



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
Version	v1.14.0 for IUCLID 6 v8
Issue date:	1 st July 2024
Language:	en

IUCLID Data Extractor is developed by the European Chemicals Agency.

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

© European Chemicals Agency, 2024

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).

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 Contentsi**
- 1. Introduction1**
- 2. Version 1.14.0 (July 2024).....2**
- 3. Version 1.13.1 (February 2024)3**
- 4. Version 1.12.1 (December 2023)5**
- 5. Version 1.11.3 (July 2023).....7**
- 6. Version 1.10.0 (January 2023).....8**
- 7. Version 1.9.0 (April 2022).....9**
- 8. Version 1.8.0 (January 2022)11**
- 9. Version 1.5.1 (09 February 2021)14**
- 10. Version 1.4.0 (19 August 2020)15**
- 11. Version 1.3.1 (29 April 2020)16**

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 1.14.0 (July 2024)

This version is a cumulative update to the previous version 1.12.1. The main changes since the previous version are:

IMP Compatibility with IUCLID 6, v8

Ref. 944688

The Data Extractor is now compatible with the latest changes of the API that makes IUCLID data available. Please use the latest Data Extractor with the latest IUCLID version for continued data extraction use.

IMP Improvements for when using an H2 Database

Ref. 954249

A more in-depth improvement has been made to the already previously provided fix to the bug that caused memory problems and errors when using an H2 database.

NEW First version with option to extract the IUCLID structure of a working context

Ref. 952663

Background: Data analysts who want to build their post-processing workflows for the extracted flat data, may want to see all the possible IUCLID fields that will result in columns and files, even though the current target dossiers or datasets would not yield them in the output, because there was no data to extract in them. Extracting empty columns and files may not be relevant for a one-off extraction job that will not be repeated later on with other target UUIDs, but as soon as the same IUCLID fields in the same working context are extracted for other target UUIDs, which contain data in the previously data-empty fields, it becomes interesting to build a post-processing workflow that can handle these additional columns and files. If the data analyst can up-front build a post-processing workflow, it will save time and effort for him/her later, because the workflow is already ready for the complete IUCLID working context.

Implementation: In this first version of the implementation the user has the option to include empty IUCLID fields in the traditional extraction outcome, by selecting a dedicated tick-box in the Data settings pane on the job configuration page. When this tick-box is selected, the extraction outcome will include all data-containing and empty IUCLID fields as columns, as long as at least one of the targeted fields contain data for a particular target UUID.

Future development: The second version of the feature is thought to list all of the target UUIDs as rows in the outcome, even if they are without data altogether. It will also be possible to extract the complete IUCLID structure of a working context already with one target UUID, even though it would not contain any IUCLID field data. The development team is also looking in general into the need to align the order of appearance of extracted columns, so that their order would stay consistent between extraction jobs and their order +be based on the order of appearance of the sections in the IUCLID working context.

IMP Improved numbering of the extracted files in the schema.map file

Ref. 945419

The numbering of how the extracted files is now based on the IUCLID hierarchy.

FIX Correctly extract and display IUCLID content with Greek alphabet characters.

Ref. 966705

IUCLID data containing Greek letters is now extracted and displayed correctly.

FIX Correctly extract non-standard inventory numbers (EC numbers)

Ref. 969942

In cases where the inventory number is in non-standard format, the Data Extractor might have failed in the past. Most likely the user encountered this issue with datasets or not yet submitted dossiers, because the format is most likely caught with the IUCLID validation assistant in the submission step. After this fix, the inventory number will not trigger an error, but will be extracted as entered in the IUCLDI file.

3. Version 1.13.1 (February 2024)

This version is a cumulative update to the previous version 1.12.1. The main changes since the previous version are:

IMP Compatibility with IUCLID 6, latest release

Ref. 939074

The Data Extractor is now compatible with the latest changes of the API that makes data available. Please use the latest Data Extractor with the latest IUCLID version for continued data extraction use.

IMP CSV Schema: Improve the assignment of unique identifiers for table numbering

Ref. 932005

The schema file (schema.csv), which the user receives among the extracted files, now has a more intuitive system for assigning unique reference numbers to the extracted flat data files, when they are referred to in the schema. In the previous DE version, the number assignment could have unexpected jumps. The work still continues with the schema file. The aim is that the higher a file is in the IUCLID file hierarchy, the smaller the assigned reference number is. Currently the assignment may not always follow this principle. The map is still useful already, because it in most cases is able sort the extracted files correctly and the user can see how the extracted files pair up and which is the foreign key column that binds the two files.

IMP Loading of existing extraction fails, when the name of the extraction job is very long

Ref. 809479, 936482

A loaded extraction job will no longer start its name with the prefix “Copy of..”. Such a prefix did not give added value to the user and when a job was loaded several times, the prefix resulted in an excessive repetition of this prefix that also affected the performance of the Data Extractor itself. This occurred especially when the free text name was long. Now a loaded job will get the same free text name as its predecessor has, but the unique numerical identifier will still distinguish it from the predecessor. Users are also recommended to use the “Description” field on the configuration page of a new job. This field allows longer free text descriptions of the job than the “Name” field does.

