



Best Practices for Publishing Biodiversity Data from Environmental Impact Assessments

Andrew Rodrigues, Dag Endresen, Rui Figueira, Cristina Villaverde, Miguel Vega,
Nick King, Asha Rajvanshi, Jo Treweek

Version dbb04d4, 2022-02-25 15:56:10 UTC

Table of Contents

Colophon	1
Suggested citation	1
Contributors	1
Licence	1
Persistent URI	1
Abstract	1
Document control	1
About GBIF	1
About IAIA	2
Cover image	2
Introduction	2
1. Benefits of sharing biodiversity data	3
2. Key principles and concepts of data publishing	4
2.1. Types of biodiversity data	4
2.2. Operating principles: Steps in the publishing process	5
Step 1: Becoming a data publisher	6
Step 2: Data capture	6
Step 3: Selecting a tool to prepare data for publishing	7
Step 4: Preparing data for publication	8
Step 5: Publishing data to GBIF	8
Step 6: Discovering and citing data through GBIF	9
Appendix A: Methods and tools web sites: sources of additional assistance	9
References and further reading	9

Colophon

Suggested citation

GBIF Secretariat & IAIA (2020) Best Practices for Publishing Biodiversity Data from Environmental Impact Assessments. Copenhagen: GBIF Secretariat. <https://doi.org/10.35035/doc-5xdm-8762>

Contributors

Andrew Rodrigues of the GBIF Secretariat, **Dag Endresen** of GBIF Norway, **Rui Figueira** of GBIF Portugal, **Cristina Villaverde** and **Miguel Vega** of GBIF Spain, and Nick King, Asha Rajvanshi and **Jo Treweek** of IAIA contributed to this version of the document.

Licence

The document *Best Practices for Publishing Biodiversity Data from Environmental Impact Assessments* is licensed under [Creative Commons Attribution-ShareAlike 4.0 Unported License](#).

Persistent URI

<https://doi.org/10.35035/doc-5xdm-8762>

Abstract

This guide aims to help practitioners, consultants and other Interested & Affected Parties (I&APs) working with environmental impact assessments to improve the curation, archiving and management of primary biodiversity data captured during EIA processes and to share data freely and openly in standardized, accessible and interoperable formats through the Global Biodiversity Information Facility (GBIF). I&APs are encouraged to share the most detailed data possible, to support knowledge about species distributions and provide baseline data for future assessment.

Document control

v1.0, December 2020

Based on an earlier publication: 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>.

About GBIF

GBIF—the Global Biodiversity Information Facility—is an international network and data infrastructure funded by the world’s governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth. Coordinated through its Secretariat in Copenhagen,

the GBIF network of participating countries and organizations, working through participant nodes, provides data-holding institutions around the world with common standards and open-source tools that enable them to share information about where and when species have been recorded. For more information, visit <https://www.gbif.org>.

About IAIA

IAIA—the International Association for Impact Assessment—is the leading global network on best practice in the use of impact assessment for informed decision making regarding policies, programs, plans and projects. IAIA is committed to the promotion of sustainability, the freedom of access to information, and the right of citizens to have a voice in decisions that affect them. IAIA promotes the free flow of complete, unbiased and accurate information to decision-makers and affected parties. This includes biodiversity information. IAIA's Best Practice Principles on Biodiversity-Inclusive Impact Assessment promote transparent approaches and the sharing of biodiversity data. IAIA actively encourages its members to encourage their clients to share data using the GBIF facility. For more information, visit <https://www.iaia.org>.

Cover image

Whale shark (*Rhincodon typus*), Australia. Photo 2009 Erik Schlogl via [iNaturalist research-grade observations](#), licensed under [CC BY-NC 4.0](#).

Introduction

Data from biodiversity baseline assessment and monitoring plays a crucial role in understanding current and potential future impacts of development on the natural environment, whether these impacts are from industrial, infrastructure, agricultural, extractive or other projects. Data about the occurrence of species in space and time are needed to underpin efforts to avoid, mitigate, restore or offset impacts on biodiversity through the Mitigation Hierarchy. Effective decision-making in this area is also crucial for aligning with international best practices and fulfilling international commitments under the [UN Sustainable Development Goals](#) and the [Convention on Biological Diversity](#).

