

D5.5: WHITEPAPER ON PARTNERING WITH OTHER INFRASTRUCTURES TO ENSURE SUSTAINABLE DELIVERY OF SYNTHESYS+ SERVICES

MARIE GROSJEAN (GBIF), TIM ROBERTSON (GBIF) AND NIELS RAES
(NATURALIS)

Grant Agreement Number | 823827

Acronym | SYNTHESYS PLUS

Call | H2020-INFRAIA-2018-2020

Start date | 01/02/2019

Duration | 54 months

Work Package | NA5: Internationalisation and engagement of new user
communities in EC priority areas

Work Package Lead | H Goodson

Delivery date | 31.01.2023



SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

Contents

| | |
|--------------------------|---|
| Summary | 2 |
| Background..... | 2 |
| Roadmap..... | 3 |
| Future development | 4 |

Summary

Deliverable 5.5 asked partners to examine how to partner with other infrastructures to ensure the sustainability of SYNTHESYS+ services, namely, access to and digitisation of natural history collections across Europe and priority areas. One of the biggest challenges to access, is simply knowing what collections exist where. In order to effectively deliver on this task, we decided to examine what needed to be developed to progress a consolidated view of what we have, in a way that would continue on after the SYNTHESYS+ project and therefore be a baseline service for continuation of other services and tools developed in this project.

Background

With digitisation of natural history collections over the past decades, their traditional roles— for taxonomic studies and public education — have been greatly expanded into the fields of biodiversity assessments, climate change impact studies, trait analyses, sequencing, 3D object analyses etc. (Nelson and Ellis 2019; Watanabe 2019). Initial estimates of the global natural history collection range between 1.2 and 2.1 billion specimens (Ariño 2010), of which 169 million (8-14% - as of April 2019) are available at some level of digitisation through the Global Biodiversity Information Facility (GBIF). With iDigBio (Integrated Digitized Biocollections) established in the United States and with the European DiSSCo (Distributed Systems of Scientific Collections) accepted on the ESFRI roadmap, it has become a priority to digitize natural history collections at an industrialized scale. Both iDigBio and DiSSCo aim at mobilising, unifying and delivering bio- and geodiversity information at the scale, form and precision required by scientific communities, and thereby transform a fragmented landscape into a coherent and responsive research infrastructure. In order to prioritise digitisation based on scientific demand, and efficiency using industrial digitisation pipelines, it is required to arrive at a uniform and unambiguously accepted collection description standard that would allow comparing, grouping and analysing natural history collections at diverse levels.



SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

SYNTHESYS+
Synthesis of Systematic Resources

a DiSSCo project

Several initiatives attempt to unambiguously describe natural history collections using taxonomic and storage classification schemes. These initiatives include One World Collection, Global Registry of Scientific Collections (GRSciColl), TDWG (Taxonomic Databases Working Group) Natural Collection Descriptions (NCD) and CETAF (Consortium of European Taxonomy Facilities) passports, among others. In a collaborative effort of DiSSCo, ICEDIG (The Global Registry of Scientific Collections (GRSciColl) is a comprehensive, community-curated clearing house of information on scientific collections, which builds on previous efforts developed by the Consortium of the Barcode of Life (CBOL) and others.

GRSciColl includes data on institutions, collections and associated staff members and spans all scientific disciplines, including earth and space sciences, anthropology, archaeology, biology and biomedicine, as well as applied fields like agriculture, veterinary medicine and technology.

GRSciColl also serves as the registry for InstitutionCodes and CollectionCodes—elements used in the DarwinCore data standard used in the biodiversity informatics community. Use of these terms enables publications and databases to point unambiguously to collections and their contents.

The main goals of GRSciColl are to:

- Improve access to information about scientific collections, their host institutions as well as contacts for key collections and curatorial staff
- Improve interoperability among databases by providing unique codes and machine-readable identifiers for institutions and collections

Roadmap

The current phase of development has opened the registry to validated users for submissions and edits. Subsequent stages will engage other partners to add functionality like linking collections to specimens, tracking DOI-based citations of collections, and supporting or enhancing various collection metadata schemas.

Staff members from institutions and collections originally contributed the information contained in GRSciColl, either directly or through partners and predecessors like the NCBI Institution table and Collection table, Biorepositories.org, or the Biodiversity Collections Index.

Maintenance and development of GRSciColl is an ongoing task, and key priorities are outlined in this road map.

As part of the roadmap development, GBIF worked with partners to establish a standardized, interoperable flow of information between this registry and other key collections infrastructures. GRSciColl runs a weekly synchronization with Index Herbariorum (IH). Following content migration from iDigBio Collections, GRSciColl now shares an integrated registry delivering consistent information with shared editing access to users of both GBIF.org and iDigBio.org.

Within this period of work GBIF has also restored community-curation functionality, enabling those working within each of the institutions and collections to help maintain up-to-date information in the



SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

SYNTHESYS+
Synthesis of Systematic Resources
a DiSSCo project

registry. Most significantly, we linked occurrences from GBIF.org with institution and collection codes and identifiers with GRSciColl.

This roadmap is outlined below. A comprehensive and dynamic version is also available on github [here](#). This builds on the [2021 roadmap](#) as well as the efforts in 2022 to build a community of editors and mediators and takes queues from the consultation and report delivered as part of 5.1, Delivering a roadmap for Integrated Implementation of Priority Component Areas.

[Establish a new public site for GRSciColl \(rebranded, outside of GBIF.org\)](#)

This build on the work mentioned in the 2021 roadmap of developing a richer user interface and the need to make GRSciColl more visible.

[Enable specimen search and browse capability to the catalogue](#)

In addition to making specimen-related occurrences linked to GRSciColl searchable on the new public website, the catalogue should also support the richer data model as it becomes available.

[Make GRSciColl compatible with Latimer Core \(TDWG Collection description standards\)](#)

GRSciColl should be compatible with Latimer Core [[REG-176](#)].

[Explore making GRSciColl a better tool to monitor data mobilisation efforts.](#)

Explore how GRSciColl could be used to generate national views and tracking trends in available specimen data (i.e. output of digitisation effort) published through GBIF.

Future development

This is a non-exhaustive list of items that are recommended to be addressed in the longer term to continue to ensure sustainable delivery of SYNTHESYS+ services through being able to search and discover collections across infrastructures, and continually add to the number and richness of those collections in a way that benefits an aggregate of infrastructures.

Explore adding DOIs or similar to institutions and collections [[REG-320](#)]

Explore synchronization with the NCBI BioCollections [[REG-307](#)]

Explore synchronization with CETAF [[REG-322](#)]

Explore integration of external identifiers [[REG-274](#)]



SYNTHESYS+ was funded by the Horizon 2020 Framework of the European Union under the H2020 Open Innovation and Open Science Research Infrastructure call.

SYNTHESYS+
Synthesis of Systematic Resources
a DiSSCo project