Technical guidance for EBRD clients on uploading biodiversity data to the Global Biodiversity Information Facility

European Bank
for Reconstruction and Development

GBIF Global Biodiversity Information Facility

Last updated on 17 June 2025

1 PREAMBLE

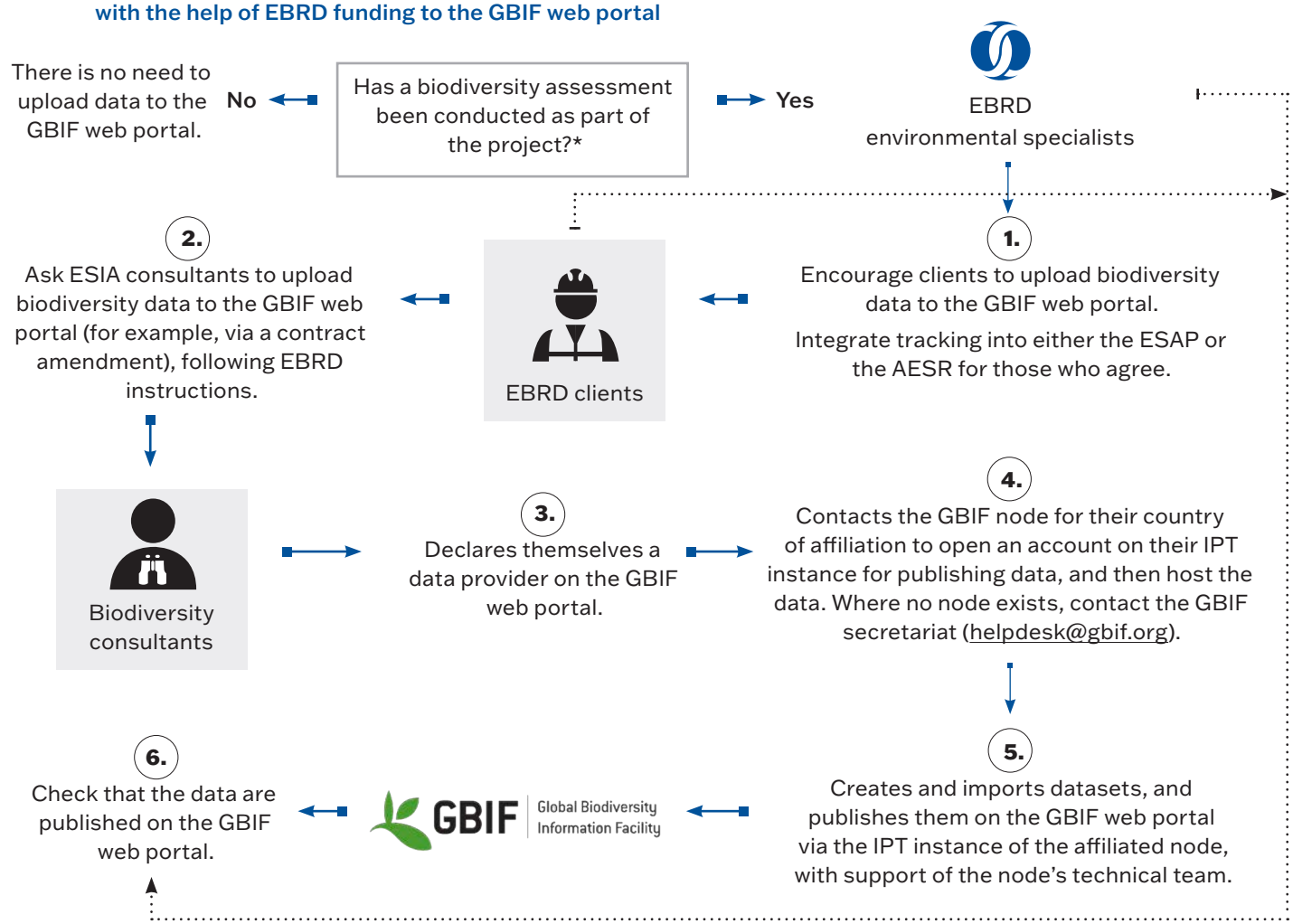
Following similar efforts made by multilateral development banks such as the Agence Française de Développement (AFD), the European Bank for Reconstruction and Development (EBRD) expects its clients and their consultants to publish any primary biodiversity data generated during the course of funded projects to the Global Biodiversity Information Facility (GBIF). The term "primary biodiversity data" refers to the observation of fauna or flora situated in time and space. This type of data, also known as "taxon occurrence data", provides evidence that a species (or some other taxon) is present in a particular location on a given date.

