



# IUCLID Data Extractor: Release Notes

## Legal Notice

The information in this document does not constitute legal advice. Usage of the information remains under the sole responsibility of the user. The European Chemicals Agency does not accept any liability with regard to the use that may be made of the information contained in this document.

**Title:** IUCLID Data Extractor: Release Notes

**Issue date:** January 2023

**Language:** en

Text Analytics is developed by the European Chemicals Agency.

IUCLID 6 is developed by the European Chemicals Agency in association with the OECD.

© European Chemicals Agency, 2023

Reproduction is authorised provided the source is fully acknowledged in the form

“Source: European Chemicals Agency, <http://echa.europa.eu/>”, and provided written notification is given to the ECHA Communication Unit ([publications@echa.europa.eu](mailto:publications@echa.europa.eu)).

If you have questions or comments in relation to this document, please send them to ECHA via the information request form at the address below, quoting the reference and issue date given above:

<http://echa.europa.eu/contact>

## European Chemicals Agency

Mailing address: P.O. Box 400, FI-00121 Helsinki, Finland

Visiting address: Telakkakatu 6, FI-00150 Helsinki, Finland

# Table of Contents

- Table of Contents .....i**
- 1. Introduction .....1**
- 2. Version 1.10.0 (January 2023) .....1**
- 3. Version 1.9.0 (April 2022).....1**
- 4. Version 1.8.0 (January 2022) .....2**
- 5. Version 1.5.1 (09 February 2021) .....5**
- 6. Version 1.4.0 (19 August 2020) .....5**
- 7. Version 1.3.1 (29 April 2020) .....6**

## 1. Introduction

Please also refer to the [Frequently Asked Questions](#) on the IUCLID website for known issues, and if relevant, workarounds and advice on how to handle such issues.

The release notes highlight changes compared to the previous published version of IUCLID Data Extractor.

In the next sections, new features and improvements are identified by the labels **NEW** and **IMP** respectively. Bug fixes are labelled **FIX**.

## 2. Version 1.10.0 (January 2023)

This version is a cumulative update to the previous version 1.9.0. Changes since the previous version are:

**FIX** Handling extractions that contain attachments with the same file name

Ref. 657388

A solution has been provided to allow extractions that contain attachments that have the same filename. A system of renaming the files has been created, which is described in the user manual.

**IMP** Revisit the result aggregation algorithm

Ref. 672217

The performance of the extractor has been improved. Particularly the normalisation step in the extraction process has become faster.

## 3. Version 1.9.0 (April 2022)

This version is a cumulative update to the previous version 1.8.0. Changes since the previous version are:

**IMP** Verify the format of the input UUID

Ref. 249389

The extractor verifies that the format of the input UUID is valid so that only UUIDs with the correct UUID format are added in an extraction job

**IMP** Normalize reset button functionality upon creating/loading/importing an extraction

Ref. 653949

The reset button now reverts the configuration page of an extraction job to the last saved status. If the configuration page has not been saved by the user, the reset button will clear the configurations altogether. We are planning to further improve the revert and clear all functions in a future update of the extractor, by splitting them in two buttons.

**IMP** Extract a report that contains the successfully and unsuccessfully processed dossiers

Ref. 656663

The user interface of a completed extraction job had already in the past allowed the user to view the input UUIDs that had been successful, not found or erroneous in the completed extraction job. The download button had though retrieved all the input UUIDs as a single list and without an indication if a UUID had been successful or not. Now the downloaded UUID list features an additional column that indicates if the UUID was successfully processed or not.

**FIX** Change default replacement delimiter character in DE default configuration

Ref. 655965

Fixes some inconsistencies in the default character delimiter settings when the user configures a new extraction job. The default column delimiter is Tab.

If the user prefers to use any other column delimiter, the user should choose “Other:” and type the preferred delimiter. NB! The use of “Other:” as delimiter writes out the delimiter in the extracted data. For instance, “\t” would not generate a tab separation but would instead write out “\t”. The user should then read the extracted data in a third-party application with this character set as column delimiter.