Gathering biodiversity data is one of the most expensive and time-consuming components of the impact assessment process. During field surveys, experts must account for the range of species and habitats, as well as migratory patterns and life cycles of species across seasons. While aligning data collection periods with project and financing life cycles can be a challenge, failure to plan repeat visits across different seasons can lead to project delays, additional costs and failure to understand differences in likely impacts over time. The scarcity of available resources in developing countries often adds further limitations in the amount of existing data available for EIAs, even in areas known to be biologically rich and diverse.

Despite its potential value beyond any given project, biodiversity assessment and monitoring data are rarely shared. Instead, valuable datasets remain archived within company databases and systems where they fail to yield returns on the significant amounts of time and money already invested in them.

GBIF—the Global Biodiversity Information Facility—is an international network and research infrastructure that makes biodiversity data freely and openly accessible to scientists, researchers, authorities and citizens anywhere. Its aim is to produce economic and social benefits and enable sustainable development by providing sound scientific evidence on biodiversity. To that end, GBIF and its global community of practice provide a suite of standards, tools and infrastructure to support the management, publication, use and reuse of primary biodiversity data. This guide updates [Publishing EIA-Related Primary Biodiversity Data: GBIF-IAIA Best Practice Guide](#) (2011), a joint publication of GBIF and IAIA, offering a current view of how EIA practitioners can make more effective use of these resources.

1. Benefits of sharing biodiversity data

Companies responsible for ordering or conducting EIAs can accrue both operational and reputational benefits by sharing biodiversity data through GBIF and similar open access data platforms. Specifically, publishing primary biodiversity data from EIAs:

- Provides long-term cost savings and improved understanding of natural heritage of project areas by leveraging existing biodiversity information from earlier data collection efforts
- Shares biological data methodically and consistently using standardized formats and conditions, aligning with best practices to improve data management, documentation and retention in large and small projects
- Reduces field survey effort through improved targeting of species and a better understanding of species ranges
- Through cumulative impacts of shared data, increases data coverage for sensitive ecosystems, habitats and sites to help detect and avoid species of conservation concern, migratory and ephemeral species in early project stages
- Offers companies low-cost leadership opportunities that significantly reduce costs and increase impact
- Increases transparency, accountability and disclosure of assessments to I&APs, including regulators and citizens
- Provides social licence to operate and a positive profile within the environmental and conservation community
- Fills data gaps in under-sampled regions of the world
- Enhances the evidence base available for reuse in decision-making and research applications related to biodiversity
- Enables tracking of the reuse of data in research and policy applications through data citations, thus returning reputational credit to companies and consultants
- Contributes to evidence needed to attain international targets, including SDGs, related to conservation, climate change, invasive species, food security, human health, and zoonotic disease management

These benefits are obtained at minimal additional cost to the process of biodiversity surveys and monitoring for EIAs, as data can be collected and prepared from the outset in formats suitable for sharing with global, national, subnational and thematic biodiversity data aggregators and

repositories. Initial investments in training staff from project sponsors and consultancies in biodiversity data skills will ensure consistent, efficient data collection and management, improve overall data quality, and thus maximize the reputational benefits.

2. Key principles and concepts of data publishing

2.1. Types of biodiversity data

Biodiversity data can encompass structured information data across any level of biodiversity—molecular, species or ecosystem. These data can be either primary biodiversity data, such as observations or collections at a specific time and place, or secondary, synthesized or interpreted data, which combine biodiversity and environmental data from different sources to present real-world interpretations, as in a species distribution map. Although EIAs tend to present a considerable amount of secondary data, this information relies in turn on large volumes of the primary biodiversity data that are the focus of this document.

The GBIF network specializes in bringing together ‘species occurrence data’ that typically includes, at a minimum, a scientific name, date and location of occurrence. Traditionally, these records have come from sources such as specimens from natural history collections, field work and monitoring surveys, but today other key sources include camera-trap images, environmental DNA (eDNA) sampling and citizen science projects.

While scientific name, date and location represent a minimum recommended level of information about an organism, occurrence records can include other useful information, such as the observation method, abundance counts, habitat structure (like height, stratification, density), abiotic characteristics (such as substrates, hydrology, climate) and associated information about land use and threats. [Learn more about data-quality requirements.](#)

GBIF publishes [four different classes of datasets](#): metadata-only, checklist, occurrence and sampling-event. These classes represent increasingly richer levels of information rather than different types of data. [Table 1](#) describes each class and examples of the type of information they include.

