

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
New standard table and fields names.....	4
Fields description and characteristics.....	5
Events table.....	5
Records table.....	7
Protocols table.....	8

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 improve 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).

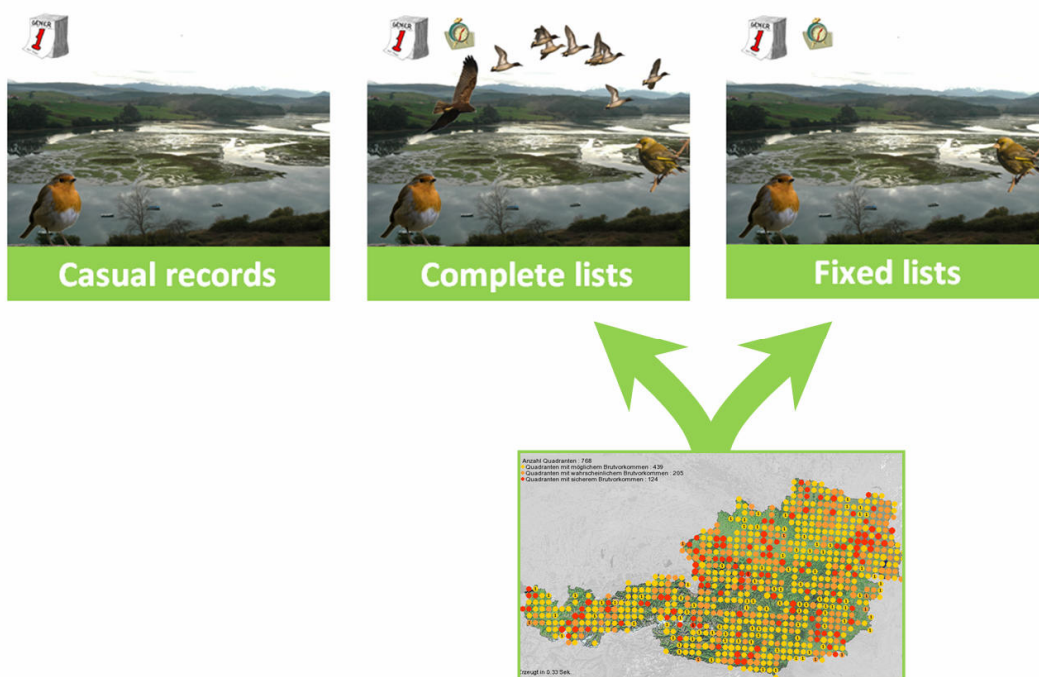
To go in that direction, the new EBP standard has been designed to be able to deal with fixed lists too. These lists are like complete ones but limited to a given group of species and/or conditions (e.g. a complete list of all waterbirds species or of all birds detected in a garden that were not flying-over).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

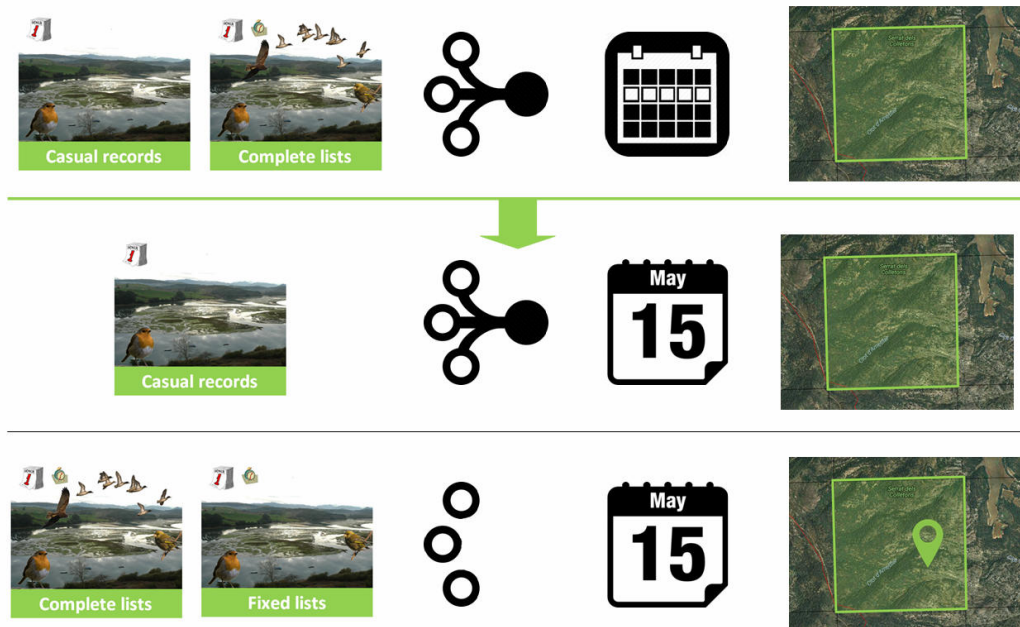
Up to now, all data was submitted aggregated by week and 10x10 km square. With the new standard, casual records will still be kept essentially aggregated, again by 10x10 km square but by date, not by week as before (the new standard will allow casual records to be stored disaggregated and given with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

