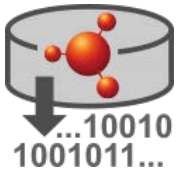


IUCLID Data Extractor: User Guide

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: User Guide
Version	v1.11.3 for IUCLID 6 v7. See the Release notes.
Issue date:	July 2023
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, 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).

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

Changes to this document

Date	Modification
07/07/2023	Updated for version 1.11.3. Compatible with IUCLID 6 version 7.
18/01/2023	Updated for version 1.10.0.
17/08/2022	Updated for version 1.9.0.
10/01/2022	Updated for version 1.8.0. Compatible with IUCLID 6 version 6.
12/02/2021	Updated for version 1.5.1.
19/08/2020	Updated for version 1.4.0.
17/06/2020	Updated for version 1.3.1.
18/12/2019	First official version for industry.

Table of Contents

Changes to this document i
Table of Contents i
Table of Figures ii
1. Introduction 1
2. User 1
3. Log in 1
4. Home page 3
5. Create an extraction 4
6. Define extraction 5
6.1. Name 5
6.2. Format 5
6.2.1. JSON 5
6.2.2. Normalised 5
6.2.2.1. Data Settings 5
6.3. Targets 6
6.4. Fields / TOC 7
6.5. Attachments 7
6.6. Table of Contents (TOC) 9
6.6.1. Table of Contents (TOC) – Composite 10
6.6.1.1. Document references 15
6.6.1.2. REACH Dossier Header 16
6.6.2. Table of Contents (TOC) – Other 17
6.6.3. Save field selection 17

6.7. Start an extraction....17
6.8. The states of an extraction and queuing....19
6.8.1. Managing the queue....21
6.9. Completed extraction....23
6.10. Failed extraction....23
6.11. View result of extraction....24
6.11.1. Normalised output....24
6.11.2. JSON output....30
6.11.3. Download extracted attachments....31
6.11.3.1. Attachments that have the same file name....32
7. Re-use an extraction....35
7.1. Select extraction to re-use....35
8. Export and import a definition of an extraction....36

Table of Figures

Figure 1: The login box....1
Figure 2: The data definitions available in IUCLID Data Extractor....2
Figure 3: Overview of the Home page....3
Figure 4: Create an extraction....4
Figure 5: Open a newly created extraction, or go back to the Home page....4
Figure 6: Default setting for Data settings under Normalised output....6
Figure 7: Targets added to an extraction....7
Figure 8: Including attached files in an extraction....8
Figure 9: Attaching a file to a whole document or within a field....9
Figure 10: Filtering TOCs by type of entity in the Composite view....10
Figure 11: Filter TOCs by legislation - example BPR....11
Figure 12: The dossier types that share the same TOC structure in REACH group 1....12
Figure 13: Selecting fields in a TOC....13
Figure 14: Fields added to an extraction....13
Figure 15: View the Field identifier and the Field type....14
Figure 16: Document references....15
Figure 17: Selecting fields from a Dossier header under the REACH regulation....16
Figure 18: Selecting fields from entities other than *Mixture*, *Substance* or *Template*....17
Figure 19: Save the definition of an extraction....17
Figure 20: Start an extraction....18
Figure 21: Viewing the queued extractions....19
Figure 22: The states of an extraction....20
Figure 23: Pause the processing of a queue....21
Figure 24: A paused queue with no data being extracted....22
Figure 25: Change the order of priority in a queue....23
Figure 26: Find the reason why an extraction failed....24
Figure 27: Download the extracted data from the entry of the extraction on the Home page....24
Figure 28: Download the extracted data from the record of the extraction....24
Figure 29: Example data in IUCLID....25

Figure 30: Example of selecting fields to extract....26
Figure 31: Example dossier UUIDs....27
Figure 32: Example Normalised output files....27
Figure 33: Example Normalised output shown in Notepad++....28
Figure 34: Open a text file in Excel....28
Figure 35: Parse TAB delimited text in Excel....29
Figure 36: Example Normalised data viewed in Excel....29
Figure 37: Download extracted attachments....31
Figure 38: Limit the size of the zip archive in which attachments are downloaded....31
Figure 39: Download attachments in batches....32
Figure 40: Files attached that have the same name and/or content....33
Figure 41: Extracted files....33
Figure 42: Filenames and checksums in the matrix file....34
Figure 43: Load an extraction for use as a template....35
Figure 44: Export and import of an extraction....36

1. Introduction

IUCLID Data Extractor is an advanced tool that extracts data from IUCLID in accordance with sets of user-defined rules. A set of rules is stored in an *Extraction*. An *Extraction* can be edited, re-used as a template for a new extraction, and exported to be shared with other users.

The user interface of *IUCLID Data Extractor* is viewed using a web-browser. If *IUCLID Data Extractor* is run on the same server as IUCLID, the default address is shown below:

`http://<address of IUCLID>/iuclidde`

If you do not know the address, ask your local administrator.

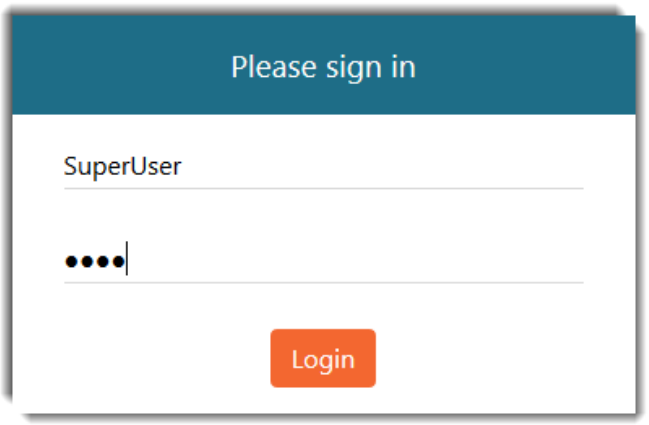
2. User

The authentication credentials and access to data in *IUCLID Data Extractor* are based on IUCLID users. If you do not have suitable user credentials, contact your IUCLID administrator.

There are two different types of user: *Analyst* and *Administrator*. An *Analyst* can extract the data to which it has access in IUCLID. This manual describes the main points of data extraction that can be performed by a user of type *Analyst*.

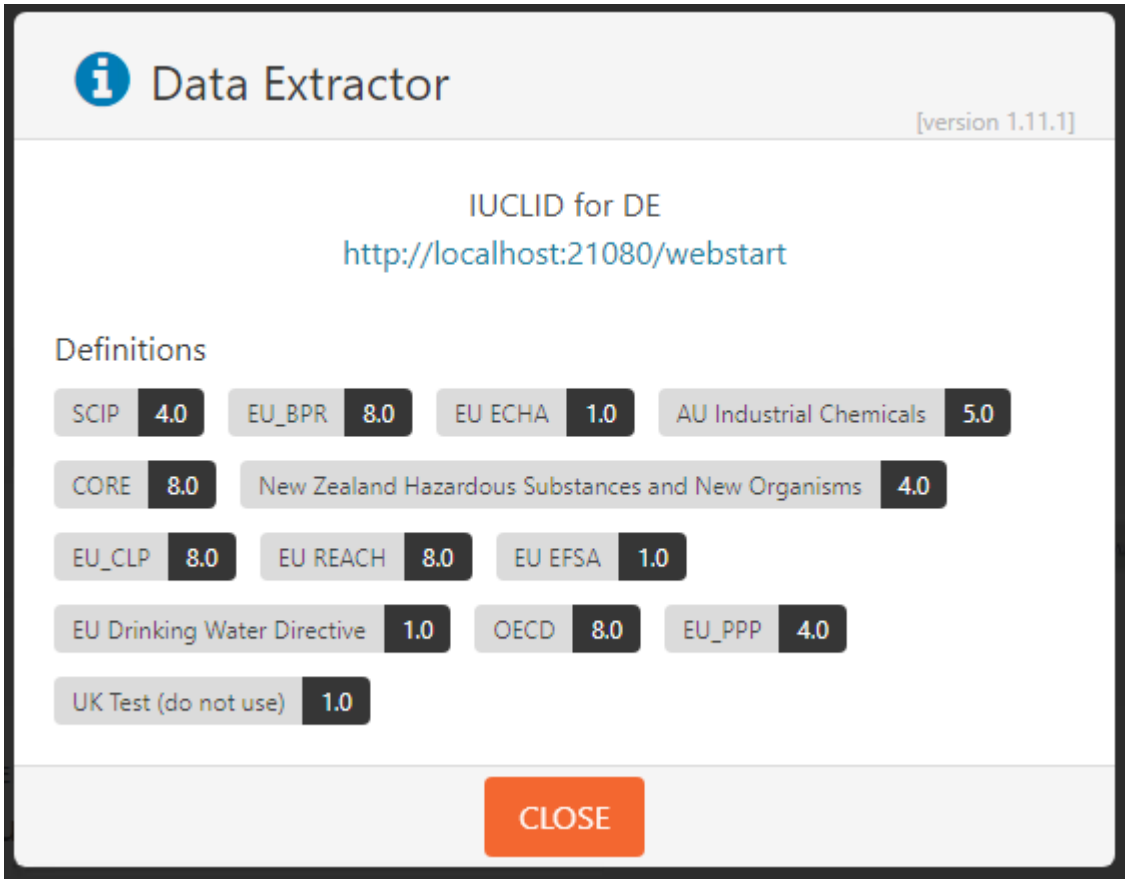
3. Log in

Figure 1: The login box



After logging in, a summary is shown of the data definitions that can be used to create extractions. Click on *Close* to enter the *Home Page*.

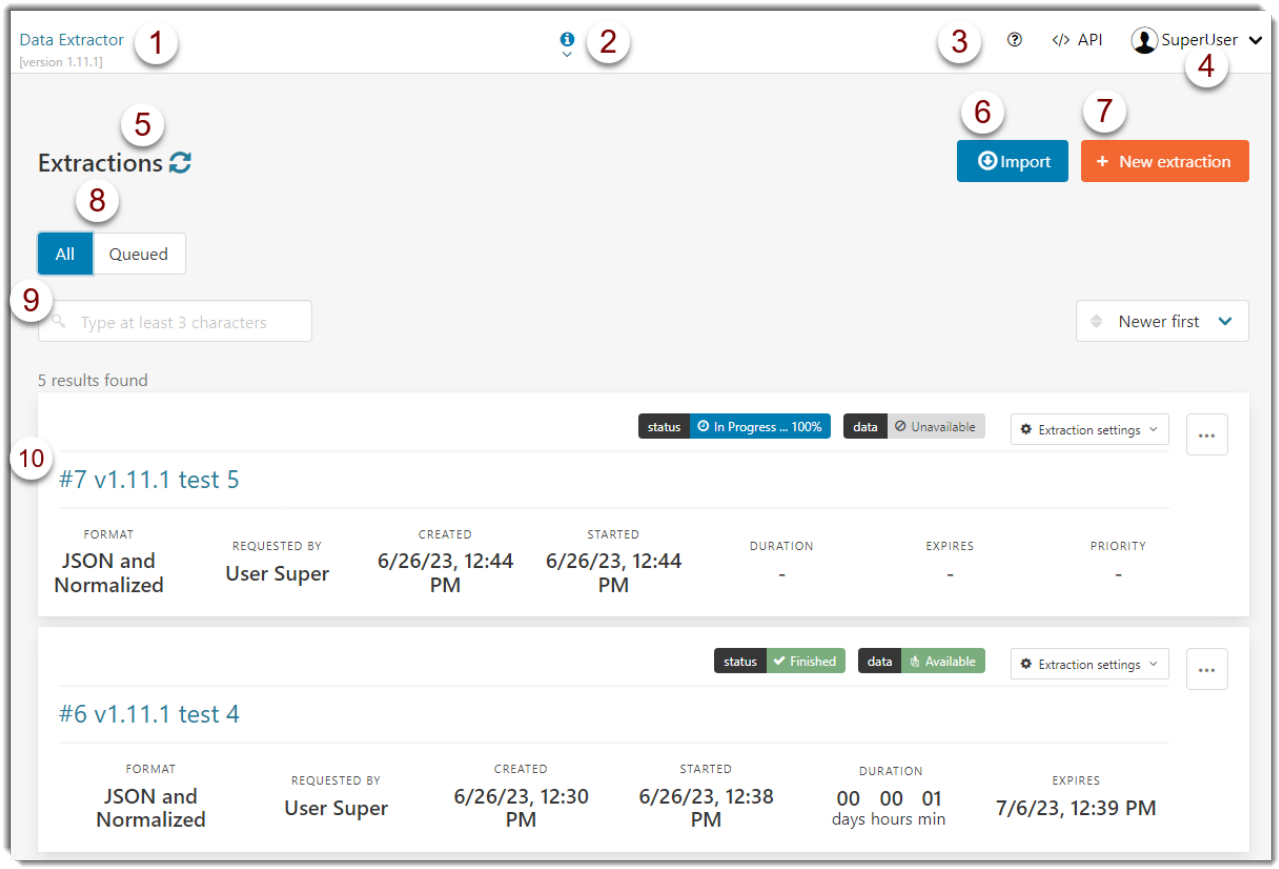
Figure 2: The data definitions available in IUCLID Data Extractor



4. Home page

The *Home page* of the interface of *IUCLID Data Extractor* contains a list of extractions and various functions related to them.

Figure 3: Overview of the Home page



Legend for Figure 3

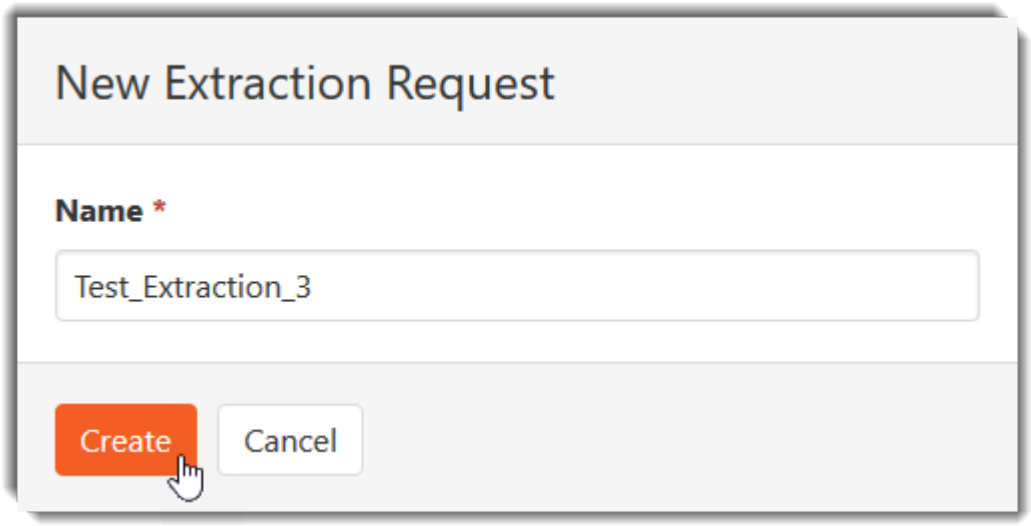
1. This is link to the home page;
2. The blue icon for info, , toggles the display of the available data definitions;
3. If you cannot see the icons in the top right of the interface, try closing the info banner (2) by clicking on the arrow underneath the icon . If they are still not available, zoom out in the browser. The question mark is a link intended to point to a user manual. To make a link to this manual, open the settings under the username and then upload the PDF file;
4. Log out, or access the *Dashboard* where there are some admin settings;
5. To refresh the list of extractions, click on the recycling icon to the right of the title **Extractions** ;
6. Import a previously exported definition of an extraction as a new extraction. This is a way of re-using previously entered data;
7. Create an extraction from scratch;

8. The list of extractions is split into two views. *All* shows the currently running extraction, and the ones that have finished. *Queued* shows the currently running extraction, and the extractions behind it in the queue waiting to start;
9. Search in the list page;
10. Access the definition of an extraction, by clicking on its name on the left in blue.

