

THE NEW EBP DATA STANDARD FORMAT

The work on the new EBP data standard was based on the main guidelines agreed at the specific workshop held in the Netherlands during the 8th EBP meeting. The present document describes the new data standard and points out the main differences with the previous one.

Table of content

Incorporating fixed lists and standard monitoring data.....	1
Improvements in terms of data disaggregation	2
New fields	3
Data visualization	4
Fields description and characteristics	4
Events table	4
Records table.....	6
Protocols table	7

Incorporating fixed lists and standard monitoring data

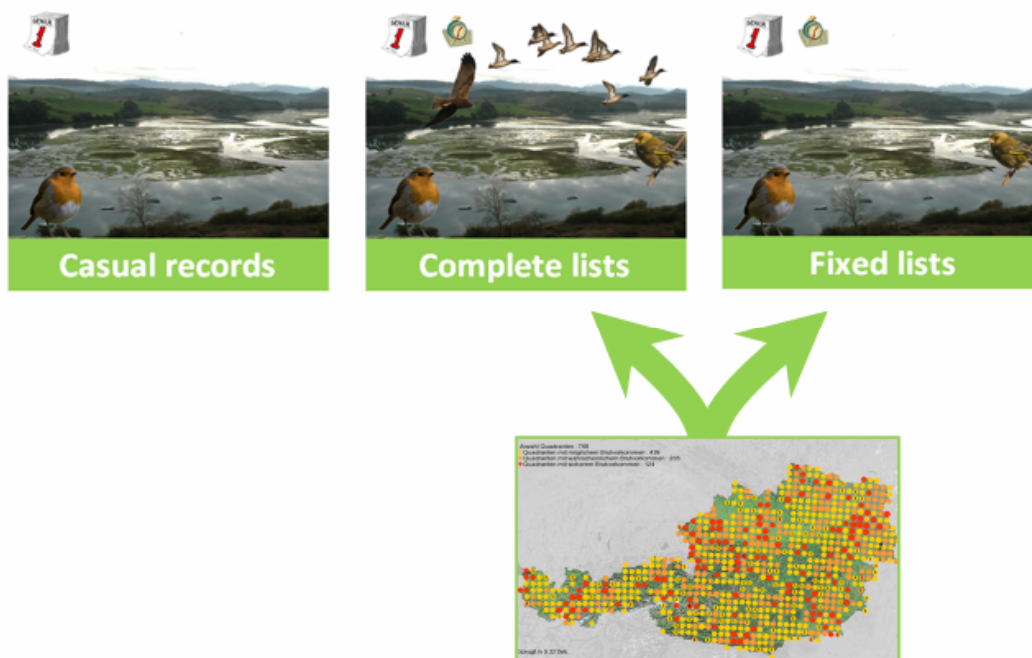
Complete lists are of utmost importance to the EBP because, unlike casual records, they contain effort-like information that greatly improves their analytical value. For this reason, one of the main objectives of the EBP project is to properly store and analyse the essential information that can be extracted from them (namely, date, timing, location, distance/area covered, and the list of detected and not detected species and the counts).

To go in that direction, the new EBP standard has been designed to be able to deal with fixed lists too. These lists are like complete ones but limited to a given group of species and/or conditions (e.g. a complete list of all waterbirds species or of all birds detected in a garden that were not flying-over). Thus, if stored/codified properly, the essential complete list-like information can also be extracted from them.

Importantly, this improvement will also make possible to properly incorporate to the EBP database the essential complete list-like information contained in standard bird monitoring schemes (since their data can easily be simplified to a basic complete or fixed list format).

Standard bird monitoring data is already collected by some local online portals and, therefore, we needed a standard that could handle it correctly. Moreover, additional monitoring data could be directly incorporated in the future in case this may be needed. Properly storing this information will certainly increase the overall quality of the EBP data but also opens new possibilities in terms of data analysis and regarding the development of further synergies with other EBCC initiatives.

Note that to be able to deal with data coming from fixed lists and standard monitoring projects in general, we needed to add a third table to the ones already existing in the former standard: the tables events and records. This third table, named protocols, will collect the details of the protocol followed (e.g. a given Common Breeding Bird Survey) and, in case of fixed lists, the definition of the list.



The new EBP data standard will be able to properly deal with casual records and both complete and fixed lists, thus allowing also the incorporation of the essential complete list-like information contained in the standard bird monitoring schemes.

