

Thursday, 16. 11. 2023

Training Session

Uploading datasets in the EU Pollinator Hub





Agenda

1. About EU Pollinator Hub
2. Data lifecycle
3. Registration
4. Dataset workflow
5. Cooperation
6. Demonstration





Why using the Hub?

- Your data will be FAIR (Findable, Accessible, Interoperable, Reusable)
- Sharing data with the community
- Receive data report related to submitted uploaded dataset
- Automatically share on EFSA knowledge junction on Zenodo
- Generate DOI (if it does not yet exist)
- Processing of data for advanced visualizations
- Combining data with other datasets to enrich it
- Generate publications with combined data



1. About EU Pollinator Hub

What we do

- Building infrastructure allowing users to bring observations and data together, making sense of them
- Transforming raw data into valuable information for decision making
- Collective approach of the pollinator community
- Operative from February 2024



Open-source platform

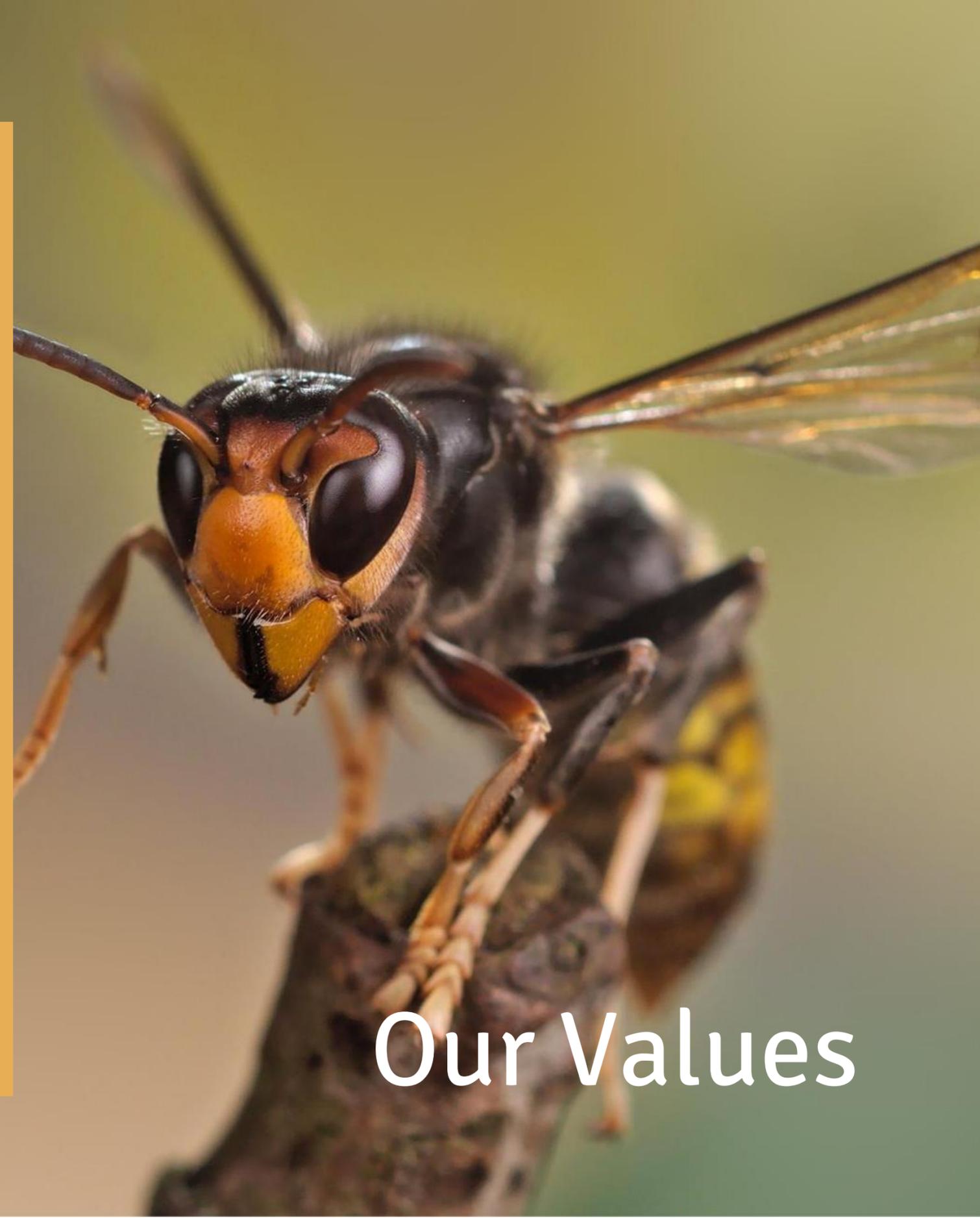
The software is open source and available to everyone

