

Guidance for private companies to become data publishers through GBIF

Rui Figueira • Pedro Beja • Cristina Villaverde • Miguel Vega • Katia Cezón • Tainan Messina •
Anne-Sophie Archambeau • Rukaya Johaadien • Dag Endresen • Dairo Escobar

Version 3.0, January 2023



Table of Contents

Colophon	1
Suggested citation	1
Authors	1
Licence	1
Persistent URI	1
Document control	1
Abstract	1
Cover image	1
1. Introduction	1
2. <i>Proposal</i> : Biodiversity Data Publication through GBIF	2
2.1. Presentation	2
2.2. What is GBIF?	3
2.3. What are the advantages of publishing data to GBIF?	3
2.4. Who publishes data through GBIF?	4
2.5. What data could the company publish through GBIF?	8
2.6. What does it take for a company to be a data publisher to GBIF?	8
2.7. Involvement of different parties in the publication process	9
2.8. How could the proof of concept be developed?	9
2.9. What are the costs for this company ?	10
References	10

Colophon

Suggested citation

Figueira R, Beja P, Villaverde C, Vega M, Cezón K, Messina T, Archambeau A, Johaadien R, Endresen D & Escobar D (2020) Guidance for private companies to become data publishers through GBIF: Template document to support the internal authorization process to become a GBIF publisher. Copenhagen: GBIF Secretariat. <https://doi.org/10.35035/doc-b8hq-me03>

Authors

Rui Figueira, Pedro Beja, Cristina Villaverde, Miguel Vega, Katia Cezón, Tainan Messina, Anne-Sophie Archambeau, Rukaya Johaadien, Dag Endresen & Dairo Escobar

Licence

The document *Guidance for private companies to become data publishers through GBIF* is licensed under [Creative Commons Attribution-ShareAlike 4.0 Unported License](#).

Persistent URI

<https://doi.org/10.35035/doc-b8hq-me03>

Document control

Version 2.0, June 2020

languageLinks:combined[]

Abstract

This document is a contribution to the [OpenPSD project: Promoting publication and use of private-sector data on biodiversity](#), a joint creation of the GBIF nodes of [Spain](#), [Portugal](#), [Norway](#), [France](#) and [Colombia](#) and the companies [EDP](#), [CIBIO](#) and [Asplan Viak](#), and co-funded by the GBIF Secretariat through its [Capacity Enhancement Support Programme](#).

Cover image

Sea fig (*Carpobrotus edulis*), Aljezur, Portugal. Photo 2020 sheborg via [iNaturalist research-grade observations](#), licensed under [CC BY-NC 4.0](#).

1. Introduction

The steps for a company to become a data publisher through [GBIF—the Global Biodiversity Information Facility](#)— are relatively simple. However, the future publisher needs to be aware of the responsibilities of GBIF data publishers and GBIF data users, as well as the different aspects related to preparation, publication and monitoring of published data.

Depending on the size of the company, it may be that the first contact with GBIF is made by staff

from the Environment or Sustainability departments. However, the decision to become a publisher will usually be made by an administrator or director, who must be informed about the meaning of becoming a GBIF publisher and what are the benefits for the company.

A [guideline document for private-sector organizations that develop Environmental Impact Assessments \(EIAs\)](#) was published by GBIF and the [International Association for Impact Assessments \(IAIA\)](#) which covers most of the benefits and technical aspects of data publication through GBIF and this document will be updated in 2020. It lacks, however, the components related to costs, license adoption and relation with different parties that are involved in data production within an organization, which might include subcontractors.

The purpose of this document is to work as a template for an internal documentation process to be presented to the company's decision-making bodies to inform and support their decision in becoming a GBIF publisher. In that regard, several aspects are addressed, such as:

- What is GBIF?
- What are the advantages of publishing data through GBIF?
- Who publishes and what data is published through GBIF?
- What steps do you need to complete in order to become a data publisher?
- How to involve all parts associated with the datasets in your publication?
- What are the costs associated with publishing?

This template proposes that a proof of concept should be developed internally. This will allow to demonstrate the steps of data publishing and working as an example for the internalization of data publishing in the company's biodiversity information processing, including the assessment of internal and external resources that need to be allocated.