This document provides practical recommendations to consultants and other stakeholders responsible for publishing such data. Publishing biodiversity data linked to EBRD projects requires the involvement of at least three or four different players:

- the EBRD
- the EBRD's client, usually a project owner
- one or more consultancy firms specialised in collecting and analysing biodiversity data
- the GBIF

As Figure 1 shows, each player has a specific role in publishing the data. The practical task of data publication is generally delegated to the consultant(s) responsible for data collection, who carry out several steps before publishing the data.

Figure 1. Process for publishing primary biodiversity data generated with the help of EBRD funding to the GBIF web portal



Definitions: ESAP: Environmental and Social Action Plan. AESR: Annual Environmental and Social Report . ESIA: Environmental and Social Impact Assessment

*For the EBRD, biodiversity data collection pertains to Category A and select Category B projects.

2 HOW TO BECOME A GBIF DATA PUBLISHER

2.1 Accepting the GBIF's Data Publisher Agreement

To publish data on the GBIF's portal, you must first register as a data provider. To do this, log into the GBIF's web portal, accept the Data Publisher Agreement's terms and conditions, and complete a short form describing your organisation, the type of data you will provide and how the data will be made available.

Table 1. Recommendations for completing the Data Publisher Agreement form

Headings	Recommendations
Information about the organisation	Provide contact details for your organisation's head office. You can also show your head office's location on the map.
Endorsing node	Keep the default setting active unless you intend to supply marine data exclusively.
GBIF projects	Keep the "No" setting selected. This is the default setting.
Contacts	In addition to your organisation's contact details, provide details for the person responsible for managing biodiversity data.
What and how	Select "Occurrence-only data". You can also select other data types if appropriate. Other types include "sampling-event" and "checklist" datasets. Next, provide a short description of all the data types you are likely to publish. If you cannot give the GBIF direct access to your data through your computer server, we recommend that you ask for help in publishing your data. Staff from your designated GBIF node (a team responsible for coordinating GBIF activities and providing technical support to data publishers in a particular country or region) can discuss data publication methods with you.

As soon as you submit the form, it is sent to your designated GBIF node. That node will assess your request before accepting it.

Once your request for registration as a GBIF data publisher is accepted, a dedicated web page will be created for you automatically on GBIF.org with the information you supplied in the form. From that page, you will also have access to the datasets you have published and to metrics for those data.

2.2 Opening an IPT instance for publishing data

Publishing your data on the GBIF web portal requires:

- a computer server that you will use to store your primary biodiversity data and to which you will provide access via the GBIF's website
- data publication software, such as the public and free Integrated Publishing Toolkit (IPT) software made available by the GBIF

If you do not have a computer server, or if it is not possible for you to share your data from your own servers, your GBIF node can host your data and open an IPT instance for publishing it on their own servers. In this case, your GBIF node will give you the URL of the IPT instance to use, together with your user log-in.

Only organisations can be endorsed as GBIF data publishers. You cannot be endorsed as an individual.

The GBIF's Node Steering Group will endorse your organisation if it is based in a country that is not a member of the GBIF.

More information on endorsing an organisation as a GBIF data publisher is available here.

You can find the contact details for the GBIF node to which you belong here.

If the country where your organisation is based does not have a GBIF node, contact the GBIF Secretariat in Copenhagen at helpdesk@gbif.org for alternative solutions.

In any case, we recommend you reach out to your GBIF node to set up and configure your data publication tools.

3 RECOMMENDATIONS FOR DATA COLLECTION AND FORMATTING

3.1 Data collection

Publishing primary biodiversity data on the GBIF web portal does not require additional information to be collected in the field. It only involves formatting your data in accordance with data standards.