Clear ownership policies

We do not own the data, the users do!

Tight integration

We are working close to beekeepers, naturalists, public institutions, scientific communities, etc.



Our Values

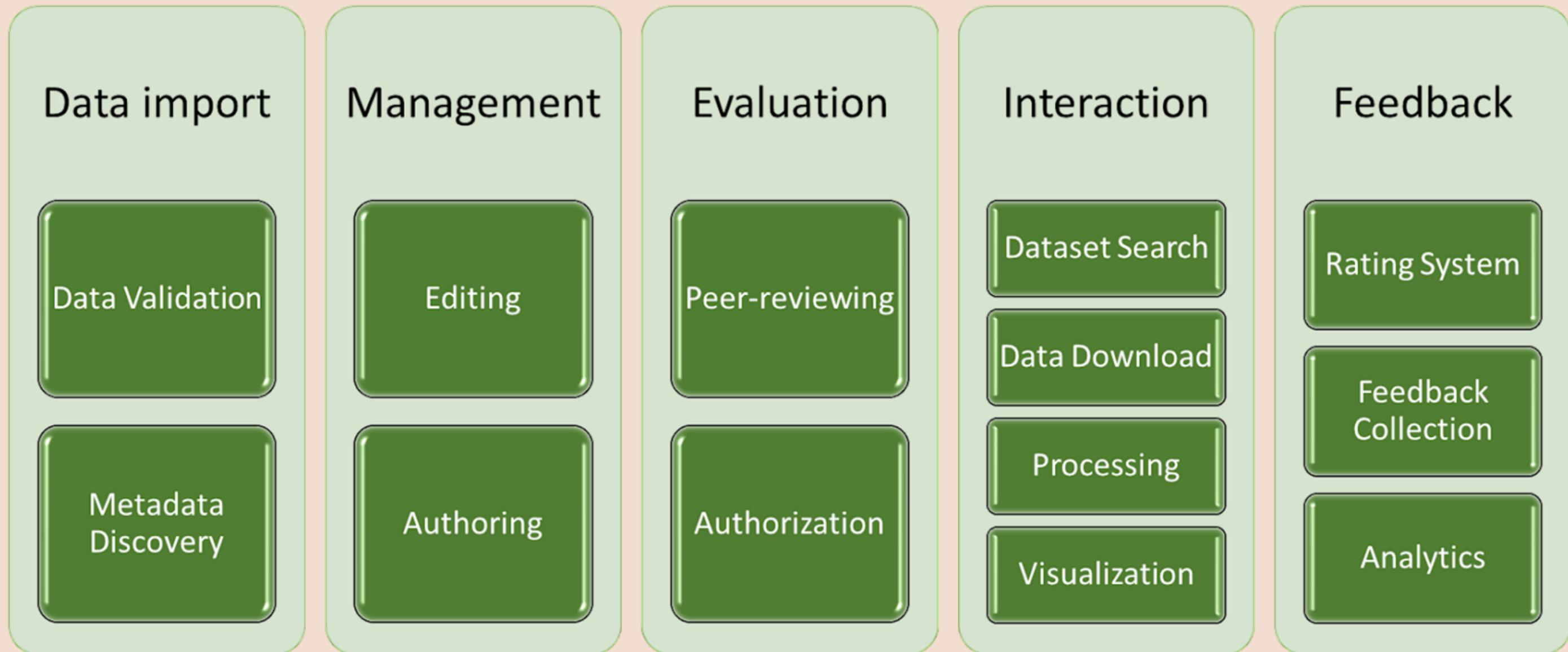


Key Features

- **Data management and publishing tool**
Dataset upload, processing and visualisations
- **Internationalisation and localisation with Dictionary**
Building a standardisation database.
- **Processing and reporting**
Building visualisation based on uploaded datasets and generating automated reports
- **Community driven cooperation Tool**
Manage datasets and build international pollinator related dictionary



