Introduction to GBIF

GBIF Secretariat

Versão 2, Aug 2022

Índice

Course description

This course provides an introduction to GBIF, the data available in GBIF's portal, accessing that data and information about engaging with GBIF and its community of practice.

Topics include:

- Information about GBIF
- GBIF-mediated data
- Data publishing
- Data access
- GBIF community of practice

This course is comprised of video and written instruction and paired with quizzes and practical exercises.

Audience

This course is designed for individuals who work in biodiversity research or policy institutions and contribute to or use data from the GBIF portal. The instruction provided is particularly useful for those who have a desire to know more about GBIF and its place in the biodiversity community and want to engage further with GBIF.

Prerequisites

There are no prerequisites for this course. This course serves as a prerequisite for other GBIF courses.

Participants should have a good command of English. While efforts are made to provide materials in other languages, instruction videos will be in English with subtitles in other languages.

About GBIF

The Global Biodiversity Information Facility (GBIF) is an international network of country and organizational Participants that exists to enable free and open access to biodiversity data from all sources and to support biodiversity science, environmental research, and evidence based decision-making. GBIF operates as a federated system of distributed data publishing efforts, coordinated through a global informatics infrastructure and collaborative network. In this module, you will learn more about GBIF.

What is GBIF?

In this video (03:19) you will learn about GBIF. It is a co-production of SiB Colombia and GBIF Spain, both GBIF national nodes, with co-funding from GBIF Spain and Instituto de Investigación de Recursos Biológicos Alexander von Humboldt (IAvH). If you are unable to watch the embedded Vimeo video, you can download it locally. (MP4 - 11.3 MB)

https://vimeo.com/661945151 *(Vimeo video)*

If you see the CC symbol on a video, you can click it to enable closed captioning/subtitles in English or other available languages.

Introduction to GBIF

In this video (07:55), Tim Hirsch, Deputy Director of the GBIF Secretariat, provides you with an overview of GBIF. If you are unable to watch the embedded Vimeo video, you can download it locally. (MP4 - 32.3 MB)

https://vimeo.com/434831655 *(Vimeo video)*

Introduction to GBIF Participant Nodes

Since GBIF's founding in 2001, the participating countries and organizations have been testing and developing models for coordinating the mobilization, management and reuse of biodiversity data at the national level or within an organization's scope. The formation of Participant 'nodes' has been central to these efforts. Designated by each Participant, these teams coordinate the needs and interests of the many stakeholders involved. In this video (07:53), we look at the variety of node models in the GBIF network to present generalized concepts relating to Participant nodes and the roles that they play in the GBIF community. If you are unable to watch the embedded Vimeo video, you can download it locally. (MP4 - 14.0 MB)

https://vimeo.com/543599833 *(Vimeo video)*

Further guidance for Nodes

- Establishing an effective GBIF Participant Node
- Online capacity self-assessment tool for national biodiversity information facilities

Countries/Economies in GBIF

Investigate how your country/economy is represented in GBIF

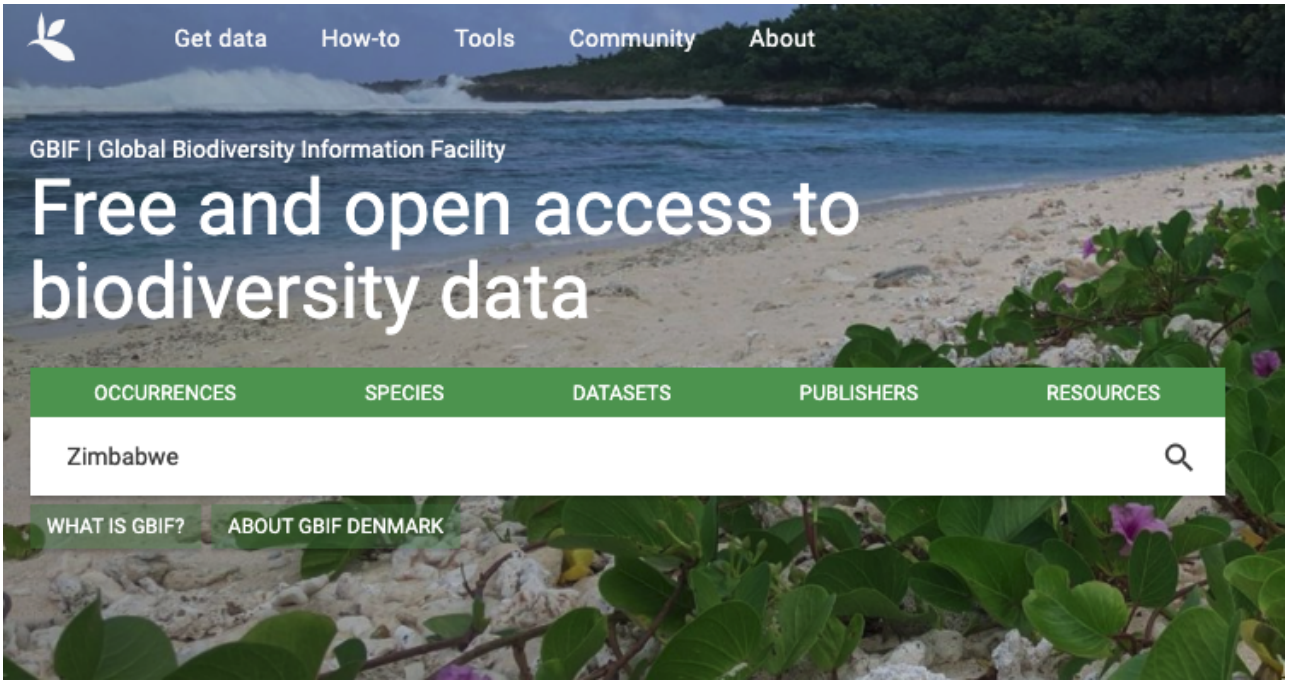


Figura 1. The GBIF website has pages dedicated to countries/economies, including for countries/economies that do not yet participate in GBIF. You can use the search box on the homepage to look up a country/economy name.

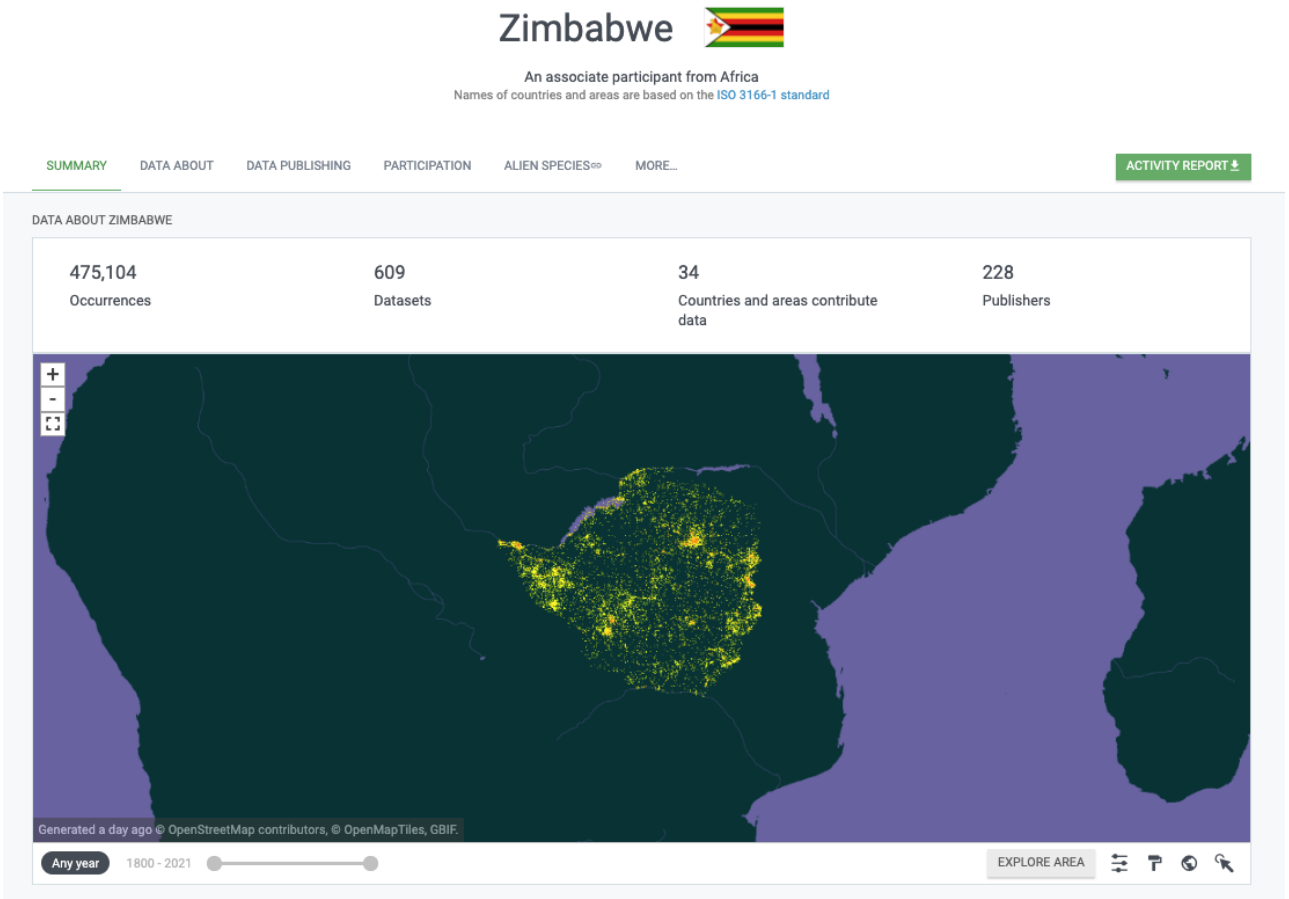


Figura 2. The tabs on these pages provide a general description of the data available about the biodiversity of the country, any data published by national institutions, as well as other relevant information on the use of data by researchers in the country.

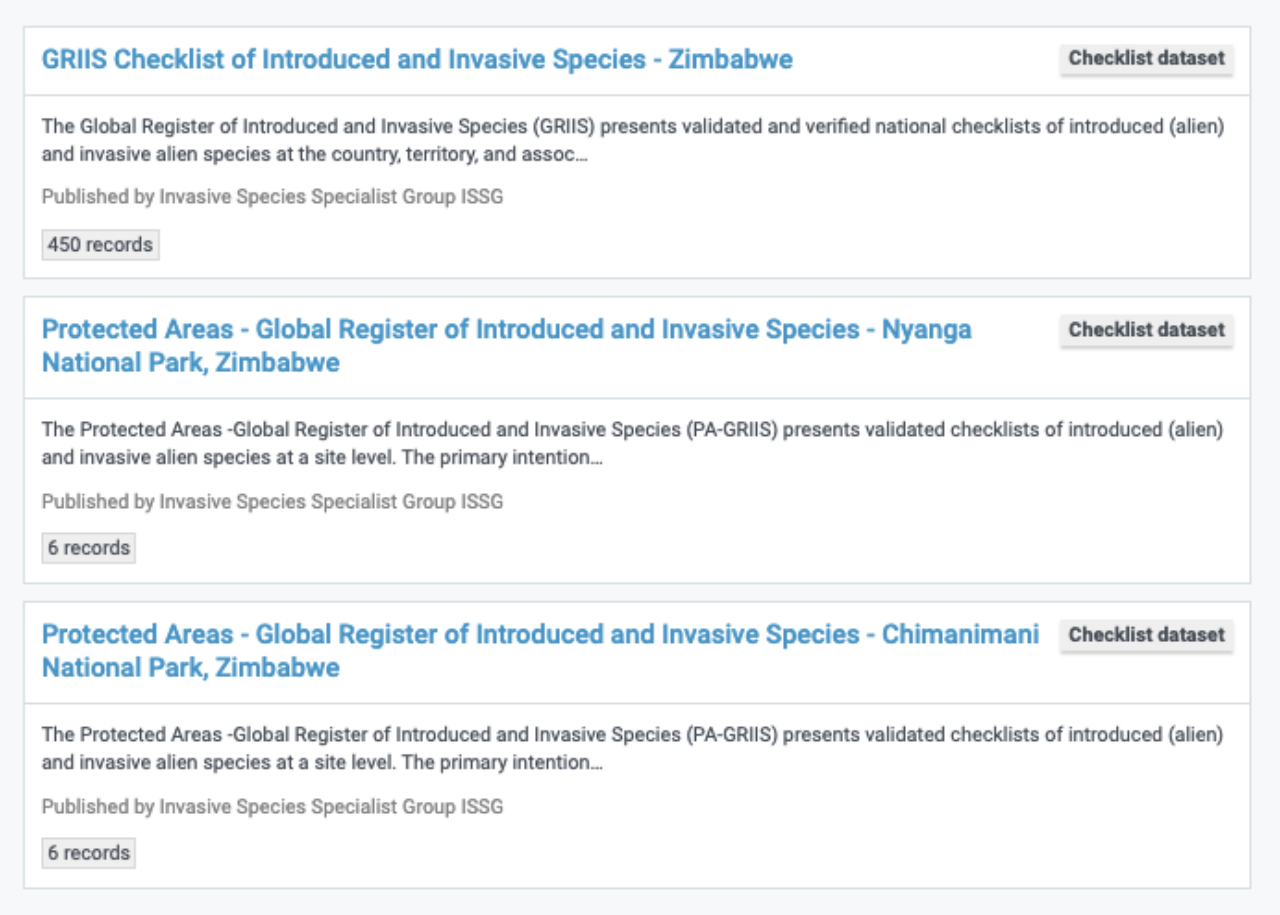


Figura 3. A recent collaboration with the IUCN Invasive Species Specialist Group means that, for many countries, a checklist of introduced and invasive alien species is available from the Global Register of Introduced and Invasive Species (GRIIS).

1. Look up your country on the GBIF website.
2. Does your country participate in GBIF?
3. How many institutions are publishing data?
4. Is a list of introduced and invasive alien species available for your country?
5. How well does the information available on GBIF reflect what you know about the biodiversity of your country?

Review

Quiz yourself on the concepts covered in this module.

