Mobilizing animal GPS tracking data to Movebank and GBIF

Project report for the NLBIF project "MOVE2GBIF"

Project information

Title	MOVE2GBIF - Mobilizing animal GPS tracking data to Movebank and GBIF
Code	nlbif2021.004
Duration	2021-07-01 - 2022-12-31
Subject	Infrastructure
Partner	Radboud University (RU), in collaboration with the Research Institute for Nature and Forest (INBO), Sovon, Movebank and the Dutch Bryological and Lichenological Society (BLWG)
Coordinators	Eelke Jongejans (RU) & Peter Desmet (INBO)
Funding	This project was realized with a financial contribution from Stichting NLBIF - Netherlands Biodiversity Information Facility. NLBIF supports Dutch organizations and researchers in putting biodiversity data online via gbif.org in an open and internationally standardized manner.
Report citation	van der Kolk H, Desmet P, Davidson S, Ens BJ, Jongejans E (2022) Mobilizing animal GPS tracking data to Movebank and GBIF - Project report for the NLBIF project "MOVE2GBIF". https://doi.org/10.5281/zenodo.7408643

Summary

[Movebank](#) is a leading global platform for animal tracking and other animal-borne (a.k.a. bio-logging) sensor data (Kays et al., 2022). It is used by over 3,500 data owners to collect, manage, share, and archive >4 billion location records describing movements of 1,252 taxa in over 7,500 user-managed studies. While Movebank is well known within movement ecology and wildlife management circles, it remains less recognized as a source of general purpose occurrence data. Like most other remotely-sensed animal occurrence data, data from Movebank are not discoverable or accessible through the [Global Biodiversity Information Facility (GBIF)](#). This lack of integration across platforms limits the potential for interdisciplinary collaboration, and for the use of remotely-sensed species occurrence data in conservation assessments and other products that rely on GBIF.

The goal of this project was to develop an open-source workflow to make animal GPS tracking data on Movebank accessible to the GBIF community. This workflow consists of best practices and software to allow semi-automated publication of Movebank data to GBIF. We developed, refined and demonstrated this workflow by publishing a set of datasets from the Netherlands Institute of Ecology (NIOO-KNAW) and the Research Institute for Nature and Forest (INBO) that represent similar and relatively straightforward methodology and quality (avian GPS tracking). In this project, we aimed to complete three work packages:

1. Publish five NIOO-KNAW bird GPS tracking studies as open data on Movebank.
2. Create best practices and software to publish Movebank animal GPS tracking data to GBIF.
3. Publish 11 bird GPS tracking datasets from Movebank to GBIF.

We published data from five Oystercatcher GPS studies in the Netherlands as open data on Movebank and [Zenodo](#). These datasets collectively contain 6 million GPS positions and 164 million acceleration measurements and are formatted as Frictionless Data Packages (Walsch and Pollock, 2017) to facilitate understanding and access. The datasets are described in an open-access data paper published in ZooKeys (van der Kolk et al., 2022).

We developed a new open-source R package "movepub" (Desmet, 2022), that automates the transformation of Movebank data to Darwin Core and metadata to Ecological Metadata Language (EML). Members of the Darwin Core Maintenance Group and Machine Observation Interest Group at TDWG reviewed the suggested transformation to Darwin Core. We created and published best practices to aid Movebank users in publishing their animal GPS tracking data to GBIF and published these on the Movebank website.

We then used the R package to transform 15 datasets to Darwin Core Archives and published these via an Integrated Publishing Toolkit (IPT) to GBIF. We also registered the marine datasets with the [Ocean Biodiversity Information Facility (OBIS)](#).

Overall, we reached or surpassed the goals of this project. The developed workflow proved successful and is already used to support other mobilization efforts. The resulting datasets, software and data paper all acknowledge NLBIF for its support and are publicly accessible.