2. Data Lifecycle





3. Registration

Simplified process for clarity and ease of access



Email

Password

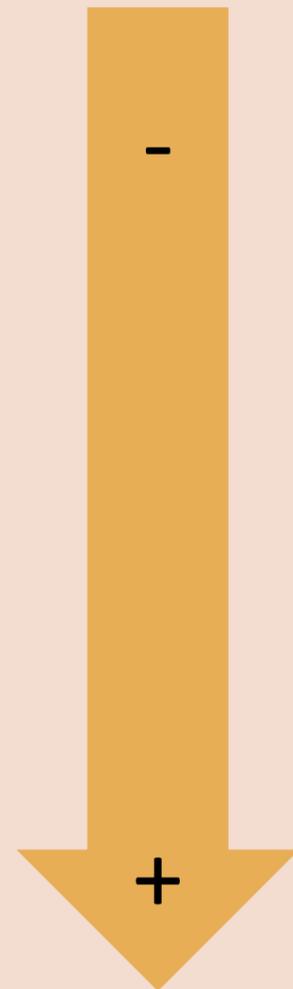
Remember me

[Register](#) [Forgot your password?](#) [LOG IN](#)



Platform interactivity

Level of
interaction



Anyone

Anonymous dataset explorer
with visualisations and downloads

Registered user

Anonymous but with additional features:
Favourites, search queries, dataset feedback and more

Non-anonymous
user

Dictionary
translator

Data
provider

Peer
reviewer



Become a Data Provider

1. Register

2. Upgrade account to non-anonymous user

3. Create a Data provider

Data provider profile is publicly displayed on the platform

The screenshot shows a web interface for creating a data provider profile. The page has a light green header with the title 'Become a Data Provider' and navigation links for 'Dashboard' and 'Dictionary'. The main form is white with a light green border. It contains several sections: 'Full Name' with a text input field containing 'Gregor Sušanj' and a subtext 'Full, registered name of the entity or person.'; 'Abbreviation' with a text input field containing 'GS' and a subtext 'Most often abbreviation is used to describe the source of the dataset. Use initials for persons.'; 'Country' with a dropdown menu showing 'Slovenia'; 'About' with a rich text editor containing a toolbar with icons for bold, italic, underline, link, unlink, list, list, table, undo, and redo; 'Featured Image' with a large empty box and the text 'Drag & Drop your files or Browse'; 'Contact information' with an 'Add another contact' button and a subtext 'By entering your contact details here, public platform users can contact you. However, unless you added it, your email address used for login will be kept private.'; and two checkboxes at the bottom: 'I understand that all information provided here will be publicly displayed, including any personal data.' and 'I accept the Data Provider terms and conditions.'. A dark green button labeled 'BECOME DATA PROVIDER' is located at the bottom right of the form.



Become a Data Provider

Types of Data Providers:

Personal (privately owned)

Entity (owned by one or more users)

The screenshot shows a web application interface for adding a Data Provider. The top navigation bar includes "Dashboard", "Dictionary", and "My Data Catalog". The user is logged in as "Gregor Sušanj" (Managing as: GS). The main content area is titled "Data Provider Details" and contains the following fields:

- Data Provider Owner:** Gregor Sušanj, gregor.susanj@zip.solutions
- Full Name:** Gregor Sušanj
- Abbreviation:** GS
- Country:** Slovenia
- About:** A rich text editor with a toolbar containing icons for bold, italic, strikethrough, link, unlink, heading, quote, list, table, undo, and redo.
- Featured Image:** A drag-and-drop area with the text "Drag & Drop your files or Browse".
- Contact information:** An "Add another contact" button.

At the bottom right, there is a "SAVE" button. A disclaimer at the bottom states: "By entering your contact details here, public platform users can contact you. However, unless you added it, your email address used for login will be kept private."



4. Dataset Upload Workflow



Insert Meta Information - Document the Dataset

General information about the dataset, data provider and tables (the real data).



