Standard Operating Procedure

Number: **012**

Name: **Data standardisation**

Title: Procedure for the standardisation of information assets on the EU Pollinator Hub

Category: Data management

Status: Draft

Version: 2.0

Date: 2025-01-29

Authors: Michael Rubinigg

This document is intended for the regulation of processes within the EU Pollinator Hub. It is published under the CC BY 4.0 licence. You may distribute, transform, adapt and build upon the material in any medium or format, including for commercial purposes, as long as you give credit to the author.

Table of contents

1 Document history 3
1.1 Current Release 3
1.2 Change history 3
2 Abbreviations 4
3 Introduction 5
3.1 Definitions 5
4 Validity 6
4.1 Area 6
4.2 Time 6
4.3 Related Documents 6
4.4 Related Tasks 6
5 Procedure 7
5.1 Responsibilities 7
5.2 Classes 8
5.3 Descriptors 15
5.4 Units 16
6 References 18
7 Annexes 19

1 Document history

1.1 Current Release

1.1.1 Drafted

Submitted on 2025-01-28 by Michael Rubinigg, quality manager.

1.1.2 Reviewed

Reviewed on 2025-01-28 by Noa Simòn Delso, general manager.

1.1.3 Approved

Approved on 2025-01-29 by Michael Rubinigg, quality manager.

1.2 Change history

Version	Date	Author	Modification	Reason
1.0	2025-01-28	Michael Rubinigg	In section 5.2.2.2 the word description was replaced by the word name.	Imprecise use of the term *description* and a modified content in the user interface.
1.0	2025-01-28	Michael Rubinigg	In section 5.2.2.2 additional rules for the creation of a class name were added.	The integration of external standards such as Darwin Core and ISO made it necessary to adapt the rules.
1.0	2025-01-28	Michael Rubinigg	Section 5.2.2.6 is added, which contains an improved description for the definition of analogies.	The integration of external standards such as Darwin Core and ISO made it necessary to adapt the rules.
1.0	2025-01-28	Michael Rubinigg	Section 5.2.3 is added.	The section contains an improved description for the process of a definition of classes.
1.0	2025-01-28	Michael Rubinigg	Section 5.4.1.3 receives additional instructions.	A software version with additional features for the management of descriptors has been deployed, which requires a review of the process.
1.0	2025-01-28	Michael Rubinigg	The name of the application was changed from *Dictionary* to *Vocabulary*.	Dictionary has been renamed to Vocabulary, which from a technical point of view is more appropriate.
1.0	2025-01-28	Michael Rubinigg	The expression *parent entities* is replaced by the term *superclasses*.	The term superclasses complies with the nomenclature of the Darwin Core standard.
1.0	2025-01-28	Michael Rubinigg	Section 5.2.2.5 is added.	A software version with additional features for the management of relations between classes has been deployed.
1.0	2025-01-28	Michael Rubinigg	Section 5.2.3.3 is expanded by stipulating the inclusion of axioms in the definition.	Improved definition of the ontology design.

2 Abbreviations

EU	European Union
EUPH	EU Pollinator Hub
IRI	Internationalized Resource Identifier
n.a.	not available
SOP	Standard Operating Procedure
UID	Unique Identifier
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
UTF-8	8-Bit Universal Coded Character Set Transformation Format
WI	Work Instructions

3 Introduction

The aim of the present SOP is to provide a procedure for the definition of classes and their translations into different languages (referred to as expressions) through a collaborative, software-based process. The software allows registered users to participate in a fully traceable process, including the possibility to submit and edit proposals, the possibility to comment on proposals and the approval of proposals by higher-level-entities (supervisors).

3.1 Definitions

Class is a data element, which describes objects, organisms, abstract concepts or processes that are sufficiently similar to be considered different from any other class in terms of their purpose.

Expression is defined as a single word, a group of words or a phrase in human language used to describe a class. Different expressions (synonyms) may be used to describe the same class in any given language.