2. Proposal: Biodiversity Data Publication through GBIF

2.1. Presentation

It is proposed in this document to involve **company** as a publisher of biodiversity data on the platform of [GBIF—the Global Biodiversity Information Facility](#).

In the context of global biodiversity decline, species occurrence and abundance data are essential tools for planning, implementing and monitoring of conservation and sustainable use strategies. This importance is recognized on a global scale by the [United Nations Sustainable Development Goals](#), namely SDG14 and SDG15, to which GBIF becomes instrumental for the assessment of the progress towards the goals. Likewise, Aichi Target 19 of the Convention on Biological Diversity (CBD) uses the publication of data through GBIF by 2020 as [one of the indicators to measure the Target's compliance](#).

Alongside universities, research institutes and governmental agencies, the business sector can play a key role here as it develops thousands of environmental impact assessments locally and worldwide, which annually produce millions of records. However, participation in this effort has been very low as data produced are never or very rarely made widely available. In this area, **company** could take a prominent position at **national / international** level, contributing with data to GBIF in an organized and systematic way.

Therefore, notwithstanding the necessary care with aspects of intellectual property and confidentiality, it is proposed to carry out a proof of concept to develop and evaluate a model of biodiversity data publication by **company**, which works as a case study for the company and to GBIF. This exercise would have several advantages for **company**, helping to consolidate its reputation at **national / international** scale as a leader in the field of sustainability and to strengthen data availability, integration and organization processes that could be used for future investments.

2.2. What is GBIF?

The Global Biodiversity Information Facility is an intergovernmental organization established in 2001 to facilitate the free sharing and open access of biodiversity data. There are currently 59 signatory countries of the GBIF Memorandum of Understanding. GBIF provides a single access point (www.gbif.org) to over one billion global biodiversity data and is the largest biodiversity network available via the Internet. Data accessible through GBIF relates to records of more than 1.6 million species collected over three centuries of exploration of Natural History and includes recent observations by citizens, researchers and automated monitoring programs. Data downloaded through GBIF were used between 2018 and 2019 in more than 1,300 scientific articles in international journals. Globally, GBIF has agreements and provide services directly to global policy-making initiatives for the assessment and conservation of biodiversity and the environment, such as the Convention on Biological Diversity (CBD), the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) and the GEO BON (Group on Earth Observation - Biodiversity Observation Network).

Paragraph with information about national GBIF node

2.3. What are the advantages of publishing data to GBIF?

Publishing biodiversity data in GBIF is an initiative that could be highly visible globally, helping to consolidate **company's** reputation as a reference in the field of sustainability. In fact, data publication through GBIF is seen as very important within the CBD and has even been chosen as one of the Aichi Targets (Target 19) indicators of the Biodiversity Decade (2010-2020). Data published through GBIF also contributes to Aichi Targets 9 (invasive species), 11 (protected areas) and 12 (threatened species). Likewise, these data will be relevant also to the assessment of the SDG 14 (Life below water) and 15 (Life on Land).

By contributing to these objectives, **company** will also be improving its environmental and social profile by increasing the return on investment initially made to obtain these data. GBIF is very keen to use and disseminate new case studies, and data mobilization from the business sector is currently an important objective of the organization. This process could favourably contribute to the organization's performance on the Dow Jones Sustainability Index and the assessment of the company in the framework of Equator Principles.

Another relevant advantage of this process would be to lay the foundations for improving the information management model collected in environmental impact assessment processes, as well as impact monitoring and compensatory measures. Collecting these data involves very costly studies, but where the added value of the information collected is often very scarce due to the inefficient way it is often managed later. Because of this, information is usually lost or difficult to access and therefore, does not contribute to internal learning processes or to the reuse of information collected through new investments or environmentally sustainable management of the infrastructures in operation.

Publishing data through GBIF in interconnection with a national GBIF Node, would provide motivation and additional tools for the management processes of this type of information, which would

potentially be scalable for the **company** group at national and international levels.