Upload Files and Data Profiling

Reviewing uploaded data for errors.



Describing Data and Building a Collective Dictionary

Describing in detail each column of the dataset (tables).



Submit for Review

Before data can be processed or downloaded, it requires review for technical and quality validation.



My Data Catalog

1. Browse own datasets
2. Browse datasets owned by other collaborators of my teams

The screenshot displays the 'My Data Catalog' interface for the EU Pollinator Hub. The top navigation bar includes 'Dashboard', 'Dictionary', 'Platform Management', and 'My Data Catalog'. The user is logged in as 'EUPH'. The main content area is titled 'All Datasets' and features a 'CREATE NEW DATASET' button. The datasets are organized in a grid:

Dataset ID	Dataset Name	Description	Access
000001	EUPH	This dataset contains the basic data created for the platform to function.	Private
000002	Countries	EUPH reference dataset containing names and codes of countries, aggregates and subdivisions of countries. The dataset contains standardised information on countries published by the Organization for Standardization (ISO) and the United Nations Statistics Division...	Private
000003	NUTS	Nomenclature of territorial units for statistics. The dataset contains the nomenclature of territorial units for statistics (NUTS), a classification set up by Eurostat as a single,...	Public
000004	Postcode	Postcodes from various countries. The dataset contains postcodes from 97 countries for internal use on the EUPH.	Public
000005	Currencies	EUPH reference dataset containing currencies. This dataset contains standardised information on current and historic currencies, funds and precious metal codes contained in the ISO 4217...	Private
000007	Units	EUPH Reference dataset containing units. This dataset contains standardised information on units.	Private
000009	Faostat	Worldwide agricultural production related to beekeeping provided by FAOSTAT. The dataset contains a selection of tables on worldwide crop and livestock production as well as macroeconomic and environmental indicators...	Public
000010	Neonic	Neonic. This dataset contains data on regular assessments of colony development in colonies exposed to different levels of Clothianidin during corn...	Private
000011	Pollination	Data on economic dependence of pollination by animals in crops as well as observations of pollinators on plant species, inon economic dependence of pollination by animals in crops as well as observations of pollinators on plant species, in particular crops. The dataset contains data on economic dependence of pollination by animals in crops as well as observations of pollinators on...	Public
000013	BeeTRIP	BeeTRIP	Private
000014	Vetcases	Reports of notifiable diseases reported for Apis mellifera	Private
000016	Varroa Monitoring Austria	Varroa Monitoring Austria	Private



Insert Meta information - Dataset

1. Who is the Data Provider?
2. Dataset Title
3. User License Definition
4. Contributors (ownership)
5. Contact Information

The screenshot shows a web application interface for creating a new dataset. The page title is "GS > Data Catalog > Create a new dataset". The form is titled "Create a new dataset" and includes the following sections:

- Create a new dataset:** A paragraph explaining that a dataset is a collection of data for a single purpose and that users should provide basic information.
- DOI:** A section explaining that a DOI is a digital identifier and providing a link to the DOI Foundation website.
- Licences:** A section explaining that the platform supports Creative Commons licenses and recommending the CC-BY license, with a link to the Creative Commons website.
- Data Providers:** A dropdown menu with "Gregor Suijanj" selected and a "Select an option" prompt.
- Short title:** A text input field with a prompt: "Provide a short title using no more than 30 characters (including blanks)".
- Long Title:** A text input field with a prompt: "Provide a long title of the dataset using up to 256 characters (including blanks)".
- DOI:** A text input field with a prompt: "Provide a Digital Object Identifier for this dataset, if it already exists."
- Licence:** A dropdown menu with "No Licence (private dataset)" selected and a "Select a licence from the list." prompt.
- About:** A rich text editor with a toolbar containing icons for Bold (B), Italic (I), Underline (U), Link (L), Unlink (U), Headings (H), Lists (ul, ol), Table (T), Undo (↶), and Redo (↷). Below the editor is a prompt: "Provide a detailed description using no more than 100 words."
- Featured Image:** A large text input field with a "Drag & Drop your files or Browse" prompt.
- Contributors:** A section that is partially visible at the bottom of the form.