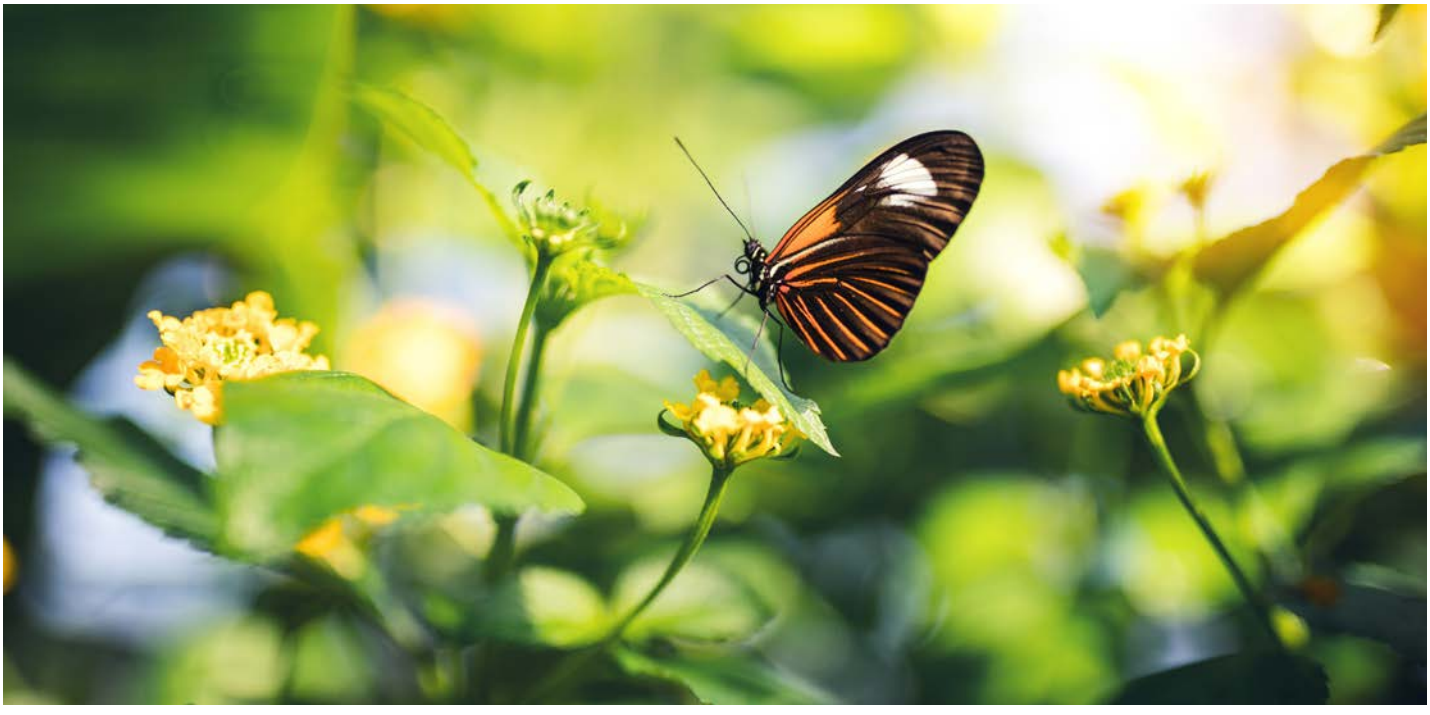
It is, therefore, recommended that you anticipate this formatting requirement when data are collected in the field, so that data can then be formatted and published efficiently. Anticipating data publication can result in many good practices that will also contribute to improving the quality of your data. For example:

- Using data-entry sheets for field observations allows you to specify the type of observation, the scientific name of the taxon, its location and the date of observation (see Appendix 1 for an example of a field data sheet).
- Mapping the most interesting surveys and observations with the help of a GPS. While using a GPS is not essential, each observation must be associated with a geographic location and published with the same location accuracy as in the field, except for species considered sensitive (see Section 3.3).
- Centralising and digitising observations in a table structured similarly to the standard defined for publishing species occurrences on the GBIF web portal (see Section 3.2). Appendices 2 and 3 contain standard guidance for sampling-event and checklist datasets, respectively, where such data are relevant and appropriate.

If the location of an observation is not referenced precisely, the geographic position may relate to the coordinates of a point located in the centre of the study area. In this case, the radius of the smallest circle that encompasses the study area from the central point will be evaluated and incorporated into the data (the coordinateUncertainty field should be completed; see Table 2).