Table 1. Steps to publishing data on GBIF.org

Dataset Class	Description	Example
Metadata dataset	Information about the dataset	Description of the methodologies used for collecting the data, date range, geographic and taxonomic scope (see example)

Dataset Class	Description	Example
Checklist Dataset	Catalogue or list of named organisms, or taxa	List of species recorded at a site, within a geographic area or sharing particular characteristics, e.g. medicinal plants, invasive alien species (see example)
Occurrence dataset	Occurrence of a species (or other taxon) at a particular place on a specified date	Scientific name, latitude, longitude, date (see example)
Sampling-event dataset	Occurrence data for a specific site that has been sampled using a specific protocol, including repeated samples over time, with abundances based on defined units and quantities, and sampling effort.	Scientific name, latitude, longitude and how the species were collected/observed and in which series of monitoring events it was recorded (see example)

2.2. Operating principles: Steps in the publishing process

GBIF provides a means of sharing biodiversity data through a publishing process that uses simple tools and follows standard procedures and protocols to make it universally accessible over the Internet. Data publishing through the GBIF network follows a series of clear steps, shown in [Figure 1](#). Each of these steps is described in more detail in the subsequent sections of this document.

This guide will help environmental assessment practitioners (EAPs), consultants and other interested and affected parties to choose the most suitable option or tool for publishing the primary biodiversity data they have gathered, as an integral part of the EIA process.

As a first step towards publishing biodiversity data, EAPs can seek assistance from the wide [network of GBIF national, regional and thematic Participants](#). A majority of these nodes encourage, coordinate and assist in biodiversity data publishing activities within their respective jurisdictions and domains. If the EAP's operations fall outside of the GBIF network e.g. in a country where there is no node, then the [GBIF help desk](#) can provide additional support.

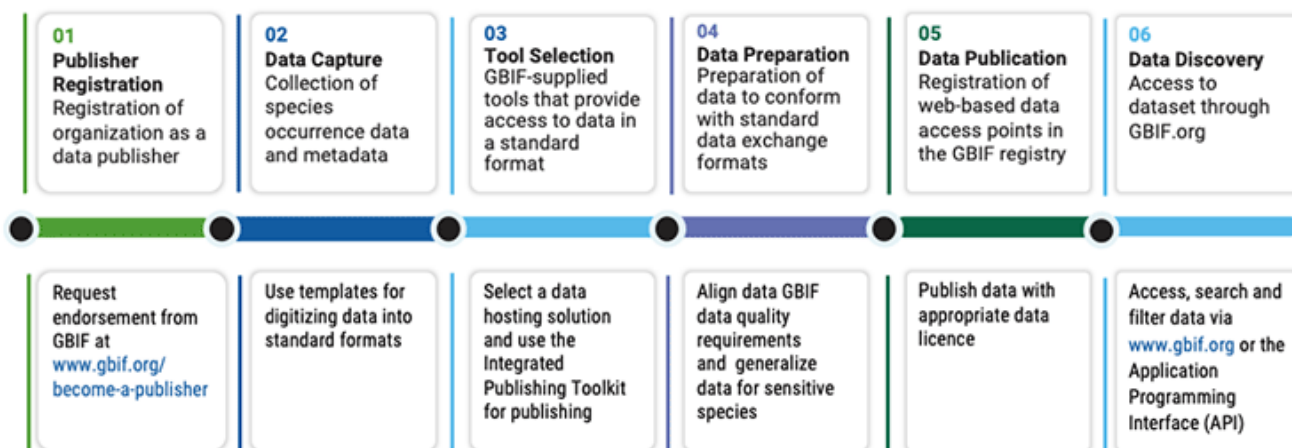


Figure 1. Data publishing workflow: steps for publishing data on GBIF.org

Step 1: Becoming a data publisher

Once an organization agrees to share EIA data, its staff must establish provisions ensuring that stakeholders in each stage of data collection, curation and management agree to the terms by which data publishing takes place and are properly acknowledged in their role. These provisions should include agreeing to the GBIF [Data Publisher Agreement](#) (the English version is valid for legal purposes) and understanding the conditions of the GBIF [Data User Agreement](#) that users of GBIF-mediated data must abide when reusing data.

