Text Analytics: 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.

European Chemicals Agency

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

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

Table of Contents

1. Introduction

Please also refer to the <u>Frequently Asked Questions</u> 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 Text Analytics.

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

2. Version 3.12.1 (December 2023)

This version includes needed technological upgrades, new fixes to improve the indexing process, and addresses needs expressed by the users.

NEW Update of Tika library in order to solve the issue of missing rtf. attachments in the index

This enables us to index all documents including images, even of low quality (<300 dpi).

Ref. 699874

NEW Query all the BPR Biocides dossiers at the same time in the Query builder

As was already the case for REACH Registrations, the users can now restrict their search to BPR dossiers only.

Ref. 804691

NEW Suggest possible values for metadata (file_registered_full_tonnage, file_registered_osii_tonnage, file_registered_tii_tonnage)

Ref. 810249

IMP Field labels suggestions are now filtered by only REACH registrations in 'Search by containing document' block.

Now, the selection of "All REACH registrations" as Submission type, limits the suggested labels to only REACH ones.

Ref. 675199

IMP -   The character separating the different values within the columns of the export file is changed from " $ " to "$"

This will facilitate data processing in the KNIME analytics platform.

Ref. 777262

IMP Update error message description - aggregation mechanism

New error message displayed: "The number of results to be processed by the aggregation mechanism is too high. Please redefine your query, to limit the results."

Ref. 699219

FIX Export does not work in aggregated mode

Ref. 929897

FIX The "exists" query does not work for external content attachment properties in capital

Ref. 933647

FIX Export in selecting one search scope item (e.g. only attachments)

Ref. 935045

3. Version 3.11.3 (July 2023)

This version accounts for the IUCLID changes, includes new fixes to improve query performance and the indexing process, and addresses needs expressed by the users.

NEW Refreshing creation of elastic search index aliases

This task increases query performance

Ref. 690342

NEW IUCLID User Groups – Implementation

By default, users will perform queries on all the user groups they have access to.

Ref. 692017

NEW Export > 150MB cannot be completed

Larger number of hits can now be exported. For data analysis, users often need to export the data. This is now possible for data sets of up to 60 000 rows

Ref. 690689

NEW Value field in 'Search by fields/attachments' tab, does not show content even if exist in the results

All fields of type "text" can be queried for a keyword, using field.value:"keyword"

Ref. 692087

IMP Enhance job "refresh user groups aliases"

This is to ensure that even when the system faces performance issues, the job which refreshes the groups aliases will be executed properly.

Ref. 695823

IMP Implementation of negations before of parenthesis

The user is now allowed to use the NOT logical operator before parentheses

Ref. 695672

IMP Assess the use of logical operators on query analyser

IUCLID 6

The OR operator cannot be used between joinLevel:relative clauses and the application will return an error.

Ref. 674480

IMP The character separating the different values within the columns to be changed from comma "," to " $ "

This facilitates data processing.

Ref. 663009

IMP Searchable properties page - Update the values – Implementation

The values for FILE_SUBTYPE (Full, OSII, or TII) have been added in the Query Builder and Search UI auto-complete functionality

Ref. 700077

FIX Query builder: values suggested for metadata are not always deleted

Ref. 689562

FIX Query builder: keyword fields (e.g. asset name) have a picklist option when the equal is selected

Ref. 808779

FIX Export does not return hierarchy

Ref. 808941

FIX Check metadata (e.g. submission number) - Missing metadata

Metadata (information coming from the submission systems) are now returned for all dossiers that should have metadata.

Ref. 696529

FIX Number of hits higher in aggregated mode than in flat mode

The number of hits in aggregated mode is now correct

Ref. 698800

FIX snapshot_uuid: alone should return an error message

It must be queried with for example field.snapshot_uuid:xxx

Ref. 662795

FIX User group aliases are pointing to both current and shadow indexes

During a shadow indexing process, only the old index should be queried.

Ref. 680017

FIX Relative query in asc and desc mode return no results(.ClassCastException: class java.util.ArrayList)

Relative queries in ascendant and descendant mode now return compatible results

Ref. 702343

FIX Asset owner JS role values are incorrectly reported

Ref. 677475

4. Version 3.10.1 (March 2023) including v3.9.1 and v3.10.0

Technological changes are the main drivers for this new version.