5. Create an extraction

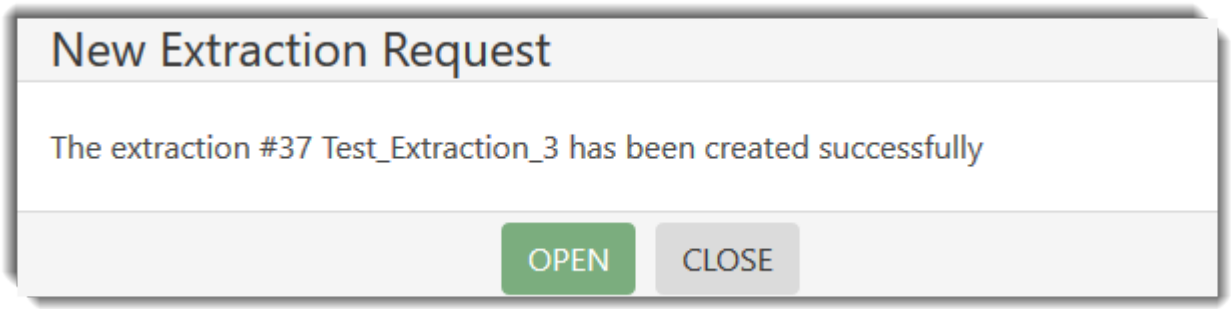
To create an extraction from scratch, click on the button labelled + *New extraction*. This does not actually extract data, but starts the process of defining a set of rules by which data will be extracted. Enter a name, as shown below.

Figure 4: Create an extraction



An extraction is created with an incremental number #N at the start of its name.

Figure 5: Open a newly created extraction, or go back to the Home page



To start defining the extraction rules straight away, click on *Open*. To save that task for later, and go back to the *Home page*, click on *Close*.

6. Define extraction

After having clicked *Open*, or after clicking on the name of the extraction in the *Home page*, the rules for the extraction can be defined. The following fields are mandatory.

- Name
- Format
- Targets
- Fields

6.1. Name

A text identifier of up to 64 characters.

6.2. Format

This is the format of the output of the extraction.

6.2.1. JSON

The IUCLID data is output in JSON format. This is a single text file in which the data is arranged in a tree-like structure.

6.2.2. Normalised

Normalised outputs data in character separated format in files that reflect the structure of the IUCLID *Table of Contents (TOC)*. The output can be opened in a spreadsheet such as Excel.

Multiple repeatable blocks are placed in the same file, numbered according to their order of appearance in the document. Fields that can contain multiple blocks/sets are indicated in the TOC by icons that have multiple sheets ( ).

6.2.2.1. Data Settings

These settings are provided to help you to create valid character separated output, and to provide some flexibility. The two main factors affecting validity of the output are:

1. the presence of new lines (LF) within the extracted text content;
2. the presence of column delimiters within the extracted text content.

These need to be replaced with something else.

Replace new lines: Replaces a LF character with the text that is in the box.

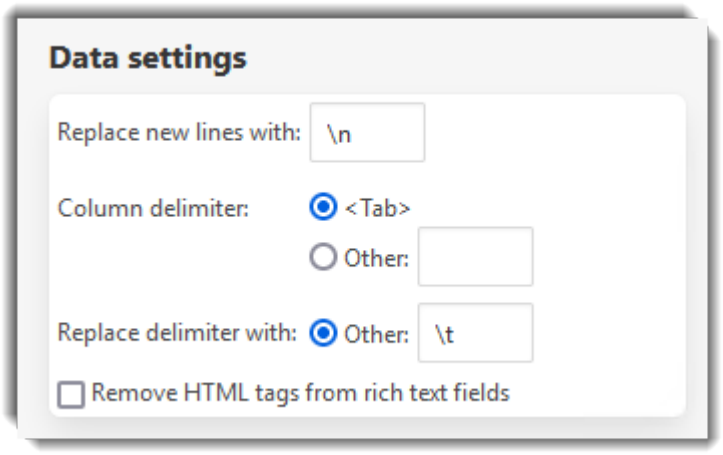
Column delimiter: sets the character that is placed between columns of output. The selection <*Tab*> inserts a TAB character. The selection *Other* inserts the value that is in the box.

Replace delimiter with: acts in text values to replace the value of *Column delimiter* given above, with something else. The selection <*Tab*> inserts a TAB character. The selection *Other* inserts the value that is in the box.

The interface has a dynamic rule that prevents **Column delimiter** and **Replace delimiter with** from having the same value.

When defining extractions, you are free to select your own values. The default settings are shown below, which output in TAB separated format, sometimes referred to as TSV.

Figure 6: Default setting for Data settings under Normalised output



If a box contains \n or \t, those characters are inserted literally, not LF or TAB. Values of \n and \t are suggested as defaults because they provide a true record in the output of what the source values contain, with a low chance of confusion between them and the actual data.

The option *Remove HTML tags from rich text fields* does just that, leaving only the text values.

Note

The TAB character cannot be typed in to a IUCLID field, but it can be pasted in, and it can exist due to formatting in rich text fields.

Tip

If you want to completely avoid having characters in the output that require processing at a later step, and you know it will not affect your own data adversely, it is possible to use space characters for **Column delimiter** and **Replace new lines**.

For example, the following text:

“This is some data

Across two lines with some tabs”

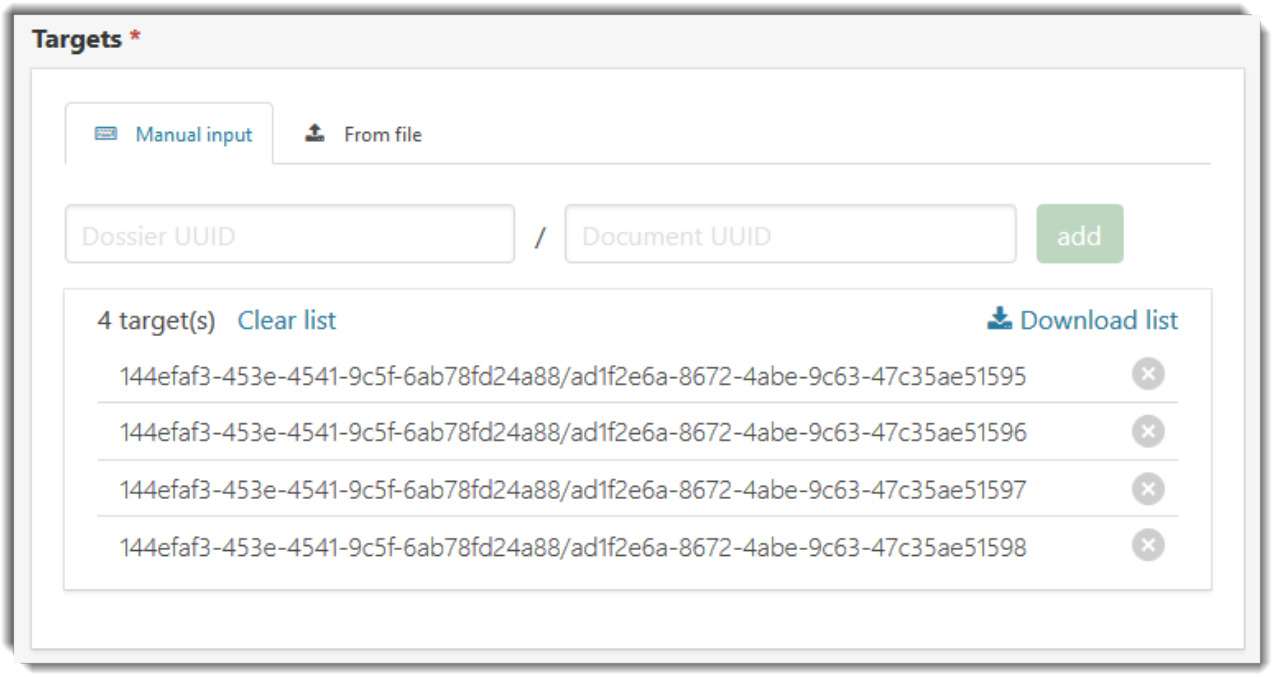
Is output as:

“This is some data Across two lines with some tabs”

6.3. Targets

A target defines the Dossier or document in UUID from which data is extracted. A target can contain a Dossier UUID, a document UUID, or both, separated by a forward slash. Values can be entered manually, or imported from a file. The format of the file is one target per line, with the Dossier UUID and the document UUID separated by a comma.

Figure 7: Targets added to an extraction



6.4. Fields / TOC

Fields are selected from the *Table of Contents (TOC)*, as described below under TOC.

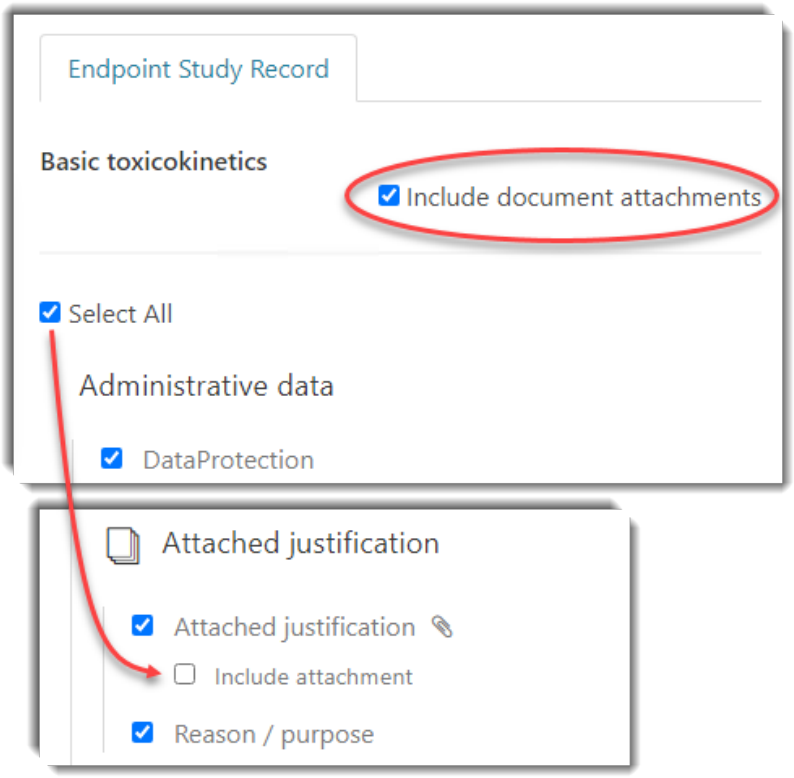
6.5. Attachments

In IUCLID a file can be attached either to a whole document, or within certain types of fields. This affects how attachments are selected for extraction.

To extract files that are attached to whole documents, tick the box at the top right of the table of contents for the selected section. This does **not** apply to attachments in fields. To extract files that are attached within a particular field, tick the box for the individual field. This applies across all documents in that section, but not to other fields. Ticking the box *Select all* does not select attachments in fields. These must be ticked separately.

In the example below, all files attached to documents in the section for *Basic toxicokinetics* will be extracted. However, the attached justification will not be extracted, even though *Select all* is ticked.

Figure 8: Including attached files in an extraction

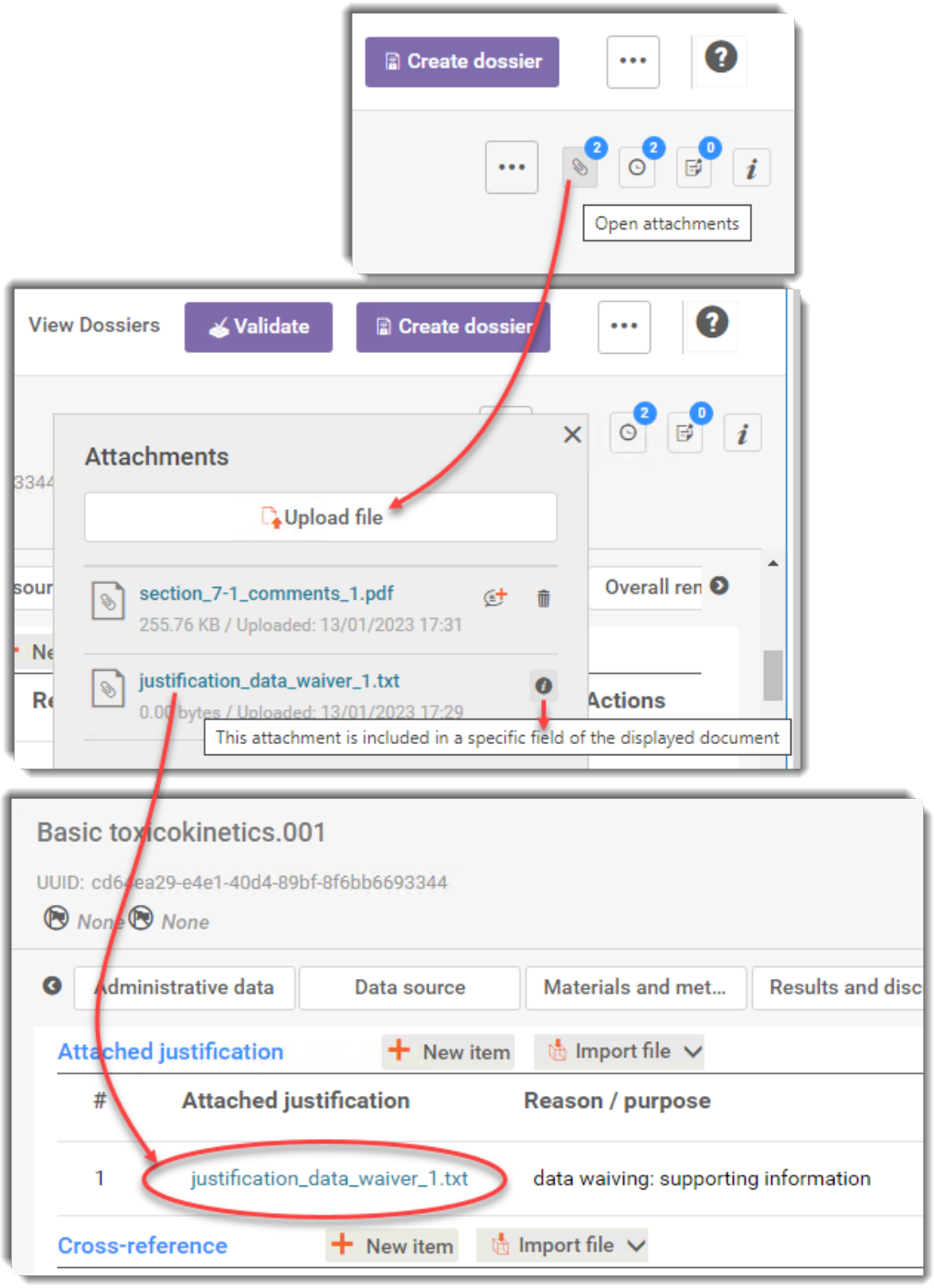


A file is attached to a whole document via *Upload file* in the menu that is opened via the paperclip icon at the top right of the record of the document.