Insert Meta information - Table

1. Name

2. How is it provided?
(Format: file, multiple files, cloud, APIs)

3. About

Typical information included in README docs.

The screenshot shows the 'Create a new table' form in the EU Pollinator Hub Data Catalog. The form is titled 'Vetcases' and contains the following sections:

- Create a new table:** A table is a collection of related data that is transferred to the EU Pollinator Hub.
- Short Title:** A clear and concise title of the table will help others to quickly grasp its content. The form includes a text input field and a note: 'Provide a short title using no more than 30 characters (including blanks)'. There is a checkbox for 'This is my own data or I have full permission to use it'.
- Incoming Licence:** The incoming licence documents under which conditions the data source in this table was made available to the data provider. The form includes a dropdown menu with 'Custom licence' selected and a note: 'Select a licence from the list.' Below this is a rich text editor for 'Custom Licence Text IN' with a toolbar containing icons for bold, italic, strikethrough, link, header, list, table, undo, and redo.
- Types of data source:** The EU Pollinator Hub currently supports four different types of data sources for data that is stored in a table. The form includes a dropdown menu with 'File' selected and a note: 'Select a data source from the list.'
- About:** Provide a detailed description of the table. Answer the following questions: What does the data measure or track? Who provided the data? The form includes a rich text editor with a toolbar containing icons for bold, italic, strikethrough, link, header, list, table, undo, and redo.
- Featured Image:** Provide an image that helps others to quickly grasp the content of the dataset. The form includes a text input field.



Upload Files and Data Profiling - Upload

1. Select File

- Currently only .csv supported
- More formats in the future

2. Details about the file

3. Define missing (null) values

Dashboard Dictionary My Data Catalog Managing as: GS

GS > Data Catalog > Vetcases > Disease References > Upload Data

Upload Data

Select a file to start the data upload process.

Supported file types

Currently the only supported file type is a comma-separated (CSV) file with included header. More supported types will be added in the future.

Character encoding

Currently the platform only supports UTF-8 encoding. Make sure when exporting data from your own database to use UTF-8 character encoding.

Disease References

This table contains 3 records with the description of honey bee diseases.

Select a file*

Drag & Drop your files or Browse

Maximum file size is :limit.

Column delimiter*

Comma (,)

Decimal separator*

Period (.)

Encoding*

UTF-8

Character sequence to be interpreted as missing values

Type value

Case-sensitive list of entries, designating null or unavailable item values. For example NULL, N/A., NaN

CANCEL UPLOAD



Upload Files and Profiling - Profiling

1. Metadata Measures
2. Descriptive Measures
3. Data Quality Measures
4. Sample Data
5. Profiling Charts

The screenshot shows a web interface for a data catalog. The top navigation bar includes 'Dashboard', 'Dictionary', and 'My Data Catalog'. The user is logged in as 'GS'. The breadcrumb trail is 'GS > Data Catalog > Vetcases > Vetcaseat'. The main content area is divided into several sections:

- Vetcaseat**: A description of the dataset, stating it was obtained from the Consumer Health Communication Platform (Kommunikationsplattform Verbrauchergesundheit, KVG) provided by the Austrian Federal Ministry of Social Affairs, Health, Care and Consumer Protection (Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz, BMSGPK). It contains information on food and consumer protection as well as animal health and animal welfare, among which outbreaks of notifiable honey bee diseases from 2015 to 2022 in Austria (AT).
- Actions**: Two buttons, 'EDIT' and 'DELETE'.
- Properties**: A list of metadata including Dataset (Vetcases), Type (File), Licence (None required), Columns (15), Rows (513), Data points (7695), Data points (72.45 KiB), File Encoding, Created (16 November 2023 12:50:29 by Gregor Sušanj), and Last modified (16 November 2023 12:50:29).
- Source Files**: A table with columns 'File name', 'Status', and 'Created at'. It shows one file: '02_vetcaseat.csv' created at '16.11.2023 12:52:43'. There are 'Describe' and 'Delete' icons for this file. Below the table, it says 'Showing 1 result' and 'Per page 10'.
- Profiling**: A section with tabs for 'METADATA MEASURES', 'DESCRIPTIVE MEASURES', 'DATA QUALITY MEASURES', 'SAMPLE DATA FROM FILE', and 'PROFILING CHARTS'. The 'METADATA MEASURES' tab is active, showing a table with columns: 'Column', 'Basic Descriptor', 'Advanced Descriptor', 'Length of characters', 'Unit', and 'Type'. The table contains two rows:

