2.2. Identifiers for collections (INFORMATION)

dhobern #1 March 27, 2020, 10:19am

> This is topic 2.2. 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
Most collections are already identified by one or more collection codes and may have existing web identifiers (URLs, DOIs, etc.) in one or more databases. The catalogue may reuse one or other of these identifiers or may help to support a standardised scheme. GRSciColl exists to assist with standardisation of collection codes and machine-readable identifiers, but several other efforts are also in place. Unique identifiers for each collection will be important to maximise cross-linkage of information and standardise citation, but other existing identifiers should ideally resolve to the same information and be recognised as synonyms for the preferred identifiers.

Other materials

The following contributed materials are particularly relevant to this topic:

- **Presentation: TDWG Collection Descriptions Data Standard Task Group**
- **Presentation: CETAF and DiSSCo Collections Registry**

Questions

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

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

---

建立《全球自然历史馆藏名录》

1.1. Directory to support the collections community (USE)

Integrated summary from 17 to 30 April 2020

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

Presentation: CETAF and DiSSCo Collections Registry

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

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

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

---

qgroom #4 April 17, 2020, 10:07am

Multiple identifier systems can work in parallel, though we should be 100% sure we can’t reuse an old one if we propose to create a new one.
One problem is likely to be the conflation of identifiers for different things. So institutional identifiers can’t substitute for collection identifiers.
There is also the problem of merging and splitting collections and what happens to the identifiers in these cases.

Ultimately, a big problem is that there is no consistent definition of a natural history collection and it is difficult to image there ever will be. Therefore, we have to rely on collections self-identifying their own collections and describing what they mean by a collection in the metadata of the description.

2 Likes

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

---

elyw #5 April 21, 2020, 3:56am

I couldn't agree more @qgroom - what is the definition of a natural history collection?
Sometimes 'collection' = curatorial department. This is convenient as a way of grouping records as they go out to an aggregator (like ALA or GBIF), it's a convenient way of providing contact information (e.g. for someone who wants to borrow bird specimens, having the contact details of people working in the Ornithology department makes the process quicker).
But sometimes 'collection' = a group of things connected together at a more granular level. These might be a person's collection (the collection of Dr Graeme Smith), a collection attributed to a donor (bequest of Dr Annie Smith) or might be a collection by virtue of the expedition or collecting trip. "The Australian seamount survey collection" - which might have specimens distributed out to a number of departments and potentially even over a number of different institutions.
Multiple identifiers seem inevitable, as does institutions self-identifying what they mean when they call something 'a collection'

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

---

mswoodburn #6 April 22, 2020, 4:51pm

Also agree strongly - from the use cases that we've looked at in for the **TDWG Collection Descriptions Data Standard task group**, the definition of a natural history collection varies widely depending on the context. The most specific we can get that fits all of them is a fairly generic 'a group of physical collection objects with one or more common characteristics', with those characteristics being defined by the context.

It may be that only the more traditional concept of a 'collection', such as those that represent a whole institution's collection or tier below (e.g. herbarium, palaeo collection) need a human-readable identifier due to their historical use in previous and current registries. More granular collections like @elyw mentions may not have the same need. However, every collection needs a machine-readable identifier, and that would have to be unique in a global catalogue.

I think the main trap to avoid (as with specimen identifiers) is conflating the purpose and requirements for human and machine-readable identifiers. The human versions need to short, memorable and informative, and ideally would be globally unique and persistent to avoid confusion, but it won't break things thing if they aren't, or if they don't exist for a collection. The machine versions need to be globally unique, persistent and resolvable, and the ones relied upon by software for unambiguous identification and data linkage.The issues tend to arise when we try to rely on the human-readable identifiers for that, or force human readability concerns on machine identifiers.

2 Likes

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

---

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

Integrated summary from 17 to 30 April 2020

Debbie #7 April 23, 2020, 3:31pm

Great points Ely. I note too that you added more about this in the What is a collection? section of this conversation **2.1. Scope for the catalogue and definition of "collection" (INFORMATION)**.