In the example below, the file named `section_7-1_comments_1.pdf` is attached to a document in the section for *Basic toxicokinetics*.

The file named `justification_data_waiver_1.txt` is attached to a field which is shown in the lower part of the figure.

Figure 9: Attaching a file to a whole document or within a field



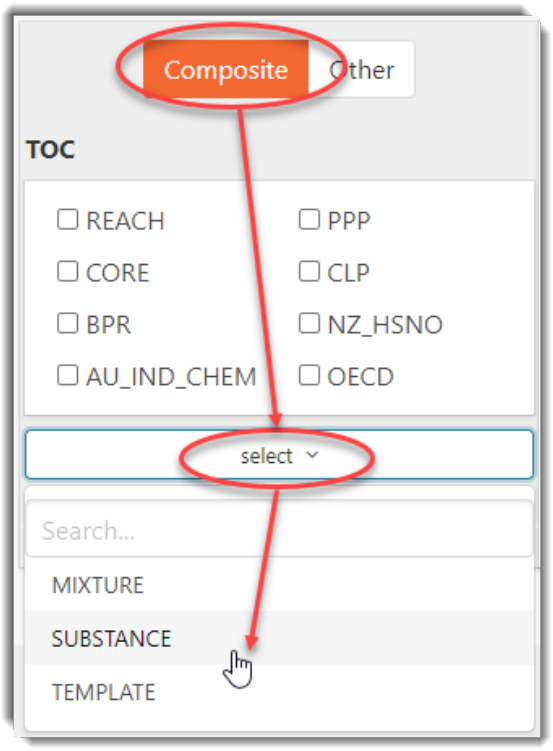
6.6. Table of Contents (TOC)

First, consider which fields are to be extracted, and where they are in the structure of IUCLID. For a field that is in an entity of type *Mixture*, *Substance* or *Template*, select the option *Composite*. For fields within other types of entity, including *Article*, select *Other*, as described later in this section.

6.6.1. Table of Contents (TOC) – Composite

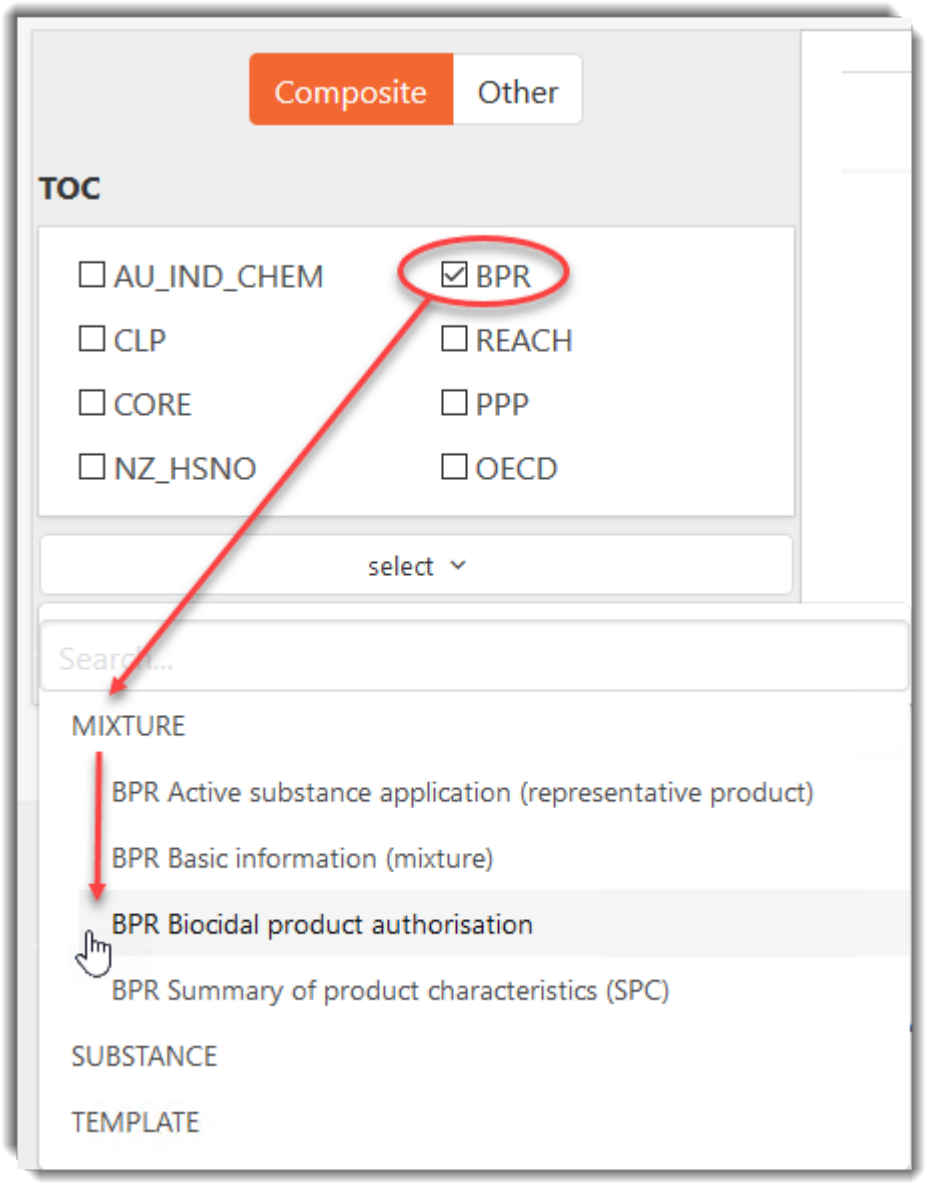
For fields in an entity of type *Mixture*, *Substance* or *Template*, select the type of source under *Composite*, and select the type you need.

Figure 10: Filtering TOCs by type of entity in the Composite view



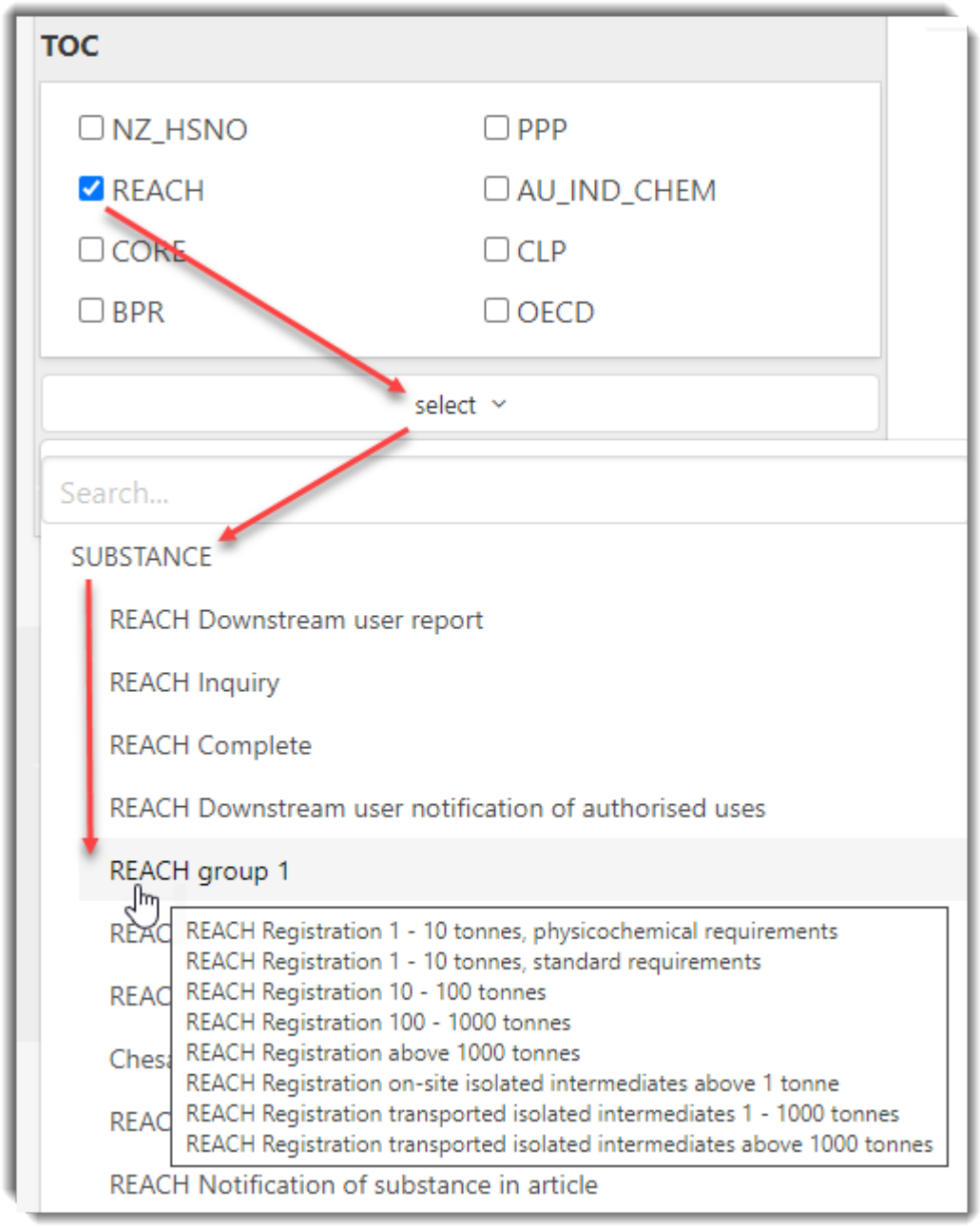
Filtering by legislation shows only the entities and TOCs that are relevant to the legislation(s). Selecting no legislations shows all the TOCs. In the example below, BPR is selected. The possible entities, *Dossier* types, and therefore corresponding TOCs for BPR are shown.

Figure 11: Filter TOCs by legislation - example BPR



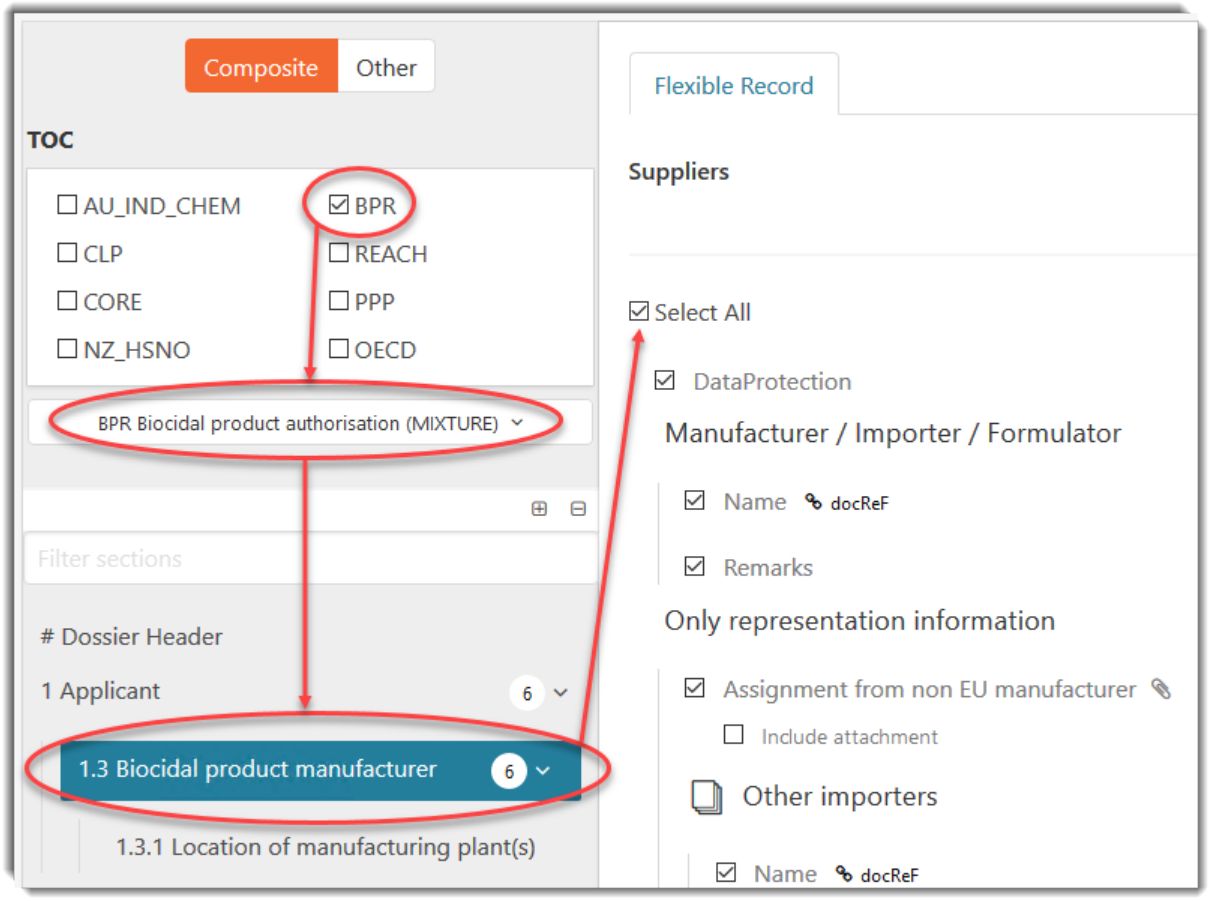
TOCs for different types of dossiers that have the same structure are grouped together per legislation. To see which dossier types are in the group, hover the mouse pointer over the name of the group. For example, REACH has two groups, the first of which is shown in the figure below.

Figure 12: The dossier types that share the same TOC structure in REACH group 1



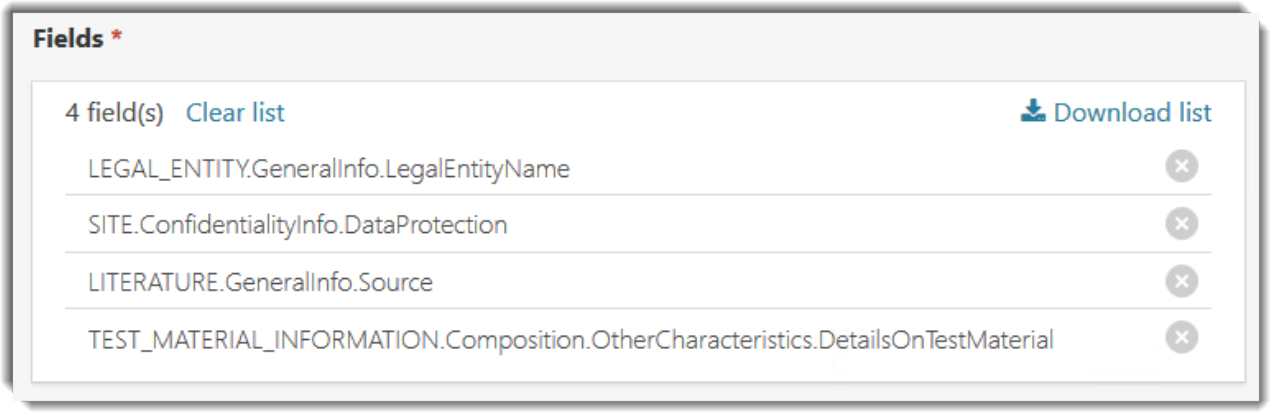
Once a TOC has been selected, it is indicated below the legislation selectors, and the TOC opens in the window on the right of the screen. To select a field, tick its box. All the fields in a document can be selected at once by ticking *Select all*, as shown below.