Column	Basic Descriptor	Advanced Descriptor	Length of characters	Unit	Type
id	Integer	n/a	1 - 3	Default unit	int
year	Integer	n/a	4 - 4	Default unit	int



Describing the Data and Building a Dictionary – Describe

For each column:

1. Provide concise description
2. Select Data type
3. Select or create a new descriptor identifying the column
4. Select or create a unit, if applicable

The screenshot shows a web interface for describing data. The breadcrumb trail is: GS > Data Catalog > Vetcases > Vetcaseat > Describe your data. The page is titled 'Describe your data' and contains the following sections:

- Describe your data:** A text area for providing a general description of the data.
- Describe the column:** A text area for providing a concise description for each column.
- Vetcaseat:** A text box containing the following text: "Data in this table was obtained from the Consumer Health Communication Platform (Kommunikationsplattform Verbrauchergesundheit, KVG) provided by the Austrian Federal Ministry of Social Affairs, Health, Care and Consumer Protection (Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz, BMSGPK). It contains information on food and consumer protection as well as animal health and animal welfare, among which outbreaks of notifiable honey bee diseases from 2015 to 2022 in Austria (AT)."
- Column list:** A horizontal scrollable list of column names: id, year, month, casenumber, disease, animalspecies, datestart, dateend, country, nuts, district.
- Column description:** A rich text editor with a toolbar containing icons for bold, italic, strikethrough, link, unlink, list, list, table, undo, and redo.
- Select a Basic Descriptor:** A dropdown menu with 'Integer' selected.
- Select or create a detailed Descriptor:** A dropdown menu with 'None selected' and a '+' button.
- Unit:** A dropdown menu with 'Select an option' selected.
- Additional Column Analytics:** A dropdown menu.



Describing the Data and Building a Dictionary – Dictionary

For descriptor or type:

1. Create a class
2. Write a definition
3. Add possible translations

The screenshot shows a web interface for proposing a new class in a dictionary. The top navigation bar includes 'Dashboard', 'Dictionary', 'Platform Management', and 'My Data Catalog'. The user is logged in as 'GS'. The current page is 'Dictionary > Classes > Propose Class'.

Propose a Class
A class is a group of unique real-world entities or objects (things, organisms, abstract concepts, processes) Which comprise elements that are sufficiently similar to be considered different from any other class.

Description
Use single words (e.g. nouns, verbs, adjectives/adverbs), groups of verbs (e.g. compound nouns, compound verbs) or phrases in English language that describe a real world entity.

Parents
The dictionary has a hierarchical structure. Classes are embeded in a system with higher-level classes (the parent) and lower-level classes (children).

Dataset
Not related to Dataset
If you don't select a specific dataset this class will be shared among all datasets.

Class description*
See on the left of this form for more information about this field.

Parents
Select an option
You can add as many parents as you want. See on the left of this form for more information about this field.

Buttons: CANCEL, PROPOSE



Submit for review

1. Review data

2. Submit for Review

Optionally peer-review

3. Publish

The screenshot shows a web interface for managing data catalog entries. The top navigation bar includes "Dashboard", "Dictionary", and "My Data Catalog". The user is logged in as "GS". The breadcrumb trail is "GS > Data Catalog > Vetcases".

The main content area displays the details for a dataset titled "000108 Vetcases". The description is "Reports of notifiable diseases reported for Apis mellifera". A note states: "This dataset contains data from various sources on the outbreak of notifiable diseases reported for honey bees (Apis mellifera L.) in Belgium and Austria from 2000 to 2022." The "Actions" panel includes buttons for "EDIT", "SUBMIT FOR REVIEW", and "DELETE".

The "Properties" section shows: "Publicly displayed" (checked), "Current status" (Draft), "DOI" (Not Assigned), "Data Providers" (Gregor Sušanj), "Licence" (Attribution 4.0 International), "Created" (16 November 2023 12:28:15), and "Last modified" (16 November 2023 13:30:00). A link "See related classes" is also present.