In summary, according to the report Digitally Transforming Environmental Assessment, when private companies publish their biodiversity data, this will result in:

- Lower cost of information search
- Reduced need for detailed field surveys
- More complete impact information at the pre-referral stages may reduce the need for a formal EIA (6-18 month time saving)
- Improved monitoring data reduce risk of operations being temporarily suspended due to a single compliance breach
- Increased confidence among environmentally aware companies for business investment and future partnerships
- Improved predictability of accessing markets
- Using open tools and data to guide environmental decisions will make the process more transparent and build trust and certainty in the process
- Increased investor and community confidence will deliver continued interest in business development granted by public bodies
- Better tools to process easily accessible data will help assess the scale of the impact, and the value, type and success of any proposed offset

2.4. Who publishes data through GBIF?

To date, GBIF includes more than 60,000 datasets on its portal published by **more than 1,700 data publishing institutions**. These publishers are mostly government agencies, natural history museums and herbaria, universities, research centres and non-governmental organizations of various kinds.

The almost complete absence of the business sector is noticeable, which limits access to a vast and important volume of biodiversity data collected by companies around the world. In total, private companies now publish at least 7,789,180 occurrence records, accounting for 0.3 per cent of **all records published in GBIF**.

Table 1 lists the most significant private sector publishers.

Table 1. Private-sector companies that publish their data through GBIF (as of August 2023)

Company	Activity sector	Country	Datasets	Occurrence records	Data citations
AGBAR	Consulting	Spain	1	103,424	42
ARC - Arctic Research and Consulting DA	Consulting	Norway	1	8,914	60
Aguas de Bogotá S.A. E.S.P.	Utilities	Colombia	1	13,280	63
Akvaplan-niva	Consulting	Norway	3	594	13
Anadarko Colombia Company	Energy	Colombia	7	1,178	41

AngloGold Ashanti Colombia S.A.S	Materials	Colombia	5	87,020	113
Asplan Viak AS	Engineering	Norway	14	3,775	349
Aures Bajo	Energy	Colombia	2	368	22
Awake Travel	Consulting	Colombia	1	8,644	9
Aigos SAS	Consulting	Colombia	3	2,404	37
Biofokus	Consulting	Norway	1	605,695	927
Biolog J.B. Jordal AS	Consulting	Norway	1	177,814	550
Biotica Consultores Ltda	Consulting	Colombia	4	1,318	143
Carbones del Cerrejón Limited	Materials	Colombia	9	197,100	178
Carsa Gold S.A.S	Mining	Colombia	1	4,159	37
Celsia Colombia S.A. E.S.P.	Energy	Colombia	5	2,290	45
Central Hidroeléctrica de Caldas S.A E.S.P	Energy	Colombia	1	1,137	23
Cerro Matoso S.A	Materials	Colombia	3	19,309	131
Chevron Australia	Energy	Australia	1	2,048	53
Compensation International Progress S.A. -Ciprogress Greenlife-	Industrials	Colombia	1	820	51
Concesión La Pintada S.A.S	Industrials	Colombia	2	0	0
Construcciones y Ambiente Conambiente S.A.S	Consulting	Colombia	3	273	43
Cunaguaro Consultores LTDA	Consulting	Colombia	1	657	34
DNV	Energy	Norway	1	2,372,473	51
EDP - Energias de Portugal	Energy	Portugal	106	1,831,557	349

ENGIE	Energy	France	5	13,888	0
Ecofact	Consulting	Norway	3	12,508	382
Econativa Consultores SpA	Consulting	Chile	1	3	7
Ecopetrol S.A.	Energy	Colombia	45	397,693	90
Empresas Públicas de Medellín E.S.P.	Energy	Colombia	39	2,151,219	113
Enel Colombia	Energy	Colombia	11	29,101	30
Equinor	Energy	Norway	1	1,017	1
Faun Naturforvaltning AS	Consulting	Norway	1	3,788	344
Federación Nacional de Cacaoteros	Agriculture	Colombia	1	17	12
Federación Nacional de Cafeteros de Colombia	Agriculture	Colombia	6	26,804	343
Grupo Energía Bogotá	Energy	Colombia	1	61,111	99
HBH Projekt spol. s r.o., Kabátníkova 5, 602 00 Brno, ČR – organizačná zložka Slovensko	Engineering	Slovakia	2	204	4
Hatovial S.A.S	Engineering	Colombia	1	1,898	118
INERCO Consultoría Colombia	Consulting	Colombia	1	1,090	132
Isagen S.A. E.S.P.	Energy	Colombia	12	41,665	285
LafargeHolcim Spain	Mining	Spain	2	35	30
Lake Tanganyika Floating Health Clinic	Health Care	Congo, Democratic Republic of the	1	132	4
Mineros Aluvial S.A.S. BIC	Mining	Colombia	1	7,307	15