Figure 13: Selecting fields in a TOC



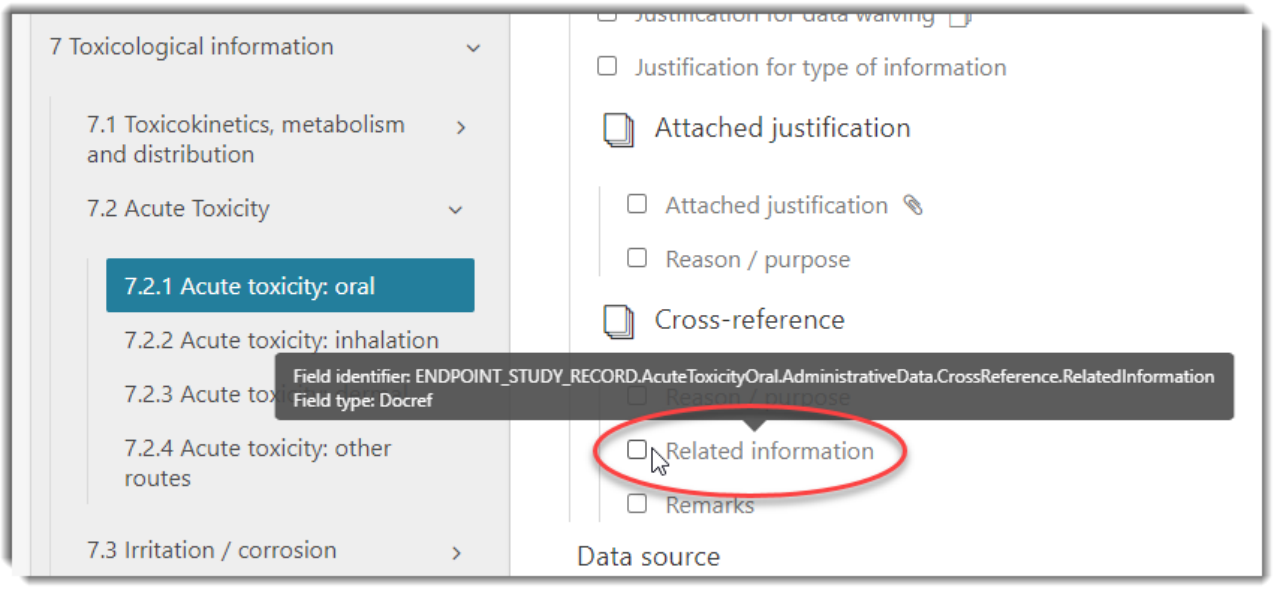
Selected fields are immediately added to the list in the record of the extraction. They can be removed from the list by either unticking in the selection view, or by clicking on the cross icon in the list.

Figure 14: Fields added to an extraction



The path to a field in IUCLID, *Field identifier*, and the *Field type* can be viewed by hovering the pointer over the field, as shown below.

Figure 15: View the Field identifier and the Field type

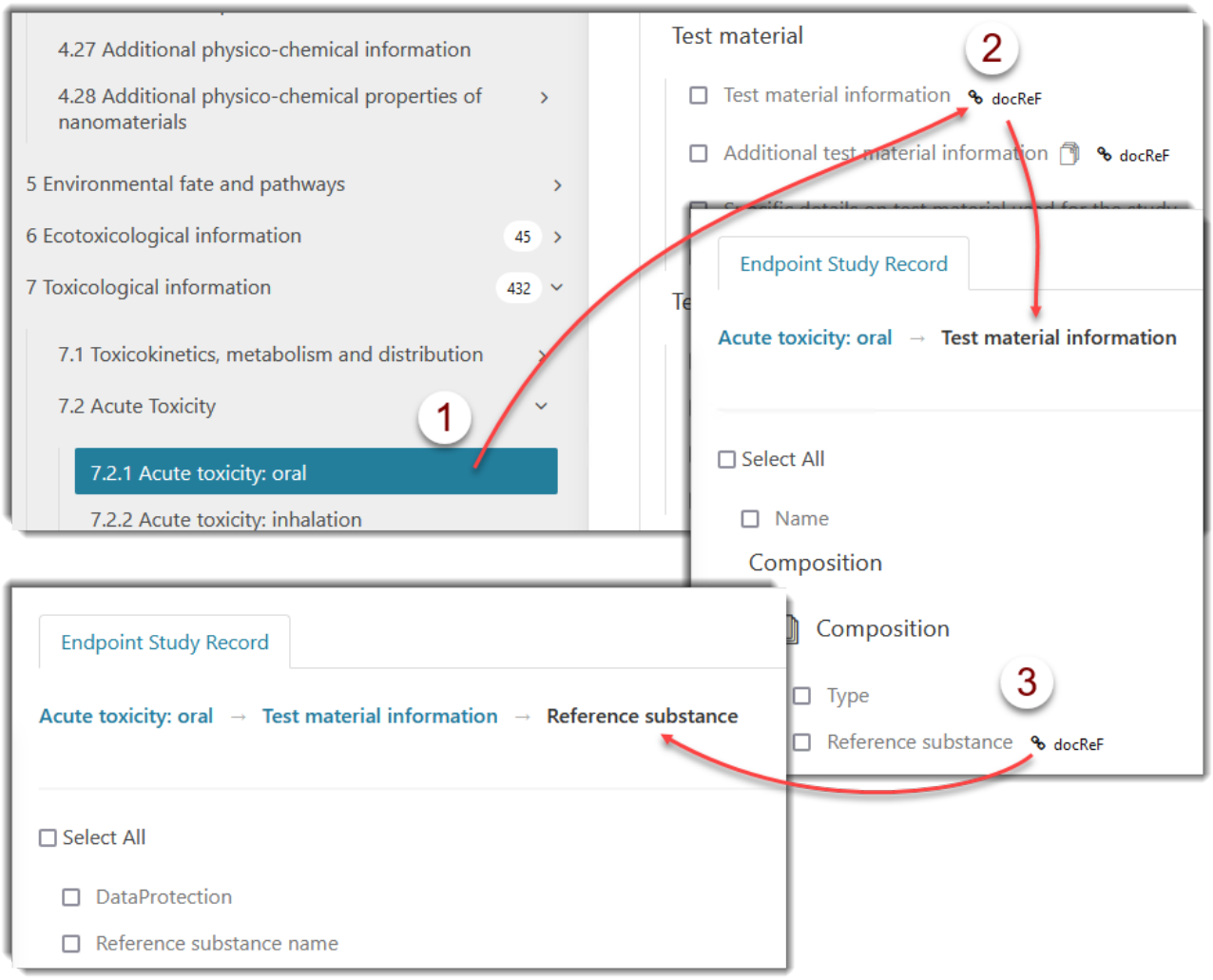


The option *Other* is used to select fields in entities of type *Article*, and entities that are referred to by the entities accessed under *Composite*, for example *Reference substance*, as shown below in section *6.6.2 Table of Contents (TOC) – Other*.

6.6.1.1. Document references

IUCLID allows for document references in which multiple documents can refer to a single entity. This re-uses data, which increases efficiency of working and consistency of data. Whilst selecting fields, the structure of these links can be followed. The links are indicated by the symbol, docReF. An example is given below.

Figure 16: Document references



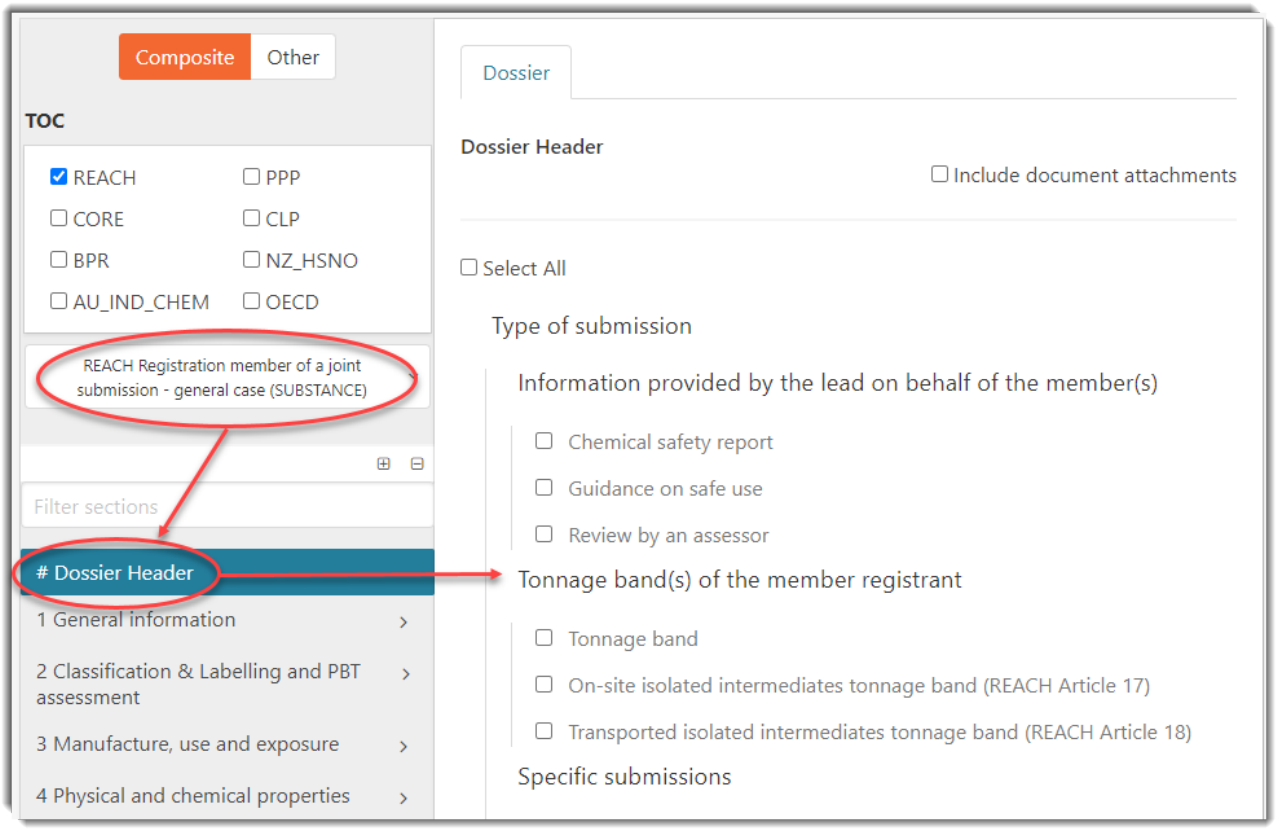
Legend for Figure 16

1. The document type selected in the *Table of contents* is for REACH in section *7.2.1 Acute toxicity: oral*;
2. The document type contains a field titled *Test material information* that can contain a reference to a separate entity of type *Test material*;
3. That entity contains a field titled *Reference substance*, which can contain a reference to a separate entity of type *Reference substance*.

6.6.1.2. REACH Dossier Header

In the REACH regulation, the fields in the Dossier header depend on the type of Dossier. Therefore, for REACH, there is a special option *Dossier Header*, which presents only the fields that are relevant to the current Dossier type. In the example shown below there are fields that are relevant only to joint submissions.

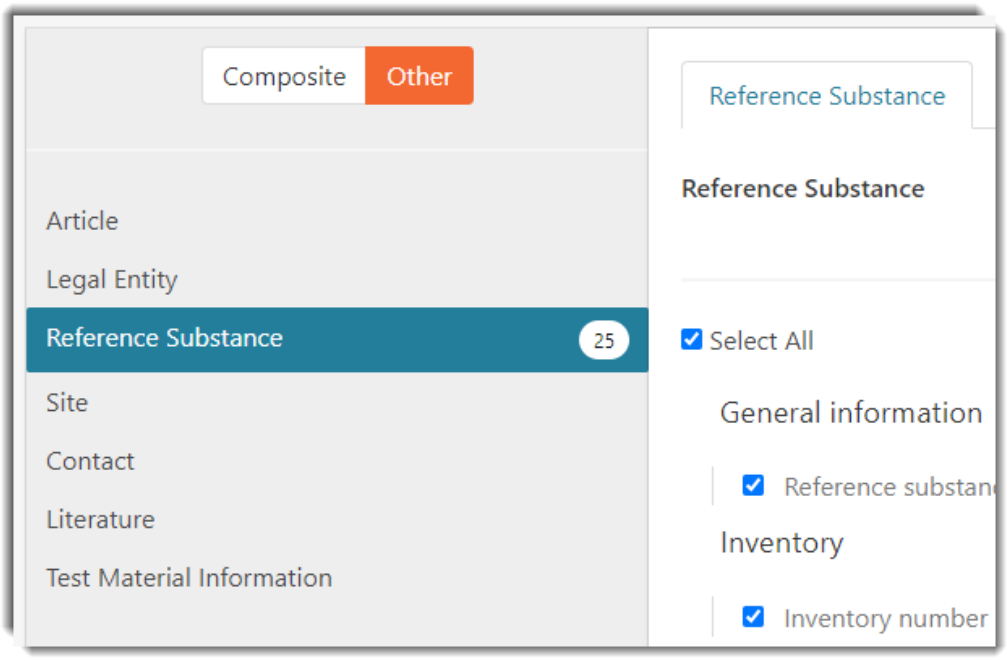
Figure 17: Selecting fields from a Dossier header under the REACH regulation



6.6.2. Table of Contents (TOC) – Other

For fields that are not directly in an entity of type *Mixture*, *Substance* or *Template*, select the type of source under *Other*, and select the type you need.

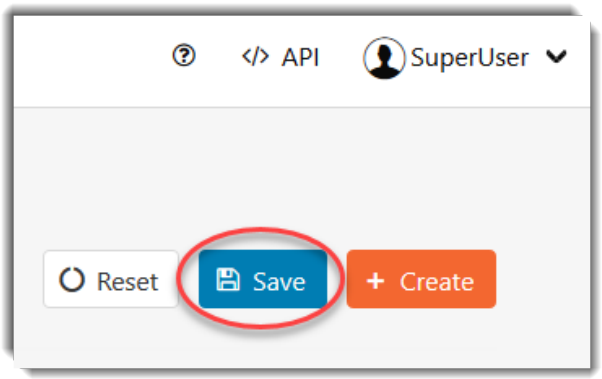
Figure 18: Selecting fields from entities other than *Mixture*, *Substance* or *Template*



6.6.3. Save field selection

To record the field selection in the definition of the extraction, click on *Save*, which is at the top right of the extraction record. This must be done if the selection is to be retained after navigating away from the definition of the extraction.

