Document: GBIF Services and Support for the Collections Catalogue

dhobern #1 April 11, 2020, 6:56am

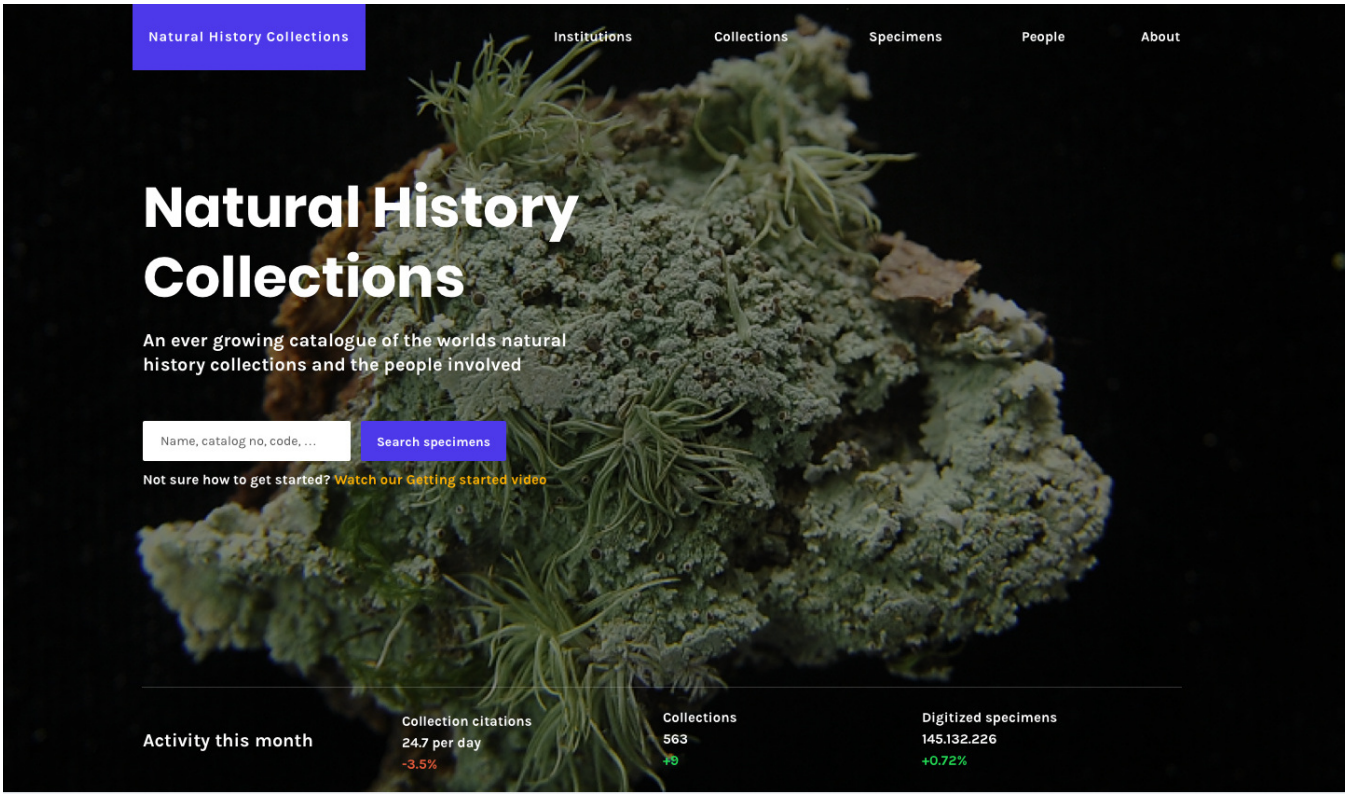
Design explorations for Collections Catalogue services

Tim Robertson, GBIF Head of Informatics

> This document is particularly relevant to discussions under the following topics:
>
> - 1.2. Locating specimens and genetic materials (USE)
> - 1.5. Increased value for data on specimens, taxonomic publications, etc. (USE)
> - 1.7. Foundation for new and enriched services (USE)
> - 1.8. Improvements to citation and visibility for collections (USE)
> - 2.5. Wider data linkages (INFORMATION)
> - 2.6. Information services relating to collections (INFORMATION)
> - 3.3. Integrated catalogue (TECHNOLOGY)
> - 3.5. Interfaces, APIs and client modules (TECHNOLOGY)
> - 4.3. Technical infrastructures (GOVERNANCE)

As explained in the presentation on the Global Registry of Scientific Collections (GRSciColl), GBIF has migrated this resource to become part of the GBIF infrastructure. It now possible to explore the future evolution of the GBIF/GRSciColl Collection Catalogue and to plan the set of services that it should provide.

To prepare for this evolution, we have sketched a design to show what we believe could be developed to build upon the feeds of information that GBIF already offers.



Please click on the snapshot above and browse the design. The snapshot links to an image with embedded links that you can use to explore the presentation presented. Please click around to see what links are available from each view.

Here are some important aspects to understand about the design shown in this sketch.