Moam Monitoreos Ambientales S.A.S	Consulting	Colombia	1	1,781	45
Monitoramento fauna e flora Mineração Vale Verde do Brasil Ltda.	Materials	Brazil	1	299	88
Multiconsult	Consulting	Norway	1	308	133
NNI Resources AS	Consulting	Norway	2	3,116	84
NaturRestaurering AS	Consulting	Norway	8	16,024	212
Nature monitoring data, Amphi Consult and Biomedica, Denmark	Consulting	Denmark	1	47,254	1
Navantia, S.A.	Industrials	Spain	6	823	18
Nocturne Environmental Surveyors Ltd	Consulting	United Kingdom of Great Britain and Northern Ireland	1	32	16
Oleoducto Bicentenario	Energy	Colombia	11	4,161	211
Parex Resources Colombia - AG Sucursal	Energy	Colombia	8	41,581	4
Pierre Fabre	Consumer Staples	France	20	4,049	112
Promigas S.A E.S.P	Energy	Colombia	12	180,848	216
Regelink Ecology & Landscape	Consulting	Netherlands	1	157,976	96
Rådgivende Biologer	Consulting	Norway	5	15,214	323
SWECO Norge AS	Engineering	Norway	1	1,139	327
Stratos Consultoría Geológica	Consulting	Colombia	2	1,084	25
TERRASOS	Consulting	Colombia	9	24,817	201

TotalEnergies	Energy	France	14	22,232	89
Veolia Colombia	Energy	Colombia	2	672	1

Table 2. Grand Totals

Datasets	Occurrence records	Data citations
418	8,719,141	7,575

2.5. What data could the company publish through GBIF?

Companies that carry out environmental impact assessments, impact monitoring and compensatory measures studies, thereby collect species occurrence and abundance data, may publish them on GBIF.

A lot of these data are collected in regions that lack sampling efforts and are less known, or have groups of organisms that are underrepresented and would, therefore, be valuable to the scientific community and to organizations such as CBD, IPBES or GEO BON.

Even data from studies in better-known regions could be of high value as they allow information gaps to be filled and improve time series representations. Thus, all data collected by **company** as part of its operation could be published in GBIF, without injury to the need to protect intellectual property issues, or transitory or permanent confidentiality of the information.

If data includes sensitive information, such as the location of threatened, sensitive or economically valuable species, it is recommended to apply [best practices for generalizing this information](#).

Thus, data collected by private companies can be published through GBIF in a relatively short period, if procedural aspects of publication are completed and the data format is adapted to GBIF standards (primarily [Darwin Core](#)). Also, the national node may be provided all technical helpdesk needed for the standardization process.

2.6. What does it take for a company to be a data publisher to GBIF?

The decision to become a publisher of biodiversity data at GBIF would first come with a decision by the **company** management bodies. After that, it is necessary to complete a set of steps that are common to any institution applying for data publishing:

- To guarantee institutional arrangements to ensure that all parties involved in the process, from management to the partners from information production, agree to data publishing and to the terms by which it takes place
- To acknowledge and agree to the [Data Publisher Agreement](#) (the English version is valid for legal purposes)
- To be aware of the [Data User Agreement](#), that GBIF data users must agree before using them
- To apply for the institution to register with GBIF as a data publisher and request the endorsement of the national node. Application for registration and endorsement is made online with [this form](#)

2.7. Involvement of different parties in the publication process

Depending on the size of the projects that originated the datasets, it is possible that the company's biodiversity data was obtained by hiring other companies or organizations that carried out the sampling work. This is the most common situation in an EIA or monitoring study, where sampling services are subcontracted. Involvement of these contractors and field technicians who have observed or identified species in the data publishing process is desirable, whenever possible. These technicians can play a relevant role, notably in reviewing data and metadata, contributing to better description and quality of the dataset. On the other hand, it is equally important for them to be recognized and accredited for their work and to associate them with their records. Another way to associate them and their organizations with the dataset is by identifying the associated parties when preparing metadata. Also, they have to be included as co-authors of the dataset and recommended citation.