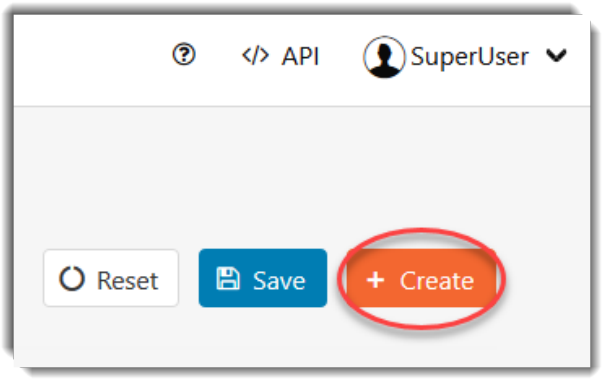
Figure 19: Save the definition of an extraction



6.7. Start an extraction

To start an extraction, within the record of the extraction, click on *Create*, which is at the top right of the extraction record, as shown below.

Figure 20: Start an extraction



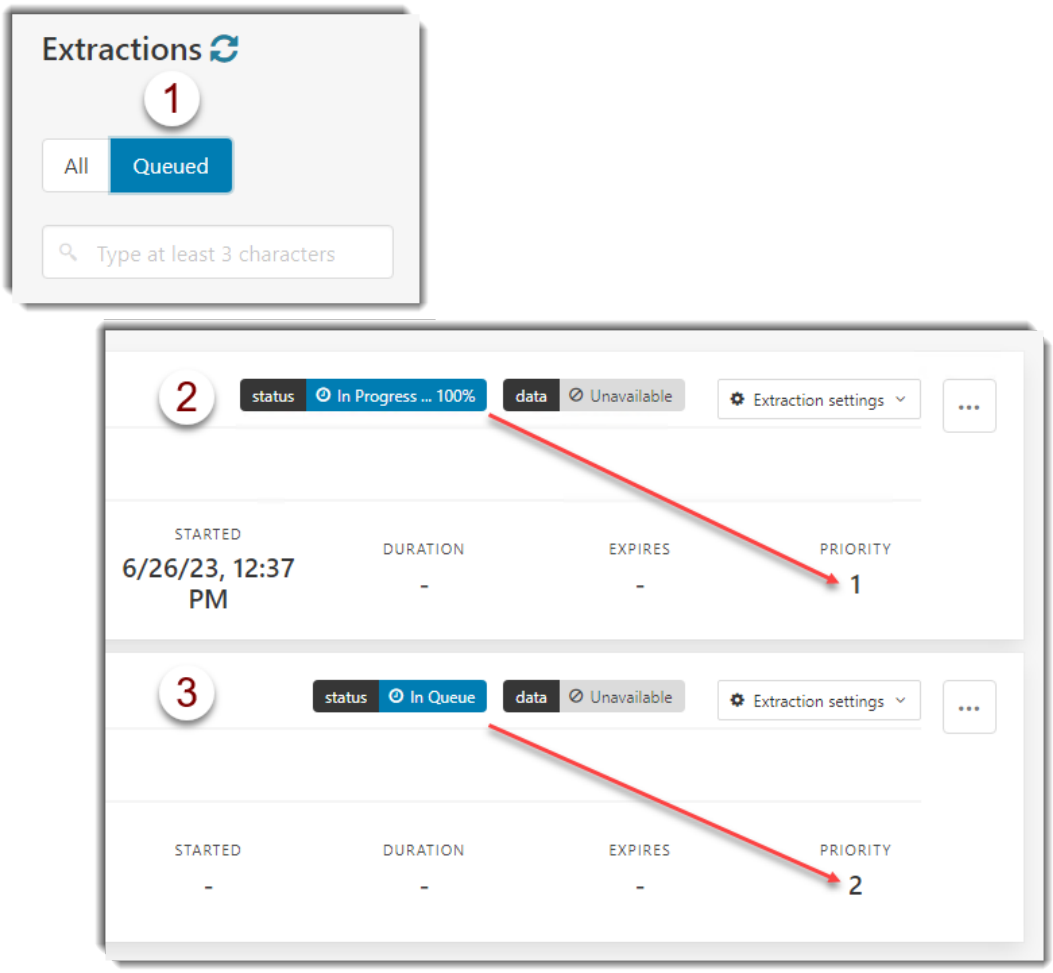
6.8. The states of an extraction and queuing

Extractions are queued. Only one extraction is performed at any one time. This prevents the system from becoming overloaded if several large extractions are started in a short period. The load on the system and the time required per extraction can be managed by dividing up the extraction work across multiple extractions. This can be done either per dossier UUIDs, or per field definitions.

If an extraction is to be run from start to finish in isolation, queuing is not relevant.

If another extraction is started before the previous one has completed, it is placed in a queue behind the earlier one. The queue can contain up to 25 extractions. The home page has two listing views. *All* shows every extraction, irrespective of its status. *Queued* shows the currently running extraction, and the extractions behind it in the queue waiting to start. An example of a queue is shown below.

Figure 21: Viewing the queued extractions



Legend for Figure 21

1. These buttons toggle between the views *All* and *Queued*;
2. The first entry in *Queued* is the entry currently being processed. It has priority one. This is also visible under *All*;

3. The next entry and the subsequent ones have a state of *In Queue* and a priority that was set by the order in which they were run. They are also visible under *All*.

When an extraction is run, its record closes automatically and the *Home page* is shown with the listing set to the view for *Queued*. The states of the extraction are illustrated below. To refresh the status and data information, click on the recycling icon .

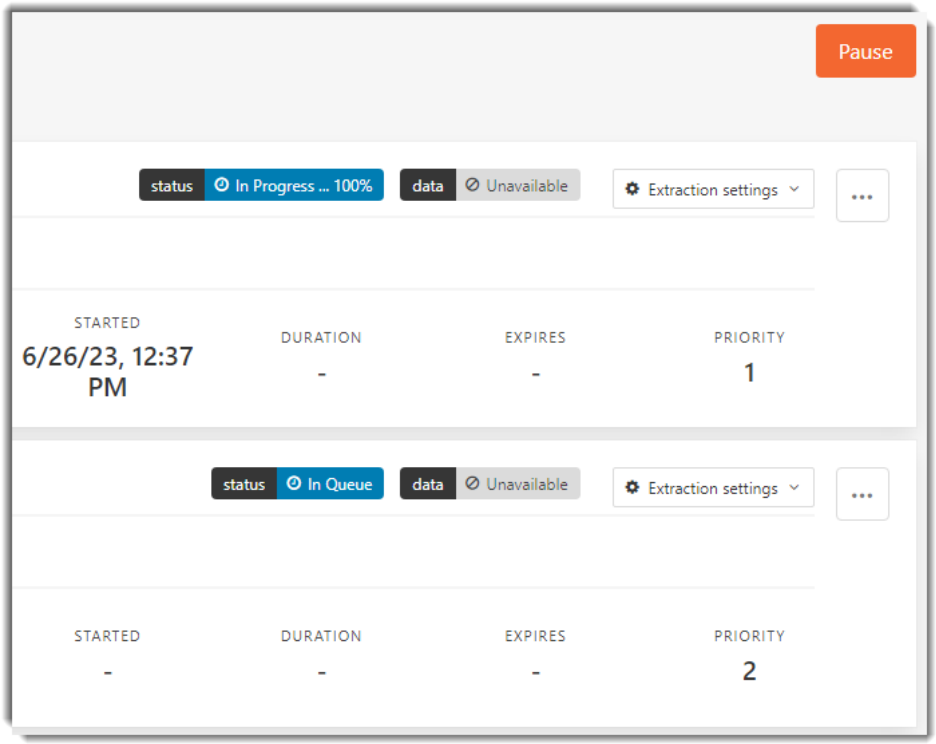
Figure 22: The states of an extraction

At first, the status is *In Queue* and the extracted data is *Unavailable* as shown above. This occurs even if there are no other extractions ahead of the extraction in the queue. Once the actual extraction begins, the status changes to *In progress* and a percentage of completeness is shown. When an extraction is completed, it is removed from the view *Queued*.

6.8.1. Managing the queue

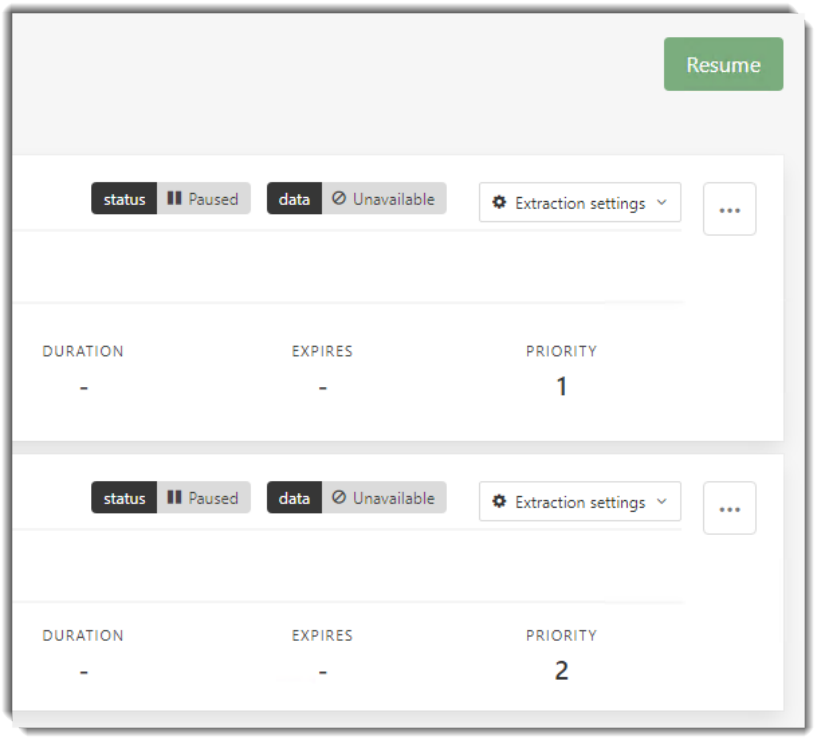
The feeding of the queued extractions into the state of *In progress*, can be paused by clicking on the *Pause* button at the top right of the *Queued* view.

Figure 23: Pause the processing of a queue



Extractions that have a status of *In Queue* are changed to a status of *Paused*. An extraction currently in progress is not affected. The orange *Pause* button changes to a green *Resume* button. When the extraction currently in progress is completed, it is removed from the *Queued* view. At this point no data are being extracted. An example is shown below.

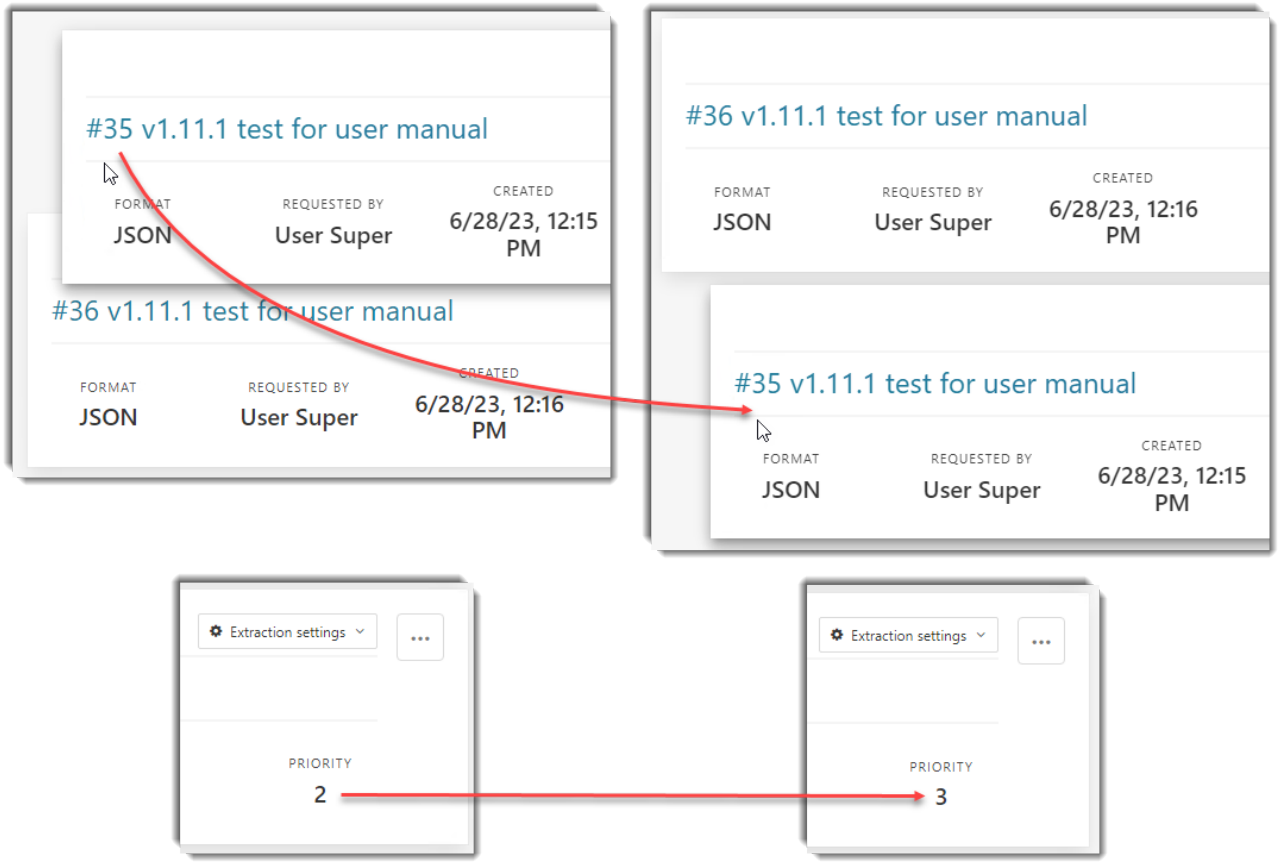
Figure 24: A paused queue with no data being extracted



To resume the processing of a queue, click *Resume*.

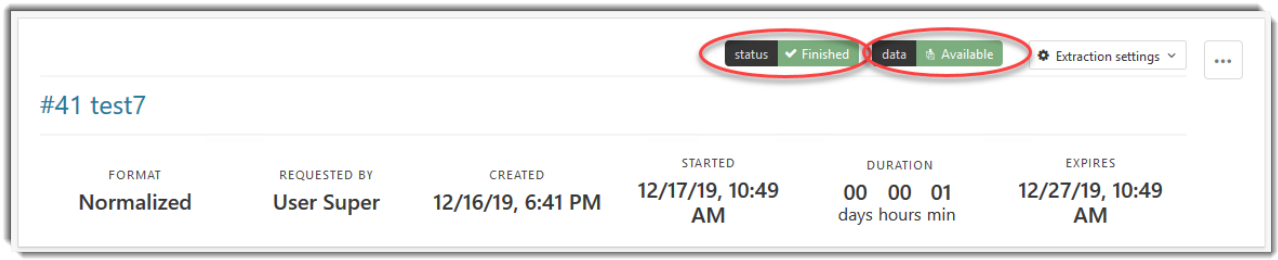
The order in which extractions enter the status *In progress*, is from the top to the bottom of the list in accordance with the value of the *Priority* as stated at the right of the record. The order can be changed by dragging and dropping an extraction in the list. Do not do this with an extraction that has the status *In progress*. In the example shown below, extraction #35 is being dragged down one place in the list of priorities from 2 to 3, behind extraction #36.