Supplying information on the location of the observations requires a basic knowledge of cartography.

It is recommended that you use a geographic information system and/or GPS position management software to identify coordinates.



3.2 Formatting data in accordance with the GBIF standards

The EBRD expects clients to publish primary biodiversity data, at least on species occurrences, so as to provide sufficient information on the location of observations in time and space.

Species occurrences must be aggregated and published in datasets and corresponding metadata (additional information detailing the conditions of data acquisition and use) must be provided. Section 4.4 gives details of the metadata to enter.

In the absence of additional EBRD requirements for data publication, each relevant project will result in the publication of at least one dataset containing all observations made in the field.

To prepare to publish your dataset, you must list your observations in a spreadsheet. The spreadsheet should contain the information set out in Table 2 (or, for sampling-event and checklist datasets, the tables in appendices 2 and 3, respectively).

There is no required format, but you can use the GBIF's template. You may need to add fields to the template manually. However, it is important to use the exact field names listed in Table 2 (or, for sampling-event and checklist datasets, the fields listed in appendices 2 and 3, respectively).

You may publish data in several datasets if you wish to supply more detail on the inventory protocols used for each data record.

Table 2. Description of the species occurrence standard used by the GBIF and expected in the publication of biodiversity data acquired through EBRD funding

	Field name	Description
Fields always required	occurrenceID	Unique identifier for the taxon occurrence. You may define this identifier yourself. The identifier must be unique within the dataset. It may be a numerical or alphanumeric identifier (for example: 2354 or Obs0015). Best practice is to use a unique and permanent Universally Unique Identifier (UUID) or a computer-generated Uniform Resource Identifier (URI).
	basisOfRecord	Nature of the observation that was the source of the occurrence. It may be, for example, a human observation or a recording made by a machine (for example: ultrasound recorder, camera trap, Argos buoy, radar).
	scientificName	Full scientific name on the lowest possible taxonomic level, including the author of the taxon description and the date the name was published (for example: *Canis lupus Linnaeus*, 1758). The taxon's scientific name can be looked up on the GBIF's website with the species matching tool (see Section 3.4 for more information).
	eventDate	Date or time interval when the taxon was observed. The date must be formatted according to the ISO 8601 standard (for example: 18 November 2020 should read 2020-11-18).
	decimalLatitude	Geographic coordinates of the occurrence in decimal degrees. For example: decimalLatitude = 48.861037 and decimalLongitude = 2.335860.
	decimalLongitude	
	geodeticDatum	Geographic coordinate system used to indicate the latitude and longitude of the occurrences. The GBIF uses the World Geodetic System 1984 (WGS84) benchmark geographic coordinate system.
Fields conditionally required	coordinateUncertainty	Location of the taxon occurrence in metres. In other words, this is the radius of the smallest circle containing the observation. **Condition:** The coordinates provided correspond to the centre of an area within which the occurrence was observed.
	dataGeneralizations	Description of a (possibly intentional) degradation of the data (blurring of the geographic position, for example). This informs the user that a more precise version of the published data may exist. **Condition:** Observing a species considered to be sensitive.
Recommended fields	individualCount	The individualCount field allows the user to specify the number of individuals observed for the taxon in question. It is also possible to quantify the taxon occurrence with other measurement systems, such as abundance-dominance coefficients, density and/or an abundance indicator. In this case, the organismQuantity and organismQuantityType fields are used to show the quantity and measurement unit used to quantify the taxon occurrence.
	organismQuantity	
	organismQuantityType	
	countryCode	Two-letter code showing the country in which the taxon was observed. That code must follow the ISO 3166-1-alpha-2 standard.
	taxonRank	Taxonomic rank of the scientific name given for the taxon occurrence in Latin or English. For example: kingdom, phylum, class, order, family, genus, species.
	kingdom	Full scientific name, in Latin, of the reign to which the observed taxon belongs. The possible values for this field are: Animalia, Archaea, Bacteria, Chromista, Fungi, Plantae, Protozoa or Viruses.
	InformationWithheld	If data have been intentionally degraded, it is useful to complete the informationWithheld and the dataGeneralizations fields to show that more precise information exists but that it has been intentionally omitted, for example: "Location information not given for endangered species".