Project results

WP1: Publish five NIOO-KNAW bird GPS tracking studies as open data on Movebank

We successfully published all tracking data from five studies on Eurasian oystercatchers (*Haematopus ostralegus*) in the Netherlands as open data on Movebank and Zenodo. Data from these studies were collected using [UvA-BiTS](#) GPS trackers — which record location and acceleration in regular intervals — and stored exclusively in a database at the University of Amsterdam, which is not publicly accessible. Before making the datasets openly accessible, we verified data quality and metadata. Data quality checks included marking GPS position outliers and ensuring that all birds that were found dead during the study period were marked as such in the UvA-BiTS database. Deployment metadata were verified and supplemented with additional biometric measurements on individual birds that were not yet included in the database.
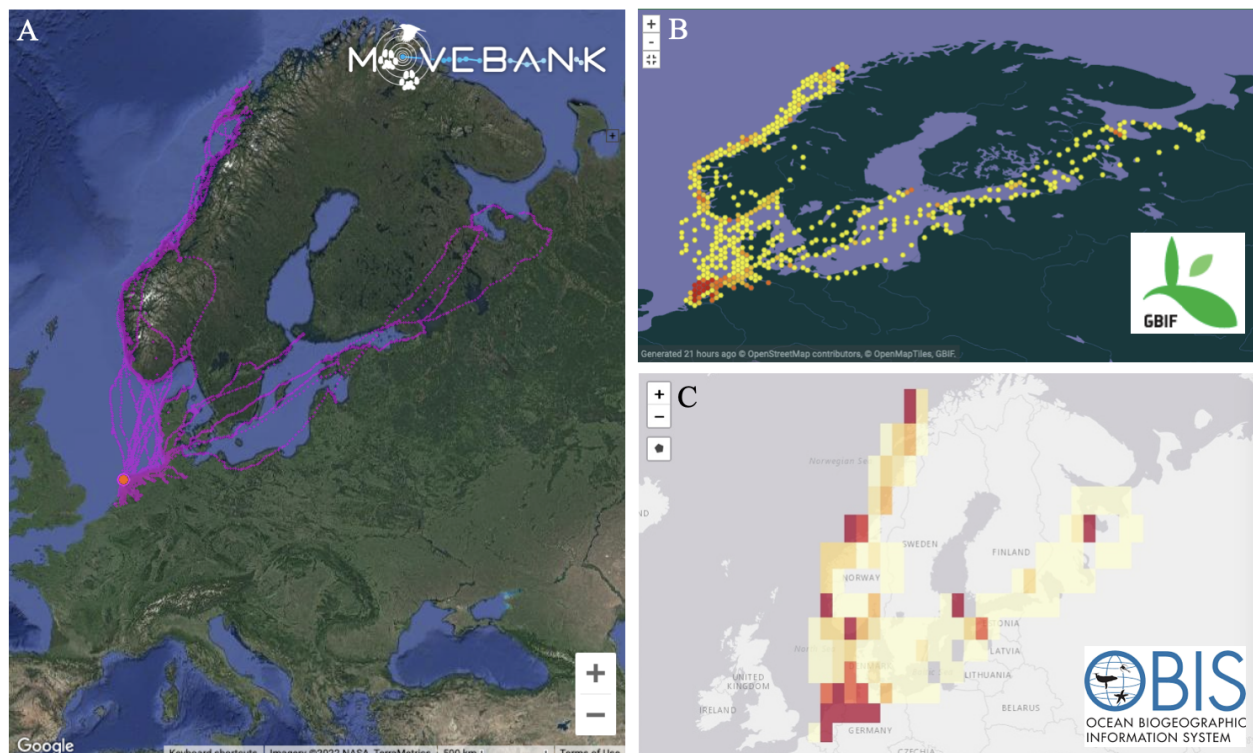

Figure 1. Maps of the data from Movebank study "O_VLIELAND - Eurasian oystercatchers (Haematopus ostralegus, Haematopodidae) breeding and wintering on Vlieland (the Netherlands)" (a) as stored on Movebank, (b) published in GBIF and (c) published in OBIS (van der Kolk et al., 2022; https://doi.org/10.5281/zenodo.5653891).

