

# OpenPSD

## Description of Use Cases of data publication by the private sector

### Purpose of the document

The purpose of this document is to describe different use cases of data publication by the private sector, identifying the motivation, procedures, solutions, difficulties, actors involved and their roles in the process. Ideally, the use cases described should cover different scenarios that could be characterised as:

- a large corporation which has not the expertise for biodiversity surveys internally, subcontracting them to other companies
- a large consulting company with expertise in biodiversity surveys that provide them as a service to other companies
- a small company which provides services of EIA surveys to contractors

### Description of the Use Case

Description of the company

Motivation for the publication of data

Actors involved

Procedures

Resources

Limitations or recommendations

# Short draft of the [EDP](#) Use Case

## Description

EDP is a Portuguese private corporation that produces, distributes and sells electricity. It is present in 14 countries and is one of the largest wind energy operators in the world. This use case refers to the implementation of a data publication strategy implemented by the company in relation to two projects involving the construction of Baixo Sabor and Foz Tua hydroelectric Dams, in the northeast region of Portugal.

The two projects involved Environmental Impact Assessment (EIA) studies, performed before the construction, and comprehensive biodiversity monitoring program implemented during and after the construction. These biodiversity surveys have been developed by subcontracted consultants, including universities and research centres, but also private companies from the EIA sector. The studies covered a wide range of biological groups, including phytoplankton, vascular and non-vascular (bryophytes and lichens) flora, and many groups of fauna, like aquatic organisms (eg., fishes, bivalves), arthropods, amphibians, reptiles, mammals and birds. There was a particular focus on species of conservation concern including for instance various species of birds of prey and bats, and wolves, among others.

## Motivations

The main motivations for the development of this publication programme have been:

- **social responsibility** - the company is indexed in the Dow Jones Sustainability Index (DJSI). Making biodiversity data open and available for research and society increases the public profile of social responsibility, and may contribute to improve the DJSI rating;
- **legal requirements** - there are legal requirements to make EIA data available to the national environment agency. This was done through a custom built GIS application which is in end-of-life, and a new solution needed to be adopted; However, the legal requirements did not involve publication through GBIF;
- **improve data management** - data was reported by each subcontractor without a defined template, which resulted in varied formats and file types. Preparing data for publication is an opportunity to standardise data, and also store data in a repository for future preservation. Also, learning best practices might help data management in future projects

# Actors involved

EDP does not have in their structure the capacity to develop EIA or biodiversity monitoring projects. Therefore, it needs to contract most of these services from external partners, in this case to university research centres and private consulting companies, which are the data producers.

Additionally, within the frame of a Portuguese research programme Invited Chairs, which joins public and private funds to attract high level researchers, a Chair EDP on Biodiversity was established in CIBIO/InBIO, University of Porto, between this university, EDP and the portuguese science funding agency, FCT. This Chair acts as scientific advisor of these projects, ensuring also the data management.

EDP owns and holds the data resulting from these studies, so it became the publishing institution through GBIF. The step to become a data publisher was taken by suggestion of the scientific adviser CIBIO/inBIO, with the support of the Portuguese Node of GBIF. The decision was taken after internal information in the company to support the decision. An example of such a document is available [here](#).

All actors are involved in the preparation of a dataset publication, as explained in the Procedures section. The main reason is to ensure that expertise involved in field studies is also present in metadata and data publication. This is also useful to ensure that credits are given to all people involved in the creation and publication of data.

## Procedures

### Registration of the data publisher

An initial procedure was developed to register EDP as a data publisher. This involved an internal information and decision step. It was produced a document similar to the one available [here](#), in support of the decision by the Direction Board of the company to authorize registering EDP as a data publisher. In this registration process, an Administrative contact from EDP was appointed, which has been pivotal in the subsequent actions towards dataset publications.

Additionally, EDP decided to use the Integrated Publishing Toolkit (IPT) provided by GBIF Portugal, which is agreed in the terms of a signed Service Level Agreement between both entities.

## Involvement of the data producer

An initial communication, by email, is made by the administrative contact of EDP to the subcontracted entity responsible for the production of the data. This email provides a brief introduction to the initiative of data publication through GBIF, including the link for the publisher page in GBIF and examples of datasets already published. It also informs the willingness of EDP in publishing the dataset through GBIF, the importance of ensuring that proper credits are given to all involved in the production of the dataset. The email also proposes a methodology for the preparation for the datasets and asks for assistance in:

- preparation of the metadata, namely identifying people involved, description of the study area, methods;
- final review of data and metadata previously to the publication.

## Publication of datasets

The next step involves data processing, metadata description and data upload to the IPT for publication, which is performed by a staff of the EDP Chair on Biodiversity. As different types of data are collected, a considerable effort has been devoted to find the best way to convert the formats provided by subcontractors into the Darwin Core format. GBIF Portugal provided support to the preparation of the data, especially on the use of Darwin Core and in data quality verification.

Regular meetings (presential, videoconferencing) take place at least every six months to assess progress in data publication, evaluate difficulties and improvements in the processing of data, and plan for the next set of databases to be published. Additionally, email communication is used for day to day support and planning. A table is maintained in a shared cloud repository to keep track of the datasets status and planning.

## Internal and external communication and outreach

Main milestones in the process of data publication by EDP were internally and externally communicated by the company:

- an internal email, informing the company about the publication of the first dataset, in April 2017;
- a news item in the intranet portal with the title “Share knowledge to protect life on Earth” (in portuguese) informs, in December 2017, about the initiative and how it is relevant for promoting sustainability and contributing to CBD Aichi targets;
- a [page](#) of the corporate public website is dedicated to the initiative, and links to the institution’s page in GBIF;
- the corporate report on sustainability mentions the initiative in relation to the EDP Chair on Biodiversity.

## Resources

Five people are involved in this initiative, one EDP staff, two EDP Chair on Biodiversity staff and two GBIF Portugal. Only one EDP Chair staff is almost fully dedicated to data publication, which will conclude the publication of all databases available for the two projects. No additional costs result from this involvement, other than already supported by the respective institutions.

Data publication is hosted at the IPT service (<http://ipt.gbif.pt>) provided by GBIF Portugal. The service is made available under the terms of a Service Level Agreement established between EDP and GBIF Portugal. See a template of this SLA [here](#). No costs are involved in this service for the publisher.

## Limitations or recommendations

The following limitations were identified in the development of this initiative:

- In the publication of a dataset, it is difficult to give proper visibility to the entities subcontracted to provide services to the contractor. There are no fields to identify these entities neither in the metadata or in the DwC data. In metadata, the name of the entity only appears as the affiliation of the people that created the dataset, or in the description of the project. But in neither case it will be in a field that is indexed by GBIF, which limits visibility in GBIF.org.