1. What is GBIF?
   - ☐ An intergovernmental network and research infrastructure
   - ☐ A collaboration among governments and international organizations
   - ☐ A network of participant nodes
   - ☐ A secretariat, based in Copenhagen, Denmark
   - ☐ All of the above

2. When was GBIF established?

   - ☐ 1992
   - ☐ 1999
   - ☐ 2001

3. Which of the following is the best description of a GBIF Participant node?

   - ☐ A GBIF national office, funded by the GBIF Secretariat
   - ☐ A team designated by a Participant country or organization to coordinate a network of people and institutions that produce, manage and use biodiversity data, collectively building an infrastructure for delivering biodiversity information
   - ☐ A regional hub for expertise in biodiversity data mobilization and data use
   - ☐ The informatics infrastructure that connects with GBIF.org to enable a Participant country or organization to publish biodiversity data

4. Which of the following is NOT a typical function of a GBIF Participant node?

   - ☐ Coordinating a community of initiatives relating to biodiversity information, including making connections to the international GBIF network
   - ☐ Promoting and supporting the mobilization of biodiversity data within the country or organization's scope so that as many sources as possibly are freely and openly available
   - ☐ Encouraging the reuse of the available data to support biodiversity-related science and support decision-making for sustainable development
   - ☐ Providing expertise on biodiversity data management and improving data quality to support users' needs
   - ☐ Maintaining a mirror website of the GBIF.org to ensure real-time backup of the GBIF data index and improve user access from within the country

5. What is a GBIF Participant?

   - ☐ The person designated by a participating country/economy/organization to manage the activities of the node to coordinate a biodiversity information facility
   - ☐ A country, economy or organization that joins GBIF by signing the Memorandum of Understanding and establishing a co-ordinated effort to support open access and use of biodiversity data, to advance scientific research, and to promote technological and sustainable development
   - ☐ The broader structure of people and institutions, coordinated by the node, that collectively forms an infrastructure for delivering biodiversity information to relevant stakeholders
   - ☐ The person designated by the participating country/economy/organization to act as its representative to the GBIF Governing Board and take part in the global-level decision making

6. What is a GBIF Head of Delegation?

   - ☐ The person designated by a participating country/economy/organization to manage the activities of the node to coordinate a biodiversity information facility
   - ☐ A country, economy or organization that joins GBIF by signing the Memorandum of Understanding and establishing a co-ordinated effort to support open access and use of

biodiversity data, to advance scientific research, and to promote technological and sustainable development

- ☐ The broader structure of people and institutions, coordinated by the node, that collectively forms an infrastructure for delivering biodiversity information to relevant stakeholders
- ☐ The person designated by the participating country/economy/organization to act as its representative to the GBIF Governing Board and take part in the global-level decision making

7. What is a Biodiversity information facility?

- ☐ The person designated by a participating country/economy/organization to manage the activities of the node to coordinate a biodiversity information facility
- ☐ A country, economy or organization that joins GBIF by signing the Memorandum of Understanding and establishing a co-ordinated effort to support open access and use of biodiversity data, to advance scientific research, and to promote technological and sustainable development
- ☐ The broader structure of people and institutions, coordinated by the node, that collectively forms an infrastructure for delivering biodiversity information to relevant stakeholders
- ☐ The person designated by the participating country/economy/organization to act as its representative to the GBIF Governing Board and take part in the global-level decision making

8. What is a Node manager?

- ☐ The person designated by a participating country/economy/organization to manage the activities of the node to coordinate a biodiversity information facility
- ☐ A country, economy or organization that joins GBIF by signing the Memorandum of Understanding and establishing a co-ordinated effort to support open access and use of biodiversity data, to advance scientific research, and to promote technological and sustainable development
- ☐ The broader structure of people and institutions, coordinated by the node, that collectively forms an infrastructure for delivering biodiversity information to relevant stakeholders
- ☐ The person designated by the participating country/economy/organization to act as its representative to the GBIF Governing Board and take part in the global-level decision making

9. Who designates the institution that hosts the GBIF Participant node?

- ☐ The Head of Delegation
- ☐ The GBIF Secretariat

GBIF-mediated data

In this module, you will learn about primary biodiversity data and the principles that GBIF follows with regards to data. You will also have a chance to review the various metrics that are available for the data within the portal.

Primary biodiversity data

In this section, you will learn how GBIF makes primary biodiversity data accessible, the accepted dataset types and how GBIF uses the taxonomic backbone to provide taxonomic information.

When we refer to primary biodiversity data, we mean the data that document where and when species have been recorded. This knowledge derives from many sources, including everything from museum specimens collected in the 18th and 19th century to geotagged smartphone photos shared by amateur naturalists in recent days and weeks.

The GBIF network draws all these sources together through the use of data standards, such as Darwin Core, which forms the basis for the bulk of GBIF.org's index of hundreds of millions of species occurrence records. Publishers provide open access to their datasets using machine-readable Creative Commons licence designations, allowing scientists, researchers and others to apply the data in hundreds of peer-reviewed publications and policy papers each year. Many of these analyses—which cover topics from the impacts of climate change and the spread of invasive and alien pests to priorities for conservation and protected areas, food security and human health— would not be possible without this.

GBIF dataset classes

We encourage data holders to publish the richest data possible to ensure their use across a wider range of research approaches and questions, but not every dataset includes information at the same level of detail. Sharing what is available through GBIF.org is valuable, because even partial information answers some important questions.

The four classes of datasets supported by GBIF start simply and become progressively richer, more structured and more complex.

- Meta-data only - datasets describing **undigitized** resources like those in natural history and other collections
- Checklist - a **catalogue** or list of named organisms, or taxa
- Occurrence - the evidence of the **occurrence of a species** (or other taxon) at a particular place on a specified date. Occurrence datasets make up the core of data published through GBIF.org
- Sampling-event - offering evidence that a species occurred at a given location and date, but also making it possible to **assess community composition** for broader taxonomic groups or even the **abundance of species** at multiple times and places.

More information on dataset classes can be found on the GBIF website.

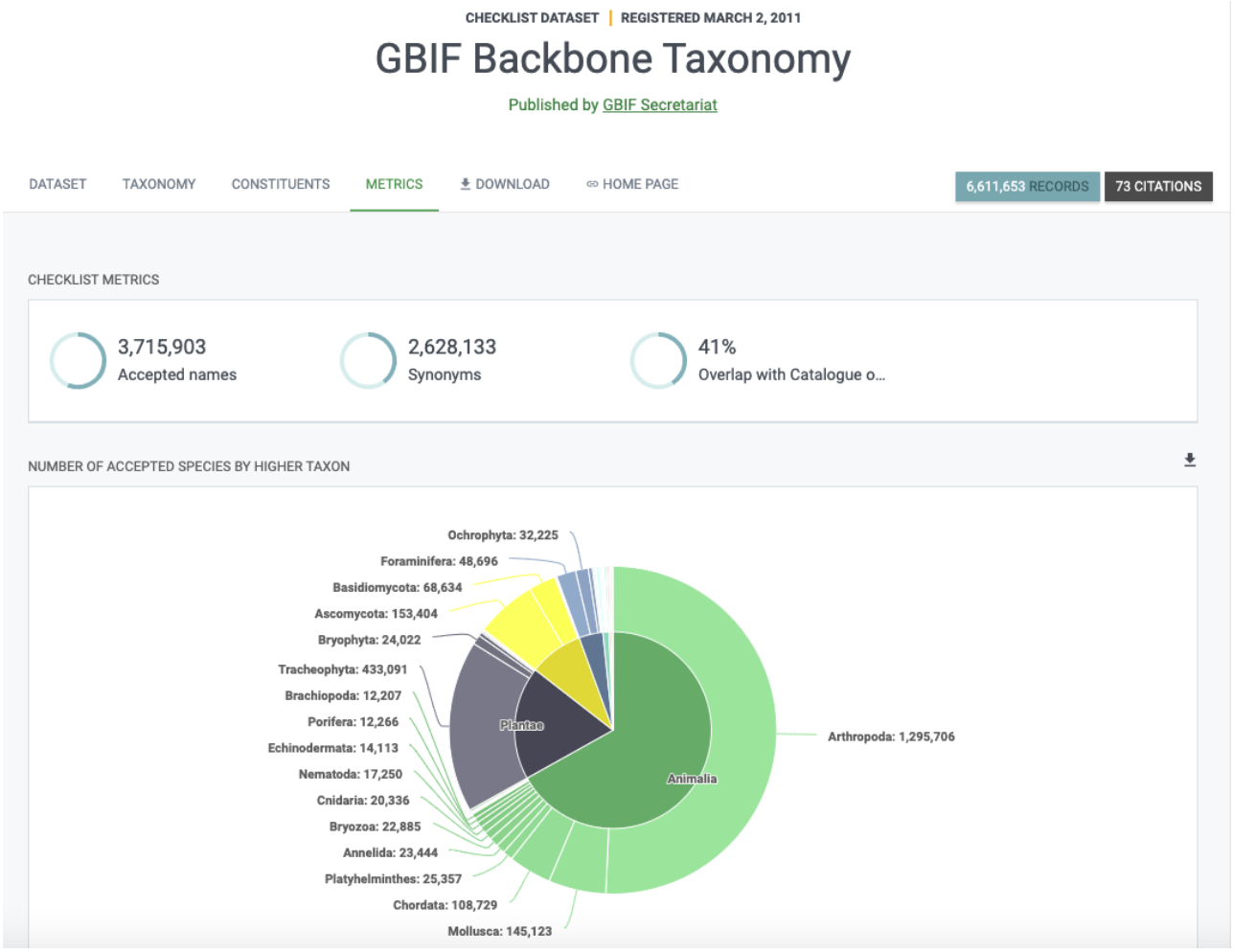
You might also want to explore how to choose a dataset type.

GBIF taxonomic backbone

What is the GBIF Backbone taxonomy?

The Backbone taxonomy is actually a GBIF dataset. But not just any dataset, it is probably the most important dataset for GBIF. On its page, it is defined as:

> a single synthetic management classification with the goal of covering all names GBIF is dealing with



Why does GBIF need a backbone?

The backbone is needed to organize the data available on GBIF. Without it, we wouldn't be able to do any taxonomic search and it would be difficult to generate consistent statistics and maps.

As you can imagine, not everyone uses the same classifications or names. This results in considerable variations in higher taxa and a large number of synonyms. The backbone aims to bring all these names together and organize them.

How is the backbone generated?

The backbone is built from other checklists. These include:

- 55 authority checklists,
- a checklist generated from the type specimens shared on GBIF,

- two large sources for stable Operational Taxonomic Units (OTUs): iBOL Barcode Index Numbers and the UNITE Species Hypothesis identifiers,
- and any checklist shared by PLAZI.org on GBIF (currently 27,054 but not all these were available when the backbone was generated).

These checklists are ordered by priority starting with the Catalogue of Life for most taxa. This order is crucial as it shapes the taxonomy.

Note that many sequence-based occurrences have no Latin names but are named using species hypotheses (UNITE: fungi) or Barcode Index Numbers (iBOL: primarily animals). This is why adding these two major sources of stable OTUs to the latest backbone taxonomy significantly improves GBIF's indexing functionality for sequence-based biodiversity data.

The information above is an excerpt from a 2019 blog post by Marie Grosjean. Read the blog post for more detail on the backbone.

Principles of GBIF-mediated data

In this section, you will learn about the principles that GBIF follows with regards to data and how data in the GBIF portal are FAIR.

Digital object identifiers

A Digital Object Identifier, or DOI, is a standard, permanent identifier that provides an actionable, interoperable, persistent link to any entity. The concept is that DOI differs from commonly used references like URL web links because it identifies an object itself as a first-class entity, not simply the place where the object is currently located.

In the context of GBIF.org, DOIs serve as stable identifiers for four different types of things:

1. datasets from the GBIF network
2. data downloads from GBIF.org
3. research articles and reports published by scientific journals, agencies and NGOs
4. materials deposited in a general-use repository

GBIF assigns DOIs to all datasets and occurrence downloads. When data is used, following DOI citation practice ensures an easy and consistent way of crediting dataset holders while also allowing for reproducibility. The DOIs will always resolve to dataset or download pages, even if the underlying data is no longer available.

GBIF started issuing DOIs on 3 February 2015. Downloads requested before this date do not have DOIs, however, if you wish to cite older downloads, you can contact helpdesk@gbif.org and we will assign DOIs as appropriate.

Standards

The data available through GBIF.org and its associated services is the result of the GBIF network of Participants and publishers applying shared rules and conventions to describe, record and structure thousands of different datasets drawn from hundreds of institutions around the world. Common

standards are the main enabler for bringing together the hundreds of millions of primary biodiversity records in the GBIF index.

Within the biodiversity domain, the group most often responsible for developing and maintaining data standards is Biodiversity Information Standards. This nonprofit scientific and educational association focuses on the development of standards for the exchange of biological and biodiversity data. Members of the biodiversity community generally refer to this group as TDWG (pronounced tad-wig)—a vestigial reminder of its earlier manifestation as the Taxonomic Databases Working Group.

Commonly use standards include:

- Darwin Core: The Darwin Core Standard (DwC) offers a stable, straightforward and flexible framework for compiling biodiversity data from varied and variable sources. The majority of the datasets shared through GBIF.org are published using the Darwin Core Archive format (DwC-A).
- Ecological Metadata Language (EML): Ecological Metadata Language is a metadata standard that records information about ecological datasets in a series of modular and extensible XML document types. All of the descriptions of datasets in GBIF.org rely on 'metadata'—that is, the information about data—using the open-source EML standard, which is administered and maintained by The Knowledge Network for Biocomplexity. Each Darwin Core Archive includes as one of its components an EML file (written in XML format).
- BioCASe/ABCD: The Biological Collection Access Service, commonly referred to as BioCASe, is an international network linking biological collections data from natural history museums, botanical/zoological gardens and research institutions. The BioCASe protocol relies on the Access to Biological Collections Data (ABCD) data exchange standard, which TDWG also administers.

Open data

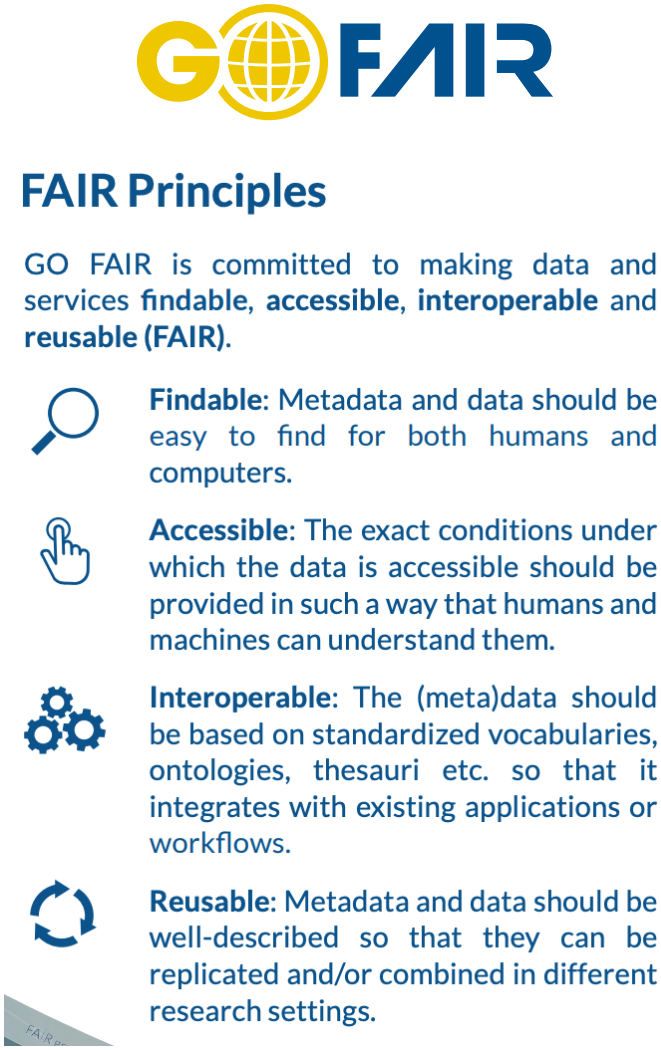
In keeping with a 2014 decision by the GBIF governing board, data publishers must assign one of the three Creative Commons options to any occurrence dataset. The Governing Board recognized the need for much greater clarity both for data publishers and users on how data may be used when shared via GBIF.org. Creative Commons is a nonprofit organization that helps overcome legal obstacles to the sharing of knowledge and creativity to address the world's pressing challenges.