NEW Resizing of the Elasticsearch cluster

This task should speed up the performance and make the cluster more efficient.

Ref. 232503

NEW Upgrade to Java 11 (installation and testing)

This technological upgrade was requested by ECHA.

Ref. 233431

NEW Update Elasticsearch to 7.16.x

This technological upgrade was requested by ECHA.

Ref. 233431

NEW Suggest possible values for metadata (implementation)

Metadata values are suggested in both Query builder and Search UI

Ref. 660498

NEW Suggest possible values for basic properties

Entity types like for example "Endpoint Study Records" are suggested in the Search UI

Ref. 660513

NEW Suggest possible values for submission type

Submission types like for example "REACH Registration above 1000 tonnes - SUBSTANCE" are suggested in the Search UI

Ref. 660513

IMP Update message to: "Copy query to Search UI"

Ref. 677100

IMP Improve performance on new "relative" functionality

Ref. 673739 (mostly done in v3.9.0)

IMP Negations (NOT or !) cannot be followed by a parenthesis (technical analysis)

Ref. 659836

IMP Upgrade angular from 11.x to 13.x

Angular is our platform for building our User Interface. Its upgrade goes hand in hand with the Elasticsearch upgrade.

Ref. 659836

IMP Rename asset_type metadata property to authorization_type for BPR files

Ref. 675705

IMP Corrections on the query builder's functionality

Ref. 675796

FIX Inclusion of complex queries on TA's statistics

All queries are now taken into account in the application usage statistics

Ref. 665590

FIX The Export functionality does not work as it should

Issues related to the export of attachments hits have been fixed

Ref. 669912

FIX Users with TA_Simple role have no access to the External Content dynamic properties

TA_Simple users can now see the External Content dynamic properties (metadata) of all External Content public groups

Ref. 671380

FIX Query builder not working as it should when adding sets of criteria

Ref. 671381

FIX joinLevel:parent queries not returning results as expected

Ref. 688355

FIX Auto-complete functionality fails when updating queries

Ref. 676649

FIX TA version 3.9.0: "relative" is wrongly suggested by the auto-complete functionality

Ref. 682564

5. Version 3.9.0 (August 2022) – Not released

NEW Relative hierarchies supporting both ancestors and descendants

This functionality replaces the "ancestor" query with the relative "descendant" query. It also includes the new relative "ascendant" query which performs the opposite of the "descendant" query.

Ref. 232503

NEW Visualization of descriptions in the Query builder UI

To avoid confusion, both description and property names are now returned when selecting a property via the Query builder

Ref. 651209

IMP No suggestion in substance name when we type letters

Possible values for the label Substance.Substance name are now suggested

Ref. 664118

IMP Export mechanism is failing

In case the export process fails, a message is exposed to the end user as part of the exported file, to inform that the data included in the produced excel file is not complete.

Ref. 648401

FIX When selecting metadata property and then a non-metadata property the "metadata" word is erroneously transferred in query box

Ref. 664772

FIX Equivalent queries give different results depending on the position of the clauses

A pair of parentheses is to be added when OR clauses are used in order to make sure that the query runs correctly.

Ref. 664883

6. Version 3.8.2 (March 2022) – Not released

NEW Support searches for empty/null fields and dossier metadata

Thanks to the "exists" functionality, it is now possible to check whether for example the value of a field, the content of an attachment or a particular metadata (EC# for example) exists.

It is also possible to query the absence of values, contents or metadata, using: NOT exists{ }

Ref. 232943

NEW Basic search by containing document: suggest possible hierarchies

IUCLID Hierarchies are now suggested under the "Query builder", using "Hierarchy" as "Entity properties". The hierarchy starts by default with the "DOSSIER" entity, followed by a list of all entities to be picked up from.

Ref. 641008

NEW Query all the registration dossiers at the same time in the Query builder

Users can now query all REACH registrations via the "Query builder" without having to select all the possible submission types of registrations. The needed regular expression is automatically added into the Query box, after selecting "Entity Properties" → "Submission type" → "like" → "All Registrations"

Ref. 647004

NEW Display a warning when the user selects 'User groups' other than the default ones

The user is now warned if she/he is not searching in the default "IUCLID User group" which is "Prod_ECHA". The following message appear: "Your search has been performed on user groups different than the default ones!"