NEW Migration to PostgreSQL process

Ref. 807430, 940339

The Data Extractor can now be installed also with PostgreSQL.

FIX Performance improvements, H2 database connection and Java heap space

Ref. 919883, 918550

Fixes an out of memory related application crash that was reported in the industry version of the Data Extractor.

FIX Login window remains on the screen, even after a successful login step

Ref. 945867

Particularly in the industry version the user was sometimes presented with the login window remaining on the screen, even after a successful login step. A workaround had been to click on the Data Extractor link in the top left corner. This fix removes the erroneous behaviour of the login window.

IMP Extract all the test material composition reference substance inventory identifiers

Ref. 683104

The feature vastly improves the amount and accuracy of the information a user can get of the test material composition reference substance inventory identifiers. Previously the user could only extract the inventory number (EC or list number) and it was always appended by “@EC” for each number.

This data was previously presented in a column called “value”, which was not meaningful for the user. The extracted file called “REFERENCE_SUBSTANCE.Inventory.InventoryEntry.csv” will now contain all the data from the IUCLID field “REFERENCE_SUBSTANCE.Inventory.InventoryEntry”. In the IUCLID view this field merges and presents several different kind of data, separated by “/”-characters. For the data analyst the contents of this field is important and useful, particularly when the different kind of data in this field are accessible separately. The data extractor will now present all of these different kind of data in this field and in separate columns. The columns have derived names from the IUCLID field name, e.g. for the inventory number column “REFERENCE_SUBSTANCE.Inventory.InventoryEntry.numberInInventory”. Like so, the data analyst has a better possibility to process bulk extractions of substances and the knowledge of the reference substance within the test material composition can be included in the data analysis.

4. Version 1.12.1 (December 2023)

This version is a cumulative update to the previous version 1.11.3. Changes since the previous version are:

NEW Generate a schema for the outcome of a normalised extraction job

Ref. 807420

DE now generates a schema file (schema.csv) of the extraction outcome with the purpose to explain the structure of the extracted data, that is, the relationship between the extracted files. The schema does not aim at explaining the IUCLID working context as such. The schema is generated in csv format, so that it can for instance be opened in Excel, where the user can view it by sorting and filtering rows.

NEW User interface improvement to show the status of target UUIDs in a completed job

Ref. 682451

The main page displays now an amended view of the status of the target UUIDs. The user can quickly see how many target UUIDs have been successful in the extraction, or not been found or if some error occurred. Previously the user identified the existence of non-successful target UUIDs, by always opening a completed job. By the introduction of this feature, the user is better informed of such a case and can enter the completed job to examine further or retry the extraction with the failed target UUIDs.

IMP Enabling correctly and informing about partially successful extraction jobs

Ref. 809960

If the user runs an extraction in both JSON and normalised format at the same time, the data extractor will now be able to handle better a rare case where one of the formats would fail. DE will finish the complete job at a reasonable number of extraction attempts and display on the main page a note about which data format was successful and make that data available

IMP Improved attachment file extraction

Ref. 924859, 915744

This is a technical improvement that makes DE delete quicker extracted files in completed jobs, so that they take less space on the disk. Attachment files are as before saved on the disk for possible reoccurring needs of extraction of the same file. Now if the targeted file is not found on the disk, DE will attempt to extract it again, without causing an error.

IMP Performance of reloading an old extraction job

Ref. 656812

The user can benefit a lot from reloading a past extraction, especially if the targeted IUCLID sections can be reused as such in a new job. Reloading was sometimes slow, mainly because it also retrieved the target UUIDs from the previous job. As of v.1.12, DE offers the user the possibility to load a previous job with or without the target UUIDs of the previous job.

5. Version 1.11.3 (July 2023)

This version is a cumulative update to the previous version 1.10.0. Changes since the previous version are:

NEW Generate and modify a sequentially executing job queue

Ref. 682132, 805086, 695875, 692179, 682209, 682600 and related 692912

The user can now build a queue of upcoming extraction jobs and DE will execute only one job at a time. When DE will execute only one job at a time, it will result in a more stable job execution and the resource burden on the hosting infrastructure will be lighter. The possibility to create a job queue will enable the user to split bigger data extraction needs into smaller jobs. A smaller extraction job can for instance target only one hazard endpoint in IUCLID at a time or if the number of target UUIDs is very large, to run many jobs for the same IUCLID fields, but for different batches of target UUIDs. In case of a targeted job for only one hazard endpoint, it is easier for the data analyst to post-process the extracted flat data. While the job queue is executing from the beginning, the user is free to change the order of the waiting jobs in the queue. Further on, a job queue will progress automatically, without the need of the user to keep an eye on when a job has finished, and the next one can be launched.

NEW Install DE with https as an alternative to http

Ref. 807991

DE now supports also https.

Known bug: Some attachment files are not extracted from IUCLID dossiers or datasets. We are aware of the issue and are currently investigating the cause and solution to the problem. The extraction job may also become slow in such a job. A fix will be provided later.

6. 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.

7. 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.

8. 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.key, 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

9. 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

10. 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

11. 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