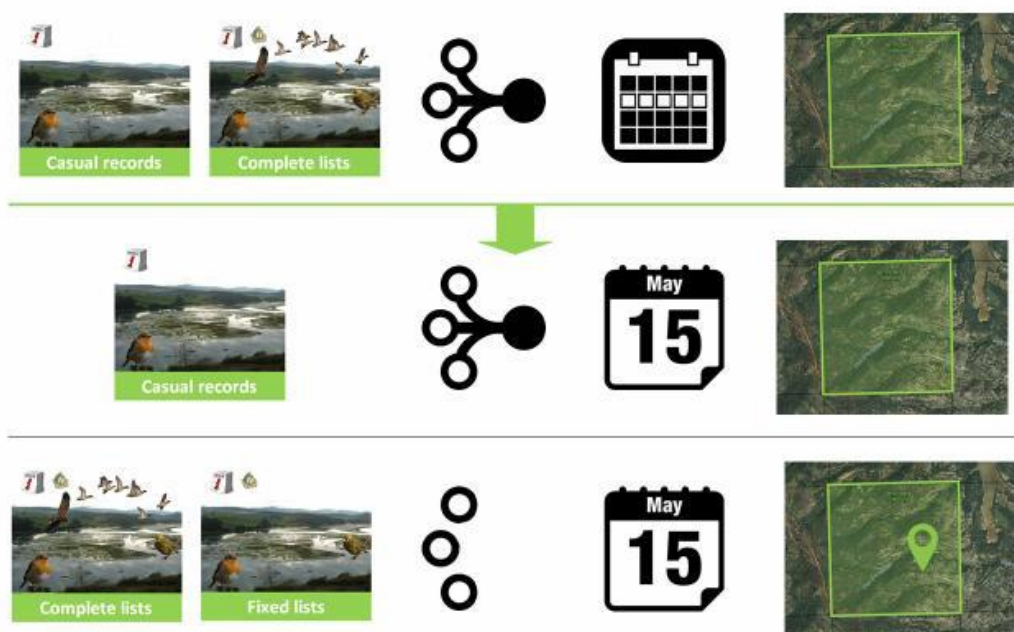
Improvements in terms of data disaggregation

Up to now, all data was submitted aggregated by week and 10x10 km square. With the new standard, casual records will still be kept essentially aggregated, again by 10x10 km square but by date, not by week as before (the new standard will allow casual records to be stored disaggregated and given with exact location if this is required in the future).

On the other hand, the data coming from complete and fixed lists will be collected as individual records (i.e. not aggregated; one species in the lists = one record) using an event identifier to assign each record to each observational event (a list recorded in a given date and location).

Regarding location, partners could choose between giving the exact original location or the 10x10 square reference. Note that unlike with casual records, complete/fixed lists whose location is given as 10x10 squares are not aggregated (i.e. no precise location is given but the coarser location reference is not used to aggregate data).

By storing complete/fixed lists individually we expect to increase significantly the analytical possibilities of the data. The possibility of submitting casual records individually will be discussed again once some trials have been conducted.



With the previous standard, all data was submitted aggregated by week and 10x10 km square. With the new one, casual records will be aggregated by date and 10x10 km square but complete/fixed lists will be submitted without any aggregation.

New fields

Finally, to further improve the analytical potential of the data, the new EBP data standard includes a few new fields. Some of them, like distance travelled, observer id and breeding code are particularly important. Moreover, given that, currently, we are only storing observations from a limited number of species (105), one of the fields of table events will be used to give the total number of species recorded in each complete or fixed list (this figure could be reconstructed if all species observations were collected).

Data visualization

It's important to note that, though the new standard improves the spatial and temporal resolution of the data submitted to the central repository, the level of resolution of the public visualizations of the data in the EBP viewer will be kept unchanged (weekly and 30-km resolution).

Fields description and characteristics

Events table

event_id

Identifier of the observational event (e.g. a given complete list).

You can, for example, use julian date and 10km ETRS89-LAEA grid code as event_id for aggregated casual records (see location_mode = A below).

data_type

C (casual records) / L (complete list) / F (fixed list).

In complete lists all species that have been detected and identified have to be recorded. Fixed lists are like complete list but only with regard to a given group of species and/or conditions (e.g. a complete list of all waterbirds species or of all birds detected in a garden that were not flying-over). Therefore, when using fixed lists their characteristics should be defined using the field fixed_list_tags of the Protocols table.

date

Date of the observational event.

time

Start time of the observational event in local time.

Leave null if unknown or location_mode = A.

location_mode

E (original exact location provided) / D (location lowered to 10x10km level —ETRS89-LAEA grid—) / A (data aggregated at 10x10km level —ETRS89-LAEA grid—).

