

Documentation to support the internal authorization process to become a GBIF publisher

The steps for a company to become a data publisher through GBIF 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 for organisations in the private sector developing EIA was published by GBIF-IAIA (<https://www.iaia.org/uploads/pdf/sp7.pdf>) 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 organisation, 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.

Proposal:

BIODIVERSITY DATA PUBLICATION THROUGH GBIF (GLOBAL BIODIVERSITY INFORMATION FACILITY)

Presentation

It is proposed in this document to involve <company> as a publisher of biodiversity data on the GBIF (Global Biodiversity Information Facility) intergovernmental platform. 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 Goals¹, 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 GBIF² as one of the indicators to measure the Target's compliance, by 2020. 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 data. 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.**

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).

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 Targets³ (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 favorably 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*³, when private companies publish their biodiversity data, this will result in:

1. Lower cost of information search.
2. Reduced need for detailed field surveys.
3. More complete impact information at the pre-referral stages may reduce the need for a formal EIA (6-18 month time saving).
4. Improved monitoring data reduce risk of operations being temporarily suspended due to a single compliance breach.
5. Increased confidence among environmentally aware companies for business investment and future partnerships.
6. Improved predictability of accessing markets.
7. Using open tools and data to guide environmental decisions will make the process more transparent and build trust and certainty in the process.
8. Increased investor and community confidence will deliver continued interest in business development granted by public bodies.
9. 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.

Who publishes data through GBIF?

To date, GBIF includes 47,478 datasets on its portal published by 1,536 registered publishers. [These publishers](#) are mostly government agencies, Natural History Museums, universities, research centers 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 company registers amount to 1,069,940, which is less than 0.08 % of [records total published in GBIF](#). Table 1 lists the most significant private sector publishers.

Table 1. Private sector companies that publish their data through GBIF (data up to the 23h of March, 2020).

Company	Activity sector	Country	Number of datasets	Records Published
EDP - Energias de Portugal	Energy	Portugal	34	327,547
Isagen	Energy	Colombia	5	18,241
Asplan Viak AS	Engineering	Norway	14	3,779
Oleoducto Bicentenario	Energy	Colombia	3	2,074
Chevron Australia	Energy	Australia	1	2,048
Total	Energy	France	1	324
Multiconsult	Consulting	Norway	1	308
Arctic Research and Consulting DA	Consulting	Norway	1	8,914
Biofokus	Consulting	Norway	1	493,796
Biolog J.B. Jordal AS	Consulting	Norway	1	177,814
Ecofact	Consulting	Norway	1	11,273
Faun Naturforvaltning AS	Consulting	Norway	1	3,788
NNI	Consulting	Norway	2	3,116
Naturrestaurering AS	Consulting	Norway	2	515
Rådgivende Biologer	Consulting	Norway	5	15,214
SWECO Norge AS	Engineering	Norway	1	1,139
INERCO	Consulting	Colombia	1	1,090
Federación Nacional de Cafeteros	Agriculture	Colombia	6	26,840
Aigos SAS	Consulting	Colombia	3	2,404
Anadarko Colombia Company	Energy	Colombia	3	282
Hatovial SAS	Engineering	Colombia	1	1,898
Moam SAS	Consulting	Colombia	1	1,781
Terrasos	Consulting	Colombia	2	7,725
Celsia	Energy	Colombia	1	35
Stratos	Consulting	Colombia	1	849
LafargeHolcim	Mining	Spain	2	35

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 information⁴. 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 the structure [required by GBIF \(Darwin Core\)](#). Also, the national node may be provided all technical helpdesk needed for the standardization process.

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:

1. 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;
2. To acknowledge and agree to the [Data Publisher Agreement](#) (the English version is valid for legal purposes);
3. To be aware of the [Data User Agreement](#), that GBIF data users must agree before using them;
4. 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](#).

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.

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 license type for the data to be published, which can be one of the following: CC0, CC-BY, CC-BY-NC⁵. 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.

What are the costs for <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

¹<https://sustainabledevelopment.un.org/sdgs>

²<http://www.bipindicators.net/numberofgbifrecordsovertime>

³Hatton T et al, 2019. Digitally Transforming Environmental Assessment. Report of the Digital Environmental Impact Assessment Working Group. The Western Australian Biodiversity Science Institute and the Western Australian Marine Science Institution, Western Australia, 32 pp.

⁴Chapman AD & Grafton O (2008) Guide to Best Practices for Generalising Sensitive Species-Occurrence Data, version 1.0. Copenhagen: Global Biodiversity Information Facility, 27 pp. ISBN: 87-92020-06-2. Available at: <https://www.gbif.org/document/80512>

⁵<http://www.gbif.org/terms/licences>