If a user notices that the required column delimiter is already in use for some other meaning in the IUCLID data to be extracted, the character may be replaced in the IUCLID data using the feature *Replace delimiter with*. Thus, the preferred column delimiter can be used without changing the meaning of the data.

We are looking further into improving the delimiter settings, including the new lines delimiter.

**FIX** Error message exposed through the UI although extraction is performed with no problems

Ref. 655187

Fixes problems where upon launching a new job, the extractor displays an error message, even though the job has started and runs correctly.

**FIX** Only the last picklist value is exported in case of multiple options

Ref. 655189

Fixes an error where the extractor would only extract the last value of a multiple selection pick-list in a IUCLID field.

## 4. Version 1.8.0 (January 2022)

This version is a cumulative update to the previous version 1.5.1. Changes since the previous version are:

**IMP** Enable compatibility with IUCLID 6 v6

Ref. 638185

**NEW** Choice of a delimiter in the extraction output data in the normalised extraction mode

Ref. 645082

The version introduces a new set of options for normalised output. It is under *Data settings* which is used when configuring a new extraction job. The user has full choice to define a preferred *Column delimiter* in the extracted data. To avoid instances of the column delimiter in the text values being interpreted as a delimiter in the output, they can be replaced with something else, using the option *Replace delimiter with*: For example, LF characters may be replaced with “\n”, or a space character.

The selection panel *Data settings* has a selection box labelled *Remove HTML tags from rich text fields*. By selecting this box, the user can remove any HTML code from the value in richtext field. Such code may not be visible in the in the web interface of IUCLID. Removing the HTML code can simplify processing of the output. An example in IUCLID of a field into which HTML can be entered is *Any other information including tables*, which is present under many different IUCLID endpoints.

**Known issue**

The default values for *Column delimiter* and *Replace delimiter with*: are set to the same value. This inactivates the *Save* button for new extractions, and the *Dashboard*. A new extraction can be saved by first changing one of the values in *Data settings* under normalised output for that extraction. For example, change *Column delimiter* to <Tab>. A more general and convenient solution is to change the default values on the *Dashboard*, which also activates its *Save* button. More detail is given in the installation manual. This is due to be fixed in version 1.8.1.

Ref. 655965

**IMP** Allow both JSON and Normalised output in the same extraction job

Ref. 249181

Previously the user had to choose between JSON and Normalised extraction formats in the data settings panel. Now there is a third option, JSON and Normalised, which enables both extraction formats in the same extraction job

**IMP** IUCLID path and label columns renamed in the CSV file

Ref. 645481

In order to increase readability of the raw extraction output files, some columns have been renamed. Columns displaying only IUCLID or label path information, have been renamed to “IUCLID path” and “Label path” respectively.

Examples:

FLEXIBLE\_RECORD.SubstanceComposition.GeneralInformation.JustificationForDeviations.path.hey, has been renamed: **IUCLID path**

FLEXIBLE\_RECORD.SubstanceComposition.GeneralInformation.JustificationForDeviations.path.Label, has been renamed: **Label path**

**IMP** New Data Extractor installer package for IUCLID version 6.0.0

Ref. 649129

**IMP** Refresh the data extractor home page at periodic intervals

Ref. 635654

Previously the user had to refresh the home page in order to monitor the progress of the extraction job. Now the home page refreshes regularly by itself.

**IMP** Enable trust store configuration for SSL calls to IUCLID.

Ref. 638282

**IMP** Update to Wildfly version 21.

Ref. 249548

**IMP** Update to Java 11.

Ref. 249533

**FIX** The attachments download file was not created. This has been fixed.

Ref. 249531

**FIX** The delimiter for the file "attachments\_matrix.csv" was always comma. This file lists the attachment files of an extraction job and it is always provided when such files are extracted. Now the column delimiter in the file is the same as in the overall extraction job.

Ref. 652831

**FIX** The progress percentage indicator for an on-going extraction job showed sometimes negative values. This has been fixed, the indicator proceeds now from 0 to 100%.

