

PTICI.MK · GBIF PUBLICATION · WHAT LEAVES THE DATABASE

ptici.mk Darwin Core Dictionary

Two Darwin Core Archives are built from the ptici.mk database and nothing else is sent. This page lists every field in each file, where it comes from, how it is transformed, how often it is filled in the current build, and what is deliberately withheld.

Build **2026-09-10 18:26** Snapshot **ptici_prodsnap** (state of ~2026-08-17)

Exporter **research/publication/build_dwca.py** Self-check **clean, 0 warnings**

Licence in EML **CC BY-NC 4.0, provisional**

DATASET A · SAMPLING EVENT

Survey sessions

ptici_sessions_dwca.zip · 3.2 MB

Every survey session that qualifies as an effort-quantified event: line transects with perpendicular distances, timed searches, and complete species lists. Each event carries its duration, and where the GPS track describes a walked survey, the cleaned track length and geometry. Occurrences are linked to their event.

event.txt Event core 1,028 events

occurrence.txt 38,134 records
Occurrence extension

eco.txt Humboldt ecological 1,028 rows
extension

eml.xml · meta.xml metadata

Protocols 828 timed searches · 171
transects · 29 complete lists

Effort 832 with validated track ·
196 duration only

Complete lists 620 events reported
complete

Dates 2024-05-24 to 2026-08-16

DATASET B · OCCURRENCE

Opportunistic records

ptici_opportunistic_dwca.zip · 0.5 MB

Point records with no usable effort context: records entered without a session, records of sessions that fail the event rules, records dated more than a day from their session, and every record of a sensitive species, generalised and unlinked from its session so the event track cannot re-localise it.

occurrence.txt 17,166 records
Occurrence core

eml.xml · meta.xml metadata

Composition 15,476 direct entries · 1,370
from non-event sessions ·
236 sensitive moved out of A
· 84 off-date

Sensitive 537 records of 16 species, 25
km cell centroids, month
dates

Dates 2017-08-19 to 2026-08-16
(96% in 2025 and 2026)

Species 285

	(43 / 447 / 538 events by year)	Observers	60, pseudonymised; blank on sensitive rows
Species	260		
Observers	34, pseudonymised		

A only in the survey-session archive **B** only in the opportunistic archive **A + B** same field in both

WITHHELD reduced or removed for sensitive species

ROUTING · HOW A DATABASE ROW ENDS UP IN A, IN B, OR NOWHERE

Dataset-construction rules

Applied in this order by the exporter. Counts are from the current build on the snapshot.

RULE	TEST	EFFECT
R1 test sessions	Session id on the explicit list, or name matching test / proba / obuka / training	Session and all its records dropped
R2 duration	Recorded duration between 10 min and 12 h	Otherwise not an event; records go to B
R3 walked track	At least 3 GPS fixes, 50 m of jump-filtered length, mean speed 0.3 to 8 km/h (timed search, complete list) or 0.3 to 5 km/h (transect)	Timed search or complete list that fails: event with duration as its only effort. Transect that fails: not an event; records go to B
R4 screening	Record rejected by the automated plausibility screen	Excluded from both archives
R5 decoupled records	Record of a non-event session, or dated more than one day from its session	Published in B without an event link
R6 sensitive species	Species flagged sensitive (18 species: raptors, Black Stork, Western Capercaillie)	Coordinates to 25 km cell centroid, month date, no observer, no elevation, no distance, unlinked, always in B

Before any rule: soft-deleted sessions and entries are never read, and the parent session of every session-bound record must itself be live.

A **EVENT.TXT · EVENT CORE · 1,028 ROWS · 15 COLUMNS**

Event fields

One row per published session. The first column repeats the event identifier as the archive core id.

TERM	WHAT IT SAYS	SOURCE IN PTICI.MK	HOW IT IS D
eventID also column 0, the core id	Stable identifier of the survey session	Session.id	Prefixed ptici.mk:s ids are never the identifier dataset vers
eventDate	Calendar day of the session	Session.date	ISO 8601 da
eventTime	Start and end clock time	Session.starting_time, ending_time	Interval sta local time a on the hand
samplingProtocol	Which of the three field protocols was used	Session.type	Fixed sente the timed-s wording doe walking bec sessions ca assisted
sampleSizeValue required by GBIF	Size of the sampling effort	Session.full_walking_path, else start and end time	Jump-filtered length in me the track pa (832 events the duration (196 events
sampleSizeUnit required by GBIF	Unit of the value above	derived	metre for tr minute for c events
samplingEffort	Human-readable effort statement	duration, cleaned track	Observer-m kilometres v is published