- CC0 - no conditions for use
- CC-BY - use with attribution
- CC-BY-NC - non-commercial use with attribution

Note that the CC-BY-NC licence has a significant effect on the reusability of data. GBIF encourages data publishers to choose the most open option they can wherever possible. It is important to note that images are not subject to the same licence that is applied to the dataset and may have more restricted terms of use. Lastly, attribution/citation is a community norm, so even if the publishers has waived conditions for use, attribution is expected.

FAIR data

Many articles from 2011-2016 documented a crisis in scientific reproducibility (see below). In 2016, the FAIR Guiding Principles for scientific data management and stewardship were published in Scientific Data. The principles were designed to improve the Findability, Accessibility, the Interoperability and the Reusability of datasets and address "an urgent need to improve the infrastructure supporting the reuse of scholarly data." Implementation of these principles began in 2018. You can read more about How to GO FAIR on GO-FAIR.org.



Data found on GBIF.org are FAIR.

FINDABLE

GBIF has requirements for metadata and datasets. All datasets are identified by Digital Object Identifiers (DOIs).

ACCESSIBLE

The GBIF Portal API provides a machine readable interface (REST + JSON) and use the Integrated Publishing Toolkit (IPT) as trusted data repository.

INTEROPERABLE

GBIF recommends using the Ecological Metadata Language (EML) for datasets and Darwin Core for occurrence data.

REUSABLE

GBIF require creative common data licenses (CC0, CC BY, or CC BY-NC). Provenance available

from the GBIF portal.

Literature references

Baker (2016) 1,500 scientists lift the lid on reproducibility. Nature 533: 452-454 (26 May 2016) doi:10.1038/533452a

Baker (2016) Reproducibility: Seek out stronger science. Nature 537: 703-704 (29 September 2016) doi:10.1038/nj7622-703a

Nature editorial (2016) Reality check on reproducibility. Nature 533: 437 (26 May 2016) doi:10.1038/533437a

Baker (2016) Statisticians issue warning over misuse of P values. Nature 531: 151 (10 March 2016) doi:10.1038/nature.2016.19503

Nosek et al. (2015) Promoting an open research culture. Science 348(6242): 1422-1425. DOI:10.1126/science.aab2374

Leek and Peng (2015) Statistics: P values are just the tip of the iceberg. Nature 520: 612 (30 April 2015) doi:10.1038/520612°

Nuzzo (2015) How scientists fool themselves – and how they can stop. Nature 526: 182–185 (08 October 2015) doi:10.1038/526182a

Hayden (2013) Weak statistical standards implicated in scientific irreproducibility. Nature doi:10.1038/nature.2013.14131

Young (2012) Replication studies: Bad copy. Nature 485, 298–300 (17 May 2012) doi:10.1038/485298a

Callaway (2011) Reports finds massive fraud at Dutch universities. Nature 479, 15 (1 November 2011) doi:10.1038/479015a

Data metrics

In this section review the various metrics available for datasets.

One of the many benefits of publishing data via GBIF is that, during the indexing process, GBIF analyses all datasets and produces metrics about them. These metrics are made available in several different ways:

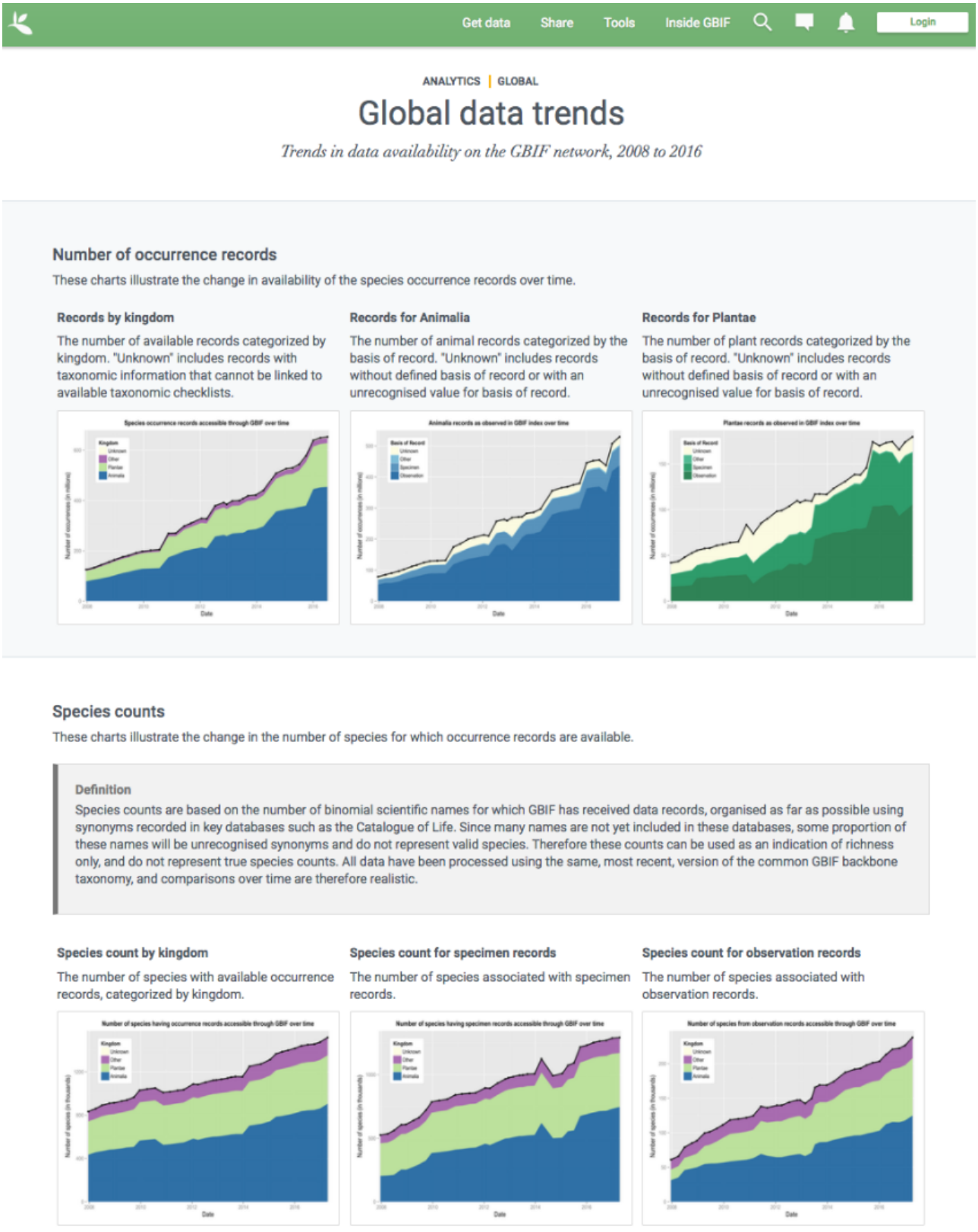
- global trends
- country pages
- dataset content statistics
- dataset download activity

Participants and publishers can use this information to improve the quality of their datasets, e.g. by addressing issues detected during the indexing process. They can also use the access statistics as evidence of real user interest in their datasets and potential use of the published data.

Global data trends

GBIF.org regularly updates analytics to provide an overview of global trends in the data from 2008 to the present. The charts illustrate trends in:

- occurrence records
- species counts
- time and seasonality
- completeness and precision
- geographic coverage for recorded species
- data sharing with country of origin



Dataset content statistics

Each dataset page includes a tab labelled ‘Metrics’. This tab gives access to graphs and tables resulting from the analysis of the contents of the dataset. This includes a summaries of:

- Taxonomic distribution (both list and chart)
- Occurrences per issue
- Occurrences per year

The charts/tables are interactive and you can click to filter and explore further. Additionally, the images can be downloaded for reporting purposes.

Data access logs

There is a third tab in occurrence dataset pages labelled ‘Activity’. In this tab you can see a list of all download requests that included records from that dataset, including their download DOI for easy tracking.

OCCURRENCE DATASET | REGISTERED 23 MAY 2016

Auckland Museum Botany Collection

Published by Auckland War Memorial Museum

Ewen Cameron • Adam Moriarty • ✉ Dhahara Ranatunga

DATASET METRICS ACTIVITY ⬇ DOWNLOAD 252,142 OCCURRENCES 8 CITATIONS

20,045 download events

5 OCCURRENCES FROM THIS DATASET

DOI 10.15468/dl.fqrgtz
Date: 7 May 2018

Occurrences: 4,267
Involved Datasets: 73

And
- Scientific name: Acacia saligna h.l.wendl.
- Has coordinate: True

RERUN QUERY SHOW

2 OCCURRENCES FROM THIS DATASET

DOI 10.15468/dl.httjwd
Date: 7 May 2018

Occurrences: 1,889
Involved Datasets: 62

And
- Scientific name: Acacia retinodes schltdl.
- Has coordinate: True

RERUN QUERY SHOW

10 OCCURRENCES FROM THIS DATASET

DOI 10.15468/dl.a3jdcm
Date: 7 May 2018

Occurrences: 1,017
Involved Datasets: 67

And
- Scientific name: Acacia podalyriifolia cunn. ex don
- Has coordinate: True

RERUN QUERY SHOW

16 OCCURRENCES FROM THIS DATASET

DOI 10.15468/dl.nye7mg
Date: 7 May 2018

Occurrences: 29,609
Involved Datasets: 116

And
- Scientific name: Acacia dealbata link
- Has coordinate: True

RERUN QUERY SHOW

17 OCCURRENCES FROM THIS DATASET

DOI 10.15468/dl.zdjvkp
Date: 7 May 2018

Occurrences: 3,514
Involved Datasets: 18

And
- Scientific name: Acacia parramattensis tindale
- Has coordinate: True

RERUN QUERY SHOW

Review

Quiz yourself on the concepts learned in this module.

1. What dataset class makes up the core of data published within GBIF?
   - ☐ Metadata-only
   - ☐ Checklist
   - ☐ Occurrence
   - ☐ Sampling-event
2. What is the taxonomic backbone?
   - ☐ A dataset
   - ☐ A management classification with the goal of covering all names in GBIF
   - ☐ Allows for taxonomic search on GBIF
   - ☐ All of the above
3. Which licenses or waivers can be applied to datasets published in GBIF?
   - ☐ CC BY
   - ☐ CC BY-SA
   - ☐ CC BY-NC
   - ☐ CC BY-NC-SA
   - ☐ CC BY-ND
   - ☐ CC BY-NC-ND
   - ☐ CC BY-NC-SA
   - ☐ CC0
4. Images are subject to the same licenses as datasets?
   - ☐ True
   - ☐ False
5. GBIF data are FAIR?
   - ☐ True
   - ☐ False

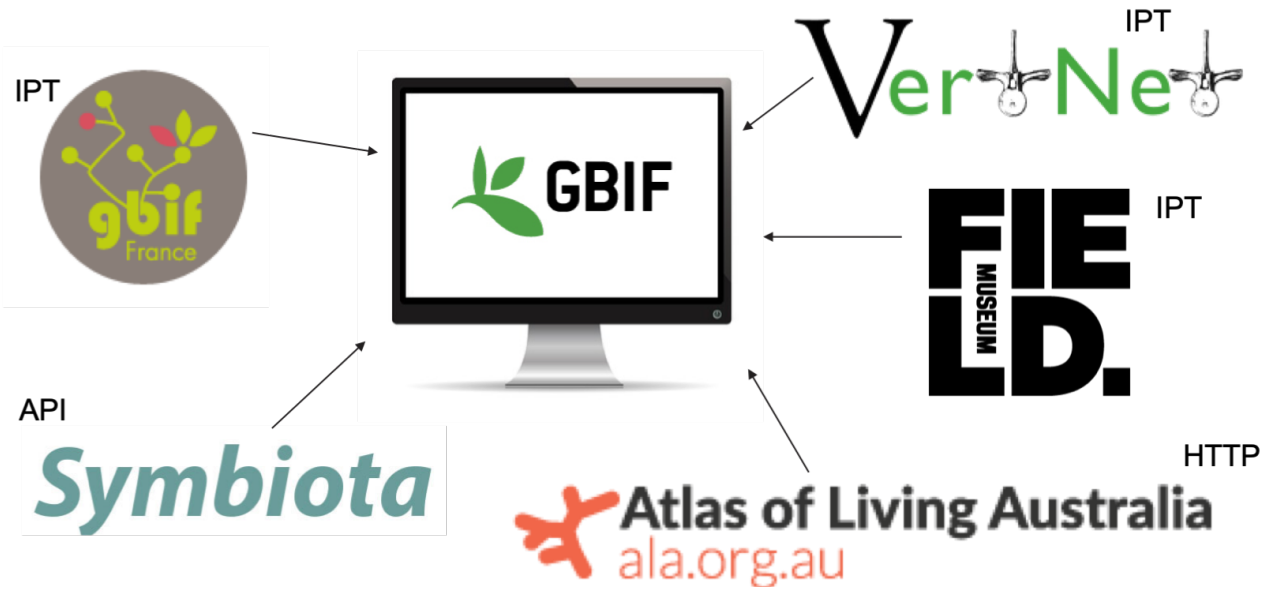
Data publishing

In this module, you will learn about data publishing, incentives for publishing data how to become a publisher, review feedback available for publishers and review information about data papers.

What is data publishing?

In this section, you will learn what data publishing is in the GBIF network.

In simplest terms, data publishing is making your biodiversity dataset(s) publicly accessible and discoverable in a standardized format.



Most organizations within the GBIF Network, otherwise known as publishers, make use of an IPT, the Integrated Publishing Toolkit, to publish their data. These publishers may choose to host their own installation of an IPT like the Field Museum, but generally prefer to find a suitable host for their data publishing activities. This might be through an established GBIF Participant Node like GBIF France or through an established thematic group like VertNet. Or, a another option, might be to make use of one of the GBIF-hosted IPTs like the BID, BIFA or regional IPTs.