Figure 25: Change the order of priority in a queue



6.9. Completed extraction

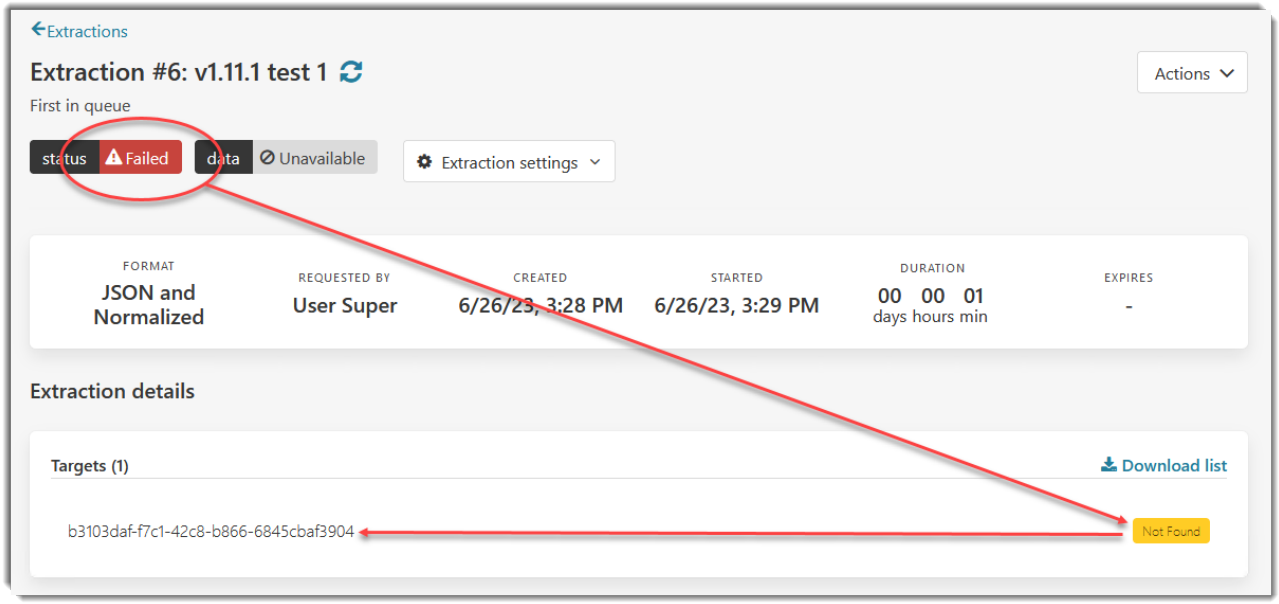
When extraction is completed, the status is "Finished" and data "Available" as shown below. Completed extractions are shown only in the view *All*.



6.10. Failed extraction

The status of a completed extraction can also be *Failed*. If that happens, to find out why, open the record of the extraction. Look for a message in yellow. A possible cause is that a dossier UUID was not present in IUCLID, as was the case in the example below where the message *Not Found* is given to the right of the Dossier UUID.

Figure 26: Find the reason why an extraction failed



6.11. View result of extraction

The extracted data is downloaded from a link under the word *Available* that is shown in the list entry on the *Home Page*, and in the record of the extraction.

Figure 27: Download the extracted data from the entry of the extraction on the Home page

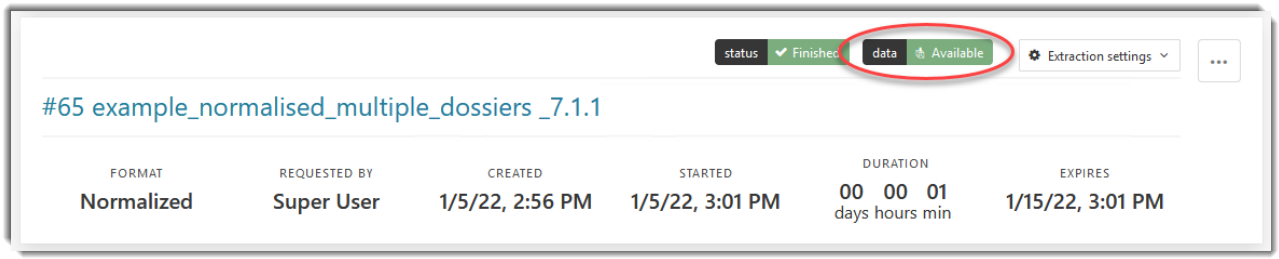
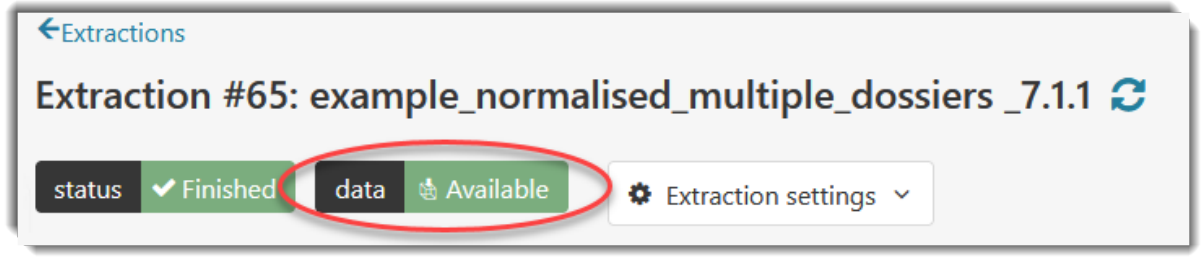


Figure 28: Download the extracted data from the record of the extraction



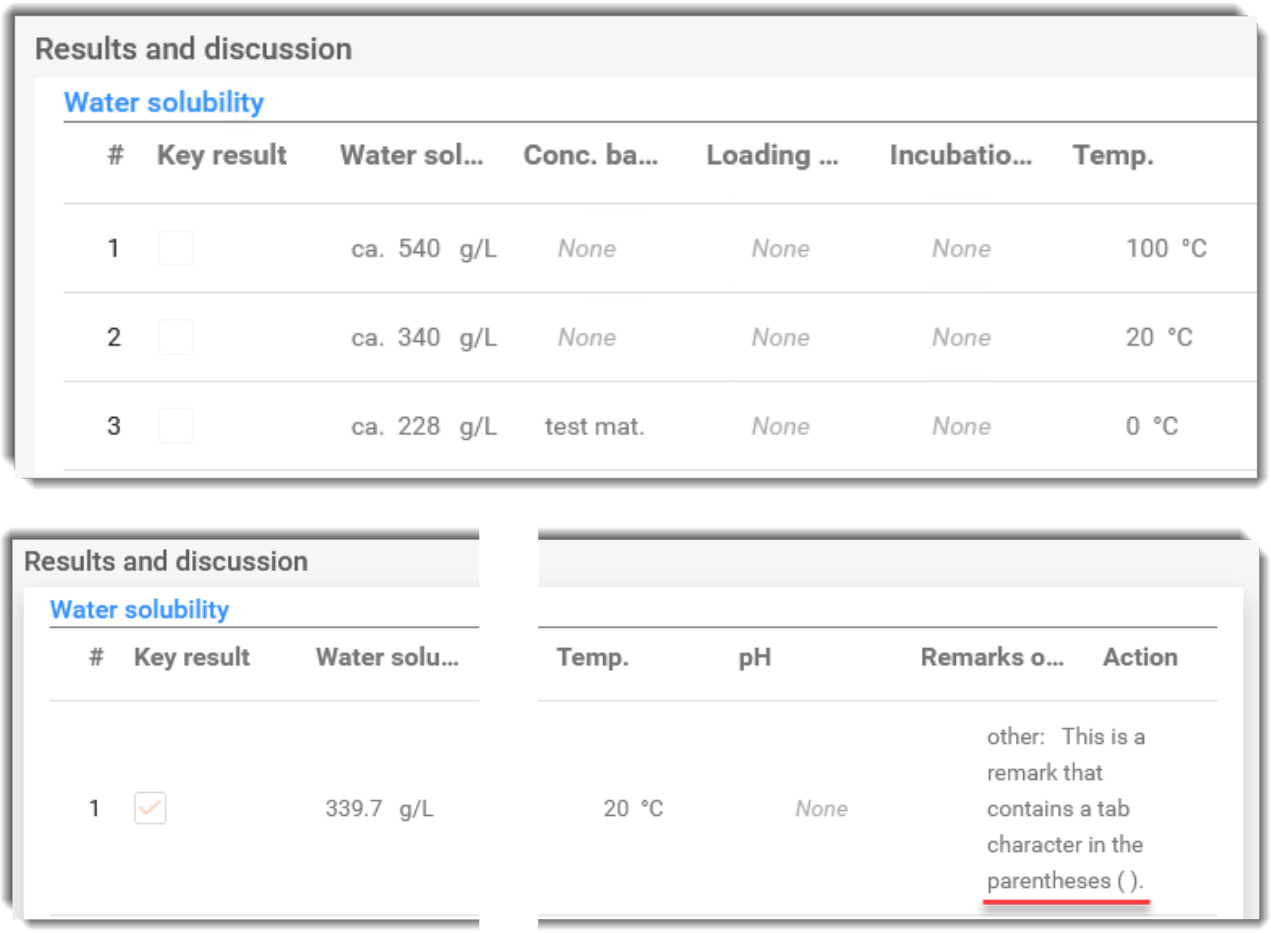
6.11.1. Normalised output

What follows is a description of Normalised output by way of example. The source in IUCLID is two dossiers, each of which contain one endpoint study record at path:

`ENDPOINT_STUDY_RECORD.WaterSolubility.ResultsAndDiscussion.WaterSolubility`

The data is shown in the IUCLID interface as:

Figure 29: Example data in IUCLID



Results and discussion

Water solubility

#	Key result	Water sol...	Conc. ba...	Loading ...	Incubatio...	Temp.
1		ca. 540 g/L	None	None	None	100 °C
2		ca. 340 g/L	None	None	None	20 °C
3		ca. 228 g/L	test mat.	None	None	0 °C

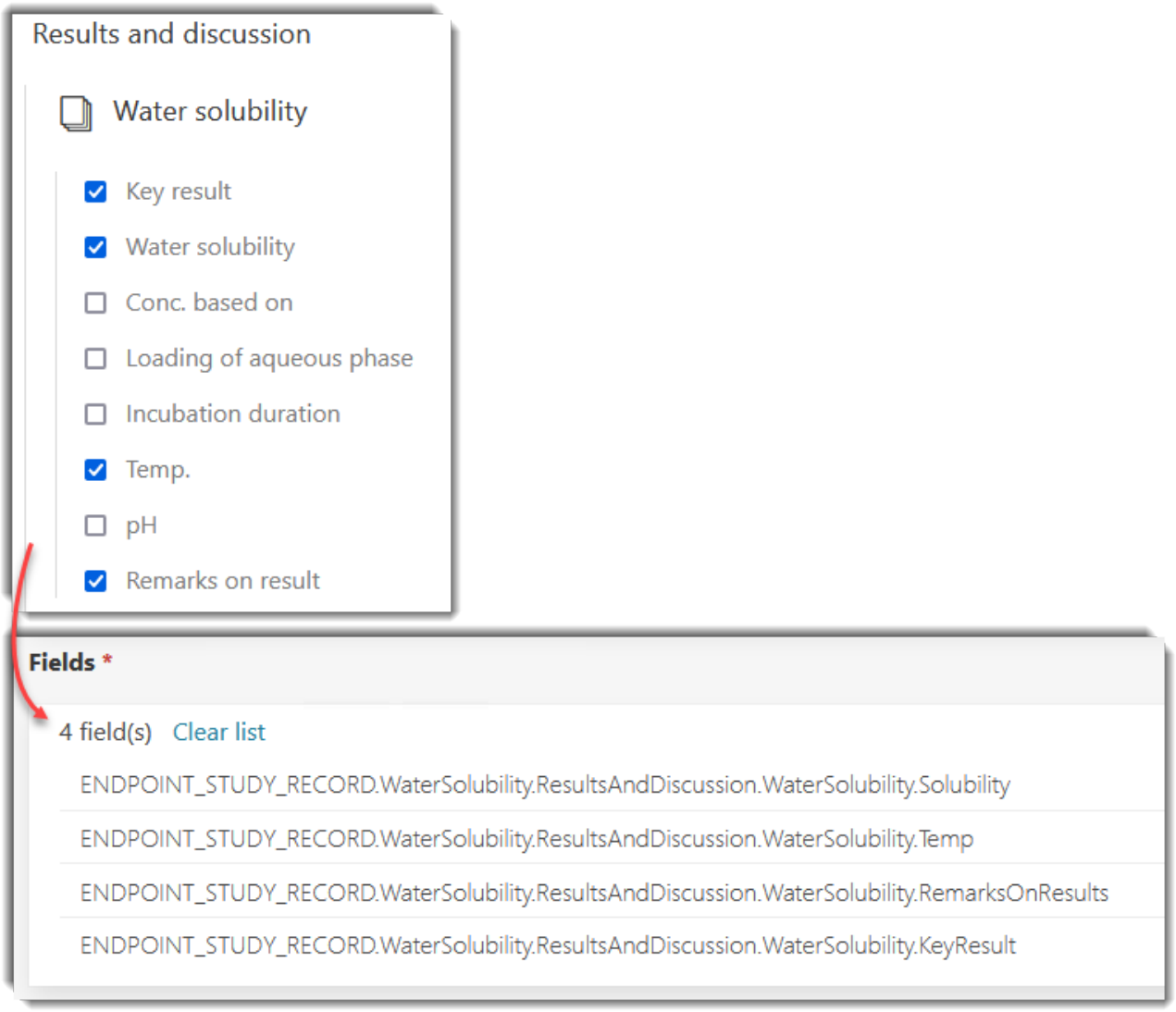
Results and discussion

Water solubility

#	Key result	Water solu...	Temp.	pH	Remarks o...	Action
1	✓	339.7 g/L	20 °C	None	other: This is a remark that contains a tab character in the parentheses ( ).	

Note how a delimiter has been deliberately entered into a text field, by pasting from a text editor. This serves as an illustration of how the *Data settings* work. The fields to be extracted are selected, and when saved, their paths are shown as below:

Figure 30: Example of selecting fields to extract



The icon in the TOC with multiple sheets ( ) next to *Water solubility* indicates a repeatable block, where more than one set of values can be recorded. This is the case for one of the endpoint study records, which has three. The number sequence is output in the second column of the output.

