BISS Biodiversity Information Science and Standards

OPEN ACCESS

Conference Abstract

GRSciColl: Registry of Scientific Collections maintained by the community for the community

Marie Grosjean‡, Morten Høfft‡, Marcos Lopez Gonzalez‡, Tim Robertson‡, Andrea Hahn‡

‡ Global Biodiversity Information Facility, Copenhagen, Denmark

Abstract

GRSciColl, the Registry of Scientific Collections, is a comprehensive, community-curated clearinghouse of collections information originally developed by the Consortium of the Barcode of Life (CBOL) and hosted by the Smithsonian Institution until 2019. It is now hosted and maintained in the Global Biodiversity Information Facility (GBIF) registry (see this news item).

GRSciColl aims to improve access to information about institutions, the scientific collections they hold, and to facilitate access to the staff members who manage them. Anyone can use GRSciColl to search for collections based on their attributes (country, preservation type, etc.) as well as their codes and identifiers. These users will find information on what the collections contain, where they are located, who manages them and how to get into contact. Furthermore, institutions can use GRSciColl to be more visible and advertise their collections, both digitized and undigitized. Plus, the ability to get an overview of institutions and collections by country can help guide some of the data mobilisation efforts by national organizations. Finally, GRSciColl is a reference for institution and collections codes and identifiers, which can enable links from other systems (as exemplified in GBIF.org) and make the information more easily available.

Engaging the community is crucial in maintaining that information. After the migration to GBIF, the first phase of development focused on data consolidation, integration with external systems, and on providing the necessary functionality and safeguards to move

from a centrally maintained to a community curated system. With all these in place, the focus is now shifting to expanding the community of data editors, and to understanding how best to serve user needs.

It can be difficult for institutions to maintain information in the various available data repositories. This is why we aim to synchronize GRSciColl with as many reliable sources as possible. In 2020, we set up weekly synchronization with Index Herbariorum, and we will be exploring synchronization with other sources such as the Consortium of European Taxonomic Facilities (CETAF) registry and the National Center for Biotechnology Information (NCBI) BioCollections database. In addition, we worked with the team at Integrated Digitized Biocollections (iDigBio) to import their collection information into GRSciColl. The data are now maintained in the GBIF registry and displayed on the iDigBio portal via the GRSciColl Application Programming Interface (API).

The GRSciColl new permission model aims to facilitate community curation. Anyone can suggest updates, and those changes can be applied or discarded by the appropriate reviewers: institution editors, country mediators, or administrators.

With these changes in place, in 2021, we reached out to the GBIF Network to increase our pool of editors. Many GBIF Node managers are now involved in the curation of GRSciColl, and we are planning to likewise include applicants for the GBIF-managed funding programs such as "Biodiversity Information for Development'' (BID) and "Biodiversity Information Fund for Asia" (BIFA). We also work with external collaborators, such as the Biodiversity Crisis Response Committee of the Society for the Preservation of Natural History Collections (SPNCH), to reach outside of the GBIF community.

Alongside the support for data integration and curation, a second important aspect is the support for data use. The information available needs to be both accessible and relevant to the community.

Specimen-related occurrences, published on GBIF.org, are cross-linked to GRSciColl entries whenever possible (see this example). As these links make use of collection and institution identifiers within individual specimen records, rather than relying on dataset entities, this procedure allows aggregation of specimen-related occurrences under their GRSciColl-registered collections and institutions, regardless of the way they were published on GBIF. This can help users and institutions get an overview of the collections digitization progress, whether through their own initiative, or from datasets contributed by other data publishers.

The Collections API is under ongoing development to provide better ways to access the GRSciColl information: more filters, a way to download the result of a search, and an API (Application Programming Interface) lookup service to find institutions and collections associated with a given code or identifier. The latter was designed to improve database interoperability.

By working together with the community, we want to ensure GRSciColl becomes and remains a tool they can rely on.

There are many ways to get involved with GRSciColl:

- Anyone can check their institution and collection entries and suggest updates or additions via the suggestion buttons in the GRSciColl interface.
- You can become a registry editor on behalf of your institution or collection.

- If you work with a National registry and are interested in sharing the data on GRSciColl, please contact us at scientific-collections@gbif.org.

- Tell us how you would like to use the registry and GRSciColl. You can contact us by email (scientific-collections@gbif.org) or via our GitHub repository.

- You can become a volunteer translator to make the GRSciColl forms accessible in more languages.

- You can follow our 2021 Roadmap and log your feedback and ideas via the GBIF feedback system or directly on our GitHub repository.

Keywords

natural history, archeology, earth planetary records, collection code, institution code, collection identifier, institution identifier

Presenting author

Marie Grosjean

Presented at

TDWG 2021

Acknowledgements