# Criteria Catalogue

Version: 1.0
Drafted: 2024-02-06 by Gilles San Martin, data scientist; Michael Rubinigg, quality manager.
Reviewed: 2024-02-06 by Noa Simòn Delso, general manager.
Approved: 2024-02-07 by Michael Rubinigg, quality manager.

The current document has been produced following the process regulated by SOP 014 and is required for the process of peer reviewing, regulated by SOP 015.

## 1 Compliance with objectives of the EU Pollinator Hub

### 1.1 Compliance with at least one of the three general objectives of the EU Pollinator Hub

1.1.1. **Pollinator Health**: Through data sharing, integration, and community-building, we strive to improve pollinators' health, which is essential for our food security and the planet's biodiversity and resilience.

1.1.2. **Beekeeping Trends**: Beekeepers closely follow up honey bee health and production. We are bringing their observations and knowledge together to provide open information about beekeeping sector trends, socio-economic conditions, and bee health.

1.1.3. **Landscape and environment**: Pollinators, farming and beekeeping closely interact with the landscape and environment. We aim to integrate all pollinator-related data and provide valuable information on landscape management for citizens and public institutions.

## 2 Data

### 2.1 General criteria

#### 2.1.1 Range of values in columns

1.1.1.1. Values are in a range that would be expected from the metadata provided.

1.1.1.2. Ranges of values provided in Percent, which are supposed to sum up to 100, fulfil this expectation.

1.1.1.3. Values are in a range compatible with the type of data.

#### 2.1.2 2.1.2. Units

2.1.2.1. All relevant columns must contain a unit from the Units Catalogue.

#### 2.1.3 Data type

2.1.3.1. Each column contains only one data type (*e.g.* numeric, character, logical).

2.1.3.2. Numeric columns do not contain text (*e.g.* comments). Only exception: string to specify missing values.

#### 2.1.4 Uniqueness

2.1.4.1. Values that are supposed to be unique on a dataset level are in fact unique.

2.1.4.2. Values that are supposed to be unique on a global level are in fact unique.

### 2.1.5 Sample

2.1.5.1. The number of samples/observations expected for a particular item corresponds to the sampling plan described in the metadata. .

2.1.5.2. Every sampling and/or observation and/or experimental level is identified by a distinct and unique identifier (alphanumeric or numeric) and/or a EUPH code, grouping related data (*e.g.* a different alphanumeric code and/or EUPH code for each site, apiary, colony, sample, observation, measurement, *etc.*).

## 2.2 Missing values

2.2.1. Missing values are clearly identified and well separated from zero values.

2.2.2. An unambiguous combination of characters is used to encode missing values. **Recommended**: *NA*, *NULL*

2.2.3. Encoding of missing values consistent in the whole dataset.

2.2.4. The same case of a string is used to encode missing values. **NOT recommended**: *Null*, *NULL*, *null*

2.2.5. The following strings are not used to encode missing values. **NOT recommended**: *empty string* (can be ambiguous).

## 2.3 Formatting and encoding

### 2.3.1 Numbers

2.3.1.1. Decimal separators must be uniform.

### 2.3.2 Date and time

2.3.2.1. All dates are formatted in the same way. **Recommended**: ISO 8601 format YYYY-MM-DD hh:mm:ss, *e.g.* 2024-02-06 22:12:12.123 (alternatively year, month, day and time may be provided in separate columns)

2.3.2.2. Data collected in more than one time zone is accompanied by the time zone information from more than one time zone. **Recommended**: ISO 8601 format

### 2.3.3 Text

2.3.3.1. Data measured on a nominal scale is encoded in a uniform way.

2.3.3.2. The same case of a string is used to encode data measured on a nominal scale. **NOT recommended**: *Male*, *MALE*, *male* for variable *sex*

### 2.3.4 Geographic coordinates

2.3.4.1. Longitude/Latitude data is provided in decimal degrees. **NOT recommended**: *degree minute second* (prone to errors)

2.3.4.2. Geographic coordinates are accompanied by a specification of the coordinate reference system used for encoding (e.g. EPSG code, PRJ4 string), either in data or in metadata. **Recommended**: *EPSG 4326*

# 3 Metadata

## 3.1 All tables and columns are clearly described.

3.1.1. The content of tables and columns is clearly described.

3.1.2. The method used for data acquisition is clearly described.

3.1.3. The method used to process raw data is clearly described.

# 4 Compliance with FAIR Principles

## 4.1 Findable

4.1.1. Data are described with rich metadata

4.1.2. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

## 4.2 Interoperable

4.2.1. (Meta)data use vocabularies that follow FAIR principles

## 4.3 Reusable

4.3.1. (Meta)data are released with a clear and accessible data usage license

4.3.2. (Meta)data are associated with detailed provenance