Ref. 651948

IMP A new window opens when clicking on the link "Searchable properties"

This prevents a formally run query to run again when clicking on this link.

Ref. 631056

IMP Query builder: Labels suggested are only either for Fields or for Attachments

IUCLID 6

In the Query builder labels are now only suggested for fields or for attachments (when labels for fields are selected no labels for attachments are suggested any longer and vice versa)

Ref. 644598

IMP Change search filter popup (UI)

User groups have been renamed "IUCLID User groups" for clarification, in the search scope. It was also re-positioned, so that it is clearer for the users that the "External content" tickbox is distinct from the "IUCLID User groups"

Ref. 649728

FIX Log4jshell TA impact analysis and actions

This covers all the tasks performed regarding the log4jshell identified vulnerability.

Ref. 655182

FIX Simple query using + operator only works one way

Now the following query works: +analyses report

Ref. 652960

FIX Results in aggregated mode are more than those in flat mode

The number of results in aggregated mode is now always correct

Ref. 654608

FIX External content results are not exported

Ref. 655681

FIX Searchable properties not showing the correct analyzer for all properties

The following searchable properties are now showing the correct analyzer for all properties:

"value.std" and "content.std" have WHITESPACE as analyser

"value.pattern" and "content.pattern" have PATTERN as analyser

Ref. 656835

FIX Equivalent queries giving different results

The position of the various query parts has no impact on the search results

Ref. 658411

FIX When there is a second parenthesis in the query it does not work

Search results are not influenced by superfluous parentheses

Ref. 659195

7. **Version 3.8.0 (February 2022)**

NEW [TA-1104] development of a fully packaged installer for industry (windows only)

IUCLID 6

An installer for Microsoft Windows is provided. The installation process now consists of providing a user in an Oracle database, configuring IUCLID, and running the installer. IDP authentication is supported.

Ref. 232402

NEW Log4jshell TA impact analysis and actions

The application was updated so that it is not affected by the Log4jshell vulnerability.

Ref. 655182

NEW Adjust TA indexing configuration for industry installations

The default configuration of the indexing queues has been revised and set to a rate that can be acceptable even for low-resource machines.

Ref. 232402

NEW Configure the user groups default values

An Administrator is able to configure the default IUCLID user group values via the "Properties" page.

Ref. 652084

NEW Add IUCLID "User Groups" information in the export file

The user can view, in the exported file "Query" sheet, the user groups being selected for the ongoing search

Ref. 639614

IMP By default, users only get results from default user groups

The user can view, in the exported file "Query" sheet, the user groups being selected for the ongoing search

Ref. 639614

IMP Revisit the use of logical operators upon search

The query language defined in Text Analytics is now stricter to protect both users and the system from syntactically erroneous queries followed by unexpected results. This means that some previously used syntaxes are not supported anymore, and they need to be written in a more verbose but syntactically correct way. Examples provided in Appendix D of the manual.

Ref. 650977

IMP Rename Basic and Advanced search

The "Basic search" window has been renamed "Query builder" and the "Advanced search" window has been renamed "Search", in order to clarify each window's specificity for the user.

Ref. 639614

FIX Label suggestions have occasionally wrong section when editing the input keyword

Ref. 647005

FIX Incorrect results when we have parenthesis AND OR in the query

Ref. 651439

FIX Reset Button in Search scope non functional

Ref. 655181

FIX Filtering misconfiguration options error message is erroneously displayed

Ref. 652245

FIX Aggregation options do not load

Ref. 652676

FIX Reset button sets the user groups to select instead of the default ones

Ref. 652760

8. Version 3.7.2 (13 December 2021)

IMP Impact assessment due to the IUCLID Oct (2021) release

Text Analytics was made to be compatible with the new format of IUCLID 6 that was released in October 2021. The new format is version 6.