The "Contacts" section indicates "No contacts is available." A "NEW TABLE" button is located at the bottom right of the main content area.

The "Tables" section lists two related tables:

- Disease References**: This table contains 3 records with the description of honey bee diseases.
- Vetcaseat**: Data in this table was obtained from the Consumer Health Communication Platform (Kommunikationsplattform Verbrauchergesundheit, KVG) provided by the Austrian Federal...



5. Cooperation

Work with peers to provide high-quality data and reduce the required labor with cooperation





Shared Data Providers

EU Pollinator Hub for teams/projects

1. Allow other users to manage uploaded datasets
2. Limit their access
3. Transfer ownership to other users

The screenshot displays a web interface for managing shared data providers. It is divided into three main sections:

- Contact Management:** A form with a "Type" dropdown menu set to "LinkedIn Profile" and a "Value" text input containing a LinkedIn URL. Below the input is an "Add another contact" button and a "SAVE" button. A note states: "By entering your contact details here, public platform users can contact you. However, unless you added it, your email address used for login will be kept private."
- Add Member:** A section titled "Add Member" with the instruction "Add a new member to your team, allowing them to collaborate with you." It includes an "Email" input field, a "Role" dropdown menu, and an "ADD" button. The role options are:
 - Administrator:** Administrator users can perform any action.
 - Editor:** Editor users have the ability to read, create, and update.
 - Member:** Members have the ability to only read.
- Members:** A list titled "Members" with the note "All of the people that are part of this data provider." It shows four members with their names, roles, and actions:
 - Gilles San Martin (Administrator) - Remove
 - Gregor Sušanj (Administrator) - Leave
 - Michael Rubinigg (Administrator) - Remove
 - Noa Simon Delso (Administrator) - Remove

At the bottom, there is a "Delete Data Provider" section with an input field.



Cooperate in dictionary

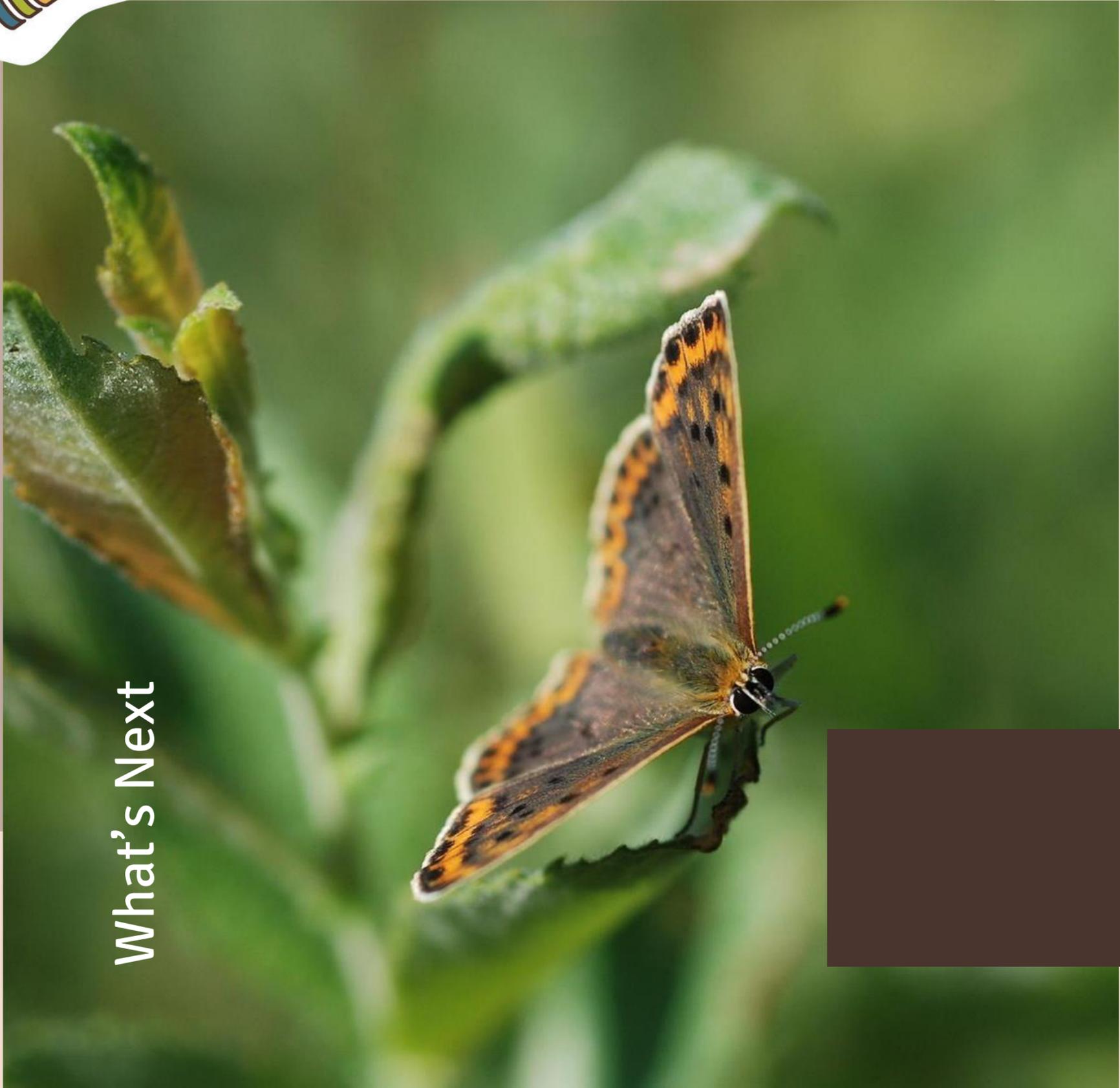
1. Standardise classes / terms used across one or more datasets
2. Translate them in other languages
3. Relate terms them to real world entities