New standard table and fields names

Events table	Protocols table
partnerID	partnerID
eventID	protocolID
dataType	projectType
protocolID	method
date	fixedlistTags
time	website
duration	description
locationMode	protocolDetails
location	EBPdataStructure
radius	citation
recordsTotal	idGBIF
observer	geographicCoverage
Records table	startYear
	endYear
recordID	
eventID	
partnerID	
speciesCode	
count	
recordsSpecies	
flyingover	
breedingCode	

Fields description and characteristics

Events table

partnerID

Identifier of the EBP partner.

Codes to be provided.

[partnerSourceID may be a better name, since this ID should identify the data source rather than the partner —e.g. for partner using more than one data source—. Therefore, an internal table with partnerSourceID and proper partnerID and name should be added.]

eventID

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

Use julian date and 10km ETRS89-LAEA grid code as eventID for aggregated casual records (see locationMode = A below).

dataType

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

Note that in complete lists all species that are detected are recorded; in fixed lists only all species from a predefined list (e.g. Meadowbirds, Waterfowl) that are detected are recorded (this list should be provided in table Protocols (see below)).

["In the rare cases where fixed lists cannot refer to a given predefined list of species their records must be provided as casual ones" has been deleted. Even if the list of species is not fully fixed, the tags in the protocol table allow to specify quite a lot of relevant information. If necessary, always such data can be used as casual.]

protocolID

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

Leave blank in case of casual records and when complete lists do not proceed from standard monitoring projects.

date

Date of the observational event.

time

Start time of the observational event.

Leave null if unknown or locationMode=A.

duration

Duration (in hours).

Leave null if unknown or locationMode=A.

locationMode

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 locationMode E or D, while casual records must be provided always aggregated at 10x10 (i.e. using locationMode A).

location

Centroid of the location in Well Known Text (WKT) in WGS84 (Spatial Reference System Identifier (SRID) 4326).

When locationMode = D/A give center of the 10x10km ETRS89-LAEA grid where the centroid of the original exact location is placed.

radius

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

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

recordsTotal

Total number of records.

Must be always >0. Note that when locationMode = E/D, the total number of records equals the number of species detected in the complete/fixed list (the total must include all species, not only those that are currently submitted to the EBP). When locationMode = A, use as total number of records the number of different combinations of observer and species recorded in the given date and 10x10 square.

[“When locationMode = A, the total corresponds to the total number of aggregated records.” has been changed for “When locationMode = A, use as total number of records the number of different combinations of observer and species recorded in the given date and 10x10 square.”. Since this is the only way to ensure some standardization in the way casual records are counted.]

observer

If locationMode = E/D → Identifier of the observer (observer ID). *Observer must be unique at the level of the partnerID.*

If locationMode = A → Number of different observers submitting observations for the given 10x10 square and date.

Records table**recordID**

Identifier of the record.

Use eventID and speciesCode as recordID for aggregated casual records (i.e. when Events table locationMode = A).

partnerID

Identifier of the EBP partner.

Codes to be provided.

eventID

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

speciesCode

Species code.

Codes to be provided.

count

If locationMode = E/D → Number of individuals counted.

If locationMode = A → 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.

recordsSpecies

Number of records of the given species.

Note that if locationMode = E/D then recordsNumber must be always 1. When locationMode = A, use as total number of records of the given species the number of different observers that have recorded it in the given date and 10x10 square. [to be homologous to Events table recordsTotal]

breedingCode

Maximum breeding code.

Codes to be provided (based on EBBA2 standard).

flyingover

Y (yes) / N (no).

Leave null if unknown or unclear (both behaviours detected) or when locationMode=A.

Protocols table**partnerID**

Identifier of the EBP partner.

Codes to be provided.

[partnerID added since protocolID will be unique at the level of each partner/data source.]

protocolID

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

projectType

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).

fixedlistTags *(only for dataType = F)*

In case the list of target species is explicit give a semicolon (;) separated list of all these species (use species codes as in speciesCode, in records table) within the tag “ESP()” (e.g. “ESP(SYLATR;ERIRUB)”.

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 (partial list no strict: other species can be reported). [the later added for cases where observers can do a more complete list if they wish —but the system cannot told them apart—.]

[to avoid having metadata somewhere else, and given that protocols table will have only a few records, I deleted the link field but added some more fields to better define the protocol metadata. These are listed below:]

website

url of the project/protocol (if existing)

description

Brief description of the protocol.

protocolDetails

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

EBPdataStructure

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

citation

Reference to the protocol.

idGBIF

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

geographicCoverage

Area covered by the protocol/project.

startYear

Start year.

endYear

Finishing year.

Leave empty if not finished.