Unit Catalog is defined as an application in the EUPH platform management tool used to manage dataset *Units*.

4 Validity

4.1 Area

The of validity of this document is limited to the process of standardisation (definition, translation) of classes in dataset *Dictionary* and units in dataset *Units*.

4.2 Time

The period of validity of this SOP begins on 2024-02-07 and ends with its revocation.

4.3 Related Documents

Document type	Number	Name
SOP	001	SOP instructions
SOP	004	Organisation chart
SOP	023	Method validation
WI	007	Citation management
WI	014	Dictionary management
WI	015	Translation management
WI	016	Dictionary application
WI	022	Discussion management

4.4 Related Tasks

Persons who have been assigned one of the following tasks, as defined in SOP-004, assume a responsibility within this SOP:

- Data standardisation

5 Procedure

5.1 Responsibilities

- The person responsible for data standardisation according to SOP-004 implements and maintains all processes described in this SOP.
- The task of carrying out certain parts of this process can be transferred to another person.

5.2 Classes

5.2.1 Software tool

- The process of standardisation of classes is carried out with the application *Vocabulary*, accessed via the EUPH Application Module.

5.2.2 Creation

5.2.2.1 Scope

- WI-007, WI-014 and WI-016 are used to execute all processes defined in section 5.2.2.
- The process defined in section 5.2.2 includes the creation, update and deletion of a class.

5.2.2.2 Assignment of draft status

- The status of a class is set to *Draft*.

5.2.2.3 Assignment of a name to a class

- Unique names of a class, fulfilling the following criteria, mut be provided:
  - If a term from an existing, external standard is used as class on the EUPH, the name and its spelling as well as its definition should be adopted in compliance with the rules laid down in this section.
  - The English language must be used for class names.
  - A class name must be a single word (*e.g.* noun, verb, adjective/adverb) or, if not possible, a groups of words (*e.g.* compound noun, compound verb) or, if not possible, a phrase.
  - A class name must be as brief as possible, linguistically correct and contain all essential information to distinguish it from other classes.
  - Words of class names, with the exception of abbreviations, acronyms and proper nouns, should start with lowercase letters.
  - If the name of a class is represented by commonly accepted abbreviations or acronyms, they can be placed within round brackets after the name of the class.
  - If it is unavoidable that two classes are given the same name, they should be supplemented by the shortest possible and generally understandable explanation within square brackets after the name of the class.

5.2.2.4 Association to a specific dataset

- A class may be associated with an existing dataset hosted on the EUPH if the provider of this dataset has an understandable interest in defining the term according to his/her own criteria.

- A class can only be associated with a specific dataset if the superclass is associated with the same dataset, with the exception of classes acting as root classes for a given set of classes related to a dataset.

5.2.2.5 Ontology design

- Each class must be assigned to one of the existing domain classes, depending on which domain class definition is most relevant in the context of the class to be integrated.
- Within a given domain, the concepts of the Basic Formal Ontology (BFO) [1, 2] must be implemented.
    - Subclasses should at least be grouped by “occurrent entities” (entities that unfold through time and do not exist wholly at a single moment, such as processes) and “continuant entities” (entities that persists through time while maintaining their identity, such as objects).
        - Continuant entities should be divided into “independent entities” of a specific domain (entities that are not a property of another entity nor do they depend on another entity), “specifically dependent entities” of a specific domain (entities that are properties of other entities or depend on at least one other entity and that that may not be copied between a number of bearers) and “generically dependent entities” of a specific domain (entities that are properties of other entities or depend on at least one other entity and that may be copied between a number of bearers)
            - For the purpose of improving clarity, independent entities may be divided into “material entities” of a specific domain (entities that always have some portion of matter as a continuant part) and “immaterial entities” of a specific domain (entities that never have any portion of matter as a continuant part).
                - For the purpose of improving clarity, material entities may be divided into “objects” of a specific domain (entities that are neither a collection of disjoint self-standing single objects, nor a proper part of an object), “fiat object parts” of a specific domain (entities that are not a collection of disjoint self-standing single objects, but a proper part of an object), and “object aggregates” of a specific domain (entities that are a collection of disjoint self-standing single objects).
                - For the purpose of improving clarity, immaterial entities may be divided into “spatial regions” of a specific domain (entities that are a spatial projection of a portion of spacetime, but whose location is not determined relative to some material entity), “sites” of a specific domain (entities whose location is