FYI, any work we do that results in enabling the community to define as needed (within scope), will also help other communities, such as library collections. The library community tried to develop a standard almost 20 years ago, and it was the at the point of this issue about grouping-the-groups in so many different ways, that they gave up on this part. See **https://github.com/tdwg/cd/blob/master/reference/papers/CollectionDescriptionFocus_BriefingPaper_2002.pdf** for their thinking (in parallel to ours) from 2002.

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

---

waddink #8 April 23, 2020, 7:33pm

We have proposed an answer to these questions in our document with 10 recommendations. As described in the paper Deb referred too, collections can be grouped on many things and yes it is hard to get to an exact definition of a natural hist collection. Nevertheless, the use cases for our community seem to require only few groupings, the distinction between collections describing in sum the total collection holdings of an institute (for discovery and access and other collections), and other collections is one. The distinction between collections curated to provide a systematic work of enduring reference from other collections that we call datasets is another. We propose to see datasets not as collections for the catalogue, e.g. a list of specimen records referred in a paper or a set of pages from a field book. We also propose to use the term natural science collections rather than natural history to make it clear that e.g. Earth sample collections are included.

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

---

waddink #9 April 23, 2020, 7:39pm

The issue of merging and splitting and persistant identifiers seems an easy one. The collection description is then no longer applicable, but the pid should keep existing. So it needs to get a tombstone description stating why it is obsolete and no longer resolving, with a reference to the new collection description or descriptions. We have given a list of preferred identifier systems with their pros and cons in our document with 10 recommendations.

1 Like

---

waddink #10 April 23, 2020, 7:51pm

This paper showcases why names or abbreviations are not enough to identify for automated search algorithms, using names and abbreviations for institutes in ROR:
https://www.tedhabermann.com/blog/2020/4/21/sometimes-a-name-is-not-enough-update

---

elyw #11 April 24, 2020, 8:23am

> dhobern:
>
> What can be done to promote use of the preferred identifiers?

Another quick comment on this specific question. Collections / institutions and the people within - collection managers, data managers, “the person in IT who exports the DwCA archive out of the collection management system” all need to know about the identifiers applied to their collections - particularly if they’re not the ones actually creating these identifiers. If ‘your’ collection appears in GRSciColl, the ALA Collectory and the GBIF Registry all with different identifiers that you’ve never seen before or had any hand in generating, why would make you use any of them? And which one would you choose?

2 Likes

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

---

Integrated summary from 17 to 30 April 2020

---

cboelling #12 April 27, 2020, 10:40am

The original collection description(s), including their identifiers, referring to the collection(s) before the split / merge might still be of value and indeed remain applicable in cases where a pointer to a collection in a historical state is needed.

This brings up the question if the catalogue should allow for entries of historical collections which have since been merged, split or transformed in other ways or if this is out of scope.

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

---

sabaj #13 April 28, 2020, 1:08am

I would argue strongly that the Catalogue should allow for entries of historical collections. Small collections may become "historical" over the course of the Catalogue's assembly!

That aside…taxonomists often refer to specimens that were once part of a collection that is now defunct, its specimens lost or transferred to other institutions. So, having a place holder or PID for Museo Laurentii Theodori Gronovii, Lugdunum Batavorum [Leiden] (also known as Museum Gronovianum, Museo Lugdunensi Batavorum) would be helpful.

Also note that a single institution may associated with both a historical collection and an active collection, but the historical collection may exist at a separate institution. How many historical collections did Napoleon "transfer" to MNHN?

For a more recent example, the historical herpetology collection of the Southeastern Louisiana University (SLU) Vertebrate Museum was transferred to Louisiana State University Museum of Natural Science in the 1990s. Since then, SLU has established a **new herp collection** that includes 20% of **historical herp collection** from yet another institution, the Tulane University Museum of Natural History.

There are plenty more examples. It would be extremely difficult for any Catalog to capture all of them. But, having PIDs for historical collections would enable links to active ones.

2 Likes

2.3. Hierarchical collection structures and subcollections (INFORMATION)

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