eventRemarks	Complete-list flag, and why a track is not published	Session.is_list_complete, R3 reason	Semicolon-j phrases; em neither appl left with no gating say s
countryCode	Country of the event	constant	Always MK
decimalLatitude decimalLongitude	Representative point of the event	cleaned track, else Session.starting_point	Centroid of track; the st duration-on Three sessio start point v placeholder coordinates
geodeticDatum	Coordinate reference	constant	EPSG:4326 coordinates
coordinateUncertaintyInMeters	Radius around the point that contains the event	cleaned track geometry	Largest dist the centric vertex, round m assumed accuracy for only events
footprintWKT	The walked route as geometry	Session.full_walking_path	Well-known longitude th 5 decimals, dropping ev longer than on events w passes R3; a track yields MULTILINE

A ECO.TXT · HUMBOLDT ECOLOGICAL EXTENSION · 1,028 ROWS · 7 COLUMNS

Effort and completeness fields

One row per event, keyed by the event identifier. These terms are indexed by GBIF since the June 2026 relaunch.

TERM	WHAT IT SAYS	SOURCE IN PTICI.MK	HOW IT IS DERIV
------	--------------	--------------------	-----------------

coreid	Link to the event	Session.id	Same value as eventid
eco:eventDurationValue	Length of the session	start and end time	Whole minutes; seconds are truncated to the nearest minute to match the stored precision. midnight crossing is not counted, nothing outside 24 min is an event
eco:eventDurationUnit	Unit	constant	Always minutes
eco:targetTaxonomicScope	What the survey looked for	constant	Always Aves
eco:taxonCompletenessReported	Whether the observer recorded every species detected	Session.is_list_complete	reportedCompleteness. If the flag is set, else notReported. Repeated non-sensitive avian sensitive records to B
eco:isAbundanceReported	Counts are recorded per detection	constant	Always true
eco:isSamplingEffortReported	Effort is stated on the event	derived	true whenever samplingEffort is present which is every event

A + B OCCURRENCE.TXT · 38,134 ROWS IN A (EXTENSION) · 17,166 ROWS IN B (CORE) · 24 COLUMNS

Occurrence fields

Identical column set in both archives. In A the first column is the link to the event; in B it repeats the occurrence identifier. Fill rates are given as A / B.

TERM	WHAT IT SAYS	SOURCE IN PTICI.MK	HOW TO USE
coreid A column 0 in B is id = occurrenceID	Which event the record belongs to	entry.session_id	Occurs in all records. In A, it is the link to the event. In B, it is the occurrence ID. In A, it is the link to the event. In B, it is the occurrence ID. In A, it is the link to the event. In B, it is the occurrence ID.

occurrenceID	Stable identifier of the record	entry table and primary key	pt be co ar
basisOfRecord	Kind of evidence	constant	Al sp
scientificName	Species as identified by the observer	Species.name	Pl w m 13 be ac
kingdom · class · taxonRank	Classification anchors	constant	Ar
vernacularName	English common name	Species.en_name	Co
individualCount organismQuantity organismQuantityType	How many birds	entry.minimum_number	M bo ir th
occurrenceStatus	Presence	constant	Al in
countryCode	Country	constant	Al
recordedBy WITHHELD	Who recorded it	entry.user_id	Ps NM la ve pu
eventDate WITHHELD	When the bird was recorded	Session.date in A, entry.date in B	In w da Se
decimalLatitude decimalLongitude WITHHELD	Where the bird was	entry.location	Ha Se

			cé EÉ cé pu A,
geodeticDatum	Coordinate reference	constant	EF ar
coordinateUncertaintyInMeters	Precision of the point	policy constant	36 17 of sc
minimumElevationInMeters WITHHELD	Elevation at the point	entry.altitude	St gr no se lo ce
identificationVerificationStatus	Whether the record passed screening	passed_approve, is_approved	ve pJ bo ur in ha ex
dataGeneralizations WITHHELD	Declares that the point was blurred	policy	Fi or ce ob
informationWithheld WITHHELD	Says what was held back and how to ask	policy	Fi or th re

dynamicProperties	Structured extras as JSON	breeding_code_id, is_overflight, distance_perpendicular_to_path	breeding_code_id, at st fl pe or se O or
occurrenceRemarks	Free-text remark	reserved	Al cc ac sc

Breeding codes present, A and B together: A1 about 15,000, A2 about 14,300, B3 about 3,100, C12 about 740, B4 about 620; 24,431 records in A and 10,562 in B carry a code. Perpendicular distances: 7,806 in A, 47 in B.

WITHHELD NOT IN EITHER ARCHIVE