Webinar: Data Uploader: converting your chemical data into IUCLID format
Questions and answers

This document is based on the questions received before and during the webinar organised on 19 October 2022. Editorial changes have been made to improve clarity and similar questions have been combined.

The European Chemicals Agency does not accept any liability regarding the use that may be made of the information contained in this document. Use of the information in this document remains the sole responsibility of the reader.

For the most up-to-date advice, contact us or refer to our support material

#	Question	Answer
1	According to the manual "With the stepwise approach the user shall be able to upload a set of entities (e.g. REFERENCE_ SUBSTANCE, LEGAL_ENTITIES, CONTACTS, SITES…)" Is it valid for the endpoint related fields as well (e.g.) just add one new study to the existing dataset?	The scenario mentioned, add an Endpoint study record to an existing Substance dataset will be included in the next release of the Data Uploader.
2	DataValidation node: is the field's size is validated according to IUCLID requirements? I have had an issue with one field that is 255 long in Uploader json and passed the validation node but for Iuclid is max 10 size field, so the node "upload" gives a truncation error. For Inventory Entry field in the json created by Data Uploader the max size is 255, it passes the validation node without problems, then the Upload node gives an error because in IUCLID is max 10 characters size.	The Data Uploader is using the IUCLID document definitions as retrieved from the IUCLID instance. Due to this, ensure that the EC number (a numerical identifier) you are uploading uses the correct convention. In order to retrieve the list of errors triggered by the DataValidation node, the user needs to check the DataValidation node's second

		outport. The "green" dot on the DataValidation node signals that the node executed successfully, and the user needs to actively check for errors. Please check Substance with an endpoint study record use case, section 2.4.2. DataValidation Node's out ports for more information.
3	Did I get that right, that there needs to be already a basic substance entry in IUCLID to upload data into it, or can this first entry also be created by the uploader?	No, you can create the substance dataset with the Data Uploader. With the current Data Uploader version, you actually need to create Substance dataset with the Data Uploader in order to correctly upload it to IUCLID.
4	Do you have an example with the Incremental Update Node? What does this mean? Is there a way to compare the data to be uploaded with the data in IUCLID and then decide what to do? Is this related to the other UUID mapper node? I am not clear about that - it seems you now have two types of uploads - one with I6Z files, one via the REST API?	Incremental data upload is the process of generating and importing to IUCLID a set of IUCLID documents and then populating these documents with additional data, through separate consecutive execution cycles of the Data Uploader KNIME workflows. The incremental upload is supported by the JSON API of IUCLID where the selected document fields along with any references to existing UUIDs are posted to IUCLID. To achieve that, during the preparation of the data (input CSV files) the user would have to include in the dataset the IUCLID UUID of the referenced document as this was generated at a previous step or existed already in IUCLID. Substance with an Endpoint Study record use case document, section 3 and Annex III provide more information about the Incremental data upload.
5	Have you tested the software (migration to IUCLID) with a large database with many fields?	So far, we tested with relatively large datasets and haven't experienced any issues.
6	I installed the package two days ago. There are a number of errors due to wrong settings and missing extensions, e.g. Counter generator, that cannot be installed. Eventually the package should be delivered without your local settings plus the required extensions?	We recommend a standard KNIME installation with the standard nodes (KNIME version 4.2.2). You can try with a fresh KNIME installation and if it doesn't help, send us

		screenshots through our contact form. (Under 'Request type', select 'Data Uploader').
7	Is it possible to upload specific application conditions like Working Condition Scenarios (WCS)? They are not pointed out in the presented Excel File.	In the Data Uploader workflow, DocumentSelection node specifically, you can select different working contexts depending on your needs. This will help you display and select fields you want to migrate data to. When you upload your created dataset in IUCLID the working context might be adapted/changed.
8	Is there a resource where to find possible picklist-values for our entries? Could these be extracted from the generated Excel-Files or are those intended for manual reference only?	Based on the list of selected fields the DocumentSelection node generates a set of Excel files corresponding to tables specified in the generated IUCLID schema. The same set of tables as generated by the DocumentSelection node (in CSV format instead of xlsx) will have to be populated with data to be uploaded to IUCLID. Tables generated by the DocumentSelection node are in Excel format to be able to display the list of IUCLID codes and phrases for columns corresponding to IUCLID fields with the [Picklist] type. The applicable list of values (codes and phrases) can be displayed by clicking on the arrow beside the column's name or by "Unhide" sheet functionality.
9	Is there any presentation for greenhorns? I do not know how and what to do. This presentation was very much confusing.	You can rewatch the presentation if it helps. In case of any issues with the installation we advise you to contact us. (Under 'Request type', select 'Data Uploader'). Our contractor will be able to provide further support. We are also updating our installation instructions document with more detailed information. You can also check the Reference Substance and Substance with an Endpoint Study Record use cases which are part of the Data Uploader package for more information.
10	Just a small clarification: the data uploader can be used with a cloud version of iuclid, but it seems that the data extractor can be used just with a local installation of Iuclid. Is it correct?	The current version of the Data Uploader cannot be connected with the IUCLID Cloud.