If you are part of groups like Symbiota or the Living Atlases Communities, they have other means to assist you with publishing your data to GBIF.

Incentives for publishing open-access biodiversity data

In this section, you will review some incentives for publishing biodiversity data.

An important part of GBIF's mission is to promote a culture in which people recognize the benefits of publishing open-access biodiversity data, for themselves as well as for the broader society.

- By making your data discoverable and accessible through GBIF and similar information infrastructures, you will **contribute to global knowledge about biodiversity**, and thus to the solutions that will promote its conservation and sustainable use.
- Data publishing enables datasets held all over the world to be integrated, revealing **new opportunities for collaboration** among data owners and researchers.
- Publishing data enables individuals and institutions to be properly credited for their work to create and curate biodiversity data, by **giving visibility to publishing institutions** through good metadata authoring. This recognition can be further developed if you author a peer-reviewed data paper, giving scholarly recognition to the publication of biodiversity datasets.

- Collection managers can **trace usage and citations** of digitized data published from their institutions and accessed through GBIF and similar infrastructures.
- Some funding agencies now require researchers receiving public funds to **make data freely accessible** at the end of a project.

Data publisher visibility and recognition

In this section, you will review frameworks for citing and tracking digital data use on GBIF.org

Giving adequate visibility and recognition to data publishers is of paramount importance to GBIF. That is why the organization has put in place the regulatory and technical frameworks needed to make citing and tracking digital data use, easier than ever before.

The GBIF Memorandum of Understanding

GBIF is a multilateral initiative established by intergovernmental agreement and based on a non-binding Memorandum of Understanding (MoU). The MoU is the official document that countries and international organizations sign in order to join GBIF.

The MoU is very clear stating that GBIF data publishers need to be acknowledged for their contributions:

```
"4. Attribution. +
GBIF seeks to ensure that the publisher/holder of data is acknowledged and requests
that such attribution be maintained in any subsequent use of the data."
```

GBIF strives to make all efforts possible to make this statement a reality.

The GBIF Data User Agreement

Prior to accessing any data using GBIF, users need to accept a data user agreement which includes very specific requirements on citation of the origin of the data accessed through GBIF. These are some of the requirements listed in the agreement:

```
"In order to make attribution of use for owners of the data possible, the identifier
of ownership of data must be retained with every data record shared onward for reuse."
```

“Users must publicly acknowledge, following the scientific convention of citing sources in conjunction with the use of the data, the Data Publishers whose biodiversity data they have used, where appropriate through use of a Digital Object Identifier (DOI) applying to the dataset (s) and/or data downloads.”

Similarly, the agreement is very specific in stating that the conditions stated in the licenses selected by the data published must be respected.

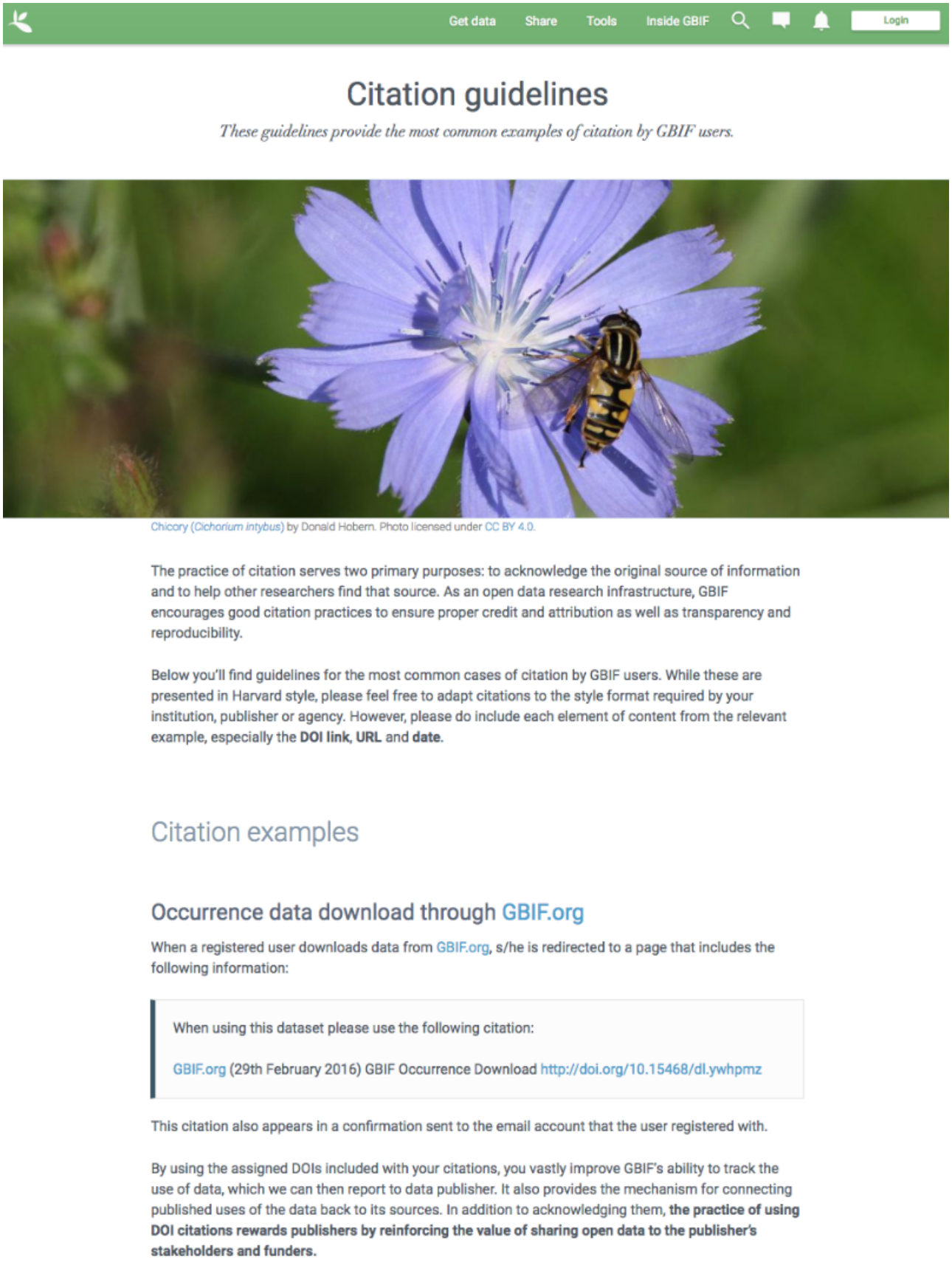
```
"Users must comply with the terms and conditions included in the licence selected by
each Data Publisher, and the licensing information included with each data download.
If any provision of this Use Agreement conflicts with the terms and conditions within
the licences selected by the Data Publisher, licences selected by the Data Publisher
shall prevail."
```

Citation

GBIF strongly encourages all users to cite data retrieved from the GBIF network. For that purpose it provides recommended citation strings on the dataset, occurrence and download pages in GBIF.org.

This is especially relevant for datasets published using the “CC-BY” and “CC-BY-NC” licenses, which include specific requirements for citing the origin of the data.

Citation and right strings are automatically generated for data publishers for each dataset when using IPT as the publishing mechanism, providing that one of the standard licenses is selected.



Data publishers must carefully select which license aligns best with any existing requirement from their institutions and from any data access policy to which they may be subject.

Data publisher page

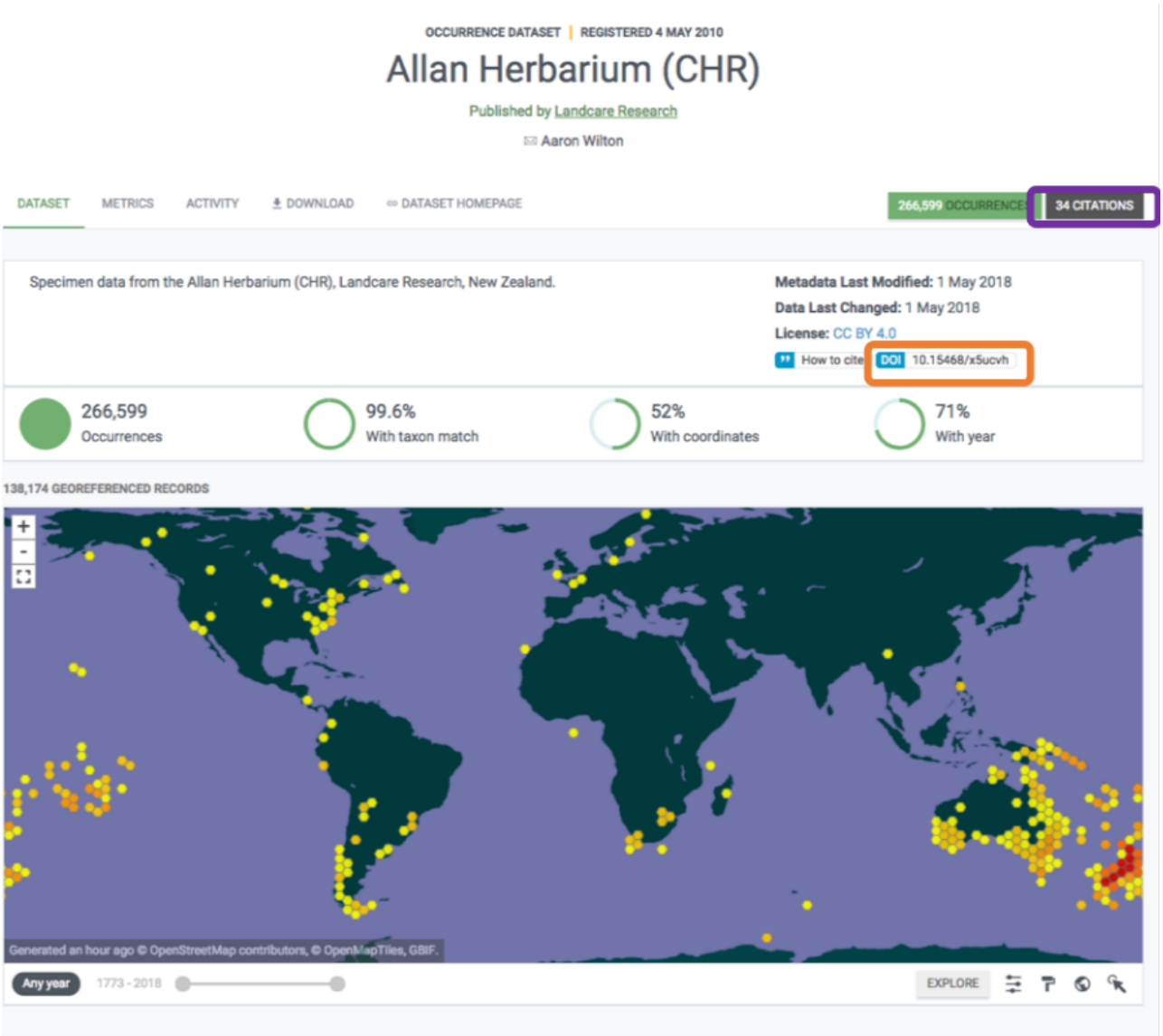
All publishers feature their own page on GBIF.org. It is important that publishers give some thought to how they want to appear on the website and provide relevant information about their institutions and

their teams at the time of registration. They should also strive to keep it up to date, as interested parties will use the contact data on that page to contact the team responsible for the data publishing.

Dataset DOI

Every time a new version of a dataset is published using an IPT, a DOI (Digital Object Identifier) is assigned. As in the case of the downloads, this identifier allows easy citation and tracking of work derived from the dataset, if the user follows good practices for source citation.

As mentioned before, you can resolve DOIs into websites like doi.org/10.xxx which will always redirect to the original source, in this case, the dataset page. You can also search for DOI using a normal web search, which will normally reveal any other resource citing use of the same DOI such as articles or public reports.



Description

Specimen data from the Allan Herbarium (CHR), Landcare Research, New Zealand.

Temporal coverage

1 January 1714 - 7 August 2014

Geographic coverages

Specimens from around world, but predominantly from New Zealand

SEARCH RESOURCES | 34 RESULTS

ALL LITERATURE

Read more about literature, how it's discovered and linked to GBIF-mediated data.

Inherent variation of functional traits in winter and summer leaves of Mediterranean seasonal dimorphic species: evidence of a "within leaf cohort" sp... Literature

Puglielli, G. Varone, L. (2018) AoB PLANTS
The covariation pattern among leaf functional traits involved in resource acquisition has been successfully provided by the leaf economic spectrum (LES). Nevertheless, some aspects such as how the leaf trait variation sources affect LES predictions are still little investigated. Accordingly, the aim...

Cistus • LMA • deciduous • evergreen • leaf cohorts • leaf economic spectrum
Journal Article Open Access Peer-Reviewed
Data used in study DOI 10.15468/dl.oqk0l5

Climatic Suitability Derived from Species Distribution Models Captures Community Responses to an Extreme Drought Episode Literature

Pérez Navarro, M. Sapes, G. Batllori, E. Serra-Diaz, J. Esteve, M. Lloret, F. (2018) Ecosystems
The differential responses of co-occurring species in rich communities to climate change—particularly to drought episodes—have been fairly unexplored. Species distribution models (SDMs) are used to assess changes in species suitability under environmental shifts, but whether they can portray populat...

SDMs • climatic suitability • dieback • drought resistance • extreme climatic events • niche
Journal Article Peer-Reviewed
Data used in study DOI 10.15468/dl.06klxe DOI 10.15468/dl.0t186i DOI 10.15468/dl.2a1bhi DOI 10.15468/dl.4hzzmk DOI 10.15468/dl.4sgdo5 DOI 10.15468/dl.6sjatk DOI 10.15468/dl.7hmtxx DOI 10.15468/dl.8cfzh0 DOI 10.15468/dl.cmlz6j DOI 10.15468/dl.enpnfp DOI 10.15468/dl.frvcwm DOI 10.15468/dl.hv3bux DOI 10.15468/dl.ilz9vf DOI 10.15468/dl.mksbqr DOI 10.15468/dl.rwpnga DOI 10.15468/dl.se9efz DOI 10.15468/dl.st4ggu DOI 10.15468/dl.treg4p DOI 10.15468/dl.uvobd2 DOI 10.15468/dl.vrhf0e DOI 10.15468/dl.x3jqrq DOI 10.15468/dl.yxkzri

Resolving relationships and phylogeographic history of the Nyssa sylvatica complex using data from RAD-seq and species distribution modeling Literature

Zhou, W. Ji, X. Obata, S. Pais, A. Dong, Y. Peet, R. ... - (2018) Molecular Phylogenetics and Evolution
Nyssa sylvatica complex consists of several woody taxa occurring in eastern North America. These taxa were recognized as two or three species including three or four varieties by different authors. Due to high morphological similarities and complexity of morphological variation, classification and d...