Note that complete and fixed lists can be provided either using location_mode E or D, while casual records must be provided always aggregated at 10x10 (i.e. using location_mode A).

location

If `location_mode = E` give original exact location as Well Known Text (WKT) in WGS84 (e.g.: POINT(3.056 41.813)).

If `location_mode = D/A` give 10x10km ETRS89-LAEA grid code of the original exact location (e.g.: 10kmE353N212).

protocol_id

Identifier of the protocol followed (e.g. a given Common Breeding Bird Survey).

Leave blank if data do not have an associated protocol.

radius

Maximum distance (in m) to the centroid of the area travelled/covered during the observational event (e.g. 500m).

Leave null when unknown and when `location_mode = A` (note that when `location_mode = D` this info is still very useful —e.g. to identify complete lists where the observer travelled too far away—).

duration

Duration (in hours).

Leave null if unknown or `locationMode=A`.

records

If `locationMode = E/D` (i.e. `data_type = L/F`) give the number of species detected in the complete/fixed list. The total must include all species, not only the EBP target ones the EBP, and should be zero when no species has been detected (i.e. the observational event exists but nothing has been observed).

If `location_mode = A` (i.e. `data_type = C`) give the number of different combinations of observer and species recorded in the given date and 10x10 square.

observer

If `location_mode = E/D` give identifier of the observer (observer ID). *Observer must be unique at the level of the EBP partner data source.*

If `location_mode = A` give the number of different observers submitting observations for the given 10x10 square and date.

state

0 (the event has been removed) / 1 (the event is new or has been modified) / 2 (the event has not been modified (unchanged))

To be provided only during “standard” data provisions.

Records table

record_id

Identifier of the record.

Use, for example, event_id and species_code as record_id for aggregated casual records (i.e. when Events table location_mode = A).

event_id

Identifier of the observational event (e.g. a given complete list).

species_code

Species code.

count

If locationMode = E/D give number of individuals counted.

If locationMode = A give maximum count of all records with counts.

Leave null if only presence is known.

Since some partners give option to use qualifiers (e.g. >=, aprox, etc), counts should be calculated on the raw numbers (e.g. using 200 for >200). Using only observations where numbers are qualified as exact numbers may reduce sample very much.

records_of_species

If location_mode = E/D records_of_species must be always 1.

If location_mode = A give the number of different observers that have recorded the species in the given date and 10x10 square (to be homologous to Events table records field).

breeding_code

Maximum breeding code in EBBA2 standard.

flying_over

Y (yes) / N (no).

Leave null if unknown or unclear (both behaviours detected)

state

0 (the event has been removed) / 1 (the event is new or has been modified)

To be provided only during "standard" data provisions.

Protocols table

protocol_id

Identifier of the protocol followed (e.g. a given Common Breeding Bird Survey).

title

Protocol name/title.

project_type

CB (Common breeding bird survey), CW (Common winter bird survey), WW (Winter waterbird count), BA (Breeding bird atlas), MC (Migration count), WA (Winter bird atlas), GS (Garden bird survey), RB (Rare breeding bird survey), BR (Bird ringing results), NF (Nocturnal flight calls survey), OT (other monitoring projects).

method

P (point counts) / L (line-transect) / M (mapping methods) / T (flexible surveys in which only time is controlled and there is no special requirement regarding the area/distance covered or speed).

website

url of the project/protocol (if existing)

description

Brief description of the protocol.

protocol_details

Details about the protocol that complement the information given in fixed_list_tags.

ebp_data_structure

Details about how the data has been "downgraded" to a complete/fixed list format.

citation

Reference to the protocol.

id_gbif

GBIF doi url to the metadata persistent (doi) of the metadata/dataset uploaded to gbif (i.e. <http://doi.org/10.15468/jsioae>).

geographic_coverage

Area covered by the protocol/project.

start_year

Start year.

end_year

Finishing year.

Leave empty if not finished.

Ongoing

True (ongoing project) / false (project already finished)

fixed_list_tags (*only for dataType = F*)

In case the list of target species is explicit give a semicolon (;) separated list of all these species, including non EBP target species, within the tag "ESP()". Use species codes as in species_code, in records table (e.g. "ESP(54105;54154;57821)").

Use the following tags to (further) define the fixed list (use a semicolon (;) to separate them; in many cases just one tag will be enough): NFO (no fly-overs), ORB (only ringed/trapped birds), OBB (only breeding birds), OWB (only waterbirds), OSB (only seabirds), ORA (only raptors), ORS (only raptors and soaring birds), OAM (only active migrants), PLN (fixed list no strict: other species can be reported).