digitisation. A comprehensive catalogue of such records could guide efforts to prioritise further digitisation, by highlighting collections with holdings of particular relevance or by assisting the development of collaborative digitisation networks like the ADBC Thematic Collection Networks.

Other materials

The following contributed materials are particularly relevant to this topic:

- **Presentation: Global Registry of Scientific Collections**
- **Presentation: Index Herbariorum - Perspectives of a Long-Standing Collection Catalog**
- **Presentation: TDWG Collection Descriptions Data Standard Task Group**
- **Presentation: CETAF Collections Registry**

Questions

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

1 Like

Presentation: Global Registry of Scientific Collections

Presentation: Index Herbariorum - Perspectives of a Long-Standing Collection Catalog

Presentation: TDWG Collection Descriptions Data Standard Task Group

建立《全球自然历史馆藏名录》

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

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》

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

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

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

建立《全球自然历史馆藏名录》

Advancing the Catalogue of the World's Natural History Collections

---

dhobern #4 April 21, 2020, 2:12am

I am interested in thinking how the catalogue could help institutions to move towards richer digital access. Some years ago, many collections at least offered online lists (not structured data) of their type materials. I haven't seen such lists in recent years. Some of those collections will instead have published Darwin Core datasets, but I think some collections may now be less discoverable than when the Internet was less crowded.

GBIF (along with other data aggregators) offer the option of publishing **four increasingly rich classes** of data:

- **Resource metadata**: Information describing a resource (e.g. a dataset), whether or not it is digitised in a machine readable form - this can be a useful discovery tool. Collection records based on the **TDWG Collection Descriptions** could be the preferred data model for this kind of advertising role.
- **Checklist dataset**: These can take many forms, but allow a list of species (or other taxa) to be shared, along with metadata describing the list. Additional Darwin Core or other terms may also be included for each species. Institutions that have lists of types or of species within their collections could easily use this model to showcase their holdings - the species or specimens in these lists could be handled as very-low-information Darwin Core records and become discoverable to researchers.
- **Occurrence dataset**: Most specimen data today is shared in this form, which supports rich information on each specimen held.
- **Sampling-event dataset**: This is rarely used by collections today, but would be a valuable extension wherever specimens or other materials were collected in a standardised way.

Every one of these four dataset types could be associated with a TDWG Collections Description. It would be wonderful if collections could be helped to use these levels as stepping-stones from being digitally undiscoverable to being fully digitised. Visibility and value are increased at each step.

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

---

建立《全球自然历史馆藏名录》

3.4. Collection management systems (TECHNOLOGY)

2.5. Wider data linkages (INFORMATION)

Integrated summary from 17 to 30 April 2020

---

tkarim #5 April 21, 2020, 3:25pm

I'm wondering if there needs to be more outreach on the types of datasets that people can share? How many collections know that they can publish resource metadata? There are a ton of untapped (undigitized and/or uncurated) specimens in "stratigraphic" fossil collections. Most paleo collections have these types of holdings. Publishing them as structured resource metadata records/sets (e.g., chrono and litho stratigraphic information, geography, collector, minimal taxonomic information) would be a great way to make them more discoverable.

3 Likes

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

---

Integrated summary from 17 to 30 April 2020

---

pzermoglio #6 April 22, 2020, 2:19am

Incentives and support needed vary regionally, given the different realities. To determine which support is needed would necessitate delineating regional roadmaps and strategies.
I would agree with **@tkarim** that not everyone knows that it is *possible* to just publish metadata. Probably some training and promotional materials should be developed and broadly distributed to help people get in loop, as the first step tends to be the most difficult one to take.

2 Likes

dhobern #7 April 22, 2020, 2:39am

Thanks @tkarim and @pzermoglio - I agree one action should be better guidance for collections on how they can progress stepwise into the world of digital access.

Debbie #8 April 23, 2020, 3:33pm

Do they have "reports" they write up internally on these types of collections already? Similar to "species inventories?"

建立《全球自然历史馆藏名录》

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

Integrated summary from 17 to 30 April 2020

tkarim #9 April 23, 2020, 4:14pm

Maybe, but not necessarily. And the types of internal reports will vary widely between museums.

clauden #10 April 24, 2020, 12:42pm

I think part of variety of 'internal' reports comes down to 'an individual with knowledge about it', that is lost when moving on, retired, or simply judged as being too much effort to bother. Guidance on stepwise progress (@dhobern), making it easier to advance–and transfer–might be appreciated.

2 Likes

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

Integrated summary from 17 to 30 April 2020

neilcobb #11 April 28, 2020, 3:03pm

One of my biggest frustrations with GBIF is that most of the information regarding a collection is out of date. There is no easy mechanism for updating information.

It is also frustrating when a collection makes multiple datasets or datasets from multiple collections, I understand the importance for allowing people to create focused datasets but I wish there was a way to have a listing for a collection that included all the data from the datasets produced by that collection or institute. The end result would be a scalable system that allows users to query for special data sets within a collection or all of the records for a collection.

I wish GBIF required all collections within an institute to be listed on an institute page. Why should the end user suffer just because collections within an institution cannot coordinate?

GrSciColl is a nightmare, there are collections listed as institutions and institutions listed as collections.

I see all of these being relevant to the world collections catalogue, solve these problems and you make the larger catalogue easier to get buy-in and easier to manage and build.

1 Like

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》

gdadade #12 April 29, 2020, 7:12am

I think the community will start using the catalogue as soon it has a strong support by the stakeholders and tools/tutorials are provided to help. Index Herbariorum is accepted in the botanical community for a long time because at some point it turned into a community convention, that herbaria must be represented there. So working on use cases and tutorials on how to provide content to the catalogue and the potential usages of this content would be very benefical to overcome obstacles.

1 Like

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 #13 April 29, 2020, 5:17pm

From **@maperlata** in the **Spanish thread**

I don't think publication of a collections catalogue would help digitization. It could influence tangentially, in the urgency for digitization derived from a potential increase of queries.

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 #14 April 29, 2020, 7:49pm

From **@ErikaSalazar** in this **Spanish thread**

Can publishing a collection record to a catalogue assist collections in moving towards full digitisation? Yes, it could, if the importance of the collection can be reflected in that record, and mostly if the needs for support can be reflected as well. For example, that in the catalogue we could see the advances (many or few) that a collection has done in digitization, so that in the event of an opportunity arising (e.g. public call for funds) or faced with the availability of funds from national or international entities, we could identify which collections would need that support.
Collections need resources for infrastructure and human capacity so that they can advance digitization processes. In parallel, knowledge transfer is needed so that the digitization work is done in a correct and standardized manner.

---

aidavasco #15 May 22, 2020, 3:07am

In Colombia is very difficult to get funds for databasing and digitalized the collections, also to maintain them.

In the Herbarium of the University of Antioquia, in Medellin, plant collection has been partially digitized thanks to a partners with Missouri Botanical Garden (I think). However I work with the Fungarium, and we have more than 12.000 collections and 90 types, but we do not have resources to digitalize .