Step 2. Data capture

Ensuring the standardization of data capture at point of collection will

- make it easier for EAPs to collect and manage primary biodiversity data
- improve the consistency and utility of data collection
- ensure that the data are collected in a consistent format, suitable for publishing using the GBIF infrastructure.

GBIF primarily relies on the Darwin Core (DwC) and Ecological Metadata Language (EML) standards, which set out the structure and format of published datasets (learn more about [applicable standards](#)).

GBIF provides [pre-configured Excel spreadsheets](#) that can serve as templates for capturing checklist, occurrence and sampling-event data. These spreadsheets are simple tools aimed at providing a common format and standard for collecting data. They use consistent terminology and can be completed with additional DwC terms to fit the data collection purpose. This standardized approach makes it easier to exchange data between users, compare it across sites, and integrate it into national and global biodiversity databases. No metadata template is provided as publishers can use the built-in metadata editor in GBIF's data publication tool—the Integrated Publishing Toolkit (IPT) (see [Step 3](#) below)—to populate the metadata. The IPT ensures that the data and metadata are in a valid XML format.

Table 2. Responses to key concerns raised by EIA data publishers

Concern	Response
Providing precise data on occurrence of sensitive species (e.g. endangered, high-value) could lead to poaching or piracy	Geographic coordinates can be generalized and other information withheld in published version of the data (see detailed guidance)
Commercial sensitivity of data during licensing period	Data publication can be delayed until project receives approval
Company faces possible reputational risk if, for example, biodiversity is damaged	Over time, increased open data on species distributions will allow for more robust and transparent assessments of site-specific damage that can provide reputational dividends
Sharing data may need government approval and buy-in	Guidelines from the Convention on Biological Diversity encourage open data sharing, and data mobilized through GBIF is an indicator of progress towards Aichi Biodiversity Target 19
Company could incur additional costs and require additional effort to monitor and share data	Costs of monitoring should already be captured within the project budget; publication is free of charge, and open-access data can provide long-term savings
Companies that don't invest in sharing data can benefit from free and open data available through GBIF more than others who contribute to its supply and maintenance	"Free riders" exist in any commons, but a large common pool resource like GBIF is not depleted by use, and parties that do participate typically build receptive capacity to understand the issues and limitations of the resource better than those who don't

Step 3: Selecting a tool to prepare data for publishing

GBIF.org does not itself host data. The system relies on each data publisher maintaining their own datasets and making them available online in a GBIF-supported format. It also relies on organizations registering datasets and providing GBIF with a stable endpoint for finding and indexing the data. GBIF recommends using the [Integrated Publishing Toolkit \(IPT\)](#) to do this. Highly skilled publishers can also use an API to register datasets programmatically (contact the [GBIF help desk](#) for more details).

Organizations may install the IPT if they have the capacity to host and maintain data on servers that always remain online, ensuring that the data that they share will have a persistent, stable point of access. An organization that either does not have this capacity or does not wish to maintain its own installation can choose one of the following options for data hosting (more details available [here](#)):

1. Data hosted at a national node (if the country is a GBIF Participant)
2. Hosted by another GBIF Participant or data publisher
3. Cloud-hosted IPTs maintained by GBIF Secretariat

The first two options provide a range of helpdesk services to potential publishers, while the final option provides very limited support to publishers. Potential publishers can request guidance from

the [GBIF help desk](#) on the most suitable option. Regardless of the hosting option selected, data publishers retain full control of the data, including the ability to correct and update datasets at any time. Data citations will always acknowledge the data publisher, irrespective of how or where the datasets are hosted.

The IPT is the most commonly-used tool and is maintained and developed by the GBIF Secretariat. IPTs can generate a Darwin Core Archive (DwC-A), the preferred exchange format, for each dataset and register them on GBIF. To use the IPT, data must already be digitized. Acceptable file types include delimited text files (e.g. text files using comma or tab-separated values) or Microsoft Excel. Database connections can also be made. If the IPT is to be hosted within the publishing institution, upon installation of the IPT, the publishing organization should register as the host. If the IPT is hosted elsewhere, the IPT administrator can add the publishing organization to the IPT using an IPT token that is issued upon endorsement of the publisher.