LGM • Nyssa • Phylogeography • RAD-seq • Refugia • SDM
Journal Article Peer-Reviewed
Data used in study DOI 10.15468/dl.mi3bzq DOI 10.15468/dl.nexm7l

Ecology and biogeography in 3D: the case of the Australian Proteaceae Literature

Pausas, J. Lamont, B. (2018) Journal of Biogeography
The key biophysical pressures shaping the ecology and evolution of species can be broadly aggregated into three dimensions: environmental conditions, disturbance regimes and biotic interactions. The relative importance of each dimension varies over time and space, and in most cases multiple dimensio...

Australia • Proteaceae • disturbance regimes • evolutionary pressures • fire ecology • plant traits
Journal Article Peer-Reviewed
Data used in study DOI 10.15468/dl.9a0rtx

The relation between global palm distribution and climate Literature

Reichgelt, T. West, C. Greenwood, D. (2018) Scientific Reports
Fossil palms provide qualitative evidence of (sub-) tropical conditions and frost-free winters in the geological past, including modern cold climate regions (e.g., boreal, or polar climates). The freeze intolerance of palms varies across different organs and life stages, with seedlings in particular...

Journal Article Open Access Peer-Reviewed
Data used in study DOI 10.15468/dl.kevers

Global database of plants with root-symbiotic nitrogen fixation: NodDB Literature

Tedersoo, L. Laanisto, L. Rahimlou, S. Toussaint, A. Hallikma, T. Pärtel, M. (2018) Journal of Vegetation Science
Plants associated with symbiotic nitrogen fixing bacteria play important roles in early successional, riparian and semidry ecosystems. These so-called nitrogen fixing plants are widely used for reclamation of disturbed vegetation and improvement of soil fertility in agroforestry. Yet, available info...

Fabaceae • Frankiaceae • Nostocaceae • Rhizobiaceae • Zygophyllaceae • nitrogen fixing clade
Journal Article Peer-Reviewed
Data used in study DOI 10.15468/dl.4nqpev

Combined multi-gene backbone tree for the genus Coniochaeta with two new species from Uzbekistan Literature

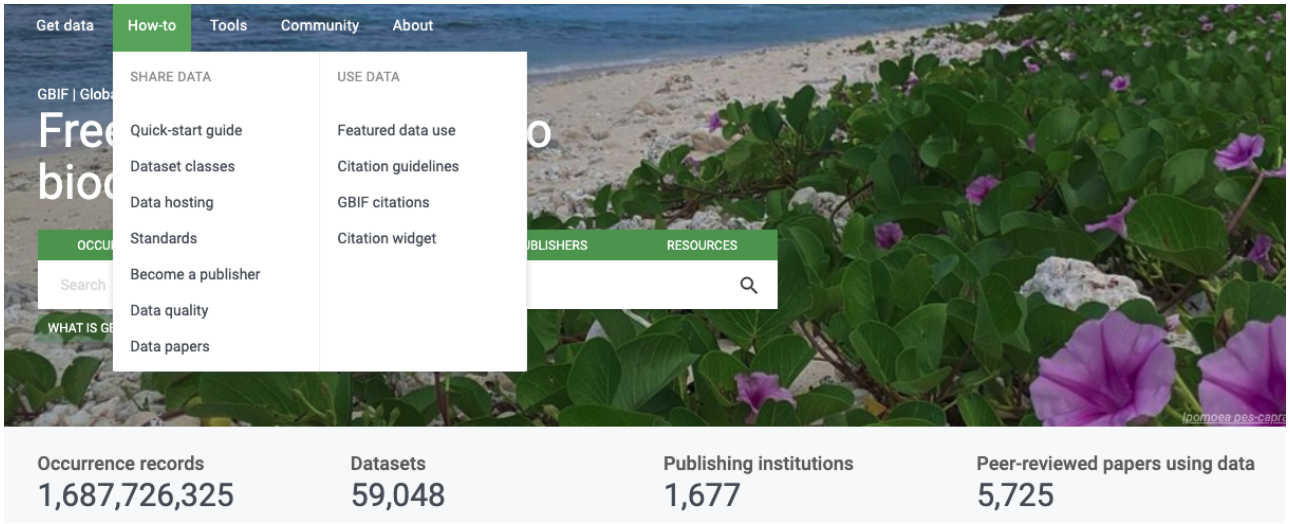
SAMARAKOON, M. GAFFOROV, Y. LIU, N. MAHARACHCHIKUMBURA, S. BHAT, J. LIU, J. ... - (2018) Phytotaxa
The genus Coniochaeta is an important ascomycete because its members live in diversified habitats and nutritional modes. In this study, two new species, C. acaciae and C. colutea, are introduced from dead branches of Acacia sp. and Colutea paulsenii Freyn (both Fabaceae) respectively from Uzbekista...

Central Asia • Coniochaetales • Fabaceae • ascomycetous microfungi • phylogenetic analyses
Journal Article Peer-Reviewed
Data used in study DOI 10.15468/dl.fr5hwn

How to become a data publisher

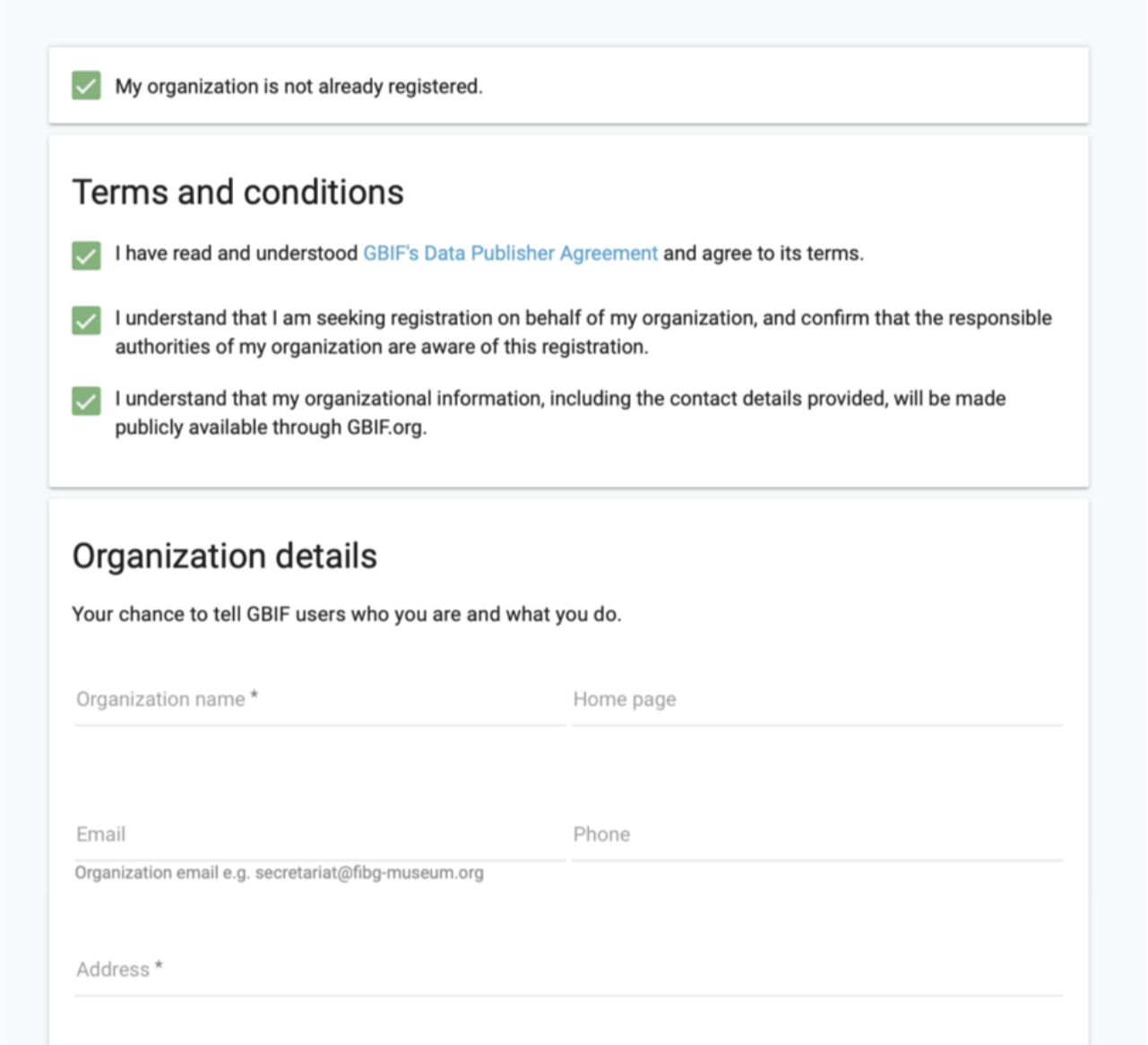
In this section, you will review the steps to become a data publisher.

Organizations wishing to share data through GBIF can register here to request endorsement as a data publisher.



Click on the How-to menu on the GBIF website to navigate to the Become a publisher page

Prospective new publishers are asked to complete an online form. The answers provided help GBIF to give proper credit and attribution for the datasets shared by the new publisher.

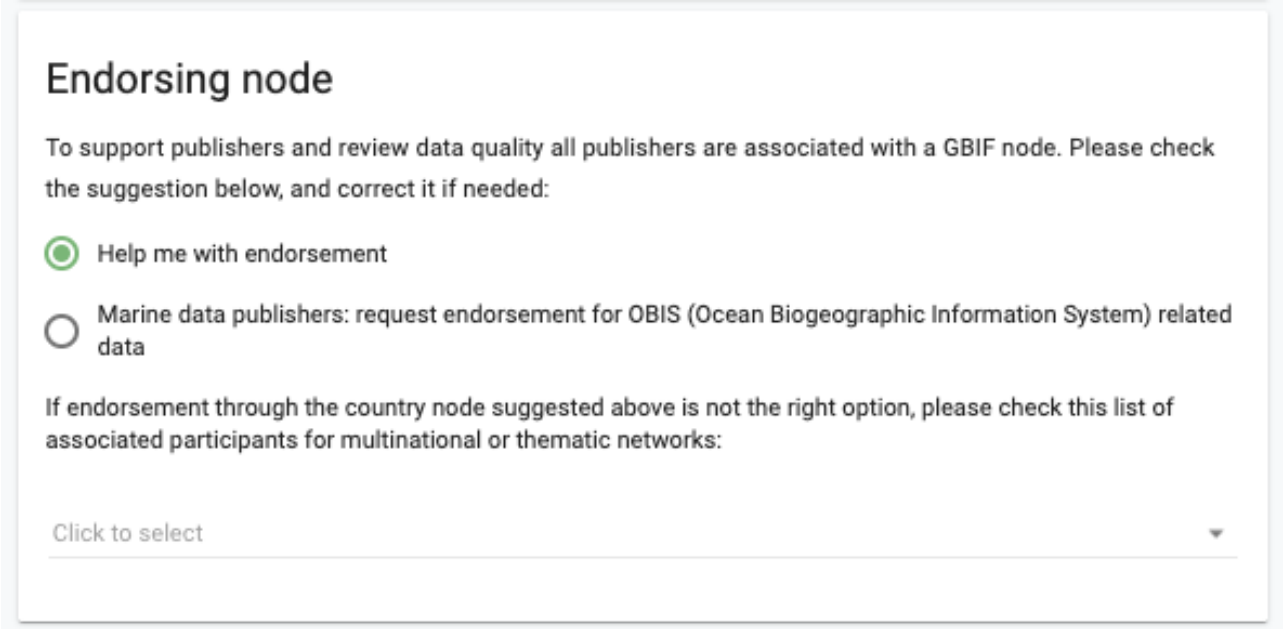


Before GBIF indexes datasets, an institution must receive endorsement as a data publisher from one of the Participant nodes that coordinate activities of the national and organizational Participants in the GBIF network. If the country is not yet participating in GBIF, endorsement is sought on the publisher's behalf from within the GBIF community.

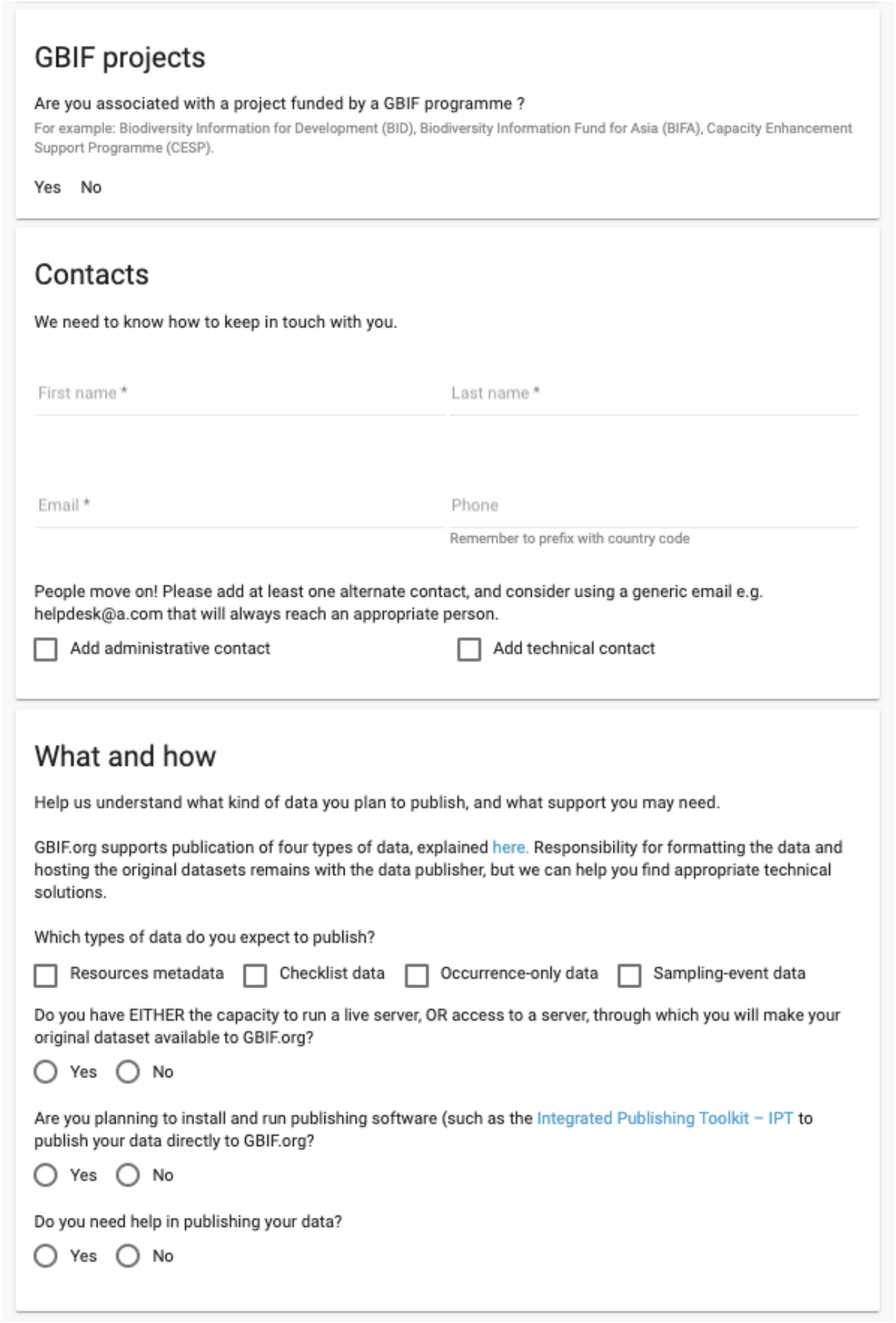
The endorsement procedure aims to ensure that:

1. Published data are relevant to GBIF's scope and objectives
2. Data hosting arrangements are stable and persistent
3. Data publishing and use are supported by strong national, regional and thematic engagement
4. Data are as open as possible, and available for sharing and reuse
5. Data publishers can respond to feedback and improve data quality

At present, GBIF and its Participants only publish data from organizations—that is, institutions, networks and societies—rather than individuals. Individuals wishing to publish data should work through their affiliated organizations to seek endorsement as a publisher.



