USING IUCLID FOR WORLDWIDE EXCHANGE OF CHEMICAL AND TOXICOLOGICAL INFORMATION

Stefan Scheer, Remi Allanou
Institute of Health and Consumer Protection, European Commission – Joint Research Centre Ispra, Italy

Keywords: database management; chemical database system; IUCLID

Abstract: A database management tool (IUCLID) has been created in order to provide with administering chemical and toxicological data sent in structured form due to existing EU legislation. This tool also offers – beyond the normal dataset administration functionality – mechanisms for data fusion, data reproduction and data deployment. Thus IUCLID is used not only by who has to receive submissions of that kind but also who has to produce such submissions. Hence this product is used by whoever is involved as stakeholder in the current legislative process, and even beyond that it has been recognized successfully. Consequently it was the worldwide acceptance that helped in promoting this software product ahead of its original purpose and to establish a network of exchange.

1 INTRODUCTION

Since 1993 the European Commission is operating a database on chemicals, called IUCLID (International Uniform Chemical Information Database). Due to EU legislation this database gets input from chemical industry in electronic form containing general chemical information, toxicological information and company-confidential information on production and uses of certain chemicals (see Heidorn et al.).

Major aim of collecting was to get an overview of those substances of concern being on the market (see Allanou et al.). Within further steps, like the EU Risk Assessment Programme, this collection database served as basic data delivery.

The management of the collection process required the development of a specific database management software which is also called IUCLID. This software is managing those submission sent by industry thus building up a local database. However, IUCLID has also been designed to modify its input and to create further or new data.

As all kinds of submissions have to be sent in electronic format, it was obvious to use IUCLID in a similar way at all the premises from which such submissions should come from. Consequently and over the years a network of collaboration has been created in which chemical data are exchanged by "speaking in the same language".

2 IUCLID AND ITS SOURCES

2.1 The Submission Concept

Basic concept concerning input to IUCLID is the concept of submission. Technically, a submission is a file that contains information on a certain chemical substance according to judgements made by the company producing a certain chemical substance.

First, a submission consists of header information which uniquely identifies the submission through three parameters: 1. the submitter coordinates, 2. the submission date, and 3. a reference to the submission object. The system will not accept two submissions having exactly the same identifiers. Thus through changing one of these parameter values, two quite similarly looking submissions could be accepted. On the other side, if a duplicate (with exactly the same parameter values) comes in, an overwrite procedure will start.

A submission can be a normal submission or a so-called template submission; they distinguish according to their reference object that, usually, is a pointer to a chemical substance definition or a template name, respectively. A template submission is a way of neutral collection of information that, as

Scheer S. and Allanou R. (2004).
USING IUCLID FOR WORLDWIDE EXCHANGE OF CHEMICAL AND TOXICOLOGICAL INFORMATION.
In *Proceedings of the Sixth International Conference on Enterprise Information Systems*, pages 541-544
DOI: 10.5220/0002615105410544

a whole, can subsequently be attached to any normal submission.

Besides of the header, a submission consists of "footer" (properties of the substance the submission refers to) information that can be

- Normal submission content (see 3.1 for that)
- A reference to the contents of an existing submission.

The three ways of referencing an existing submission are determined as follows:

Type 1: An incoming submission has the same parameter values as a submission that already resides in the database. In this case the system will detect the apparently redundant information and launch an overwrite procedure.

Type 2.1: An incoming submission has a "footer" that is already part of another submission residing in the database. In this case the system checks the existence of the other submission. Obviously, as it is a pointer to contents written by a different submitter, contents of that kind cannot be further edited; however, the contents can be viewed.

Type 2.2: Optionally an incoming submission has "footer" information pointing to a generally re-usable submission (a "template submission"). In this case, too, the system checks the existence of that "template". All information out of it will be added to the submission; however, these parts are flagged accordingly and cannot be further edited.

Interestingly, while the Type 2.1 case of referencing allows for submitting complete dossiers with a minimum of effort, the Type-1 case had been implemented in order to let organisations update their originally sent information themselves. Implementation of type 2.2 references especially intend to re-use separately collected information.

2.2 The Local Database

As IUCLID in also stands for having a local database available, a dataset administration module is needed to manage the submissions as datasets. Hence this module lists up all available datasets primarily identified by the reference object (see 2.1) and makes them accessible. The currently selected dataset will also display the additional two submission identifiers – submitter coordinates and creation date. In addition, further information on "Type 2" references will be displayed if pertinent.

Depending on the user's rights a currently selected dataset can be edited (read & write permission) or viewed (read permission only). In any case the system will make accessible the various chapters where all the input is captured. The availability of chapters for reading or writing depends also on the current view that has been initially chosen by the user. Limited views will not display certain chapters. According to various chemical awareness programmes specific views can be chosen, though, the user can define own views as well.

3 INFORMATION CAPTURING

3.1 Capturing Information

Any type of information is kept in chapters with fields. Thus the system aims at structuring as much as possible incoming or new information. An entry field can be one of these types:

- Glossary-type
- Text-type

In comparison with a text-type entry field which allows the user to write any ASCII text as input, a glossary-type entry field forces the user to choose from a pre-defined list of (glossary) values that pops up when a double-click in such a field is made.

The capturing of any additional information is done in two ways: 1. a glossary value of type "other: " which, once chosen, allows to add additional text-type input, 2. so-called "freetexts" keep context-dependent pieces of information.

The latest version of IUCLID offers a variety of freetext types, such as "RM" for a remark, "RE" for specifying a literature reference, etc. Worthwhile to be mentioned is the possibility to attach external documents (and their contents) to a chapter through an "AD"-type freetext; its use allows the uploading of an external file. In principle, as many freetexts as applicable could be added to a record.

3.2 Multiply collected Information

The straight approach of chapter divisions in IUCLID is further broadened by collecting information "in parallel". This is called the record principle. