

# Biodiversity data sharing

## Why and how Proteus Partners should share data through GBIF

### Key Messages

- This document explains why Proteus Partners should share biodiversity data through the Global Biodiversity Information Facility (GBIF) and then provides step-by-step guidance on how to do so.
- Open-access data sharing through GBIF feeds into global research and policy on biodiversity, including, for example, assessments for the IUCN Red List of Threatened Species. It can also provide operational benefits to companies and support compliance with emerging disclosure requirements.
- The simplest way for Proteus Partners to share data is using the existing GBIF infrastructure, working with one or more GBIF country-level facilities ('Participant Nodes'). This brief gives ten steps to follow when sharing data through GBIF in this way.
- Barriers to company data sharing, including technical and business barriers, are increasingly being overcome. Several companies, including Proteus Partners, have already shared biodiversity data through GBIF.
- Both UNEP-WCMC (through Proteus) and the GBIF Network can provide technical support to help Proteus Partners navigate the data sharing process.

## Data sharing overview: Why and how?

### Why share biodiversity data?

Data on biodiversity are collected by Proteus Partners as part of site management activities such as impact assessments and monitoring. This includes presence of species at and around sites, through direct observations and supported by technology like camera traps, drones, and environmental DNA (eDNA). The biodiversity data collected by Proteus Partners could benefit global conservation efforts, however at present it is rarely accessible for this purpose.

Sharing of biodiversity data collected by the private sector can strengthen existing datasets and resources. The value of making data available is recognised in the Kunming-Montreal Global Biodiversity Framework (e.g. target 21)<sup>1</sup>. Investor requirements are also driving biodiversity data sharing, for example financial institutions using the Equator Principles are required to encourage clients to share project biodiversity data (EP4 Principle 10)<sup>2</sup>, and the French Development Agency (AFD) now requires

any primary biodiversity data generated under financed/funded projects to be published<sup>3</sup>. Furthermore, it is integral to the Proteus Strategy 2021-2025 that the Partnership contributes towards open data, including enabling *"observation data collected by the private sector to feed into global datasets which then become much more useful as a result"*<sup>4</sup>.

Businesses themselves can benefit from sharing their biodiversity data. This includes operational benefits from adopting international best practices in data management, and allows companies to demonstrate leadership, increase transparency, and illustrate biodiversity values that can be maintained or restored on sites. Around 60 companies, including some Proteus Partners, are already sharing their biodiversity data through the Global Biodiversity Information Facility (GBIF).

## What is GBIF?

GBIF is an intergovernmental network and infrastructure that aims to increase the accessibility of data on biodiversity. It provides a global facility for organisations to share their biodiversity data, collating information from natural history collections, the public sector, non-governmental organisations, academia, citizen science networks, and the private sector. Free and open access to GBIF-mediated data is provided through an online platform<sup>5</sup>.

The GBIF platform accepts data on species occurrence – the presence of a species at a specific place and time – and collates it in a standardised format. The minimum information required is the scientific name of a species (or barcode sequence in the case of eDNA) and the date and location at which it was recorded. Further information can also be provided about the observation method, abundance, habitat characteristics, and a range of other parameters. Occurrence records may be associated with specific sampling events, which enables changes over time and species absences to be analysed.

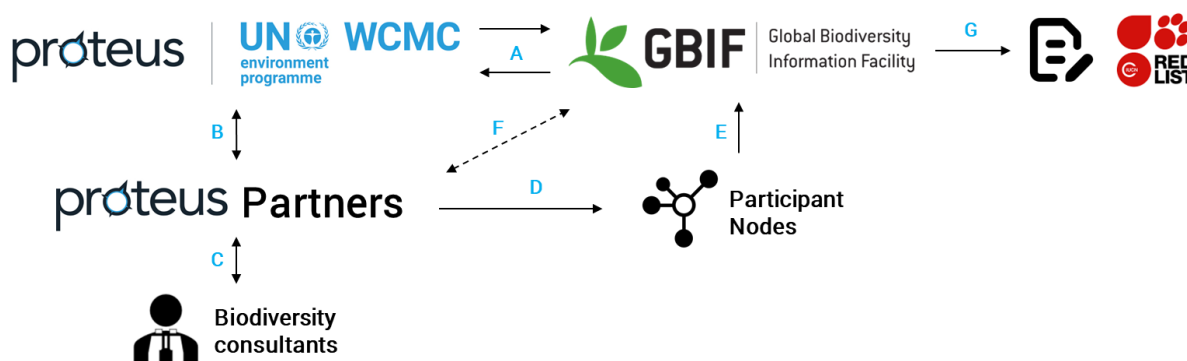
GBIF-mediated data is often used to strengthen other global biodiversity datasets. This includes