The answers provided will also help users to understand more about the provenance of data shared through the GBIF network.



GBIF projects

Are you associated with a project funded by a GBIF programme ?

For example: Biodiversity Information for Development (BID), Biodiversity Information Fund for Asia (BIFA), Capacity Enhancement Support Programme (CESP).

Yes No

Contacts

We need to know how to keep in touch with you.

First name *

Last name *

Email *

Phone

Remember to prefix with country code

People move on! Please add at least one alternate contact, and consider using a generic email e.g. helpdesk@a.com that will always reach an appropriate person.

Add administrative contact

Add technical contact

What and how

Help us understand what kind of data you plan to publish, and what support you may need.

GBIF.org supports publication of four types of data, explained here. Responsibility for formatting the data and hosting the original datasets remains with the data publisher, but we can help you find appropriate technical solutions.

Which types of data do you expect to publish?

Resources metadata Checklist data Occurrence-only data Sampling-event data

Do you have EITHER the capacity to run a live server, OR access to a server, through which you will make your original dataset available to GBIF.org?

Yes No

Are you planning to install and run publishing software (such as the Integrated Publishing Toolkit – IPT to publish your data directly to GBIF.org?

Yes No

Do you need help in publishing your data?

Yes No

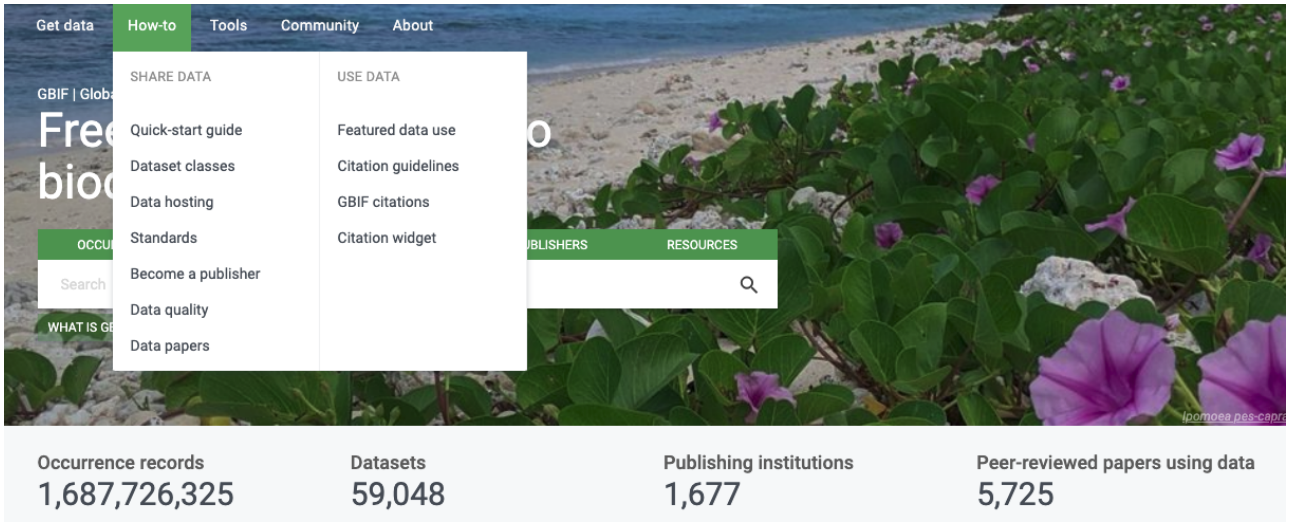
Be sure to search existing publishers before registering a new one to make sure the publisher is not already registered.

BID and BIFA projects are required to register at least one data publisher (or provide evidence of an already registered publisher) by specific milestone dates.

GBIF data quality requirements

In this section, you will review GBIF's data quality requirements.

Publishers play an essential role not simply in sharing datasets, but also in managing their quality, completeness and usefulness and ensuring their integration and value within GBIF's global knowledge base.



Click on the How-to menu on the GBIF website to navigate to the Data quality page

To share data through GBIF.org, publishers typically have to collate or transform existing datasets into a standardized format. This work may include additional processing, content editing and mapping a dataset's content into one of the available data transfer formats, as well as publication through one of the available data publishing tools, such as GBIF's free, open-source Integrated Publishing Toolkit (IPT).

Once published, GBIF's real-time infrastructure 'indexes' or 'harvests' new datasets, integrating them into a common access system where users can retrieve any and all data through common search and download services. As datasets are indexed, GBIF.org performs additional checks, interpretation and conversion routines to ensure that data are interoperable and comply with minimum standards of data formats, data quality and fitness for use. Many criteria for quality and usability of data, however, are best and most easily handled when addressed at their source: the individual dataset.

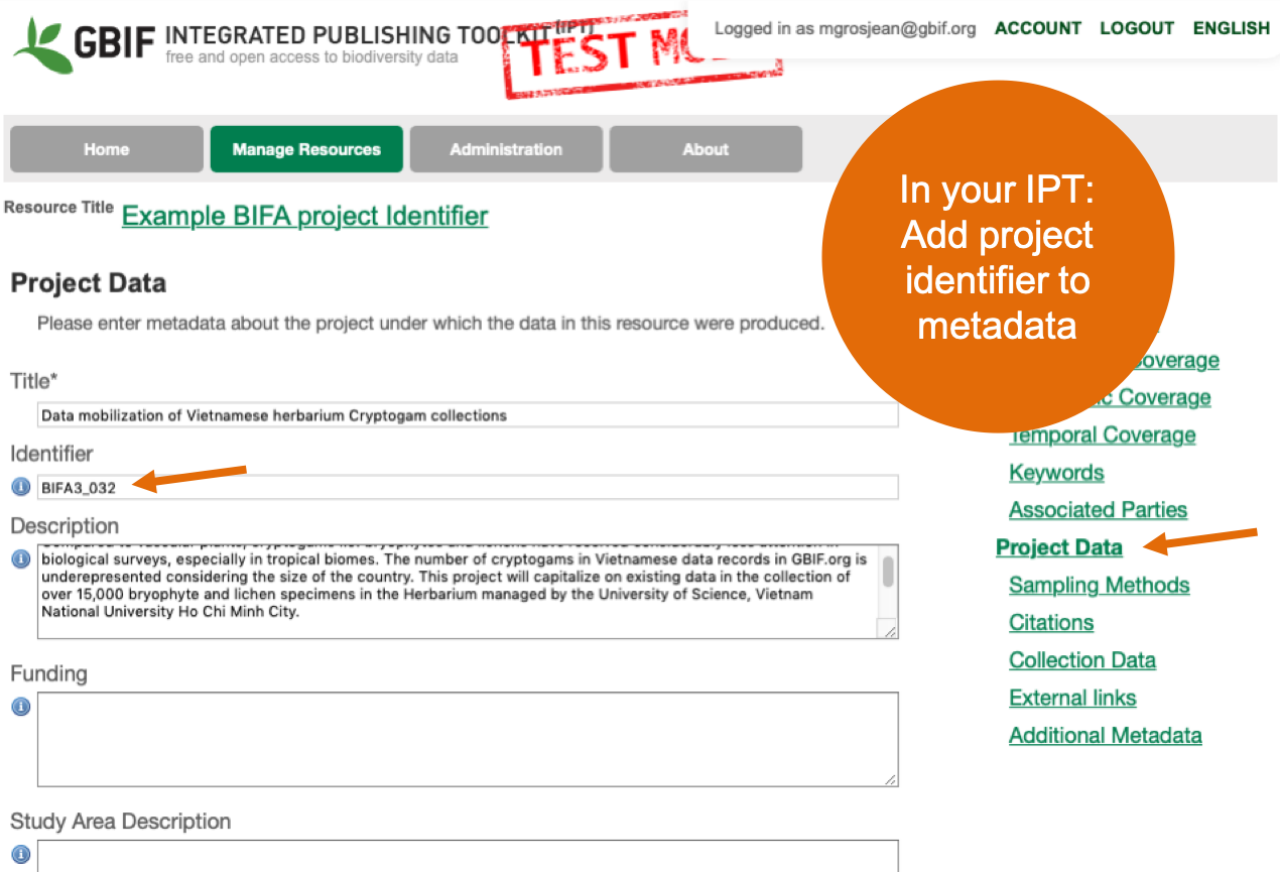
Publishers thus play an essential role not simply in sharing datasets, but also in managing their quality, completeness and usefulness as well as ensuring their integration and value within GBIF's global knowledge base. Learn more about data quality requirements and recommendations for:

- Occurrence-only datasets
- Checklists
- Sampling-event datasets

In practice, we encourage those responsible for publishing data to get acquainted with the expected data formats and content requirements as early as possible in the process (see also the pre-configured GBIF Excel templates with required and recommended terms for occurrence-only datasets, checklists, and sampling-event datasets, all available with example data). Doing so will save

a lot of effort that may be needed at later stages, for example, in adding data conversions, capturing information for required or strongly recommended fields, or performing and addressing final pre-publication data-quality checks.

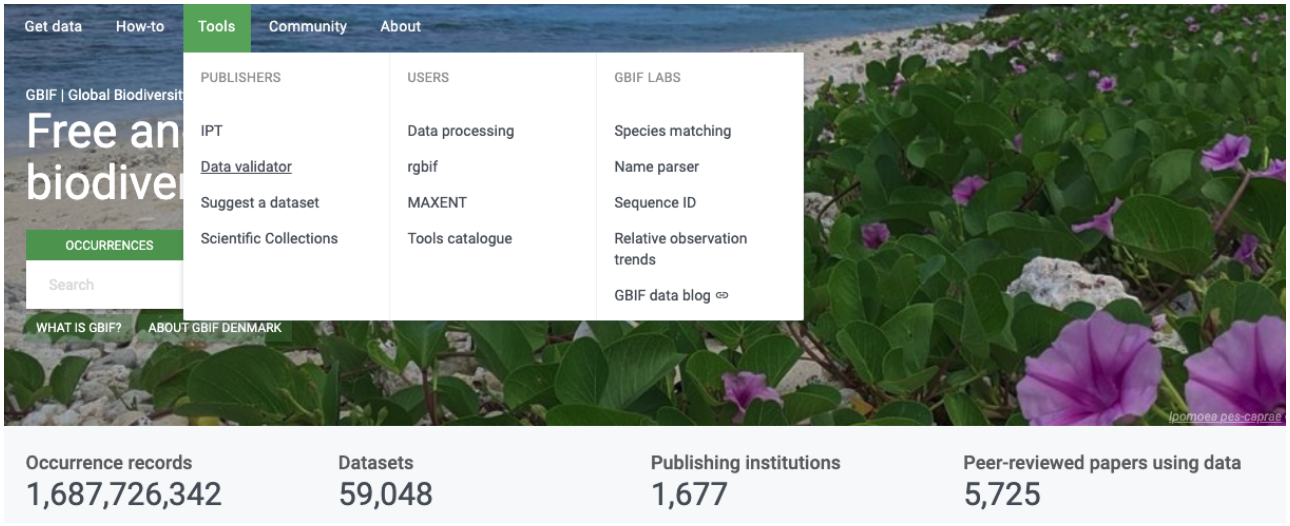
BID and BIFA projects are required to include their projectID on published datasets as part of the dataset metadata. This allows datasets to be linked to project pages.



Improve published data quality

In this section, you will learn how to use the GBIF data validator.

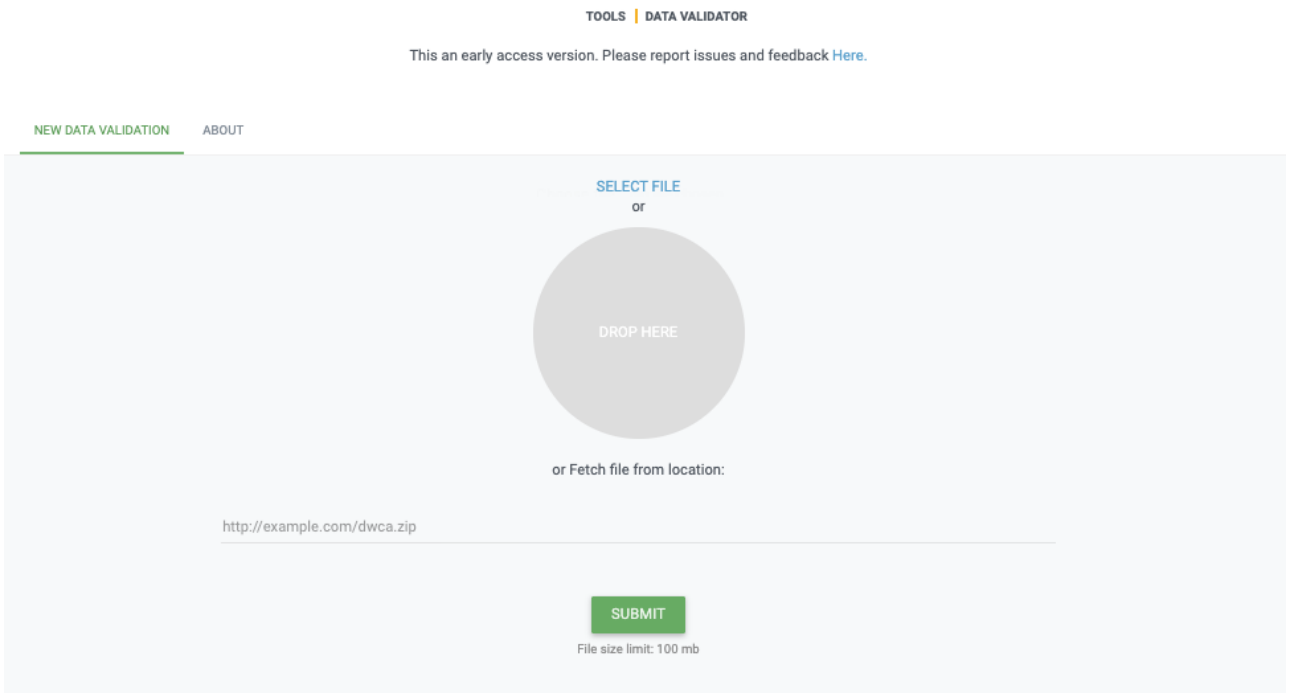
The GBIF data validator is a service that allows anyone with a GBIF-relevant dataset to receive a report on the syntactical correctness and the validity of the content contained within the dataset. By submitting a dataset to the validator, you can go through the validation and interpretation procedures usually associated with publishing in GBIF and quickly determine potential issues in data - without having to publish it.