We then uploaded data and metadata to Movebank as five separate studies, using a [custom script](#) that downloads and transforms UvA-BiTS data in the Movebank format. We also deposited the data on Zenodo for long term archival. These deposits were formatted as Frictionless Data Packages (Walsch and Pollock, 2017) by adding a `datapackage.json` file,

which documents the fields and relationships of all data files. Downsampled subsets of the data were also published on GBIF and OBIS (see next work packages, Fig. 1). All published versions of the datasets link back to the Movebank studies, where current contact information and lists of related publications can be maintained to support future collaborations and uses of the data.

The datasets on Movebank, Zenodo, GBIF and OBIS were described in a data paper that is published in ZooKeys (van der Kolk et al., 2022). Links to each dataset on each platform are provided in the data paper. In addition to the five Dutch studies, the data paper also describes data from one Belgian GPS tracking study on Oystercatchers. Combined, the 6 datasets contain 6.0 million GPS positions, 164 million acceleration measurements and 7.0 million classified behaviour events (i.e., flying, walking, foraging, preening, and inactive) that are now open accessible via several portals. *– 1a. Data quality control, 1b. Metadata description, 1c. Upload data to Movebank, 1d. Archive data on Zenodo, 1e. Data paper* [1]

WP2: Create best practices and software to publish Movebank animal GPS tracking data to GBIF

We created and published best practices on the Movebank website to aid others in publishing their animal GPS tracking data to GBIF and OBIS. We also developed an open-source R package "movepub" that can transform data to Darwin Core and metadata to EML following these best practices.

In deciding the transformations between Movebank attributes (MPIAB, 2022) and Darwin Core terms (https://dwc.tdwg.org/terms), we chose mappings that are clear, widely applicable to GPS tracking datasets on Movebank, and in support of the new GBIF data model. When in doubt, we consulted members of the Darwin Core Maintenance Group and Machine Observations Interest Group at TDWG. Key features of the Darwin Core transformation are described here.

The **R package "movepub"** (Desmet, 2022) was published as open source software on GitHub at https://github.com/inbo/movepub. This package provides two main functions:

- `add_resource()` aids in formatting data files downloaded from Movebank as a Frictionless Data Package. The function should be used with the "frictionless" R package (Desmet and Oldoni, 2022) and looks up the title and definition for each data field in the Movebank Attribute Dictionary (MPIAB, 2022).
- `write_dwc()` transforms Movebank-formatted data, along with DataCite metadata accompanying the published data package, to a Darwin Core Archive.

A website and tutorial (i.e., vignette) accompany the package containing step-by-step instructions and examples.

---

[1] All project objectives as described in the proposal are referenced as small italic text in the relevant project results.

The **best practices** are published on the Movebank website at https://www.movebank.org/cms/movebank-content/archiving-biodiversity. This page references the results of the MOVE2GBIF project and describes the project's general approach to data preparation and transformation to Darwin Core. In addition, it provides background and context for movement ecologists that are not already familiar with Darwin Core, GBIF or OBIS. This explanation highlights the potential benefits to treating animal tracking data as species occurrence data and opportunities for future integration and interdisciplinary collaboration. – *2a. Best practices to publish Movebank animal GPS datasets to GBIF, 2b. Software to make Movebank deposits frictionless, 2c. Software to transform Movebank animal GPS data to DwC*

WP3: Publish 11 bird GPS tracking datasets from Movebank to GBIF

Eleven bird GPS tracking datasets converted to Darwin Core format, as described above, were published using the INBO IPT and **registered with GBIF as well as OBIS**. They comprise five previously unpublished datasets (WP1) and six datasets published to Movebank and Zenodo, but not GBIF or OBIS (Table 1). These publications serve as a complete demonstration of the MOVE2GBIF workflow. Four additional bird GPS tracking datasets of the INBO (not described in the project proposal) were published using the same methodology. Collectively, the 15 datasets comprise 4.4 million occurrences on GBIF. – *3a. Test workflow and publish datasets to GBIF*

Table 1. 15 bird GPS tracking datasets (re)published to Zenodo, GBIF and OBIS as part of the MOVE2GBIF project. Datasets published in addition to the 11 described in the project proposal are indicated in italics.