updating species assessments in the IUCN Red List of Threatened Species and identifying Key Biodiversity Areas. For example, in the July 2022 update of the IUCN Red List, GBIF-mediated data was used in 25% of the new and updated species assessments<sup>6</sup>. GBIF-mediated data is also used for many other scientific, policy, decision-making, and biodiversity conservation purposes.

## How can Proteus Partners use GBIF?

GBIF operates through a network of organisations coordinated by a Secretariat. The network is made up of formal GBIF members called ‘Participants’, which are mostly countries and international organisations. Each Participant establishes a ‘Node’ that is responsible for coordinating its GBIF-related activities. Nodes work with institutions, called ‘Publishers’, that contribute data to GBIF. Proteus Partners can contribute to GBIF as data Publishers. Figure 1 shows the different organisations Proteus Partners will work with to share data.

In conjunction with GBIF, Proteus Partners may also share data through national, regional, and/or thematic repositories, leveraging the same workflows.



**Figure 1. Pathway for Proteus Partners to share data through the Global Biodiversity Information Facility (GBIF) and contribute towards improving global biodiversity datasets.** UNEP-WCMC and GBIF collaborate to increase the accessibility of biodiversity data (A), and under the Proteus Strategy 2021-2025 UNEP-WCMC and Partners promote data sharing from the private sector (B). As part of their biodiversity management activities, Proteus Partners collect and format data to align with GBIF standards, which may involve working with consultants (C). To share data, Proteus Partners will become a GBIF Publisher through a Participant Node (D) and upload it to the GBIF online platform (E). The GBIF Secretariat has a helpdesk that may also provide direct support (F). GBIF-mediated data can then feed into other global biodiversity datasets (G).

## Overcoming barriers

An increasing number of businesses are overcoming barriers and sharing data through GBIF. Table 1 lists some common perceived barriers and approaches through which they can be addressed. Detailed guidance on how to share data is provided later in this document, including further information on overcoming several of these barriers.

**Table 1. Approaches to address common barriers to sharing biodiversity data.**

| Potential barrier                                | Response  |
|--|---|
| Sensitive biodiversity information               | Follow guidance provided by GBIF on identifying sensitive species and obscuring location data <sup>7</sup>                            |
| Sensitive commercial information                 | Delay and/or avoid publication of data that compromises commercial sensitivities  |
| Accusations of greenwashing                      | Ensure communications around data sharing are accurate and nuanced, with consistency on why data is/isn't shared                      |
| Uncertainty around data quality and formatting   | Follow GBIF guidance on types of data that are suitable to share and formatting requirements <sup>8,9</sup>                           |
| Aligning consultant work with GBIF requirements  | Where possible, instruct consultants to collect data aligned with GBIF requirements from the outset of projects                       |
| Financial costs of data sharing                  | Publication is free, and staff time requirements are small <sup>10</sup> with potential data management co-benefits                   |
| Lack of national capacity to accept data         | Companies have options to work with different GBIF Nodes, and a helpdesk coordinated by the GBIF Secretariat can support where needed |
| Reputational risks if biodiversity is damaged    | Greater transparency will reward high performance, benefitting leading companies  |
| Companies can use GBIF without contributing data | GBIF is not depleted by use, and companies that share data can gain operational benefits  |
| Lack of precedent                                | Around 60 companies have shared data to GBIF already, and further data sharing offers low-cost leadership opportunities               |

## Support available

Proteus Partners should take internal ownership of their company data sharing activities. Several sources of advisory support are available. Through Proteus, UNEP-WCMC can support Partners looking to share data to navigate the process. The GBIF Participant Node(s) that Partners engage with will also be able to provide guidance. In addition, the GBIF Secretariat coordinates a help-desk facility, and is looking to upscale its capacity to support as more data from the private sector is shared to GBIF.

An increasing number of consultants are also gaining experience of sharing biodiversity data to GBIF, and some are beginning to offer data sharing support as part of their services. NatureMetrics, an eDNA consultancy for example, has committed to launching a service for clients to share data to GBIF<sup>11</sup>.

Step-by-step written guidance for implementing data sharing is provided in the next section of this document. Other sources of written guidance are also available. This Proteus Technical Briefing covers many of the same topics as other guidance documents, focusing on the information most relevant for Proteus Partners, but more detail on some aspects is provided elsewhere, including:

- [Guidance from the GBIF Secretariat](#) on companies becoming data Publishers, which provides more information on registering as a data Publisher and gaining internal buy-in
- [Guidance from the GBIF Secretariat](#) on best practices for sharing data from Environmental Impact Assessments, which provides more information on different data sharing options
- [Guidance from the Data4Nature initiative](#) on sharing in the context of the French Development Agency requirements
- [Guidance for clients of Equator Principle financial institutions](#) on data sharing in the context of the Equator Principles.

Considerable practical experience will be developed within the Proteus Partnership as more Partners share data to GBIF. Partners can share progress and learnings within the Partnership, to leverage collective knowledge in support of greater uptake and flow of data to

GBIF. Partners looking to share data are encouraged to contact the UNEP-WCMC Proteus team ([proteus@unep-wcmc.org](mailto:proteus@unep-wcmc.org)), both to access available support and identify opportunities for sharing learning and insight.

## Detailed guidance: How to share data

A series of steps to follow when sharing data to GBIF are explained below. This is for company environment specialists to use with support from data teams. Under each step, the guidance outlines a standard approach that most Proteus Partners will follow. Under some steps there are alternatives to the approach described. Where this is the case, links to sources of further information are provided in the text.

### Prepare for data sharing

#### Step 1. Review Data Publisher Agreement and Data User Agreement

Organisations publishing data to GBIF are required to agree to the conditions of a Data Publisher Agreement, outlined [here](#). As the first step before sharing data, review the Data Publisher Agreement to ensure your company can agree to these terms. It is also recommended to familiarise yourself with the Data User Agreement, explained [here](#). While agreeing to the Data User Agreement is not required to submit data to GBIF, it provides context on how data your company provides can be used.



#### Step 2. Secure company agreement

Before sharing data to GBIF you will need to secure support for your plans to publish data on behalf of your company, through the appropriate internal channels. Further information to support gaining internal buy-in is provided by GBIF [here](#). When you request to become a data publisher (step 3) GBIF will ask you to confirm that the responsible authorities in your company are aware of the registration.



#### Step 3. Register as a data publisher

You can register to share data to GBIF through the online form [here](#). GBIF is organised into 'Nodes' that coordinate activities of 'Participants'. GBIF Participants are formal members of the GBIF intergovernmental initiative, consisting mainly of countries and international organisations (a full list of GBIF Participants is available [here](#)). To share data through GBIF your company will need to receive endorsement as a data publisher from one of GBIF's Nodes. Endorsement is given based on the criteria [here](#), which all Proteus Partners are expected to fulfil. The registration form asks for details about your organisation, contact information, the kind of data you plan to publish, and what support you need. This guides identification of an appropriate GBIF Node to work with, which will often be a country Node.



#### Step 4. Set up data hosting arrangements

When you register to become a data publisher your GBIF Node can provide you with a company account within GBIF's 'Integrated Publishing Toolkit'. You can log in to your account from [ipt.gbif.org](http://ipt.gbif.org). The Integrated Publishing Toolkit provides an online platform through which you can upload data to GBIF. A live demo of using the Integrated Publishing Toolkit is available [here](#). It is expected that most Proteus Partners will upload data to GBIF using their sub-account linked to a Participant Node in the Integrated Publishing Toolkit. Alternative data hosting arrangements are also possible, for example you can install and maintain your own instance of the Integrated Publishing Toolkit, or use an API to publish data. If you wish to explore these alternative arrangements, further information is available [here](#).

## Identify data to share

### Step 5. Identify data to share

Once registered as a publisher you can begin uploading data to GBIF. GBIF accepts four classes of datasets, 1) resources metadata, 2) checklist data, 3) occurrence data, and 4) sampling-event data. These dataset classes are explained in more detail [here](#) – note that they describe increasing amounts of information rather than different types of data. Proteus Partners will often be sharing occurrence data – that is information on the presence of species at a specific location and time (e.g. from biodiversity baseline surveys). Proteus Partners may also have sampling-event data that includes information such as the methods used, relative abundance of species, and changes over time. Identify the data you will share and the associated data class. You may plan to share pre-existing data that your company already has and/or data that it is planning to collect in the future. For pre-existing data, some reformatting may be required to align with GBIF's requirements (see step 8). For data yet to be collected, it is recommended to familiarise yourself with the format required by GBIF to align with this from the outset.



### Step 6. Consider sensitive species

When publishing data to GBIF, responsibility for restricting access to sensitive data lies with you. Sensitive data could include information about species with small, threatened or economically profitable populations that may be jeopardised by their locations being more widely known (for example precise location data for species that are poached). It may be necessary to intentionally reduce the accuracy of data for sensitive species, for example by rounding geographic coordinates. GBIF provides detailed guidance on identifying and generalising data for sensitive species, available [here](#), which you should follow.



### Step 7. Identify appropriate data license

GBIF allows you to select one of three Creative Commons licenses to apply to your dataset. You will be asked to select the license when publishing the data through GBIF's online portal (step 10), but it can be important to consider this earlier in the process. The three licenses that GBIF allows you to apply are: 1) [CC0](#) which dedicates the data to the public domain without any restrictions; or 2) [CC BY](#) under which the data can be used with appropriate credit to the provider (i.e. your company); or 3) [CC BY-NC](#) under which the data can be used with appropriate credit to the provider but not for commercial purposes or by commercial entities (such as other companies). To support open access data sharing

with appropriate attribution, we encourage Proteus Partners to share data to GBIF under the CC BY license wherever possible.

## Collect and publish data

### Step 8. Collect/format data for publication

All data in GBIF is formatted according to consistent rules and conventions. This enables data from many different organisations to be collated into useable information. GBIF uses the Darwin Core Archive format. GBIF provides Excel templates with headings to assist appropriate data formatting – the template for occurrence datasets can be downloaded from [here](#), and the template for sampling-events datasets can be downloaded from [here](#). Further information about the data formats used in GBIF is available [here](#). When sharing existing data from previous work you may need to re-format using these templates. If you are about to begin collection of new data you should consider using the Darwin Core format from the outset. When formatting your data remember to consider sensitive species based on the guidance from GBIF (see step 6).



### Step 9. Check data formatting

GBIF provides a Data Validator tool that you can use to check datasets prior to publication. This tool can highlight potential issues such as duplicate records, incomplete fields, and inconsistencies in formatting. It assists in troubleshooting before you try to upload data to GBIF. You can access it [here](#). GBIF also provides a Species Matching tool to align the scientific names of species in your data with those used in GBIF, which you can access [here](#).



### Step 10. Publish data

Upload your data to GBIF through your company's Integrated Publishing Toolkit account (see step 4). Log in to your account and select 'Create New Resource' under 'Manage Resources'. Then fill out the online form with the requested information. If you have completed the steps above, you will have all the input information required. To make your data publicly accessible ensure to select 'Public' in the 'Visibility' section of the Integrated Publishing Toolkit form.

Once your data are published, you can track its use through your Integrated Publishing Toolkit account. This provides summary metrics on your dataset, user download activity, and traceable literature citations.