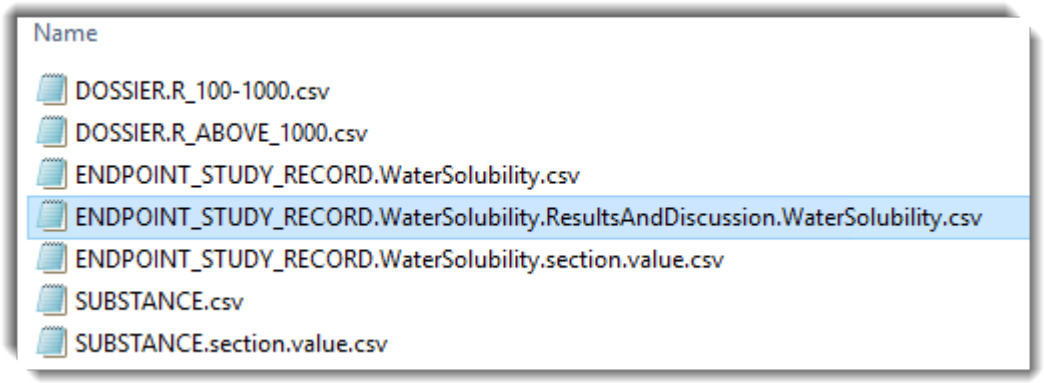
The dossier UUIDs are shown below. Note that they are visible later in the example, in the first column of the output.

Figure 31: Example dossier UUIDs

The default *Data settings* are used.

The output is a zip file that contains the files with extension csv. Note that these are not comma separated values. The delimiter is TAB.

Figure 32: Example Normalised output files

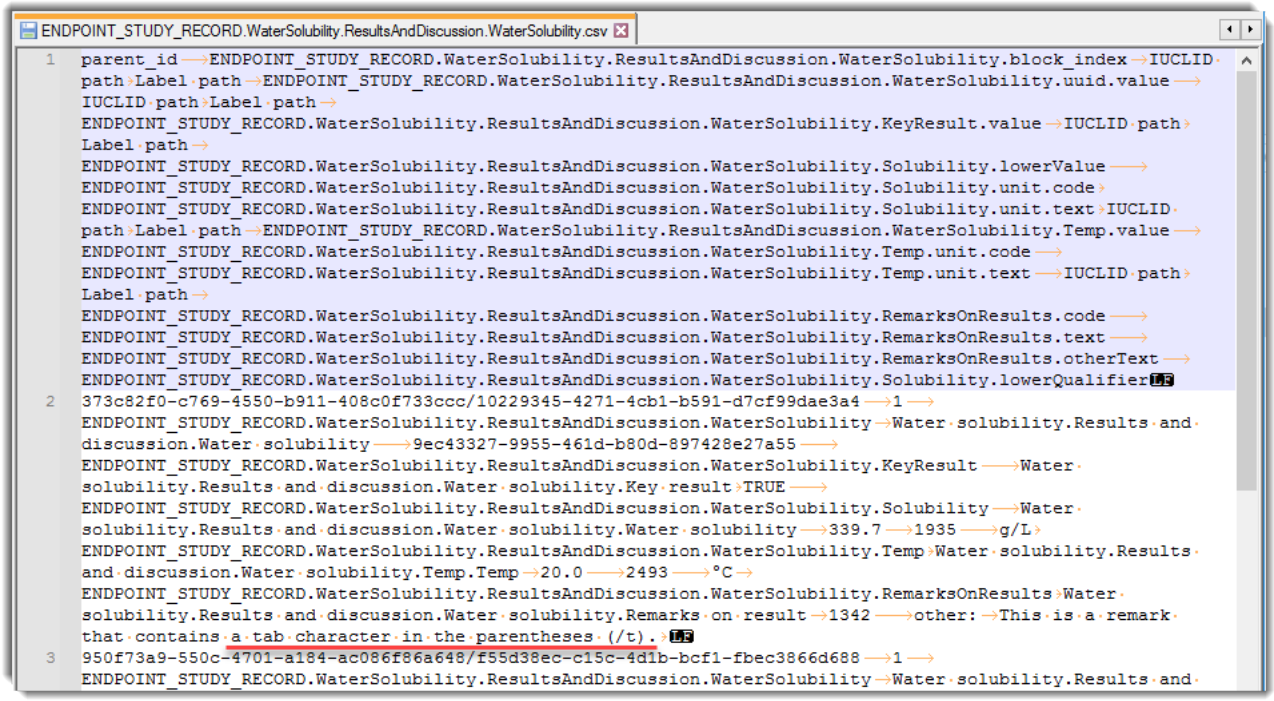


The data values that are shown in IUCLID in the screenshots above are in the file named:

`ENDPOINT_STUDY_RECORD.WaterSolubility.ResultsAndDiscussion.WaterSolubility.csv`

Below the file is shown in Notepad++.

Figure 33: Example Normalised output shown in Notepad++

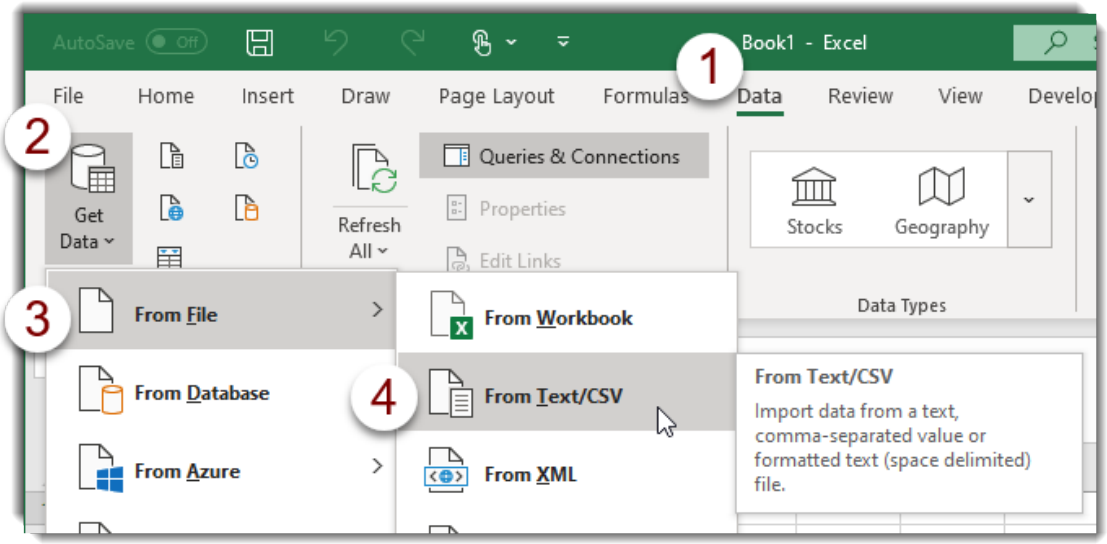


We can see that the file consists of lines terminated with the linefeed character (LF). The first line contains headers. The orange arrow is the TAB character, which is the column delimiter. As expected, an instance of the delimiter has been replaced by /t, which is underlined in red.

To open in Excel, from a workbook, follow the sequence shown in the screenshot below:

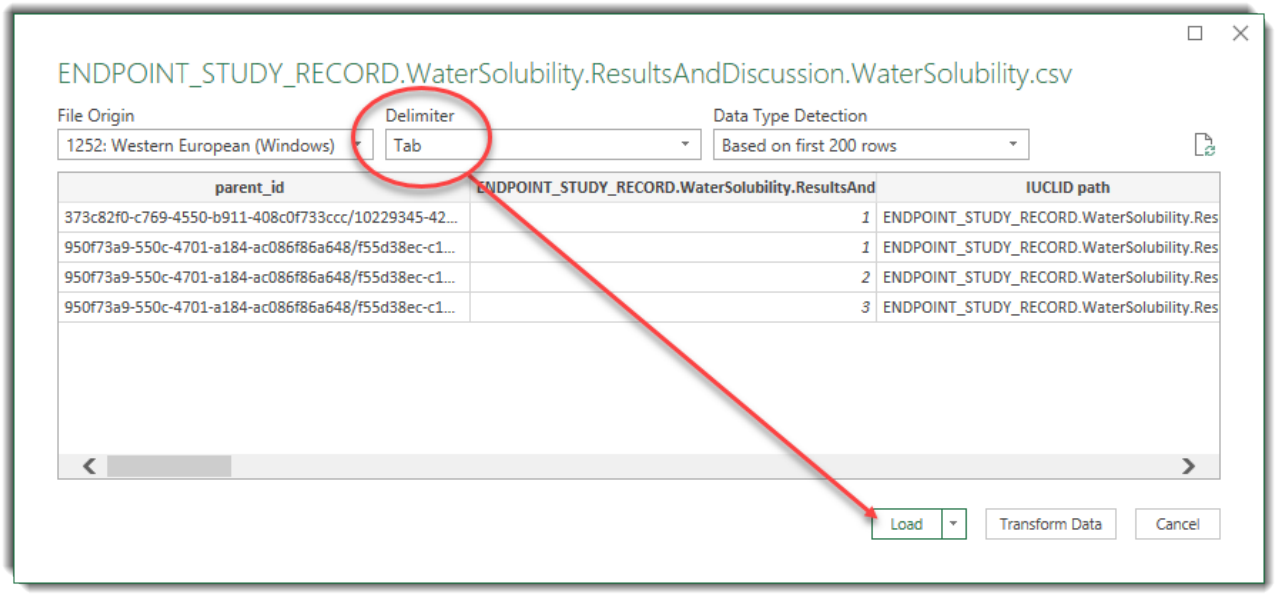
Data > Get Data > From File > From Text(CSV)

Figure 34: Open a text file in Excel



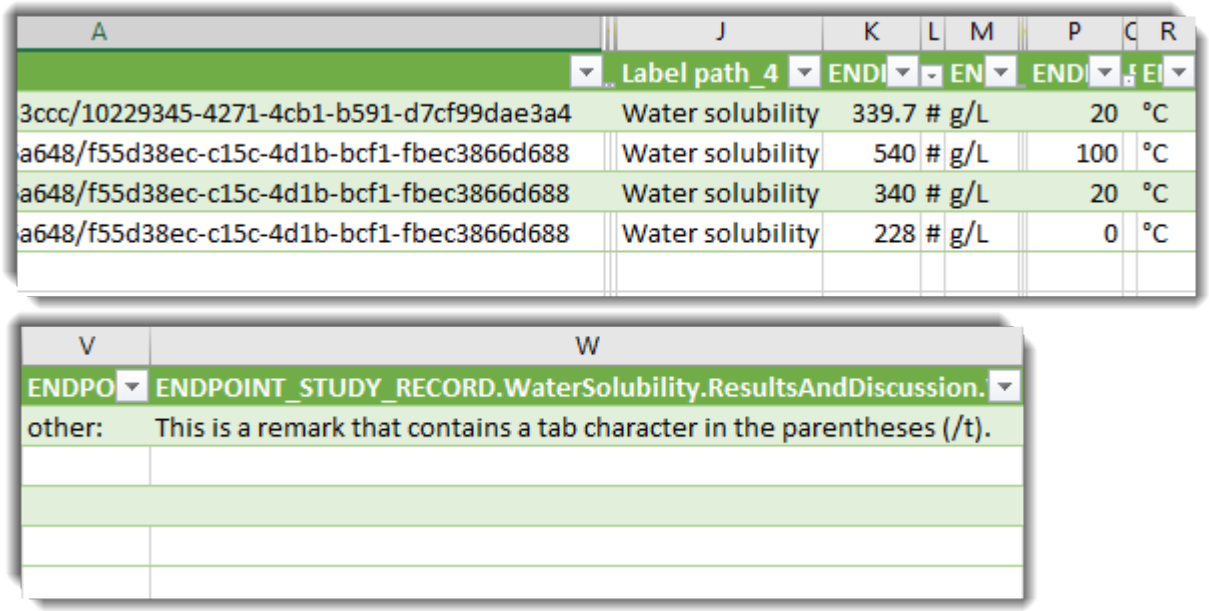
Check that the delimiter is TAB, and then click on Load.

Figure 35: Parse TAB delimited text in Excel



The first column (A) contains the document and dossier UUIDs. The data shown above has been highlighted in the view below. Note the dossier UUIDs. There are three rows for one of the endpoint study records because it contains three rows in the data table for *Water solubility*.

Figure 36: Example Normalised data viewed in Excel



A	J	K	L	M	P	C	R
	Label path_4	END		EN	END		E
3ccc/10229345-4271-4cb1-b591-d7cf99dae3a4	Water solubility	339.7	#	g/L	20		°C
a648/f55d38ec-c15c-4d1b-bcf1-fbec3866d688	Water solubility	540	#	g/L	100		°C
a648/f55d38ec-c15c-4d1b-bcf1-fbec3866d688	Water solubility	340	#	g/L	20		°C
a648/f55d38ec-c15c-4d1b-bcf1-fbec3866d688	Water solubility	228	#	g/L	0		°C

V	W
ENDPO	ENDPOINT_STUDY_RECORD.WaterSolubility.ResultsAndDiscussion.
other:	This is a remark that contains a tab character in the parentheses (/t).

6.11.2. JSON output

What follows is an example of *JSON* output, using the same extraction definition as the example for *Normalised*. It is beyond the scope of this manual to give a full description of JSON. However, the code for the data in the repeatable block for the second of the endpoint study records in the example is shown below.


```
[{"block_index":"1",
"uuid":{"path":{"key":"ENDPOINT_STUDY_RECORD.WaterSolubility.ResultsAndDiscussion.WaterSolubility",
"label":"Water solubility.Results and discussion.Water solubility"},
"value":"9ec43327-9955-461d-b80d-897428e27a55"},

"KeyResult":{"path":{"key":"ENDPOINT_STUDY_RECORD.WaterSolubility.ResultsAndDiscussion.WaterSolubility.KeyResult",
"label":"Water solubility.Results and discussion.Water solubility.Key result"},
"value":true},

"Solubility":{"path":{"key":"ENDPOINT_STUDY_RECORD.WaterSolubility.ResultsAndDiscussion.WaterSolubility.Solubility",
"label":"Water solubility.Results and discussion.Water solubility.Water solubility"},
"lowerValue":339.7,
"unit":{"code":"1935","text":"g/L"}},

"Temp":{"path":{"key":"ENDPOINT_STUDY_RECORD.WaterSolubility.ResultsAndDiscussion.WaterSolubility.Temp",
"label":"Water solubility.Results and discussion.Water solubility.Temp.Temp"},
"value":20.0,"unit":{"code":"2493","text":"°C"}},

"RemarksOnResults":{"path":{"key":"ENDPOINT_STUDY_RECORD.WaterSolubility.ResultsAndDiscussion.WaterSolubility.RemarksOnResults",
"label":"Water solubility.Results and discussion.Water solubility.Remarks on result"},
"code":"1342","text":"other:",
"otherText":"This is a remark that contains a tab character in the parentheses (\t)."}}]
```


Note how the TAB character in the *Remark* is replaced by \t. The following replacements are made:

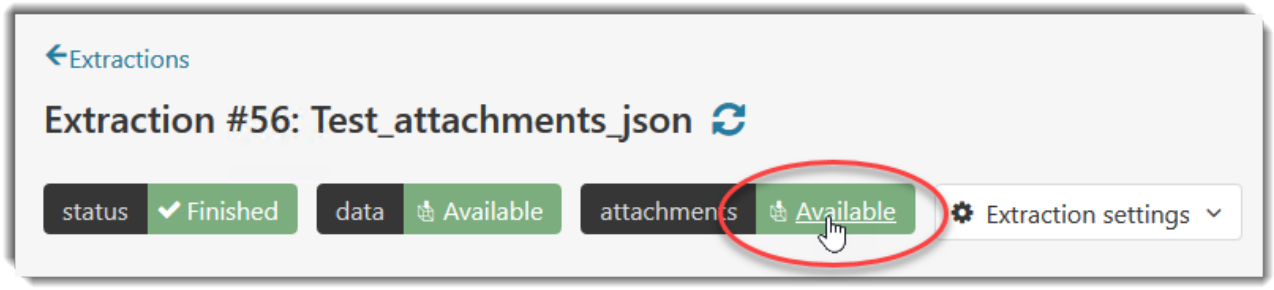
Table 1: Character replacements in text values in output in JSON format