Click on the Tools menu on the GBIF website to navigate to the Data validator page

How does it work?

You start by uploading the dataset file to the validator, either by 1) clicking SELECT FILE and selecting it on your local computer or 2) dragging the file from a local folder and dropping it on the Drop here icon. You can also enter the URL of a dataset file accessible from the internet. This is particularly useful for larger datasets. Once you hit the Submit button, the validator starts processing your dataset file. You will be taken straight to a page showing the status of the validation.



Depending on the size of your dataset, processing might take a while. You don't have to keep the browser window open, as a unique job ID is issued every time a new validation process is started. If your dataset is taking too long to process, just save the ID (bookmark the URL) and use it to return at a later time to view the report. We'll keep the report for a month, during which you can come back whenever you like.

Which file types are accepted?

- ZIP-compressed Darwin Core Archives (DwC-A) (containing cores Occurrence, Taxon, or Event)
- Integrated Processing Toolkit (IPT) Excel templates containing Checklist, Occurrence, or

Sampling-event data

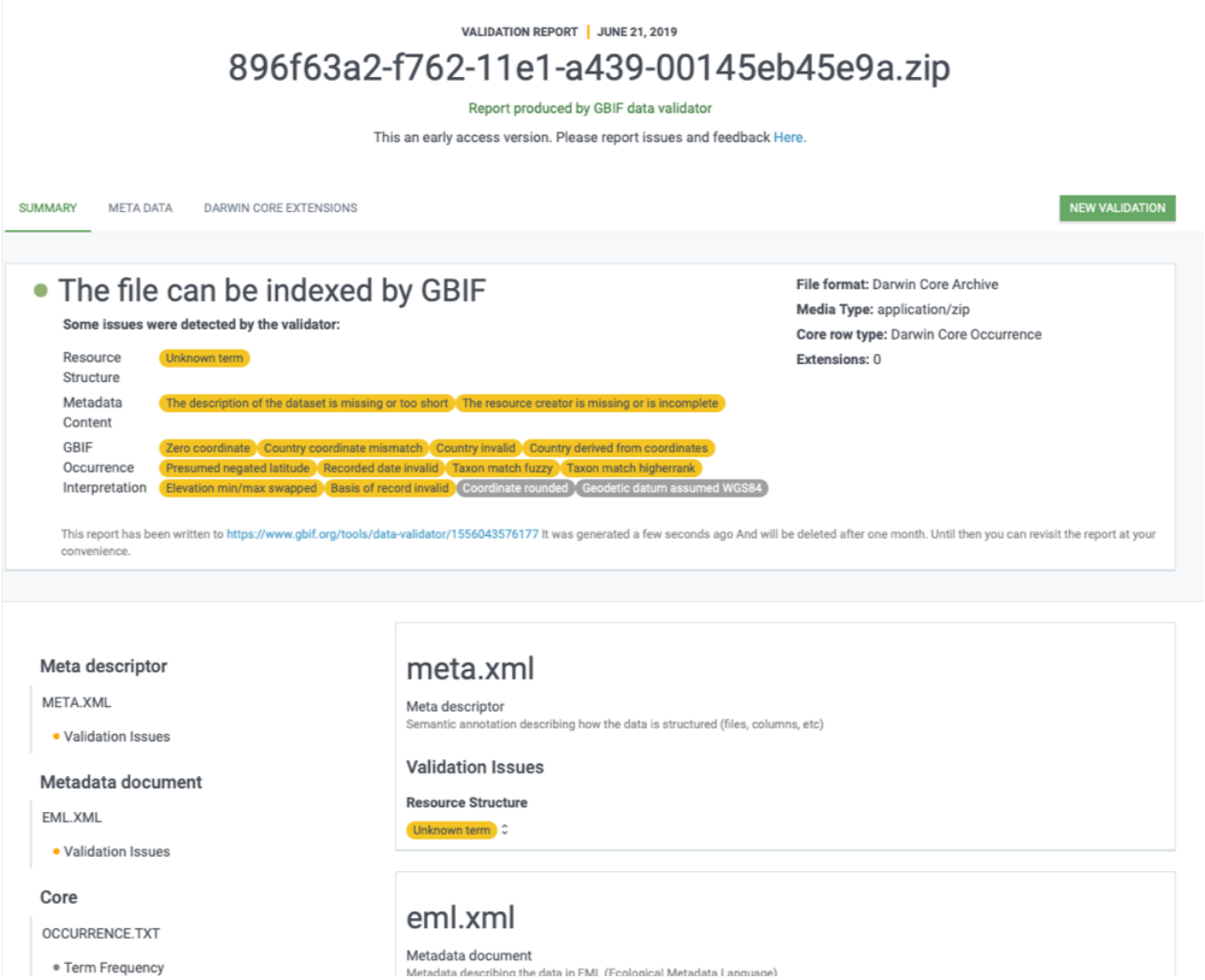
- Simple CSV files containing Darwin Core terms in the first row

What information is provided from the validation report?

Once processing is done, you will be able to see the validation report containing the following information:

- a summary of the dataset type and a simple indicator of whether it can be indexed by GBIF or not
- a summary of issues found during the GBIF interpretation of the dataset
- detailed break-down of issues found in metadata, dataset core, and extensions (if any), respectively
- number of records successfully interpreted
- frequency of terms used in dataset

You will also be able to view the metadata as a draft version of the dataset page as it would appear when the dataset it published and registered with GBIF.



I've got the validation report - now what?

If the validator finds that your dataset cannot be indexed by GBIF, you should address the issues raised by the validation report before you consider publishing it to GBIF. Even if your dataset is indexable by GBIF, you should still carefully review any issues that may be the result of e.g. conversion errors, etc. which could affect the quality of the data. If you find and correct any error -

from a single typo to large systematic problems - feel free to resubmit your dataset as many times you like.

Review

Quiz yourself on the concepts learned in this section.

1. What does data publishing mean in the context of GBIF?
   - ☐ Exporting a csv file of your cleaned data that you can share with your colleagues
   - ☐ Writing an article describing your data, and the protocol(s) you used to collect, capture and clean them
   - ☐ Making your biodiversity dataset(s) publicly accessible and discoverable in a standardized format
2. Which of the following are incentives for publishing data?
   - ☐ contribute to global knowledge about biodiversity
   - ☐ holding onto my data until it is perfect
   - ☐ new opportunities for collaboration
   - ☐ make data freely accessible
3. How do you become a publisher in the GBIF network?
   - ☐ email the GBIF helpdesk and wait for endorsement
   - ☐ fill out the *Become a publisher* form on GBIF.org and wait for endorsement
4. There are no requirements for publishing your data on GBIF.org
   - ☐ True
   - ☐ False
5. What is the GBIF data validator?
   - ☐ a tool to publish my data to GBIF
   - ☐ a tool to turn my data into XML
   - ☐ a tool to check my data for issues

Data access

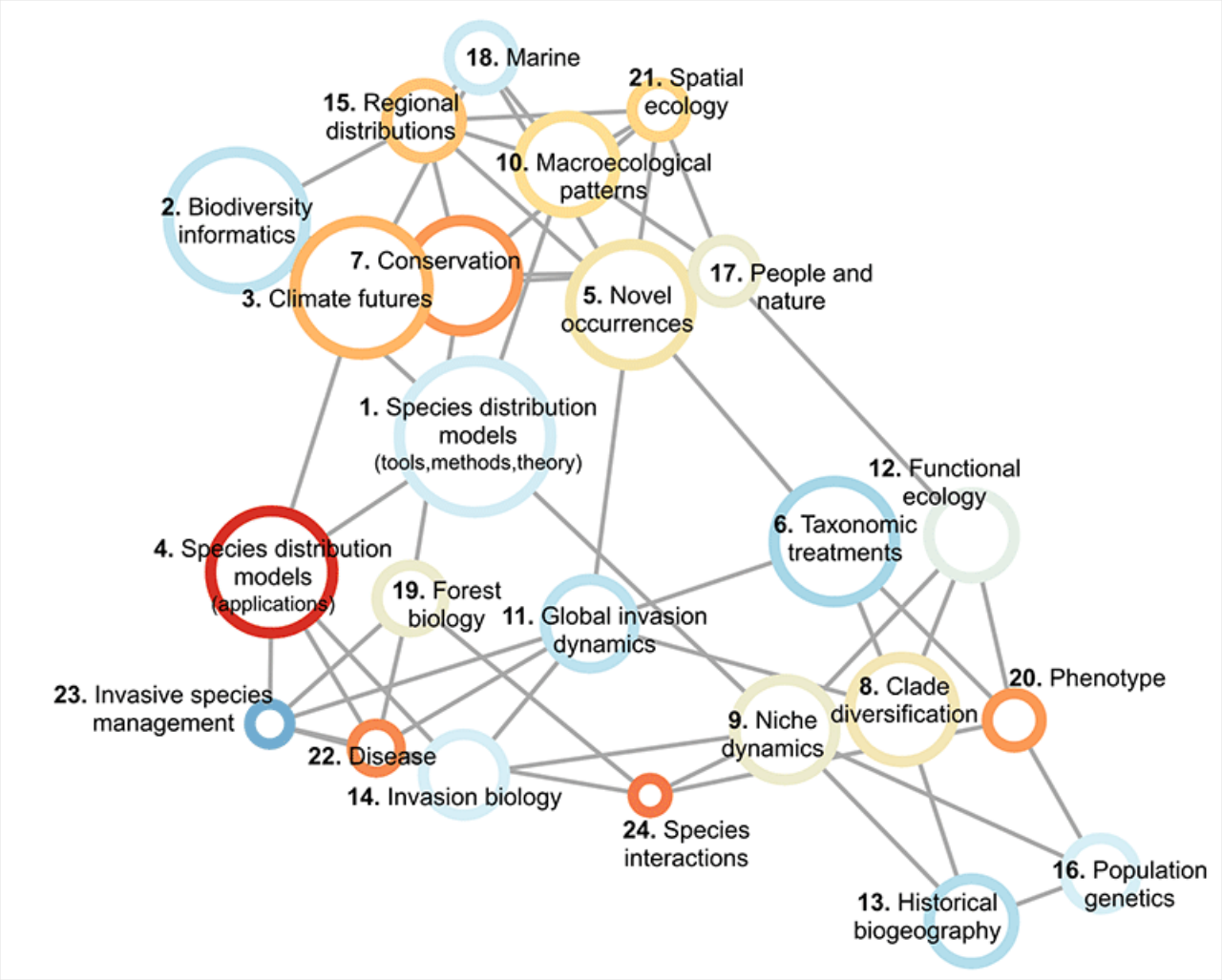
In this module, you will learn how GBIF's data is used and accessed. You will also learn about the handling of data quality and GBIF's data quality flags.

How is GBIF-mediated data used?

The GBIF literature tracking system has identified over 5,000 uses of GBIF mediated data, most of which are in peer-reviewed articles. The majority of these uses are in the field of ecology, but others

relate to climate change, conservation, human health and agriculture. A systematic review of the use of GBIF-mediated data by Heberling et al. (2020) showed:

- Both data availability and data use have increased over time.
- Data integration facilitates global research and access.
- Uses of GBIF-mediated data span disciplinary boundaries.
- The scientific areas using GBIF-mediated data are conceptually diverse and change in prevalence over time.
- Globally integrated datasets enable researchers to ask both basic and applied questions at taxonomic, temporal and spatial scales that would be otherwise impossible.
- The synergistic roles of observation- and specimen-based biodiversity data highlight the value and need for deeper integration with phylogenetic, environmental, phenotypic, ecological and genetic sources of data.



Structural topic model results from 4,035 studies that used GBIF-mediated data published between 2003 and 2019.

GBIF-mediated data is also used for monitoring the state of biodiversity and progress towards achieving the targets of the Convention on Biological Diversity. The increase in availability of GBIF occurrence data is one of the indicators for tracking progress towards the achievement of Aichi Biodiversity target 19 and GBIF is a key data source in the creation of a number of other indicators, including the Species Status Information Index, Species Habitat Index and the Biodiversity Habitat Index.

While the utility of GBIF-mediated data is clear, the wide variety of sources of data accessible through GBIF, spanning museum collections, citizen science, metagenomics, among others, means that not all GBIF-mediated data will be fit for every use. Key components of using GBIF-mediated data are understanding how to access the specific data that you need from what is available in GBIF and understanding some of the common data quality issues that affect the data so as to facilitate processing of the data before analysis.

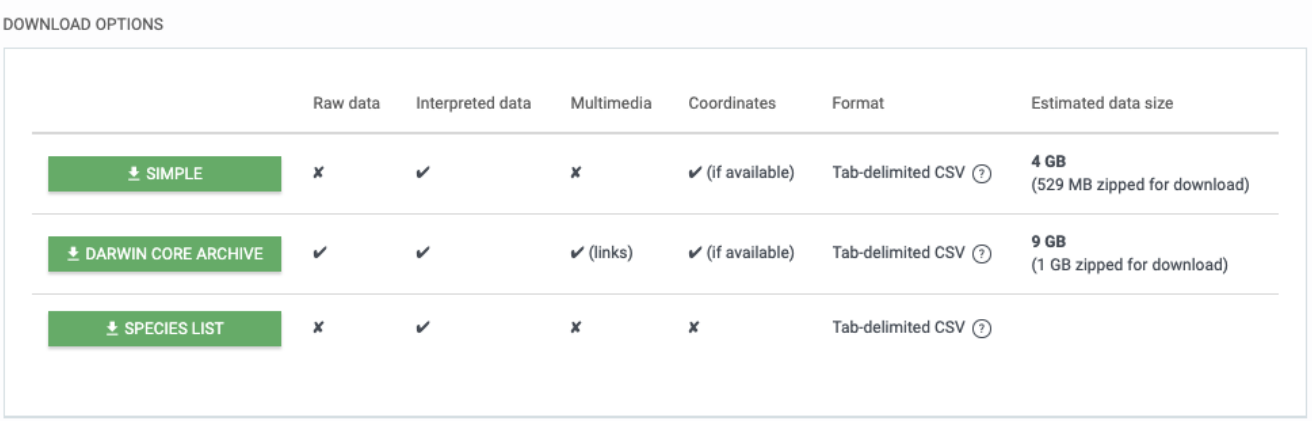
Accessing GBIF-mediated data

There are two main points of access to GBIF-mediated data: GBIF.org and the Application Programming Interface (API) services. Using the website requires no programming experience and allows for quick and easy search, filter and download functions for GBIF-mediated data, as well as a range of additional tools and metrics that are not available through API services. API services allow continued access to GBIF-mediated data through other systems and can be the basis for the development of tools that allow for the interrogation of the data. Examples include a number of R packages, such as rgbif and CoordinateCleaner, as well as more specialized tools that allow for more specific use cases, such as GeoCat for Red List assessments.