Ref. 249480

#### **Known errors under assessment:**

Only the last picklist value is exported in case of multiple options

Ref. 654039

IUCLID fields with picklists that allow multiple selection, are not fully extracted. Only the last value is extracted. Such picklists can be found for instance in IUCLID under REACH at:

- section 3.5.3 -> Uses at industrial sites -> Technical function of the substance during use
- section 13 -> CSR contains

## 5. Version 1.5.1 (09 February 2021)

**IMP** Line breaks and tabs are removed from extracted text data. The following actions are carried out:

1. Replacement of line breaks in Windows format (`[\r\n]+`) by a single space character each.
2. Replacement of tab characters (`[\t]+`) by a single space each
3. Quotation marks are not added around columns. Quotation marks that are part of a text value are unaffected.
4. Columns are separated by tabs.

Ref: IUCDEXT-372

**IMP** IUCLID Data Extractor is compatible with the format of IUCLID 6 v5.

Ref: IUCDEXT-483

**IMP** Support has been added for Oracle 19c.

Ref: IUCDEXT-490

**FIX** A definition of an extraction exported from v1.4.0 as a file in JSON format, cannot be imported back in to v1.4.0. This has been fixed.

Ref: IUCDEXT-504

**FIX** When large sets of fields are selected for many targets (> 20 K) and the output is Normalised, OutOfMemory errors occur. This has been fixed and tested for large sets of data.

Ref: IUCDEXT-511

**FIX** Where the data submission type was part of a group of types, not all header data were exported. For grouped submission types, for example REACH group 2, a user is now able to select the required dossier header fields, which are extracted for all the submission types belonging to that group.

Ref: IUCDEXT-512

**FIX** Where there were multiple rows within a literature reference, when the *Normalised* format was used, only one of the rows was extracted. JSON was not affected. This has been fixed.

Ref: IUCDEXT-517

## 6. Version 1.4.0 (19 August 2020)

**IMP** Group identical submission trees under TOC picklist. When selecting fields, the TOC used in the display of the fields can be selected from the TOCs available. The selection of TOCs has been made more compact by grouping together identical TOCs.

Ref: IUCDEXT-327, IUCDEXT-329

**IMP** How to handle SCIP article notification in Data Extractor. The entity of type *Article* which is used for SCIP notification dossiers, has been added to field selection under *Other*.

Ref: IUCDEXT-476

**IMP** Retrieve logged in user roles without the use of IUCLID admin user. During installation of Data Extractor (step 7) there is no longer any need to enter the credentials for the IUCLID user SuperUser.

Ref: IUCDEXT-489

## 7. Version 1.3.1 (29 April 2020)

**FIX** Administration dashboard page. Non-admin users could see the admin dashboard. This has now been restricted to only admin users.

Ref. IUCDEXT-428

**FIX** Extracted outcome lacks some IUCLID fields. This was identified as occurring in cases where documents contain nested blocks and tables with many rows. This has been fixed.

Ref. IUCDEXT-467

**FIX** H2 sql error during normalisation. This is a non-blocking error that can occur where the extraction has a large number of targets. A warning message has been added.

Ref. IUCDEXT-441

**FIX** IUCLIDDE Industry - Login to the extractor is possible with incorrect password. This has been fixed by the introduction of IDP in to the authentication mechanism.

Ref. IUCDEXT-479

**FIX** Label and value is missing from json and csv.

Ref. IUCDEXT-429

**IMP** Mechanism to upload the manual or to link a manual in an external link. The functionality to upload a help file and/or provide a URL for it has been added to the admin dashboard.

Ref. IUCDEXT-408

**IMP** Security assessment findings. The security of the IUCLID Data Extractor was reviewed. A number of improvements were introduced including the introduction of IDP, which also fixed the bug as referenced above IUCDEXT-479.

Ref. IUCDEXT-443

**IMP** Support IUCLID IDP connection in the industry installation process. IDP was introduced, as referred to above in IUCDEXT-479 and IUCDEXT-443.

Ref. IUCDEXT-451

**IMP** Impact assessment of IUCLID changes (6.4.0). Added the dossier type SCIP.

Ref. IUCDEXT-444