Standard Operating Procedure

Number: **008**

Name: **Data cleansing**

Title: Procedure for the cleaning of data to be integrated on the EU Pollinator Hub

Category: Data management

Status: Approved

Version: 1.0

Date: 2024-02-07

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 Detection and reporting of inconsistencies 7
5.2 Resolution of inconsistencies 8
6 References 9
7 Annexes 10

1 Document history

1.1 Current Release

1.1.1 Drafted

Submitted on 2023-06-29 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 2024-02-07 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

3 Introduction

One of the aims of the EUPH is to store cleaned analytical datasets. This document describes how to clean dataset for this purpose. This includes the detection of inconsistencies as well as the fixing of incorrectly formatted data, the correction of incorrect data, removal of duplicates, replacement of missing data.

3.1 Definitions

Inconsistencies are data which do not meet the requirements of the EUPH, or which are not in agreement with the expectations on the data.

4 Validity

4.1 Area

The of validity of this document is limited to the cleansing (comprising detection, reporting and resolution) of data in circumstances that are not defined by other SOP.

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	Document ID	Title
SOP	001	SOP instructions
SOP	002	WI instructions
SOP	003	DR instructions
SOP	004	Organisation chart
SOP	006	Dataset preparation
SOP	007	Data profiling
SOP	015	Peer reviewing
SOP	017	Dataset integration
SOP	018	Data exploration
SOP	019	Data visualisation
SOP	020	Issue 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 preparation

5 Procedure

5.1 Detection and reporting of inconsistencies

5.1.1 Responsibilities

- The person responsible for data preparation, as defined by SOP-004, or a person appointed by him/her for this purpose, is responsible for the management of the processes defined in this section.
- The person responsible for data preparation, as defined by SOP-004, assures that all individuals who carry out activities covered by this SOP are familiar with the processes described in the SOP.

5.1.2 Detection of inconsistencies

5.1.2.1 Inconsistencies detected in processes covered by Standard Operating Procedures

- During the execution of processes related to data management, in particular Data preparation, as according to SOP-006, Data profiling according to SOP-007, Peer reviewing according to SOP-015, Dataset integration according to SOP-017, Data exploration according to SOP-018 and Data visualisation according to SOP-019 the person responsible for the execution of the respective process has to actively look for inconsistencies in the data.

5.1.3 Reporting of inconsistencies

5.1.3.1 Reporting to the data provider

- The person responsible for the data preparation, as defined by SOP-004, informs the data provider in due time of the inconsistencies that were detected.

5.1.3.2 Reporting of inconsistencies not defined by Standard Operating Procedures

- Inconsistencies are reported as defined by SOP-020.

5.2 Resolution of inconsistencies

5.2.1 Responsibilities

5.2.1.1 Resolution

- The data provider is responsible for the resolution of the reported inconsistencies.
- The data provider may request assistance by the person responsible for data preparation, as defined by SOP-004.

5.2.1.2 General work instructions for data cleansing methods

- The person responsible for data preparation, as defined by SOP-004, is responsible for releasing and revisioning of Work Instructions containing general methods for data cleansing. These methods are published in the product directory of the EUPH.

5.2.2 Reporting

- All procedures used to resolve inconsistencies and the amount and type of data affected by these procedures are specified in the Dataset Report.

6 References

There are no sources in the current document.

7 Annexes

There are no annexes in the current document.