What is available to me?

Through the search functions on the website, users can access data that can either be directly downloaded through GBIF or accessed from the original sources following links that GBIF provides.

Remember that as a data user you should read and agree with the terms of the GBIF Data User Agreement that include correctly citing the use of GBIF-mediated data.


DOWNLOAD OPTIONS

	Raw data	Interpreted data	Multimedia	Coordinates	Format	Estimated data size
SIMPLE	✘	✔	✘	✔ (if available)	Tab-delimited CSV	4 GB (529 MB zipped for download)
DARWIN CORE ARCHIVE	✔	✔	✔ (links)	✔ (if available)	Tab-delimited CSV	9 GB (1 GB zipped for download)
SPECIES LIST	✘	✔	✘	✘	Tab-delimited CSV	

The data available to you are:

- Primary biodiversity data - occurrence, checklist and sampling event data that is provided to users through the one of the 3 download formats:
  - **Simple**: Tab delimited CSV. Only contains the data after GBIF interpretation. No multimedia included.
  - **Darwin Core Archive**: The Darwin Core Archive (DwC-A) contains both the original data as the publisher provided it and the GBIF interpretation. Links (but not files) to multimedia included.
  - **Species list**: Tab delimited CSV with the distinct list of names in the search result and as a map visualization of the data.
- A range of metrics are provided for countries and regions, data publishers, datasets and data searches that provide taxonomic breakdowns, trends in data collection and highlight data quality issues. For countries, these metrics can be also be downloaded in the form of a PDF activity

report.

- Searchable database of publications that have used GBIF-mediated data.

Searches can be performed on the occurrences, species, datasets, publishers and resources, and each search function carries a set of filters that allow for more refined searching and additional data associated with the data, for example, images, can be found in tabs associated with the search.

Handling data quality

Determining the precision and accuracy of the data for use is a key step in determining the usefulness of the data for any intended purpose. While GBIF can support the identification of some quality issues that arise from within the data publishing workflow, handling some quality issues requires additional expert knowledge. The two most common issues for which this may be required are:

- **Data gaps** - sampling across taxonomic groups and geographic regions is not equal and users may need to take into account sampling bias in their analyses before the data can be used effectively.
- **Taxonomic misidentification** - some taxonomic groups may require additional information to ensure that taxa have been correctly identified such as images, videos and audio recordings that accompany data or collector information.

GBIF Flags for Data Quality Issues

During the indexation process, GBIF assigns issues and flags to data for common data quality issues. These most frequently occur from data entry errors or missing data fields whose interpretation can be automated centrally by GBIF. These interpretations are classified as

- **Excluded** - where the original data couldn't be interpreted, so is excluded in the interpreted fields.
- **Altered** - where the original data is modified in the interpretation process to be indexed in GBIF.org.
- **Inferred** - where an empty field is inferred using other record information.

Be aware that if you are filtering for data quality issues, you should reverse the filter to exclude those data that have been flagged with that issue. You can also see the verbatim data i.e. the non-interpreted data in a Darwin Core Archive if you would like to validate the interpretation process.

How can I improve data quality?

Data publishers have the responsibility for improving the quality of the data, and as a user, you play a key role in identifying where there are errors. If you should find an error in the data, you should contact the publisher directly using the contact details that GBIF provides on the publisher page. GBIF also provides the ability for users to log data quality issues using the "Feedback and questions" button on the menu bar of GBIF.org.

Review

Quiz yourself on the concepts learned in this section.

1. How can you access GBIF data?

   - ☐ GBIF.org search interface
   - ☐ GBIF API
   - ☐ rGBIF

2. Which file formats are available for downloads of data?

   - ☐ simple
   - ☐ XML
   - ☐ Darwin Core Archive
   - ☐ species list

3. What kind of flags does GBIF apply to data to alert you to the quality?

   - ☐ altered
   - ☐ amended
   - ☐ translated
   - ☐ excluded
   - ☐ interpreted
   - ☐ inferred

Community of practice

In this module, you will learn about GBIF's community of practice, capacity enhancement and funding opportunities and how to stay connected with GBIF.

Community of practice

In this video (06:44), you will review participation in GBIF, and take a closer look at the four key ways that volunteers contribute to the GBIF community of practice: as mentors, trainers, biodiversity open data ambassadors and translators. If you are unable to watch the embedded Vimeo video, you can download it locally. (MP4 - 14.4 MB)

▶ https://vimeo.com/460207962 *(Vimeo video)*

If you are undertaking a biodiversity project and would like to request support from a mentor with data mobilization, management or use, please contact mentors@gbif.org.

We encourage you to consider whether you can contribute to the GBIF community of practice in any of these roles. We would welcome your participation!

Further information on how to participate in the community of practice is available on GBIF.org:

- Mentors and trainers

- Biodiversity open data ambassadors
- Translators

You are also welcome to contact us at info@gbif.org to discuss how you can contribute.

Capacity enhancement and funding opportunities

GBIF is a growing global network of Participant countries, economies and organizations, each with different priorities and capacity assets and needs. Capacity enhancement is recognized as essential to underpin the sustainable performance of the GBIF Participant network and all its members, independent of their level of development.

Broadly speaking, much of GBIF's work contributes to capacity enhancement. The Participant nodes ensure capacity development at the national level, often providing training for individuals and institutions within their networks. Collaboration between Participants through GBIF's governance structures enables further international and regional exchange and partnerships.

Training participants at the Node Management workshop in Trinidad 2019 by Mélianie Raymond (licensed under CC-BY 4.0)

The GBIF Secretariat supports the work of the network, which contributes to capacity development, in particular by coordinating activities, engaging volunteers in the community of practice and supporting the development of core training materials. The Secretariat's role in capacity enhancement activities leans more toward facilitation than implementation: providing guidance, documenting best practices, collating training material, coordinating activities, and creating or aligning opportunities. The Secretariat plays a further role in enhancing capacity by providing funding opportunities for collaborative projects in the network, often with a focus on mobilizing biodiversity data, supporting the use of GBIF-mobilized data and strengthening national biodiversity information facilities.

You can find more information about GBIF's funding programmes on our website.

Engaging with GBIF

Please stay in contact with GBIF to keep up with events, opportunities, new features and other news.

Sign up for GBIF's newsletter and to mailing lists, follow GBIF on Twitter or Facebook, and keep up to date with news on the website.

GBIF organizes regular community webinars that are open for anyone to join.

Review

Quiz yourself on the concepts learned in this module.

1. Who does most of the translations of GBIF materials, including the GBIF.org website and training materials?
   - ☐ A specialist company with experience in working with biodiversity translation
   - ☐ Volunteer translators from GBIF's community of practice
2. How can I get support with a biodiversity data mobilization project?
   - ☐ Write to info@gbif.org
   - ☐ Request a volunteer mentor to provide remote support
   - ☐ Undertake the online data mobilization course
   - ☐ All of the above
3. What is a biodiversity open data ambassador?
   - ☐ Biodiversity professionals who promote the principles and best practices of open data sharing and use
   - ☐ Experts designated by the Heads of Delegation to represent a GBIF Participant country in scientific fora
4. GBIF's materials are not available in my language. What should I do?
   - ☐ Write to info@gbif.org to request the translation
   - ☐ Sign up to be a volunteer translator and contribute to the translation myself
   - ☐ Share information on how to be a volunteer translator with others to help in the translation effort
   - ☐ All of the above

Course complete

You have now completed the course.

As a reminder, this course serves as the prerequisite for other GBIF courses. You only need to take

the course one time, but you can of course refer back to the material at any time.

Please complete this form to register your completion.

Glossary

API

Application Programming Interface. A set of clearly defined methods of communication between various software components.

BID

Biodiversity Information for Development. An EU funded project co-ordinated by GBIF whose aim is to increase data mobilization capacity in the Africa, Caribbean and Pacific regions.

BIFA

Biodiversity Fund for Asia.

CC Licences

Creative Commons. These are a series of licenses set up by the Creative Commons organization that enable sharing and reuse of creativity and knowledge through the provision of free legal tools. Three of them can be assigned to GBIF-shared datasets: CC0, CC BY and CC BY-NC.

DwC

Darwin Core is a biodiversity data standard, maintained by TDWG & widely used within the GBIF community and partners. It is a set of standardized terms (or field names) and their definitions, which are used to share biodiversity information.

DOI

Digital Object Identifier. A persistent identifier or handle used to uniquely identify objects. DOIs are in wide use mainly to identify academic, professional, and government information, such as journal articles, research reports and data sets, and official publications.

DwC-A

Darwin Core Archive. A compressed (zipped) file containing all the information needed to share with GBIF, for a particular resource. Each zip contains three types of files:

1. the actual data, in one or more text files: occurrence.txt/event.txt/measurmentoffact.txt etc
2. a mapping file: rtf.xml
3. a metadata (EML) file: eml.xml When you publish using the IPT, it creates a Darwin Core Archive, which is shared with GBIF. Also, when you download data from the GBIF website you can choose a DwC-A format as well.

IPT

Integrated Publishing Toolkit. It is a free and open source web application (software) for publishing biodiversity data. The software itself lives on a server (either at your institution or elsewhere) that must have access to the internet 24/7. It is used to create and handle Darwin Core Archive files that can be shared and used by anyone including GBIF.

Data Publishing

With regards to GBIF we have a very specific definition of data publishing. It refers to making biodiversity datasets publicly accessible and discoverable, in a standardized form, via an access point, typically a web address (a URL).

Resource

A Resource is the collective term used to refer to a particular dataset and its metadata once it has been uploaded to an IPT instance.

TDWG

Taxonomic Databases Working Group, now renamed Biodiversity Information Standards.

Appendix: Solutions

This appendix contains the answers and additional information to all of the review quizzes.

About GBIF

What is GBIF?

GBIF is all of these:

- An intergovernmental network and research infrastructure
- A collaboration among governments and international organizations
- A network of participant nodes
- A secretariat, based in Copenhagen, Denmark

When was GBIF established?

- 2001

Which of the following is the best description of a GBIF Participant node?

- A team designated by a Participant country or organization to coordinate a network of people and institutions that produce, manage and use biodiversity data, collectively building an infrastructure for delivering biodiversity information

Which of the following is NOT a typical function of a GBIF Participant node?

- Maintaining a mirror website of the GBIF.org to ensure real-time backup of the GBIF data index and improve user access from within the country

What is a GBIF Participant?

- A country, economy or organization that joins GBIF by signing the Memorandum of Understanding and establishing a co-ordinated effort to support open access and use of biodiversity data, to advance scientific research, and to promote technological and sustainable development

What is a GBIF Head of Delegation?

- The person designated by the participating country/economy/organization to act as its representative to the GBIF Governing Board and take part in the global-level decision making

What is a Biodiversity information facility?

- The broader structure of people and institutions, coordinated by the node, that collectively forms an infrastructure for delivering biodiversity information to relevant stakeholders

What is a Node manager?

- The person designated by a participating country/economy/organization to manage the activities of the node to coordinate a biodiversity information facility

Who designates the institution that hosts the GBIF Participant node?

- The Head of Delegation

GBIF-mediated data

What dataset class makes up the core of data published within GBIF?

- Occurrence

What is the taxonomic backbone?

The GBIF taxonomic backbone is all of these: * A dataset * A management classification with the goal of covering all names in GBIF * Allows for taxonomic search on GBIF

Which licenses or waivers can be applied to datasets published in GBIF?

- CC BY
- CC BY-NC
- CC0

Images are subject to the same licenses as datasets?

- False

GBIF data are FAIR?

- True

Data publishing

What does data publishing mean in the context of GBIF?

- Making your biodiversity dataset(s) publicly accessible and discoverable in a standardized format

Which of the following are incentives for publishing data?

- contribute to global knowledge about biodiversity
- new opportunities for collaboration
- make data freely accessible

How do you become a publisher in the GBIF network?

- fill out the *Become a publisher* form on GBIF.org and wait for endorsement

There are no requirements for publishing your data on GBIF.org

- False

What is the GBIF data validator?

- a tool to check my data for issues

Data access

How can you access GBIF data?

You can access GBIF data with all of these methods.

- GBIF.org search interface
- GBIF API
- rGBIF

Which file formats are available for downloads of data?

- simple
- Darwin Core Archive
- species list

What kind of flags does GBIF apply to data to alert you to the quality?

- altered
- excluded
- inferred

Community of practice

Who does most of the translations of GBIF materials, including the GBIF.org website and training materials?

- Volunteer translators from GBIF's community of practice

How can I get support with a biodiversity data mobilization project?

All of these choices are possible:

- Write to info@gbif.org
- Request a volunteer mentor to provide remote support
- Undertake the online data mobilization course

What is a biodiversity open data ambassador?

- Biodiversity professionals who promote the principles and best practices of open data sharing and use

GBIF's materials are not available in my language. What should I do?

All of these options are possible:

- Write to info@gbif.org to request the translation
- Sign up to be a volunteer translator and contribute to the translation myself
- Share information on how to be a volunteer translator with others to help in the translation effort

Colophon

Suggested citation

Raymond M, Rodrigues A & Russell L A. Introduction to GBIF course First edition. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/ce-fcmk-aq49.

Authors

Mélianie Raymond, Andrew Rodrigues and Laura Anne Russell

Contributors

GBIF Secretariat staff.

Translators

French

- Maxime Coupremanne
- Patricia Mergen
- Sophie Pamerlon
- Carole Sinou

Spanish

- Leonardo Buitrago
- Victor Chocho
- Camila Plata
- Anabela Plos
- Miguel Vega
- Paula Zermoglio

Licence

Persistent URI

https://doi.org/10.35035/ce-fcmk-aq49

Document control

First edition, May 2021