determined relative to some material entity that has three dimensions), and "continuant fiat boundaries" of a specific domain (entities whose location is determined relative to some material entity that has zero, one, or two dimensions).

- For the purpose of improving clarity, spatial regions may be further subdivided into "zero-dimensional spatial regions" of a specific domain (entities), "one-dimensional spatial regions" of a specific domain (entities that have one dimension), "two-dimensional spatial regions" of a specific domain (entities that have two dimensions), and "three-dimensional spatial regions" of a specific domain (entities that have three dimensions).

- For the purpose of improving clarity, specifically dependent continuant entities may be divided into "realisable entities" of a specific domain (entities that require some further process to be realised) and "quality entities" of a specific domain (entities that do not require any further process to be realised).
  - For the purpose of improving clarity, realisable entities may be divided into "dispositions" of a specific domain (entities that do not exist, because the bearer is in a specific situation) and "roles" of a specific domain (entities that exist because the bearer is in a specific situation).
    - For the purpose of improving clarity, dispositions may be further subdivided into "functions" of a specific domain if the entity exists (by evolution or design) to realise specific processes.
  - For the purpose of improving clarity, quality entities may be further subdivided into "relational qualities" of a specific domain if the quality entity is specifically dependent on two or more non-identical entities.

- For the purpose of improving clarity, occurrent entities may be further subdivided into should be divided into "processes" (entities that inhabit the universal spatiotemporal region and have a proper temporal part), process boundaries (entities that inhabit the universal spatiotemporal region but have no proper temporal part), "spatiotemporal regions" (entities that are a portion of the universal spatiotemporal region) and "temporal regions" (entities that are a projection of the universal spatiotemporal region).
  - For the purpose of improving clarity, temporal regions may be divided into "zero-dimensional temporal region" of a specific domain (temporal regions that have

zero dimensions) and “one-dimensional temporal region” of a specific domain (temporal regions that have one dimension).

- If it is necessary to structure classes within a domain for the purpose of improving clarity the elements of another ontology should be used, based on its:
    - relevance to the respective purpose;
    - frequency of use;
    - compliance with FAIR principles [3].
- Each class must be linked to at least one superclass so that the class is a subset of the superclass and the relationship is of the type “subclass of”.
    - The vocabulary must be structured in such a way that there are as few “subclass of” relationships as possible for the individual classes.
    - The relationship between a class and a superclass should be describable with the expression “is a”.
- Each class can be linked to an unlimited number of other classes via an unlimited number of additional, predefined functional relationships, thereby forming an axiom, provided that the definition of the relationship corresponds exactly to the meaning of the axiom.
    - The relationship “member of” may only be used for relationships between a component and an aggregate of these components (such as a collection).
    - The relationship “part of” may only be used for relationships between a component and a single integrated whole in which each component has a function.

5.2.2.6 Analogies

- The creation of classes with the same definition should be avoided.
- In the following cases it is permissible to create classes with an equivalent definition:
    - If there is a justified technical reason why a term associated with a specific dataset or data provider must be included in the vocabulary. It this case, the class associated with a specific dataset or a specific data provider is defined as “analogous to” any existing class.
    - If a class with the same definition exists in an external ontology. In this case the class is defined as “analogous to” the class in the external ontology.

5.2.2.7 Approval of a class

- A new class is released when it complies with all requirements previously defined.
- The status of a class is set to *Approved*.
- The validity of a class starts at the date and time of its release.