1. GBIF would ensure that any implemented version of this catalogue has full multi-lingual support and a friendly API that enables developers to reuse the information and build new applications.
2. The catalogue would integrate authoritative information and data feeds linking **Institutions**, **Collections**, **People**, **Literature** and **Specimens**.
3. **This example** illustrates a possible **Collection** search, based on a search index that combines collection metadata with available digitised specimen information.
   - The **Collection detail view** shows that **Collection Description standard** information would be available in addition to **any other (non-standard) documentation** such as a spreadsheet containing an inventory of holdings.
   - The Collection can be linked to **People** associated with the collection and a dashboard showing **researcher interactions with the collection data**.
   - A range of **Citations** can be presented, including those from literature, taxonomic treatments and data use.
   - Information on **Funders** is important to meet instiutional obligations and to showcase the scientific significance of the collection.
   - Many collections (~3000 from 90+ countries) already publish ~200M digital specimen records through **GBIF.org** using open standards. This catalogue proposes to integrate **specimen search**, **mapping capabilities**, image galleries and download tools as services tied to the Collection view. This will lower the technical threshold for any collection to showcase its holdings and should make many more collections explorable online.
4. The **Specimen detail view** illustrates an intention to move beyond simply representing “occurrence records”, as portals like **GBIF.org** typically do today. This example shows how data linkages and data-clustering algorithms can help to identify and bring together closely related specimens (e.g. duplicates distributed across several herbaria), links to DNA sequence data (which may be published through a different database), images, measurements, citations and other information.

Use this topic to **ask questions, share ideas or discuss features** related to this design. If there is enough interest, we will hold a Zoom session to explore it in more detail.

1.9. Support for national and regional needs and applications (USE)

1.9. Support for national and regional needs and applications (USE)

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

---

Adelantando el Catálogo de Colecciones de Historia Natural del Mundo

---

3.2. Community catalogues (TECHNOLOGY)

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》

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》

Advancing the Catalogue of the World's Natural History Collections

---

Integrated summary from 17 to 30 April 2020

建立《全球自然历史馆藏名录》

1.2. Locating specimens and genetic materials (USE)

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

---

pzermoglio #2 April 21, 2020, 11:09am

First, I really like how this looks!
Some comments as I went through the different features:

- The getting started video placeholder got me thinking: for which kinds of users? That is, to navigate the site or to contribute to it? or both? Who would we expect to interact with the portal and how? (related to **Topic 1.1**)

- Search box currently says Search specimens. I guess I would rather put the focus on the collections upfront, and specimens as secondary to it.

  Related, I think it will be important to understand which are the limits, and the overlaps, between the NHC site and what we can find in e.g. GBIF. Not sure about those borders, and not sure about a best way to communicate them to users (both searching users and provider users). We probably want to emphasize that

both resources do not compete, but rather complement each other. One way to consider would be to have links on any given specimen page saying "view in GBIF". (Related **Topic 2.5**.)

That leads me to the question on how are the specimens fed into this NHC portal. Are they meant to be managed completely separated from what's already available through aggregators? Or, is it the idea to harvest e.g. all preserved specimens already in GBIF and deploy them grouped by collection? From the description above, I understand that the latter would be the case, as much as possible. If so, info about some specimens may not be in any aggregator yet. Would the path be to publish through an aggregator --> harvest --> show in NHC portal? or would there be an alternative direct path into the NHC?

- I'd be very interested in understanding how the clustering would work. I can imagine different features to cluster on, like the ones mentioned in the description, and e.g.: collection by taxon (same or close taxa in the same collection), specimens by collectors, etc. Depending on what one would want to do, clustering may need to be based on different criteria. Could clustering have filters/criteria associated to choose from? That is, will the user be able to choose parameters that would change the algorithm parameters, or even the algorithms themselves?
- Collections. When I look at the collections list I would like to see value highlighted. For instance, a little (colorful) icon representing that a collection contains very old records, or endemic species, etc. Indicators of the singularities of any given collection that makes it unique (Somewhat related **Topic 1.4**)
- I really like the map of collections that meet certain criteria. Visually very appealing. (maybe something obvious, but I would expect each little dot to display an info box with minimal data and a link to the collection's page).
- I also really like the collectors and identifiers page, with the activity graphs. May be beyond, but I would like to be able to search by collector on a map over time.

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

---

mhoefft #3 April 21, 2020, 1:07pm

Hi Paula
I'm happy to hear you like it! I do not believe any of your questions have an answer yet.

Data sources
The focus, when thinking about this, was:
What could we do with the data already mediated through the GBIF network if linked with GRSciColl (in some version) and TDWG Collection Descriptors (in some version). That isn't to say that it wouldn't benefit from more/other datasources/aggregators/links than those three.

But if GRSciColl is considered an anchor point, GBIF would then try to link mediated data and provide services to do those links: e.g. "I have this code, please give me a DOI for the collection".

The GBIF data streams (e.g. occurrences, treatments, citations and hopefully specimen clusters) could just be one way to enrich the somewhat dull inventory that GRSciColl currently is.

Others that cared to link to GRSciColl, could enrich the collections as well. E.g. rich interactive dashboards based on NCDs could be another (developed and indexed by someone somewhere?). CETAF have rich lists of publications by institution - if we shared identifiers, that could be another datastream that could enrich a collection focused site.

Overlap with GBIF
`GBIF.org` will still exist, but there are som many ways to look at this data, and having focused views (such as a collections view) seems useful. `GBIF.org` is suffering from having much, rich and diverse data and a diverse audience. By focusing on specific audiences - here the NHC - we believe we can present the data better.

Currently we have views for presenting occurrence data (table, map, metrics, gallery). But the data graph as a whole could benefit from focused views (tracking data, sampling data and - of course - collection data).

1 Like

---

waddink #4 April 23, 2020, 8:10pm

The mockups seem to rather be a service on top of a collections catalogue, showing what you can do when combining collection descriptions with other information, than the bare catalogue itself. What I miss in the mockups and should be part of a collection descriptions catalogue is detailed provenance information who contributed what to a collection description.

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