Work Instructions

Number: **023**

Name: **Data profiling**

Title: Using the data profiling tool on the EU Pollinator Hub

Status: Approved

Version: 1.0

Date: 2024-01-29

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

1.2 Change history 3

2 Abbreviations 4

3 Introduction 5

4 Validity 6

4.1 Area 6

4.2 Time 6

4.3 Related Documents 6

5 Procedure 7

5.1 Data profiling 7

6 References 9

7 Annexes 10

1 Document history

1.1.1 Drafted

Submitted on 2024-01-28 by Michael Rubinigg, quality manager.

1.1.2 Reviewed

Reviewed on 2023-01-29 by Noa Simòn Delso, general manager.

1.1.3 Approved

Approved on 2023-01-29 by Michael Rubinigg, quality manager.

1.2 Change history

n.a.

2 Abbreviations

EU	European Union
EUPH	EU Pollinator Hub
n.a.	not available
WI	Work Instruction

3 Introduction

This documents contains instructions for using the data profiling tool on the EU Pollinator Hub.

4 Validity

4.1 Area

The of validity of this document is limited to the use of the data profiling tool on the EU Pollinator Hub.

4.2 Time

The period of validity of this document begins on YYYY-MM-DD and ends with its revocation.

4.3 Related Documents

Document type	Document ID	Name
WI	024	Access application
WI	027	Upload data

5 Procedure

5.1 Data profiling

5.1.1 Perform profiling on data in an existing data table

5.1.1.1 Log into the EUPH account

- Log in to the EU Pollinator Hub application as described in WI-024.

5.1.1.2 Select data provider

- Select the data provider for which you intend to create the dataset in the dropdown on the right-hand side of the header.

5.1.1.3 Select the data provider's data catalog

- Select ***My Data Catalog*** in the menu.
- Select the dataset from the Data Catalog where the data table on which data profiling should be performed is located.

5.1.2 Create a new data table

- Proceed as described in WI-027.

5.1.3 Profiling

5.1.3.1 Access profiling data

- Select the data table from which the profiling information should be acquired.
- Got to the ***Profiling*** section of the table dashboard.
- Select  to add or remove specific columns from the dashboard by checking the appropriate checkbox.
- Select the tab ***Metadata*** in the table dashboard to receive information on the structural analysis for each column of the data table.
- Select the tab ***Descriptive measures*** in the table dashboard to receive information on the content analysis for each column of the data table.
- Select the tab ***Data quality measures*** in the table dashboard to receive a graphical summary on the most important quality measures for the content for each column of the data table.
- Select the tab ***Profiling charts*** in the table dashboard to receive a graphical summary on the most important results from structural and content analysis for each column of the data table.

- Select the tab ***Sample data from file*** in the table dashboard to receive a small sample (24 records) of the data in the data table.

6 References

Use IEEE 2006 reference style

There are no sources in the current document.

7 Annexes

There are no annexes in the current document.