Standard Operating Procedure

Number: **005**

Name: **UID assignment**

Title: Assignment of Unique Identifiers (UID) to data assets, classes and data providers on the EU Pollinator Hub

Category: Data management

Status: Approved

Version: 1.0

Date: 2023-06-30

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 Related Documents 6
4.3 Time 6
4.4 Related Tasks 6
5 Procedures 7
5.1 Responsibilities 7
5.2 UID of data assets 7
5.3 UID of classes 9
5.4 UID of data providers 10
6 References 11
7 Annexes 12

1 Document history

1.1 Current Release

1.1.1 Drafted

Submitted on 2023-06-05 by Michael Rubinigg, quality manager.

1.1.2 Reviewed

Reviewed on 2023-06-30 by Noa Simòn Delso, general manager.

1.1.3 Approved

Approved on 2023-06-30 by Michael Rubinigg, quality manager.

1.2 Change history

n.a.

2 Abbreviations

EU	European Union
EUPH	EU Pollinator Hub
n.a.	not available
SOP	Standard Operating Procedure
UID	Unique Identifier

3 Introduction

This document describes the procedure of issuing persistent Unique Identifiers (UID) to information assets on the EU Pollinator Hub (EUPH).

3.1 Definitions

Data provider is a physical or juridical person having signed a data provider agreement with the EU Pollinator Hub.

Data table is a related set of data elements forming a contextual unit.

Dataset is a related set of data tables forming a contextual unit.

Descriptor is a unique name for an attribute of a set of values (also referred to as columns) in a data table. Its uniqueness may be limited to a dataset (local descriptor) or apply to all datasets on the EUPH (global descriptor).

Unique identifier (UID) is an identifier, which is guaranteed to be unique within the EU Pollinator Hub and remains with the asset throughout its entire life cycle.

4 Validity

4.1 Area

The of validity of this document is limited to the assignment of persistent Unique Identifiers (UID) to information assets on the EU Pollinator Hub.

4.2 Related Documents

Document type	Number	Name
SOP	001	SOP instructions
SOP	004	Organisation chart

4.3 Time

The period of validity of this SOP begins on 2023-06-30 and ends with its revocation.

4.4 Related Tasks

- Software development

5 Procedures

5.1 Responsibilities

- The person responsible for software development according to SOP-004 implements and maintains all processes described in this SOP.

5.2 UID of data assets

5.2.1 Assignment of a UID to a data asset

5.2.1.1 Creation of an UID for a data asset

- The UID of a data asset is composed of three blocks, separated by a baseline dot.
  - The first block, referred to as dataset block, codes the dataset.
  - The second block, referred to as table block, codes the data table.
  - The third block, referred to as descriptor block, codes the descriptor.
- The dataset, data table and descriptor block are composed of two components:
  - Letters, consisting of five capital letters, which are retrieved from the name of a data asset as follows.
    - Non-Latin characters are transliterated to Latin characters.
    - All numbers, blank spaces and special characters are removed from the name to produce a crude extract.
    - If the crude extract contains less than five letters, then letters are added from the English alphabet in increasing alphabetic order, starting with the letter A, until the crude extract contains five letters.
    - If the crude extract contains more than five letters, then vowels are successively removed, starting at the left side of the string, until either the string contains five letters, or the fifth letter of the string is reached. All characters following the fifth letter are discarded.
  - A numeric component, consisting of the record identifier of the data asset.
  - The alphanumeric and numeric components of a block are merged in the order specified here.
- If a data asset is not associated to another asset, then blocks representing this asset receive the value 0.

5.2.1.2 Integration of an UID of a data asset into standardised communication protocols

- The UID of a data asset must be integrated in the URL.

5.2.1.3 Integration of an UID of a data asset into metadata assets

- The UID must be added to each metadata asset which is part of a given data asset.

5.3 UID of classes

5.3.1 Assignment of a UID to a class

5.3.1.1 Creation of an UID for a class

- The UID of a class is referred to as EUPH code. It is the unique record identifier of a class and is assigned when a new class is created.

5.4 UID of data providers

5.4.1 Definitions

5.4.2 Assignment of an UID to a data provider

5.4.2.1 Creation of a UID for a data provider

- A unique identifier (UID) is created, which contains a concise and accurate description of the data provider it describes, and which is sufficiently distinct from existing UID of data providers as to avoid confusion.
- The UID of a data provider shall consist of a maximum of 255 alphanumeric characters, including the following signs:
  - (space, ASCII character code 32)
  - _ (underscore, ASCII character code 95)

6 References

There are no sources in the current document.

7 Annexes

There are no annexes in the current document.