3.4. Collection management systems (TECHNOLOGY)

dhobern #1 March 27, 2020, 10:39am

> This is topic 3.4. in the *Technology* 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 natural history collections maintain data on their specimens in a collection management system (CMS) such as **Specify**, **Symbiota**, **EMu**, **DarWIN** or **BRAHMS**. Some of these tools could develop to interface directly with the collection catalogue, providing up-to-date metadata and metrics.

Other materials

The following contributed materials are particularly relevant to this topic:

- **Presentation: Index Herbariorum - Perspectives of a Long-Standing Collection Catalog**
- **Presentation: The Specifications of Earth Science Collections (CETAF)**

Questions

- What present or future requirements are there for interfaces directly between CMS platforms and the collection catalogue?
- Are there special opportunities that should be considered?
- Could CMS platforms become a source of metadata for institutional collections within a global catalogue?

Presentation: The Specifications of Earth Science Collections (CETAF)

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

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

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》

Collections Catalogue - Daily Summaries

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》

---

dhobern closed #2 April 7, 2020, 8:33am

---

dhobern opened #3 April 16, 2020, 2:00pm

This topic was automatically opened after 9 days.

---

elyw #4 April 21, 2020, 3:24am

The missing question for me in this one is: are the vendors/developers of the various collection management systems willing to engage in this conversation and make changes according to developing standards? I'm sure some will be responsive, others less so.
It might also be another of those questions where vendors and even institutions ask - why would I store collections descriptive data in my collection management system? This was the challenge Australia faced in developing the Museum Metadata Exchange - all of the information had to be compiled before it could be published. No-one in each institution needed the information for internal purposes so it had never before been written or gathered into one place.

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

dhobern #5 April 21, 2020, 5:35am

Thanks @elyw . It will only make sense for these functions to be included within collection management systems if the following things are true:

- There are **benefits that come to the institutions** from having good current information accessible on their collections. Maybe there are parallels here with **ORCID**. As ORCIDs come more and more to be a useful tool for researchers to integrate and organise information on their work, it becomes more essential to have one. If collection records come to be important linkage points for information and services that matter to collections, they will become mission-critical.
- There are **efficient publication pathways for collection records** that can readily be integrated into the workflow of collection management systems - management of a TDWG Collection Descriptions record could make this true.

emeyke #6 April 21, 2020, 3:20pm

The short answer to @elyw is (and I cannot speak for others) - we definitely want to follow the standards and implement the solution based on these discussions and technical specs.

2 Likes

tkarim #7 April 21, 2020, 3:42pm

In terms of metadata, I also think this is a great idea. If this could be automated and pulled periodically from the database it is one less thing a collection manager has to do or remember to update.

In terms of metrics, there was a presentation at the SPNHC 2019 meeting about Arctos interfacing with some external sources and then providing metrics on things like publication and use for specimens. I would love to see something like that for Specify. It would make compiling annual reports easier and also make it easier to demonstrate the impact of our collection.

2 Likes

dhobern #8 April 21, 2020, 10:52pm

Thanks @emeyke and tkarim - I'm glad to see the interest in this idea. The **Arctos** example is a really excellent one. It shows how a collection management system can be a tool for curating information from the side of the knowledge graph - ownership and responsibility for data owned by the institution/researcher,

combined with the ability to mesh with BOLD and ecological data systems. I know that some other collection management systems (including **EarthCape** and **Specify**) have similar interests.

An interlinked question is how we know which of the potentially many records describing a given collection is a) the most authoritative and b) the most current. This ties in with other topics: **3.1. Pathways and tools for publishing collection records (TECHNOLOGY)** and **4.1. Ownership of information for each collection (GOVERNANCE)**.

1 Like

---

pzermoglio #9 April 22, 2020, 2:29am

I would like to point out that the assertion:

- *"Most natural history collections maintain data on their specimens in a collection management system (CMS) such as **Specify**, **Symbiota**, **EMu**, **DarWIN** or **BRAHMS**."*

may not be exactly true in many parts of the world. Unless you can call Excel a collection management system…

For those using CMSs, interfaces with the catalogue would be great and save people a lot of time and effort. But we should not neglect those using simple spreadsheets, for whom other solutions would be needed (the beauty of the IPT model, allowing just uploading a simple file).