		We are planning to enable this possibility in the future.
11	Not sure if I missed something from the presentation, but was there anything additional to what is already explain in the two use-case manuals?	Reference Substance and Substance with an Endpoint Study Record use cases were the basis for this webinar. In the manuals you will find explanation in more details.
12	Thank you very much for all your clarifications provided. However, from my personal point of view it would be very much practical for the users if the data model of the Data Uploader would always follow the OECD HT not as it is implemented now (depending on the fields selection)	OECD HTs do not have a data model (they are only a set of requirements for specified endpoints). What gives them a data model is IUCLID, which is the reference-application which implements the OECD HTs. The Data Uploader is using the IUCLID structure to organise the transformation of data. The OHT specifications mentioned above come directly from the latest IUCLID format (6.6 at the moment). The headers and field positions are the same as in IUCLID.
13	The current visualisation schema introduced in the version 1.0 is very difficult to visualise on any normal screen, if you have more tables that you have in your example. It is nice to hear that you are working on the visualisation. It would be very helpful to have dynamic tables that a user can move on the screen to see the relations better.	Thank you, we are working on it.
14	The installation manual cites different version of java Zulu to enable the visualisation (jre and jdk) in different parts of the manual so it is very hard to understand which one should be used. page 11 ""fx-jre8.0.332-win_x64' in a desired directory" page 12 the fx jdk version is mentioned In my case only jdk version worked properly (the visualisation plugin was enabled)	The recommended version is fx-jre. We will improve Installation Instructions documents. Please note that fx-jdk can also be used but you need to ensure all paths are correct.
15	The KNIME version you support is two years old. What is the reason for not supporting one of the latest versions?	We are currently working on a new release compatible with KNIME 4.6.
16	The structure/data model of input tables from IUCLID Data Uploader is different from the output of IUCLID data extractor (for which data model is simpler). When are you planning to bring the Extractor software to Knime? Are you planning to harmonise the data models used by these two software? The Extractor works fine also without Knime nodes but that data structure that you get as an output is a little different (higher level of aggregation) from those of the Uploader. Maybe it would be just nice to harmonise the data structure of both.	We are planning to develop Data Extractor also as KNIME nodes in the future. We are currently analysing how to bring both solutions closer together. Based on the outcome of this analysis further steps will be decided.
17	Topic: Data Schema (json) || Is it possible to detect all columns in a table that are foreign keys?	Yes, you will be able to detect all columns in the Data Schema (json). An example is "foreignKeyTo": "LITERATURE".

18	When are you planning to release the next version of IUCLID Data Uploader? Are you planning to provide a full compatibility with the workflows written with the current version? Are you planning to keep the data model/csv tables/data model the same in the next release?	Data Uploader's next release planned for end of 2022/beginning of 2023 will be compatible with KNIME 4.6. There will be some improvements, but the data model will be the same.
19	Will it be possible in the next version of data uploader to upload data into existing IUCLID datasets (e.g. by using the UUID)? This would be an important feature to consider.	We are currently analysing the so-called incremental upload and will be implementing various scenarios in Data Uploader's future releases. Substance with an Endpoint Study record use case document, section 3 and Annex III provide more information about the Incremental data upload.
20	Will there be a visual instruction on the generation of the csv files and mapping in the future? If it is the "main part" shouldn't an example be part of the webinar?	Substance with an Endpoint Study record and Reference Substance use cases included in the Data Uploader package available for download provide information about the Data Uploader workflows, nodes configuration, their outputs as well as tips for data preparation (Annexes III and IV). Use cases are complemented with Data Uploader workflows and respective Data preparation workflows.
21	Your example does not include MaterialsAndMethods table. If a field of this table (e.g. GLP) will be selected, the whole structure of the EndpontStudy schema will not be valid any more. For a large workflow it is quite hard to change many foreign keys just because one field has been added. When the table MaterialsAndMethods is missing, the other tables' names, e.g., Endpoint.MaterialsAndMethods.Methods,SpeciesStrain are confusing because it seems to be a child of non-existing tables. I understand that it is the user responsibility as you said to check the foreign keys. I just wondering why adding just one even non so important field as GLP compliance the existing workflow should require so big modifications in the tables' relations. It is not very user friendly...	We advise our users to check "Visualise IUCLID schema" that represents the IUCLID schema as a graph where the nodes of the graph are the tables in the schema and the edges their in-between relationships. The name of each table in the graph is the name of the relevant table (xls file) generated through the DocumentSelection node. The information displayed under each table in the graph is its name, the primary key (in blue), the foreign key (in red) and all the user selected fields that are marked as required (in black) in the document definitions. The full list of the selected fields and their attributes are available in the JSON representation of the IUCLID schema.

		The data schema generated by the DocumentSelection node is practically a specification of the full set of data tables required to store the input data. Columns, including their metadata (data type, mandatory or not, allowed values etc) are defined for each table. The number of tables and their contents is determined by the IUCLID document and field definitions, available via the IUCLID API.