Character name	Representation	Replacement in JSON
Tab	TAB	\t
New line	LF	\n
Double quotes	“	\”
Backslash	\	\\

6.11.3. Download extracted attachments

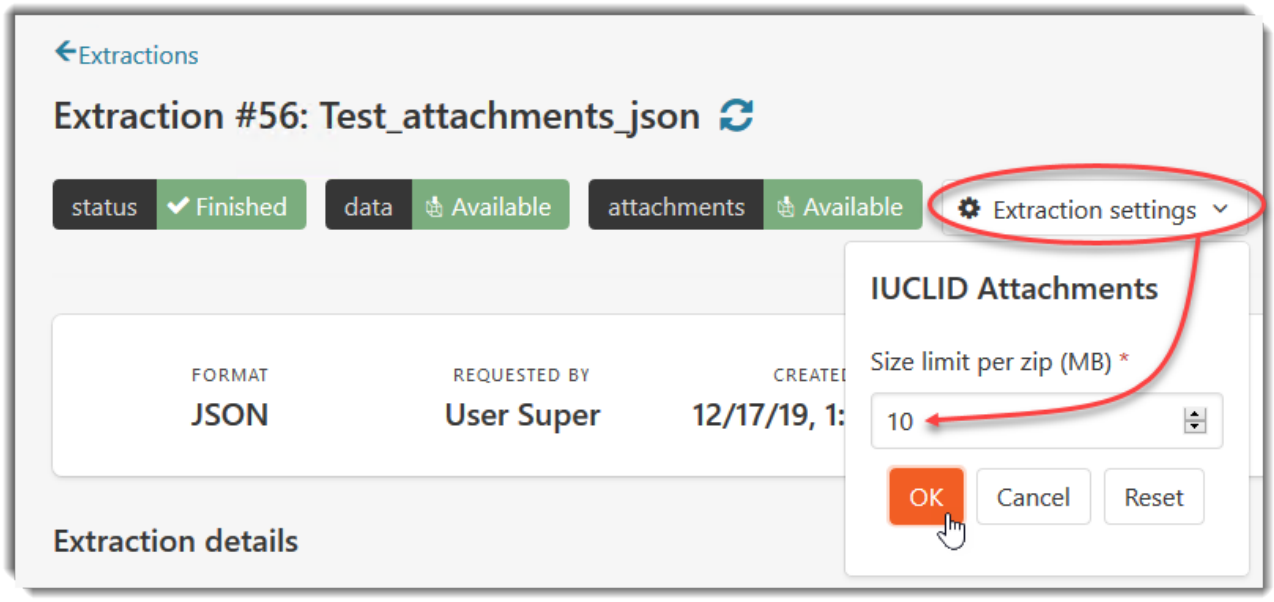
Where IUCLID attachments have been extracted, they are available for download under their own button with a link labelled *Available*.

Figure 37: Download extracted attachments



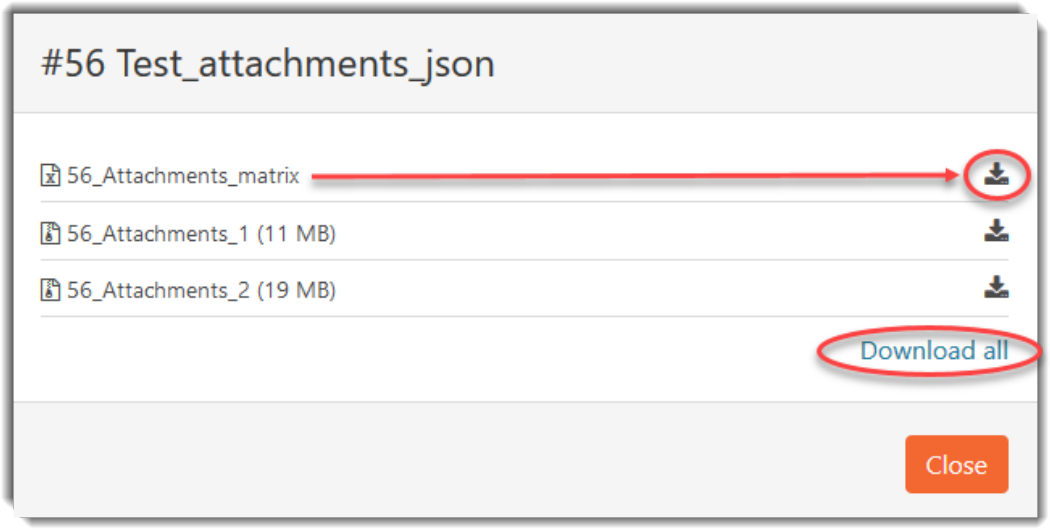
The attachments are downloaded as a single zip archive, unless that would exceed the size limit of the archive that is set under *Extraction settings*, as shown in the example below.

Figure 38: Limit the size of the zip archive in which attachments are downloaded



In the example, the attachments do not all fit in to a 10 MB archive, so they are made available in batches in zip archives, as shown below.

Figure 39: Download attachments in batches



The file named `<N>_Attachments_matrix.csv` contains a list of the attachments, and states from where within IUCLID they were extracted. The icon on the right, downloads the batch, or the matrix file. The option *Download all*, downloads the attachments and the matrix file in a single archive, thereby overriding the limit defined in the settings.

6.11.3.1. Attachments that have the same file name

This section describes the behaviour of an extraction in which there are attachments that have the same filename. The attachments can be in one or more dossiers.

Note that when a file is uploaded into IUCLID as an attachment, IUCLID calculates its MD5 checksum, which is a unique code that depends on the content of the file. IUCLID uses this to know whether files are identical. Only one copy of a file is stored per instance of the IUCLID database. Wherever the file is attached, there is a link to the single copy. Changing the name of a file does not change its content, so the checksum is not changed.

In an extraction, the attachments are output to a single folder in a zip archive. The filesystem does not allow duplicate filenames in a folder. Therefore, where the extraction would result in more than one attachment with the same filename, renaming is applied as follows.

If both the filenames and the checksums are the same, a single file is output that is renamed by appending to the name, the number of files affected. If the checksums are not the same, each file is renamed, by having a sequence number appended to the name. In both cases the number is preceded by the underscore character. The result is one or more files named in the form `<filename>_<N>` where N > 1.

Example

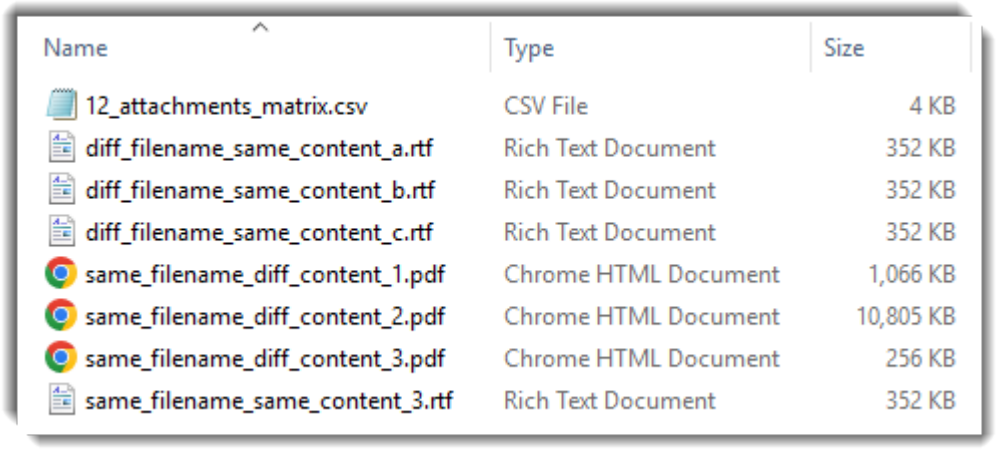
The files in the table below are included in an extraction. It does not matter where they are attached, and in which dossiers. Three have the same name, and three have different names. The final three have the same name as each other, but different content from all the others.

Figure 40: Files attached that have the same name and/or content

Content	Filename
A	diff_filename_same_content_a.rtf
A	diff_filename_same_content_b.rtf
A	diff_filename_same_content_c.rtf
A	same_filename_same_content.rtf
A	same_filename_same_content.rtf
A	same_filename_same_content.rtf
B	same_filename_diff_content.pdf
C	same_filename_diff_content.pdf
D	same_filename_diff_content.pdf

The extracted files are shown below.

Figure 41: Extracted files



Note how there are seven files derived from nine attachments. This is because three files had identical content, and were replaced by one file named:

`same_filename_same_content_3.rtf`.

The files with the same name but different content have been renamed in a sequence of 1, 2, 3:

`same_filename_diff_content_1.pdf`

`same_filename_diff_content_2.pdf`

`same_filename_diff_content_3.pdf`

The attachments with differing filenames are extracted to their own files irrespective of the checksum. There is no renaming:

`diff_filename_same_content_a.rtf`

`diff_filename_same_content_b.rtf`

`diff_filename_same_content_c.rtf`

The extraction results in some files with duplicate content, but no content is lost, and the location of the attachments can be seen in the matrix file. The filenames and the checksums in the matrix file are shown below.

Figure 42: Filenames and checksums in the matrix file

Name	Name (Unique)	Checksum
diff_filename_same_content_a.rtf	diff_filename_same_content_a.rtf	37731f9fe54cfd0652ddffe3bf9e552d
diff_filename_same_content_b.rtf	diff_filename_same_content_b.rtf	37731f9fe54cfd0652ddffe3bf9e552d
diff_filename_same_content_c.rtf	diff_filename_same_content_c.rtf	37731f9fe54cfd0652ddffe3bf9e552d
same_filename_same_content.rtf	same_filename_same_content_3.rtf	37731f9fe54cfd0652ddffe3bf9e552d
same_filename_same_content.rtf	same_filename_same_content_3.rtf	37731f9fe54cfd0652ddffe3bf9e552d
same_filename_same_content.rtf	same_filename_same_content_3.rtf	37731f9fe54cfd0652ddffe3bf9e552d
same_filename_diff_content.pdf	same_filename_diff_content_1.pdf	a73f45d88db4df24b63efc8d755011d9
same_filename_diff_content.pdf	same_filename_diff_content_2.pdf	ac60bb477aa98eaa6bf200f652aafd9b
same_filename_diff_content.pdf	same_filename_diff_content_3.pdf	c2a8c5f3a7db924a456e22ed00ee46d6

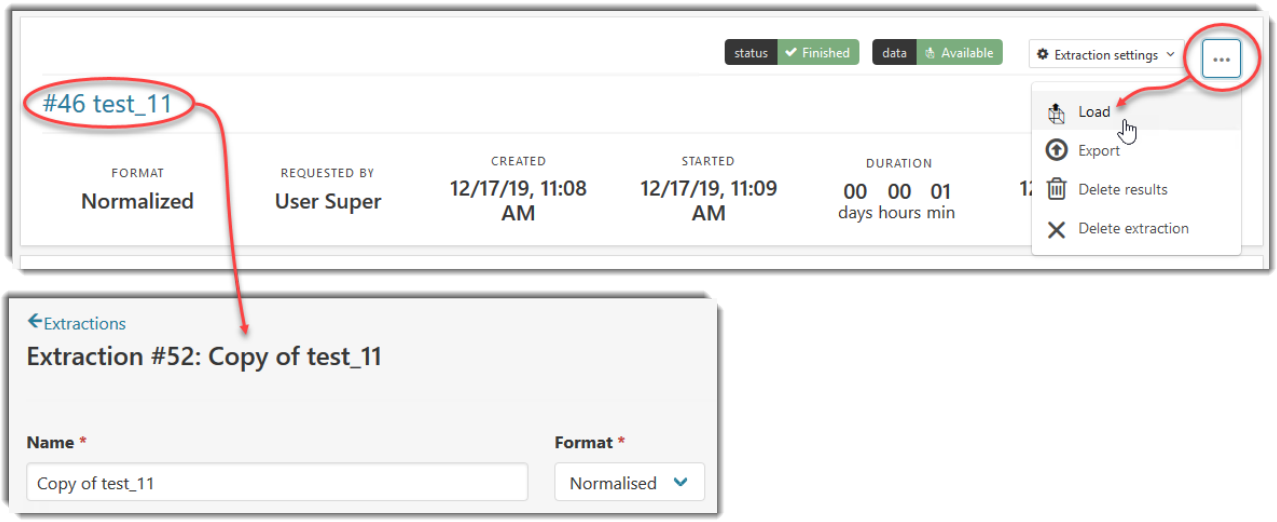
7. Re-use an extraction

An extraction can be used as a template for a new extraction. The definition of the extraction is copied into a new extraction using the option *Load*. The new extraction can then be renamed, edited as required, and run.

7.1. Select extraction to re-use

The option *Load* is available from a menu that is shown in the list item for the record on the *Home page*, and in the record of the extraction. The menu is a three-dot icon, as shown in the example below:

Figure 43: Load an extraction for use as a template



After *Load* has been selected, a new extraction is created and opened with all the same parameters, except the name has "Copy of" prepended to it, and the extraction number takes the most recent value.

8. Export and import a definition of an extraction

The definition of an extraction can be exported and imported. This allows extractions to be re-used, archived, and shared between users. The format of the export is JSON. Extracted data is not part of the export/import file.

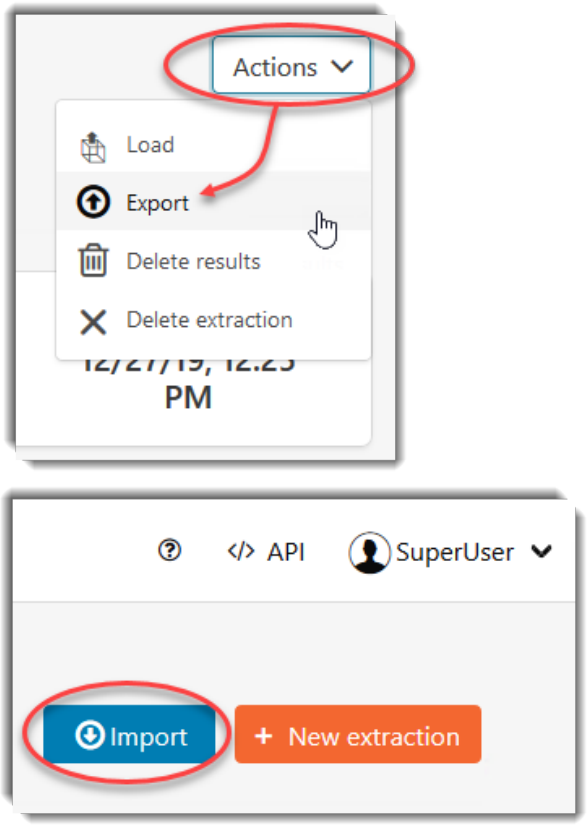



Figure 44: Export and import of an extraction