2.8. How could the proof of concept be developed?

The proof of concept regarding data publication in GBIF could be developed involving the following steps:

1. Development of the company's internal processes leading to the decision to publish data on GBIF on an experimental basis.
2. Application for the company's registration in GBIF as a data publisher.
3. To build a case study on the company's involvement as a data publisher with the regional Node, their country and with the international GBIF, in order to give visibility to the process worldwide and encourage the involvement of other companies as publishers of biodiversity data.
4. Selection of an initial dataset to be published through GBIF, resulting from studies carried out by the company. This should provide a good representation of the taxonomic groups' diversity and data typologies, in order to assess different kinds of potential problems related to the organization and availability of information.
5. Definition of information type to be published and any restrictions on its publication, e.g. due to the presence of sensitive species, confidential information, data pending validation by government institutions, etc. Occurrence data (i.e. observation or collection of a given species at a certain place and date) or abundance data may be published.
6. Establishment of agreements with data producers (i.e. the institutions and staff hired by the company to collect data for the purposes of the studies) to safeguard intellectual property rights.
7. Formatting of data to be published according to the Darwin Core standard used by GBIF to prepare databases for publication.
8. Selection of a [Creative Commons licence](#) for the data to be published, which can be one of the following: CC0, CC-BY, CC-BY-NC. Depending on their characteristics, one of these licenses may be assigned.
9. Publishing data and metadata for each dataset to the GBIF portal. Information publishing options will be evaluated, in all cases using a technology platform developed by GBIF: the Integrated Publishing Toolkit (IPT). GBIF Nodes maintain an IPT, which they make available for hosting publisher datasets from their countries. It is also possible for the company to install and maintain its own IPT. In both cases, the datasets publisher is always the institution, not the Country Node, and the institution is responsible for managing the data (e.g. change, update) autonomously.
10. Monitoring the use of published data for a period of one year after its publication in GBIF. This will be done through statistics provided to the publisher regarding data transfer. In addition, the use

of data in scientific publications will be monitored, which is facilitated by assigning a globally unique Document Object Identifier (DOI) to each dataset registered via GBIF and to each dataset downloaded through GBIF.

2.9. What are the costs for this company?

Apart from the dedication time provided by **company** staff involved in preparing the proof of concept, there are no additional costs for the company. The necessary work may be supported by the GBIF National Node, that has the knowledge and infrastructure necessary to facilitate this publication. When the National Node makes its IPT facility available for hosting and publishing data, it is recommended that this service is framed by the Service Level Agreement between the GBIF Node (as a service provider) and the company (as the user of the service). This service also has no associated costs. In addition, the GBIF Node could provide training on data publishing through GBIF, contributing to the capacity of the **company** in the fields of biodiversity information management and data quality.

References

- Cadman M, Chavan V, King N, Willoughby S, Rajvanshi A, Mathur V, Roberts R & Hirsch T (2011) Publishing EIA-Related Primary Biodiversity Data: GBIF-IAIA Best Practice Guide. Fargo, N.D., USA: IAIA Special Publication Series No. 7. Accessible at <https://www.iaia.org/uploads/pdf/sp7.pdf>
- Hatton T, Lockwood N, Moran C, Smyth E & Vogel P (2019) Digitally Transforming Environmental Assessment: Report of the Digital Environmental Impact Assessment Working Group. Perth, Australia: Western Australian Biodiversity Science Institute. Available at: https://wabsi.org.au/wp-content/uploads/2019/10/Digitally-Transforming-EIA_Working-Group-Report-FINAL-1.pdf
- Chapman AD (2020) Current Best Practices for Generalizing Sensitive Species Occurrence Data [Community review draft]. Copenhagen: GBIF Secretariat. <https://doi.org/10.15468/doc-5jp4-5g10>
- Secretariat of the Convention on Biological Diversity. (1st April 2020) Strategic Plan for Biodiversity 2011–2020 and the Aichi Targets - “Living in Harmony with Nature”. Downloaded from: <https://www.cbd.int/doc/strategic-plan/2011-2020/Aichi-Targets-EN.pdf>
- S&P Global Inc. (1st April 2020) Dow Jones Sustainability Indices. <https://www.robecosam.com/csa/indices/djsi-index-family.html>