You can find more information on field descriptions of species occurrences on the GBIF web portal.

The data in each data record may be entered in a language of your choice. The language used to describe the data and metadata must, however, be specified in the data of each dataset.

This information is set out in step 4 of the data publication process (see Section 4.4).

The GBIF also allows you to group your species occurrences together into sampling events. Your observation lists can then be associated with different surveys for which you can publish specific additional information. The EBRD expects clients to use these groupings where relevant and appropriate, as they enhance the value and usability of your data records. A table outlining the sampling-event standard which the GBIF uses, and which is expected in the publication of biodiversity data acquired through EBRD funding, is available in Appendix 2. More information on sampling-event datasets can also be accessed on the GBIF's website.

In addition, by using data checklists, you can provide a catalogue, quick summary or baseline inventory of a set of named organisms or taxa. Although the EBRD does not require checklists, in certain cases they may enhance insights and the usefulness of your data. A table outlining the checklist data standard used by the GBIF and expected in the publication of biodiversity data acquired through EBRD funding is available in Appendix 3. More information on checklist datasets can also be accessed on the GBIF's website.

3.3 Degradation of the geographic location of sensitive data

To avoid compromising the survival of local species populations under threat of intentional removal or destruction (poaching, for example), you may purposely degrade the location accuracy of observations of sensitive species.

The simplest way to degrade the location is to round the geographic coordinates up. Rounding can be adapted depending on the sensitivity of the species. For the most sensitive species, you could, for example, round the geographic coordinates up to the nearest whole numbers. For species that are less sensitive, you could round the coordinates to one or two decimal places.

Table 3. Relationship between the number of decimal places used for the geographic coordinates and the level of degradation of the data location at the equator

Number of digits after the decimal point to round the geographic coordinates	Value in decimal degrees	Equivalence in metres in the field / margin of error in the location of the observations
0	1.0	111.32 km
1	0.1	11.13 km
2	0.01	1.11 km

When you intentionally degrade the location accuracy of an observation, you must describe the nature of the degradation in the dataGeneralizations field – for example: "Rounding of geographic coordinates to nearest whole numbers". In addition to the dataGeneralizations field, it is also useful to complete the informationWithheld field to show that more precise information about the location exists but that it has intentionally not been given.

Beyond the polar circles, blurring the location of sensitive species by degrading the geographic coordinates becomes less useful because the meridians draw closer together.

For more information on degrading the location of sensitive species, refer to this best practice guide.

3.4 Checking the validity of the datasets

For your datasets to be published correctly, the observation list files that you import into the IPT software must comply with the standards used by the GBIF.

To check that your Excel files are properly formatted, we recommend that you use the following two tools. These tools allow you to highlight any possible mistakes that you might have made in your files, so that you can correct them more easily before publication.

Table 4. Tools for checking the validity of your datasets

Step	Tool name and link to the tool	Function	Remark
1	GBIF species matching tool	To check the validity of the taxon's scientific name and to find the taxon names accepted by the GBIF.	Input must be in CSV format.
2	GBIF data validator	To check your datasets are compatible with the data standards used by the GBIF.	Input is a data package generated by the IPT – DwC-A or a CSV containing Darwin Core terms.

4 DATA PUBLISHING ON THE GBIF

Publishing datasets on the GBIF web portal is done through the IPT web application. When you use an IPT instance made available by your GBIF node, you access the application using the URL, user account and password that your GBIF node has attributed to you.

When using the IPT application, access the user manual by clicking on the link at the bottom of each screen.

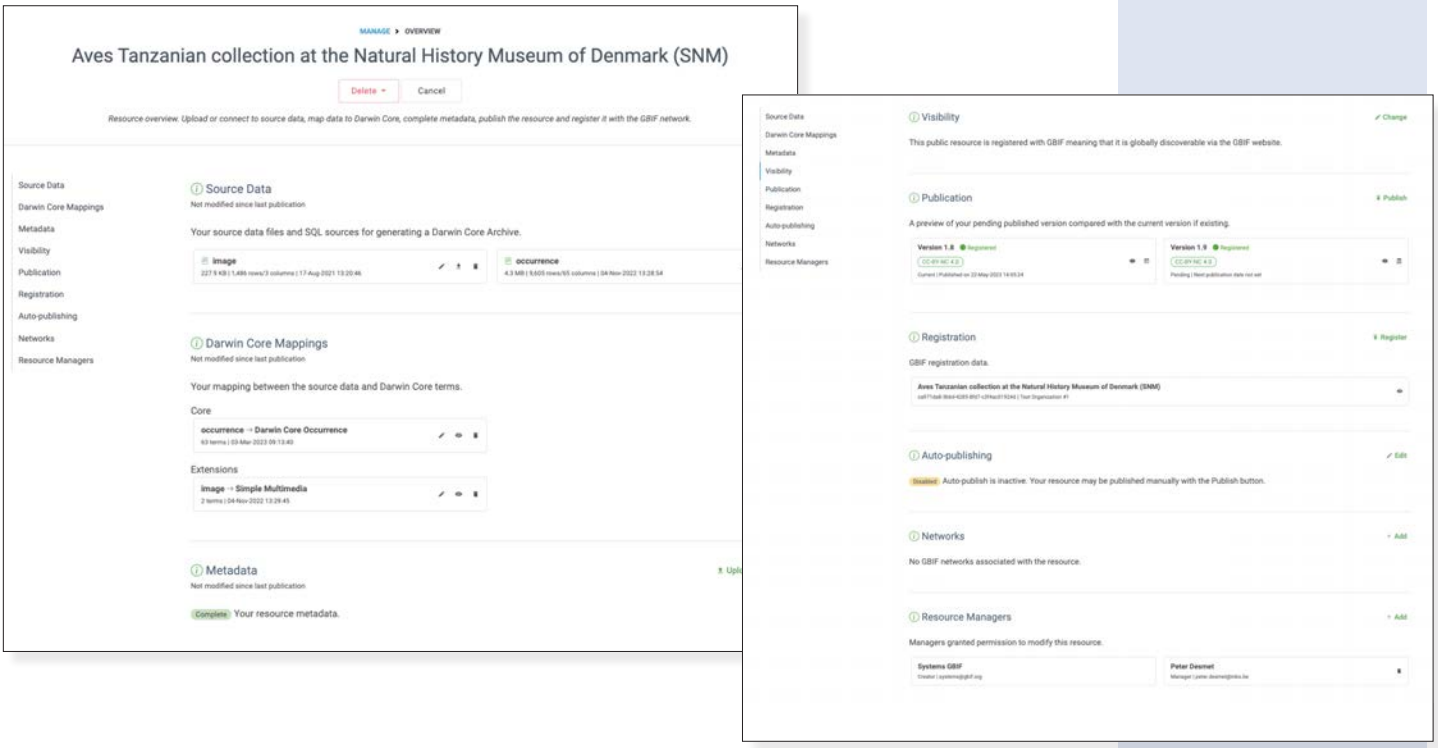
To use the IPT application correctly, we recommend that you consult the IPT web page and watch the demonstration video. You can also request a presentation on the application or a training session from your GBIF node.

When using the IPT application, access the user manual by clicking on the link at the bottom of each screen.

4.1 Step 1: Create a data record

Once you have checked and confirmed that your data comply with GBIF standards, you are ready to publish them through the IPT application. All you have to do is create a new "occurrence", "sampling-event" or "checklist" resource from the IPT "Manage Resources" tab.

Figure 2. Management page of a dataset in the IPT application



4.2 Step 2: Import data into the dataset

When you have created a dataset on the IPT, you can import the Excel spreadsheet with your data from the "Source Data" section. When you import your data, another window opens where you can edit the format of your source data if necessary. At that stage, we recommend that you preview the data you have imported to check that the formats used have been correctly integrated.

To preview the imported data, click on the eye icon to the right of the "Analyse" button in the editing window.

4.3 Step 3: Match imported data with the Darwin Core standard

Darwin Core is the standard data file format the GBIF uses. This step, therefore, is intended to ensure that imported data files are compatible with Darwin Core. If some data fields are missing or are not automatically recognised, you can add fields with pre-defined values or match some fields manually with the Darwin Core standard.

To simplify this step, we recommend that you adhere closely to the data fields listed in Section 3.2. For sampling-event and checklist datasets, please follow the data fields in appendices 2 and 3, respectively.