Ref. 638184

9. Version 3.7.0 (06 September 2021) – Not released

IMP Improvement of the Basic and Advanced search

The Query Suggest and the Search tabs were replaced by the tabs Basic Search and Advanced Search. Basic Search allows structured Boolean logic to be created in a way that is limited, but forces the correct syntax. Queries created under Basic Search can be copied over to the Advanced Search for refinement. The Advanced search includes a syntax checker, and an auto-complete function that offers selections of terms to insert that are syntactically correct in the current context.

Ref. 639537

IMP Prevent users from executing queries when the state of the system may result in incomplete results.

If it is known in advance that the state of the system would result in incomplete results, certain queries cannot be run.

Ref. 233324

IMP Searchable fields - type of analyser used

Column added to right that states the type of analyser used.

Ref. 233585

IMP Upgrade Wildfly to v19.1

Wildfly is the application server in which TA runs. It was upgraded to v19.1.

Ref. 635328

10. Version 3.6 (06 March 2021) – Not released

IMP Change language codes from uppercase to lowercase

For compliance with ISO 639-1, the language codes were changed from uppercase to lowercase.

Ref. 233187

IMP respect fully IUCLID security model (implementation)

Searching in specifc Security Groups was added under *Search scope*. For more information about Security Groups, see the user manual that is available in the interface of IUCLID. See the section on Instance Based Security (IBS).

Ref. 233392

11. Version 3.5.0 (07 December 2020) – Not released

IMP Update Oracle 12c to 19c

Compatibility with Oracle database version 19c was introduced.

Ref. 233389 (TA-1623)

IMP Improvements of the export functionality (handling of hierarchies)

Within Dossier metadata for the export of either fields or attachments, it is possible to select hierarchies and level of hierarchy.

Ref. 233392 (TA-975)

IMP Industry version (IDP), implement user roles

Compatibility with user authentication that uses IDP was introduced.

Ref. 233380 (TA-1615)

12. Version 3.4.0 (19 August 2020)

IMP Improve the handling of the hierarchies in the index and the query language (implementation)

The syntax of `joinLevel` was adapted to enable a search for a parent document by its location within a hierarchy.

Ref. TA-1399

IMP Searchable properties should be collapsible

The way in which the searchable properties are displayed was re-organised and the terminology was made more consistent with the rest of the interface.

Ref. TA-1636

IMP Select all button in export pop-up

A function was added to the export window to allow all options to be selected as once.

Ref. TA-1645

IMP Improve the visibility of field labels in the results section

The field labels on the results page were moved from below field values to above them, and the font was changed to contrast with that of the values.

IUCLID 6

Ref. TA-1383

IMP Performance tuning of Text Analytics

The performance of Text Analytics was analysed which allowed changes to be made to improve it.

Ref. TA-1533

13. Version 3.3.0 (21 June 2020)

FIX Duplicate rows returned in the export file and other remarks on snapshot_uuid

Ref. TA-1484

FIX Exception upon uploading external content with mode 2 - Upload works OK

Ref. TA-1488

IMP Information Disclosure and Verbose error messages. The response to bad HTTP requests has been made generic and thus more secure.

Ref. TA-1443

IMP Searches fail - IUCLID Submission types incompatibility. Introduced compatibility with the dossier type SCIP.

Ref. TA-1576

IMP Update Wildfly version. The version of Wildfly is now 13.0.0. The old version was 10.1.0.

Ref. TA-1502

13.1. Indexing

IMP Indexing when IUCLID Server is down. Confirmed that stopping and starting the IUCLID application does not affect indexing.

Ref. TA-1564

IMP Functionality to gracefully stop the indexing process. For large numbers of dossiers, indexing can take so long that it is necessary for it to be paused whilst maintenance activity is carried out on the server. The ability to do that has been added to the administrator interface. See the sub-tasks below.

Ref. TA-1427

13.2. Security

FIX User can view IUCLID content although he has no access to IUCLID when user is suspended

Ref. TA-1267

IMP Security assessment findings (part 2)

Ref. TA-1553

Sub tasks of TA-1553

IMP Security Enhancement on HTTP Response Headers

Ref. TA-1440

IMP Insufficient access control - IUCLID end point record hierarchies

Ref. TA-1549

IMP Insufficient access control - Upload of new help file. Access to this functionality can now be limited per user.

Ref. TA-1548

IMP Retrieval of content for user that has no IUCLID account

Ref. TA-1467