## Next steps

Proteus Partners are encouraged to share data through GBIF. Partners who are not yet sharing data should start exploring how they can do so, both within their business and by discussing with the UNEP-WCMC Proteus team ([proteus@unep-wcmc.org](mailto:proteus@unep-wcmc.org)). To begin, you may wish to identify one (or a small number of) dataset(s) from a single site to pilot the process and build internal expertise, and over time share more data for

more sites. Later in 2023, Partners who are already publishing data will have the opportunity to share their experiences with the wider Partnership through a peer-to-peer Data Forum. Given the momentum for open data and recognition of the value of data collected by the private sector, we expect sharing of biodiversity data by businesses to increasingly become standard practice.

## References

---

- <sup>1</sup> Convention on Biological Diversity (2022). Kunming-Montreal Global Biodiversity Framework. Available [here](#) [accessed 14 March 2023].
- <sup>2</sup> Equator Principles (2020). EP4. Available [here](#) [accessed 14 March 2023].
- <sup>3</sup> Data4Nature (2021). Practical recommendations guide for publishing primary biodiversity data. Available [here](#) [accessed 14 March 2023].
- <sup>4</sup> UNEP-WCMC (2020). Proteus Strategy 2021-2025. Available [here](#) [accessed 14 March 2023].
- <sup>5</sup> GBIF | Global Biodiversity Information Facility (2023). Available [here](#) [accessed 14 March 2023].
- <sup>6</sup> GBIF (2022). News: GBIF-mediated data used in one quarter of all new and updated IUCN Red List assessments. Available [here](#) [accessed 14 March 2023].
- <sup>7</sup> Chapman AD (2020). Current Best Practices for Generalizing Sensitive Species Occurrence Data. Copenhagen: GBIF Secretariat. Available [here](#) [accessed 14 March 2023].
- <sup>8</sup> GBIF (2023). Dataset classes. Available [here](#) [accessed 14 March 2023].
- <sup>9</sup> GBIF (2023). Data standards. Available [here](#) [accessed 14 March 2023].
- <sup>10</sup> GBIF (2023). Data4Nature: Share to protect. Available [here](#) [accessed 14 March 2023].
- <sup>11</sup> NatureMetrics (2022). New & Articles: NatureMetrics launches Nature Transparency service with the Global Biodiversity Information Facility to supercharge global understanding of Nature Performance. Available [here](#) [accessed 14 March 2023].

---

**Citation:** UNEP-WCMC & GBIF (2023). Biodiversity data sharing: Why and how Proteus Partners should share data through GBIF. Cambridge, UK.

**Authors:** Alex Ross (UNEP-WCMC) and Tim Hirsch (GBIF)

**Acknowledgements:** With thanks to reviewers including Stacey Baggaley, Matt Kaplan, Matt Jones and Neil Burgess

**Available online at:** [www.proteuspartners.org](http://www.proteuspartners.org)

**Contact:** Alex Ross ([alex.ross@unep-wcmc.org](mailto:alex.ross@unep-wcmc.org)), Proteus ([proteus@unep-wcmc.org](mailto:proteus@unep-wcmc.org))

Copyright: 2023 United Nations Environment Programme

The UN Environment Programme World Conservation Monitoring Centre (UNEP-WCMC) is a global Centre of excellence on biodiversity. The Centre operates as a collaboration between the UN Environment Programme and the UK-registered charity WCMC. Together we are confronting the global crisis facing nature.

**Copyright release:** This publication may be reproduced for educational or non-profit purposes without special permission, provided acknowledgement to the source is made. Reuse of any figures is subject to permission from the original rights holders. No use of this publication may be made for resale without permission in writing from UN Environment Programme. Applications for permission, with a statement of purpose and extent of reproduction, should be sent to the Director, UNEP-WCMC, 219 Huntingdon Road, Cambridge, CB3 0DL, UK.

**Disclaimer:** The contents of this report do not necessarily reflect the views or policies of the UN Environment Programme, contributory organisations or editors. The designations employed and the presentations of material in this report do not imply the expression of any opinion whatsoever on the part of the UN Environment Programme or contributory organisations, editors or publishers concerning the legal status of any country, territory, city area or its authorities, or concerning the delimitation of its frontiers or boundaries or the designation of its name, frontiers or boundaries. The mention of a commercial entity or product in this publication does not imply endorsement by the UN Environment Programme.

<http://www.unep-wcmc.org/>