Learn more about the Darwin Core data standard here.

4.4 Step 4: Edit the metadata

The metadata enable you to describe specifically the conditions for how the data record will be produced and used. The usability of the data record depends on this information, so it should not be overlooked. The GBIF requires certain metadata to be entered, such as the name of the data record, its description and contact details for the data producer. Metadata of this kind is a prior condition for publication of the data. The EBRD also requires certain metadata to be entered according to pre-defined rules.

This is the case, in particular, for the conditions of data use. In the "Basic Metadata" section, you are obliged to opt for one of the two least-restrictive levels of rights for the "Data Licence" field, namely the "Public Domain (CC0)" or "Creative Commons Attribution (CC-BY)" licence.

This is also the case for project information within the framework where you have collected the data. In the "Project Data" section, it is essential to provide a project title and a short project description – including the name of the project owner, meaning the EBRD client – in the relevant fields. In the "Identifier" field, you are obliged to enter "EBRD". In the "Funding" field, you are obliged to give the names of the institutions funding the project, including the EBRD.

The IPT application allows you to amend or complete a data record that you have already published, if necessary. In this case, it is useful to specify the amendments made when publishing a new version of a dataset.

4.5 Step 5: Publish the data

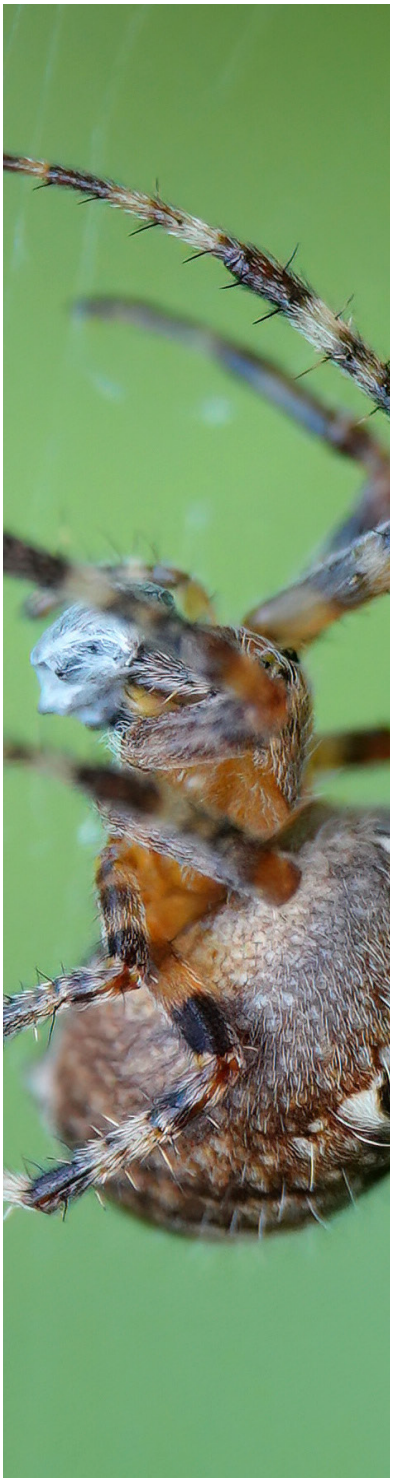
By default, the data that you publish on the GBIF web portal are private and can only be accessed by people with the role of dataset "Manager". Before publishing your datasets, we recommend you make them accessible to the public by clicking on the "Public" button in the "Visibility" section of the IPT "Manage Resources" tab.

Once your dataset has public status, click on "Publish" in the "Published version" section to go to the final step in the publication process.

5 MORE INFORMATION

The GBIF can support you throughout the process described in this guide. In addition to the wealth of information supplied on the GBIF web portal, you can also send your questions directly to private-sector@gbif.org.

The GBIF regularly organises training sessions for data publishers, and provides access to an e-learning platform and free, downloadable training materials. For more information, see here.