4 Likes

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

---

Integrated summary from 17 to 30 April 2020

---

trobertson #10 April 22, 2020, 7:44am

> pzermoglio:
>
> For those using CMSs, interfaces with the catalogue would be great and save people a lot of time and effort. But we should not neglect those using simple spreadsheets, for whom other solutions would be needed (the beauty of the IPT model, allowing just uploading a simple file).

Thanks **@pzermoglio** . This is very much in line with our thinking on the **sketch** we put together. Going beyond even the IPT model which requires someone to install a server, having some simple web forms to fill in, along with the ability to upload Excel files with standardised field heading could really lower the technical threshold for many to have an online search portal. Excel and databases like FileMaker are commonplace and we need a technical solution that allows easy participation. I propose we consider offering that natively in the catalogue itself.

2 Likes

---

emeyke #11 April 22, 2020, 10:40am

@trobertson if you are talking about just collections catalogue and maintaining *own* collection record, than Excel upload seems like a bit of overkill. Simple web form for creating/updating the record should do it.

1 Like

---

trobertson #12 April 22, 2020, 11:18am

Thanks @emeyke , Sorry that was not clear. I meant web forms for descriptive content and excel for specimen records.

1 Like

---

emeyke #13 April 22, 2020, 11:33am

Well this is interesting and I might have missed a big discussion on that. Are we talking about direct upload of occurrence datasets via csv/excel files into GBIF? We can take this elsewhere as it seems to be drifting off the current topic.

1 Like

Rich87 #14 April 22, 2020, 5:57pm

One aspect where a link between a CMS and a Global Catalogue might be useful is to provide metadata on digitization progress in a collection. As an example, Index Herbariorum entries now have this optional set of fields (data shown for NY Botanical Garden):

	Num. of Specimens	Num. Databased	Num. Imaged
Algae	221000	155365	110822
Bryophytes	700000	448769	452485
Fungi/Lichens	700000	685635	441587
Pteridophytes	300000	223911	206980
Seed Plants	6000000	2686457	2228503

Drawing similar data directly from an institution's CMS, if possible, should reduce the effort needed in keeping this type of information at least somewhat current.

4 Likes

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

---

Integrated summary from 17 to 30 April 2020

---

abentley #15 April 22, 2020, 9:31pm

Specify software certainly is. We have most of the relevant tables and fields that could accommodate this information an successfully publish it through existing IPT and xml infrastructure.

1 Like

---

trobertson #16 April 23, 2020, 7:58am

> Rich87:
>
> One aspect where a link between a CMS and a Global Catalogue might be useful is to provide metadata on digitization progress in a collection. As an example, Index Herbariorum entries now have this optional set of fields (data shown for NY Botanical Garden):

Thanks, **@Rich87** . Please know GBIF are actively working on bringing this into GRSciColl and it is available in the API already. For example, you can see these counts in the `collectionSummary` field at the bottom of the response for **NY Botanical Garden**. Today this is populated only for records sync'ed from IH records (happens automatically now) but will be expanded for more collections and moved into the user interface.

This is limited of course, and more expressive descriptors should be available when metadata in (N)CD Standard is provided.

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

---

trobertson #17 April 23, 2020, 8:06am

> abentley:
>
> Specify software certainly is. We have most of the relevant tables and fields that could accommodate this information an successfully publish it through existing IPT and xml infrastructure.

Thanks, @abentley . I have noticed a few mentions on threads of the IPT (GBIF's publishing tool) and the associated dataset metadata which is **an extension of the Ecological Metadata Language** (EML) to bring in elements of the Natural Collections Descriptions standard from that time.

Given the existence of this operational infrastructure today, I wonder if we should consider exploring a revision of the EML profile in use by GBIF and others, along with promotion of `metadata-only` resource sharing as @dhobern described **in this thread**.

2 Likes

---

waddink #18 April 23, 2020, 8:15pm

Relevant here is also MIDS (minimum information about a digital specimen) that is under development and describes digitization status, planned to be included in TDWG CD.

2 Likes

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

---

Rich87 #19 April 26, 2020, 11:30pm

Thanks waddink.

I am very curious about the MIDS standard - where could I get more information about that standard? I think this is a very pertinent issue that likely is of interest to quite a few people in our community.