5.2.2.8 Update of a class

- If any change in the parameters of an approved class has to be applied, the manager starts an update process of a class.
- The validity of an existing class ends at the date and time at which an update with the same UID is released.
- A class is updated and released as described in section 5.2.2.2 through 5.2.2.6.

5.2.2.9 Deletion of a class

- The manager of the process may delete a given class if keeping the class in the vocabulary is no longer considered useful, if it is not a superclass of another class and if it does not act as analogue of another class.
- The status of a class is set to *Expired*.
- The validity of a class ends at the date and time of its deletion.

5.2.3 Definition

5.2.3.1 Scope

- WI-007, WI-014, WI-016 and WI-022 are used to execute all processes defined in section 5.2.3.
- The standardisation process defined in section 5.2.3 definition of a class.

5.2.3.2 Assignment of draft status

- The status of a class definition is set to *Draft*.

5.2.3.3 Definition

- The English language must be used for class definitions.
- A definition must be non-circular, concise (it contains no redundant elements such as lists of examples or explanations of usage), linguistically correct and contain all essential information to distinguish it from other classes.
- Any axiom defined for the class should be incorporated into the definition in language that is as short, simple, easy to understand and correct, and the individual parts of the axiom should be linked to the URI of its elements.
- References should be attached to the definition confirming the correctness of the axiom.
- A term within the definition of a class, which exists as class in the EUPH vocabulary, should be linked with the URI of this class.
- A term within the definition of a class, which is defined by an external resource may be linked with the URL of this external resource.

- Any commonly recognised unique identifier of a class should be added to the description of the class, preceded by a text that is linked to the URI of the relationship “denoted by”. If more than one identifier is used, the identifiers should be organised in an unordered list.

5.2.3.4 Citation

- Text that is taken verbatim from an external source must be edited as such in the definition of a class.
- External sources used for the definition of the class must be specified in the list of references.

5.2.3.5 Approval of a definition

- A definition is released when the manager of the discussion process, as defined in section 5.2.2.1, in collaboration with those participating in the process, has decided that the discussion process on its definition has produced a satisfactory result.
- The status of a definition is set to *Approved*.
- The validity of a definition starts at the date and time of its release.

5.2.3.6 Update of a definition

- If any change in the parameters of an approved definition has to be applied, the manager starts an update process of a class.
- The validity of an existing definition ends at the date and time at which it is released.
- A class is updated and released as described in section 5.2.3.2 through 5.2.3.5.

5.2.3.7 Deletion of a definition

- The manager of the process may delete a definition if a better definition has been approved.
- The validity of a definition ends at the date and time of its deletion.

5.2.4 Translation

5.2.4.1 Scope

- WI-007, WI-015, WI-016 and WI-022 are used to execute all processes defined in section 5.2.3.
- The translation process includes the creation, update and deletion of expressions for classes in human languages, as defined in section 5.2.3.

5.2.4.2 Assignment of draft status

- The status of an expression is set to *Draft*.

5.2.4.3 Approval of a translation

- An expression is released when the manager of the standardisation process, as defined in section 5.2.2.1, in collaboration with those participating in the process, has decided that the discussion process on its definition and its association with other expressions has produced a satisfactory result to which a majority of the parties agreed upon.
- The status of an expression is set to *Approved*.
- The validity of an expression starts at the date and time of its release.

5.2.4.4 Update of a translation

- If any change in the parameters of an approved expression has to be applied, the manager starts an update process of an expression.
- An expression is updated and released as described in section 5.2.4.2 through 5.2.4.3.

5.2.4.5 Deletion of a translation

- The manager of the process may delete a given expression, if keeping the class in the vocabulary is no longer considered useful, and if it is not assigned as synonym by any other expression.

5.2.5 Public participation

5.2.5.1 Inclusion into the process

- WI-016 is used to execute all processes defined in section 5.2.5.
- Users with an approved account on the EUPH may participate in the process of definition and translations of classes.