APPENDIX 1: EXAMPLE OF A FIELD DATA-ENTRY SHEET FOR PUBLISHING DATA ON THE GBIF WEB PORTAL

Project name			
Name of survey			
Date of survey	The date and, if possible, time of the survey		
Location of survey	Description, GPS position and/or cross-reference with a location map		
Inventory protocol	Name or description of the protocol or, failing that, the method used to gather the data		
Conditions of survey	Important information for interpreting the survey (meteorological conditions for a bird inventory, for example)		
Species observed	**Number of individuals**	**Additional comments**	**GPS point identifier**
		Additional information, especially for species under threat of extinction (reproduction sign, for example)	Particularly for observations of threatened species

APPENDIX 2: SAMPLING-EVENT DATA STANDARDS

Table 5. Description of the sampling-event data standard used by the GBIF and expected in the publication of biodiversity data acquired through EBRD funding, where reporting of sampling-event data is relevant and appropriate

	Field name	Description
Fields always required	eventID	A unique identifier for the sampling event, linking individual occurrences to a specific event. This field is also used to cross reference events to document them (for example, when events are in a time series involving frequent resampling) or to synchronise sampling across a wider area.
	eventDate	The date or time interval when the sampling event as a whole took place, following the ISO 8601 date-time standard.
	samplingProtocol	The name or description of, or reference to, the method or protocol used during the sample event.
	samplingSizeValue & samplingSizeUnit	A numeric value and the corresponding unit for the value, specifying the size of an individual sample in the sampling event.
Recommended fields	countryCode	A two-letter standard abbreviation for the country, territory or island where the sampling event took place. This must follow the ISO 3166-1-alpha-2 standard.
	parentEventID	A cross-reference to the eventID of a broader event, for example, a long-term monitoring project that the specific event is a part of, or a general vegetation survey of a larger area that is comprised of multiple sub-plots. To reference a parent event, it needs to be specified as a separate entry, typically within the same dataset, and carry its own eventID. Reference the eventID of the parent event in the sample event record to specify the relationship between the two entries.
	samplingEffort	The measure of effort expended during a sampling event.
	locationID	An internal or external reference that links to a set of data describing the sample event location, if available.
	decimalLatitude & decimalLongitude	The geographic latitude or longitude, respectively, in decimal degrees. Where coordinate values are available, both fields need to be supplied, together with the geodetic datum.
	geodeticDatum	The coordinate system and set of reference points upon which the geographic coordinates are based.
	coordinateUncertaintyInMeters	The horizontal distance from the given decimalLatitude and decimalLongitude in metres, describing the smallest circle containing the whole location.
	footprintWKT	An alternative area description, specifying the location of the sample event in well-known text (WKT) markup language.
	occurrenceStatus	A qualifier for individual occurrence records, marking a taxon as either present or absent at a location during the sampling event. Note: this applies to associated occurrence data, not to the sample event itself.

You can find more information on the fields associated with sampling-event datasets on the GBIF's website.

APPENDIX 3: CHECKLIST DATA STANDARDS

Table 6. Description of the checklist data standard used by the GBIF and expected in the publication of biodiversity data acquired through EBRD funding, where reporting of checklist data is relevant and appropriate

	Field name	Description
Fields always required	taxonID	A unique identifier for the taxon that allows the same taxon to be recognised across dataset versions, and through data downloads and use.
	scientificName	The full scientific name, including author and year, where applicable. In the context of a checklist, the scientific name is the core data element of a taxon list or hierarchy that the dataset is intended to collate and publish.
	taxonRank	The taxonomic rank of the scientific name supplied.
Recommended fields	kingdom	The full scientific name, specifying the kingdom under which the scientific name is classified.
	parentNameUsageID	If higher taxon names are supplied as separate entries in the list, enter the taxonID of the next available parent (higher-ranked) entry in the checklist dataset. This supports the representation of the dataset as a hierarchy, such as for publishing a taxonomy.
Share if available	vernacularName	See this list of vernacular names. When supplied, also add the language of the name, using ISO 639-1 language codes.

You can find more information on the fields associated with checklist datasets on the GBIF's website.

CONTACTS

European Bank
for Reconstruction and Development

GBIF Global Biodiversity Information Facility

Five Bank Street, London E14 4BG, United Kingdom,
www.ebrd.com

GBIF Secretariat in Copenhagen (Denmark)

Email : helpdesk@gbif.org
Phone : +45 35 32 14 70

Find us on social media
@ebrd

European Bank
for Reconstruction and Development

GBIF Global Biodiversity Information Facility