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

---

emeyke #20 April 27, 2020, 8:43am

> **deliverable_d2.3_icedig_-_design_of_a_collection_digitisation_dashboard_v1.0.pdf**
>
> 1834.21 KB

Defintions are here: 2.5 Digitisation classification

1 Like

Debbie #21 April 28, 2020, 4:59am

I'm in hopes we have a sort of "software summit" to gather developers and some of the key users of these software platforms to have a decadal meeting (where have we come in 10 years, where do we want to go). In this meeting, we'd talk about needs / path to integrating TDWG DQ standards into CMS systems, interoperability needed / or opportunities for more interoperability. Example: Agents tables with Wikidata. And, we could move toward (or at least discuss feasibility) "metadata" tables in CMS to reduce the burden of creating, mapping, exporting, updating, publishing these data. This would also have the added benefit of simplifying annual reporting. As you point out @elyw , often this data must be compiled anew each time it's needed locally, let alone regionally or globally. The Global Aggregated understanding of what we have, is still far from complete, and very difficult to estimate. As we continue to digitize, our estimates will continue to improve.

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

Debbie #22 April 28, 2020, 5:01am

Thanks for joining us Rich! If you have questions about MIDS, lots of folks would be happy to discuss with you such as Elspeth Haston at RBGE. @waddink too, and me 🙂

Debbie #23 April 28, 2020, 5:09am

Hi Tim,
This is an interesting idea to explore as a possible STEP 1.
If we added CD fields to it, I could see it working.

My main observation is that many of the EML fields (being free text), often result in confusing data.

Sometimes collections describe their dataset in the EML file. Other times they describe their entire collections (which may not all be included in the dataset). And sometimes, a mix. This makes it difficult to understand what's being described in the EML w/o looking at the associated data records.

IF we could use the EML (+ CD fields) and then
a) make it clear we're looking for metadata
b) make it possible to link to any specimen-record-level datasets being shared

This could be a first pass at a workable system I think.

Even better then, once collections have metadata tables in their databases, then a view of the table data could be linked automatically to the IPT.

How might one link a collection's wikidata page into this vision of using the IPT for EML metadata files?

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

---

waddink #24 April 28, 2020, 9:54am

@rich87 , MIDS is still work in progress, you can find the latest draft here: **https://docs.google.com/document/d/1fpfn_bh2htvJl1bfDrdxd6hIZPdZnuTq9KK-cetmWRE/edit#heading=h.svg0ouedud33**. Comments are welcome!

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 #25 April 28, 2020, 11:47am

Among the various ways of providing data to the catalogue that have been mentioned in this and other topics of this consultation the integration with CDMS would be my preferred route to go.

I would like to be able to manage collection information (that is then channeled to this catalogue or other destinations) seamlessly with information on the specimens that make up these collections. Once a collection is defined in terms of the specimens that are its constituent parts or other criteria I would want automatic updates of information aggregated on the level of the collection based on the specimen-level information and I would like to be able to configure automatic publication of that collection-level data.

I would not want to compile, manage and export data on the collections' level with a second independent system.

This might be especially useful for the wider data linkages based on aggregated specimen data and corresponding services disussed in topics **2.5** and **2.6**, respectively.

Also, when collections are defined according to intensional criteria, e.g., all specimens collected by a particular collector, then such integration would automatically pick up on updated knowledge about the specimens - we expect that digitization will expose a lot of such hitherto unknown specimen properties.

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

---

neilcobb #26 April 28, 2020, 4:01pm

It is not quite that simple at least for Symbiota, it functions like Specify, Arctos, EMu etc. but it also functions like Vertnet or SpeciesLink as a primary aggregator. You can either enter data directly into Symbiota or use Excel and upload a CSV version or batch upload an EMu dataset via IPT.

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》

WUlate #27 April 29, 2020, 5:17am

From our experience developing an integrated Biodiversity Information Management System for Collections of different types, one of the requirements was to support the Collection Management, particularly the annual

estimation of specimens in the different collections. This is important because the collections are considered National Heritage in some of our countries and specimens are never given away, they are lent in perpetuity, so they need to be tracked. But this can't be accurately calculated or tracked unless all the operation was digitized, (which was never the case), there was always something in the backlog, something being prepared, something from the wet collection being mounted, something not there yet, so estimations at the Catalogue level are always necessary. Apart from the challenge of creating and maintaining totals for collections that couldn't be added, considering those loans/duplicates sent to other institutions was vital in determining how realistic the estimations were and country reports were dependant on this information at the Catalog level.

1 Like