Title	Movebank	Zenodo	GBIF	OBIS
BOP_RODENT - Rodent specialized birds of prey (Circus, Asio, Buteo) in Flanders (Belgium)	1278021460	DOI	GBIF	NA
CURLEW_VLAANDEREN - Eurasian curlews (Numenius arquata, Scolopacidae) breeding in Flanders (Belgium)	1841091905	DOI	GBIF	OBIS
H_GRONINGEN - Western marsh harriers (Circus aeruginosus, Accipitridae) breeding in Groningen (the Netherlands)	922263102	DOI	GBIF	NA
HG_OOSTENDE - Herring gulls (Larus argentatus, Laridae) breeding at the southern North Sea coast (Belgium)	986040562	DOI	GBIF	OBIS
LBBG_JUVENILE - Juvenile lesser black-backed gulls (Larus fuscus, Laridae) hatched in Zeebrugge (Belgium)	1259686571	DOI	GBIF	OBIS
LBBG_ZEEBRUGGE - Lesser black-backed gulls (Larus fuscus, Laridae) breeding at the southern North Sea coast (Belgium and the Netherlands)	985143423	DOI	GBIF	OBIS
MEDGULL_ANTWERPEN - Mediterranean gulls (Ichthyaetus melanocephalus, Laridae) breeding near Antwerp (Belgium)	1609400843	DOI	GBIF	OBIS
MH_ANTWERPEN - Western marsh harriers (Circus aeruginosus, Accipitridae) breeding near Antwerp (Belgium)	938783961	DOI	GBIF	NA

MH_WATERLAND - Western marsh harriers (Circus aeruginosus, Accipitridae) breeding near the Belgium-Netherlands border	604806671	DOI	GBIF	NA
O_AMELAND - Eurasian oystercatchers (Haematopus ostralegus, Haematopodidae) breeding on Ameland (the Netherlands)	1605803389	DOI	GBIF	OBIS
O_ASSEN - Eurasian oystercatchers (Haematopus ostralegus, Haematopodidae) breeding in Assen (the Netherlands)	1605797471	DOI	GBIF	OBIS
O_BALGZAND - Eurasian oystercatchers (Haematopus ostralegus, Haematopodidae) wintering on Balgzand (the Netherlands)	1605798640	DOI	GBIF	OBIS
O_SCHIERMONNIKOOG - Eurasian oystercatchers (Haematopus ostralegus, Haematopodidae) breeding on Schiermonnikoog (the Netherlands)	1605799506	DOI	GBIF	OBIS
O_VLIELAND - Eurasian oystercatchers (Haematopus ostralegus, Haematopodidae) breeding and wintering on Vlieland (the Netherlands)	1605802367	DOI	GBIF	OBIS
O_WESTERSCHELDE - Eurasian oystercatchers (Haematopus ostralegus, Haematopodidae) breeding in East Flanders (Belgium)	1099562810	DOI	GBIF	OBIS

References

Desmet P. 2022. movepub: Prepare Movebank Data for Publication. R package version 0.1.0. https://github.com/inbo/movepub

Desmet P and Oldoni D. 2022. Read and Write Frictionless Data Packages. R package version 1.0.2.9000, https://cran.r-project.org/package=frictionless, https://doi.org/10.5281/zenodo.5815355

Kays R, Davidson SC, Berger M, Bohrer G, Fiedler W, Flack A, Hirt J, Hahn C, Gauggel D, Russell B, et al. 2022. The Movebank system for studying global animal movement and demography. Methods Ecol Evol. 13(2):419–431. https://doi.org/10.1111/2041-210X.13767

MPIAB (Max Planck Institute of Animal Behavior). 2022. Movebank Attribute Dictionary. Natural Environment Research Council. http://vocab.nerc.ac.uk/collection/MVB/current, accessed January 2022.

van der Kolk H, Desmet P, Oosterbeek K, Allen AM, Baptist MJ, Bom RA, Davidson SC, de Jong J, de Kroon H, Dijkstra B, et al. 2022. GPS tracking data of Eurasian oystercatchers (Haematopus ostralegus) from the Netherlands and Belgium. ZooKeys. 1123:31–45. https://doi.org/10.3897/zookeys.1123.90623

Walsch and Pollock. 2017. Data Package. Version 1. https://specs.frictionlessdata.io/data-package/