5.2.5.2 Exclusion from the process

- Users who severely interfere with the standardisation process or violate legal regulations must be denied access to the EUPH.

5.3 Descriptors

5.3.1 Global descriptors

5.3.1.1 Application

- The process of standardisation of global descriptors is carried out with the application *Descriptors*, accessed via the EUPH Platform Management tool.

5.3.1.2 Initiation of the process

- The person responsible for the management of this processes may at any time initiate the creation, update or deletion of a global descriptor.
- The person responsible for the management of this process creates a global descriptor not contained in the list of Descriptors if a formal request has been made.

5.3.1.3 Creation of a global descriptor

- A global descriptor with an equivalent definition of an existing descriptor may not be created.
- If the global descriptor is part of an existing, external standard, the name and its spelling as well as tits definition should be adopted in compliance with the rules laid down there.
- If the global descriptor term is not part of an existing, external standard, a unique name fulfilling the following criteria, mut be provided:
    - The English language must be used for global descriptor names.
    - A global descriptor name must be a single word (*e.g.* noun, verb, adjective/adverb) or, if not possible, a groups of words (*e.g.* compound noun, compound verb) or, if not possible, a phrase.
    - Characters from the 8-Bit Universal Coded Character Set Transformation Format (UTF-8) must be used, except for blank spaces.
    - A global descriptor name must be as brief as possible and contain all essential information to distinguish it from other global descriptors.
    - Lowercase letters must be used, with the exception of names consisting of more than one word, in which case the second and each subsequent word must begin with a capital letter.
- A global descriptor must be linked to a namespace abbreviation.
    - If the global descriptor is part of an existing, external standard which comes with a namespace abbreviation, this namespace abbreviation must be used.
- A global descriptor must be linked to a class.
    - A global descriptor receives the definition of the class to which it is linked.

- If the class associated which a global descriptor is part of an existing standard and the term of this standard has an URI/IRI, then the resource identifier of this class must be used for the descriptor.

- If a global descriptor shall be used with a particular unit, then an existing unit from the Unit Catalog may be associated with the descriptor.
- If a global descriptor shall be used with a particular data type, then an existing data type may be associated with the descriptor.
- The validity of a descriptor starts at the date and time of its release.

5.3.1.4 Update

- The validity of an existing descriptor ends at the date and time at which an update with the same UID is released.
- A descriptor is updated as described in section 5.3.1.1.

5.3.1.5 Deletion

- The status of a descriptor is set to *Deprecated*.
- The validity of a descriptor ends at the date and time of its deletion.

5.4 Units

5.4.1 Management of the Units Catalog

5.4.1.1 Application

- The process of standardisation of units is carried out with the application *Units*, accessed via the EUPH Platform Management tool.

5.4.1.2 Initiation of the process

- The person responsible for the management of this processes may at any time initiate the creation, update or deletion of a unit in the Units Catalog.
- The person responsible for the management of this process creates a unit not contained in the Units Catalog if a formal request has been made.

5.4.1.3 Creation

- The validity of a unit starts at the date and time of its release.
- Algorithms for the conversion of a given unit to a base unit are validated according to SOP-023 (Method validation).

5.4.1.4 Deletion

- The status of a unit is set to *Deleted*.
- The validity of a unit ends at the date and time of its deletion.

6 References

[1] ISO/IEC 21838-2:2021, *Information technology - Top-level ontologies (TLO) - Part 2: Basic Formal Ontology (BFO),* Geneva, Switzerland: International Organization for Standardization / International Electrotechnical Commission, 2021.

[2] C. Emeruem, S. Wang and M. Keet, "BFO 2.0 Classifier," 2021. [Online]. Available: https://bfo-classifier.github.io. [Accessed 28 02 2025].

[3] GFISCO, "FAIR Principles," GO FAIR International Support and Coordination Office, [Online]. Available: https://www.go-fair.org/fair-principles/. [Accessed 24 04 2023].

7 Annexes

There are no annexes in the current document.