The screenshot shows a web application interface for a dictionary class 'honey'. The top navigation bar includes 'Dashboard', 'Dictionary', and 'My Data Catalog'. The user is logged in as 'GS'. The main content area is titled 'Dictionary > Classes > honey' and features tabs for 'List View', 'Tree View', 'Translator', and 'Encyclopedia'. The 'List View' tab is active, displaying the class 'honey' with a 'START DISCUSSION' button. The class description states: "Honey is the natural sweet substance produced by Apis mellifera bees from the nectar of plants or from secretions of living parts of plants or excretions of plant-sucking insects on the living parts of plants, which the bees collect, transform by combining with specific substances of their own, deposit, dehydrate, store and leave in honeycombs to ripen and mature." A reference is provided: "EU 2014 Jun. Council Directive 2001/110/EC of 20 December 2001 relating to honey. [2023-02-08] External link". The class was proposed by Michael Rubinigg. The 'Class Relationships' section shows 'Parents: primary hive product' and 'Children: See children'. The 'Contributors' section lists 'Class Contributors' (Author: Michael Rubinigg), 'Definition Contributors' (Michael Rubinigg), and 'Translation Contributors' (Michael Rubinigg, Carla Stoyanova, Gregor Sušan, Irene Guerra, Awad Hassan, Claudia Paşca). The 'Class Translations' section is a table with columns for 'Votes', 'Language', 'Expression', 'Contributed By', and 'Status'. The table lists translations for Arabic, Bulgarian, English, French, and German, all with 'Approved' status and a 'START DISCUSSION' button.

Votes	Language	Expression	Contributed By	Status
<input type="checkbox"/>	Arabic	عسل noun	Awad Hassan	Approved
<input type="checkbox"/>	Bulgarian	Мед noun,masculine,singular	Carla Stoyanova	Approved
<input type="checkbox"/>	English	honey noun,neuter,singular	Michael Rubinigg	Approved
<input type="checkbox"/>	French	miel noun,masculine,singular	Michael Rubinigg	Approved
<input type="checkbox"/>	German	Honig noun,masculine,singular	Michael Rubinigg	Approved



What's Next



6. Demonstration

Using the platform to upload, describe and publish a dataset.



Thank you



About

<https://pollinatorhub.eu/>

Contact

<https://pollinatorhub.eu/contact/>
info@bee-life.eu

Register and use

<https://app.pollinatorhub.eu>