Step 4: Preparing data for publication

To share data through GBIF.org, publishers must collate or transform and describe existing datasets into a [standardized format](#). This work may require additional processing, content editing and mapping the content of a dataset into one of the available formats. 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. GBIF provides guidance on the [data quality requirements and recommendations](#). The [GBIF Data Validator](#) is a tool that lets publishers check datasets prior to publication and makes recommendations on how datasets can be improved and cleaned by flagging, for example, duplicate identifiers, incomplete fields and recognized inconsistencies in formatting.

Publishers should use a precautionary approach and seek input from specialists on the publishing of precise locations of sensitive species, for example threatened or valuable species, when there are concerns that doing so could enable poaching or other threats to the species population. For a thorough discussion of this topic, see [Current Best Practices for Generalizing Sensitive Species Occurrence Data](#).

Step 5: Publishing data to GBIF

The GBIF IPT supports automatic registration in the GBIF network (see the [IPT manual](#)). If publishers are using an IPT, GBIF registers datasets when publishers click the 'register' button. Data should be published as soon as possible following the EIA. However, if there are concerns about commercial confidentiality or other time-sensitive issues, publication may be delayed or embargoed until the completion of a licensing process.

To publish data to GBIF, publishers must assign one of three Creative Commons licences to a dataset:

- [CC0 1.0](#), for data made available for any use without any restrictions
- [CC BY 4.0](#), for data made available for any use with appropriate attribution
- [CC BY-NC 4.0](#), for data made available for any non-commercial use with appropriate attribution.

Note that the CC-BY-NC license has a significant effect on the reusability of data, and that GBIF does

not consider non-commercial use restrictions to be enforceable. GBIF encourages data publishers to choose the most open option possible.

Step 6: Discovering and citing data through GBIF

Once datasets are registered, GBIF indexes them to facilitate access to the data by users. Each dataset has its own page ([example](#)) and can be found using the search function on the website and on the publisher's page ([example](#)). The indexing process allows the search and discovery of records from all published datasets, showing, for example, all records of a particular species or groups of species in a given geographical area.

Because search results mix records from different datasets, the GBIF Data User Agreement requires appropriate citation of data regardless of the licence applied to any individual dataset. Through the use of [Digital Object Identifiers \(DOIs\)](#), GBIF tracks data reuse and provides publishers with key metrics for downloads, which appear the 'activity' tab of each dataset page ([example](#)) and for documented citations in other research and assessments, linked from both [dataset](#) and [publisher](#) pages. Publishers can use this information to demonstrate the value of their contribution to science and society through sharing data from EIAs.

Appendix A: Methods and tools web sites: sources of additional assistance

- [Getting started: overview of data publishing in the GBIF network](#)
- [GBIF IPT Manual: How to publish biodiversity data through GBIF.org](#)
- [GBIF types of datasets, including spreadsheet templates](#)
- [GBIF data quality Requirements](#)
- [Current Best Practices for Generalizing Sensitive Species Occurrence Data](#)
- [Darwin Core Quick Reference Guide](#)
- [Ecological Metadata Language Guide](#)
- [GBIF Participants](#)

References and further reading

- 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. IAIA Special Publication Series No. 7. <https://www.iaia.org/uploads/pdf/sp7.pdf>
- Chapman AD (2005) Principles of Data Quality. Version 1.0. Copenhagen: GBIF Secretariat. <https://doi.org/10.15468/doc.jrgg-a190>
- Chapman AD (2020) Current Best Practices for Generalizing Sensitive Species Occurrence Data. Copenhagen: GBIF Secretariat. <https://doi.org/10.15468/doc-5jp4-5g10>.
- King N, Rajvanshi A, Willoughby S, Roberts R, Mathur VB, Cadman M & Chavan V (2012) Improving access to biodiversity data for, and from, EIAs – a data publishing framework built to global

standards. *Impact Assessment and Project Appraisal* 30(3): 148-156.

<https://doi.org/10.1080/14615517.2012.705068>.

- Rajvanshi A, Mathur V & Iftikhar UA (2007). Best-practice guidance for biodiversity inclusive impact assessment: a manual for practitioners and reviewers in South East Asia. CBBIA-IAIA Guidance Series. Rapid City, N.D., USA: International Association for Impact Assessment.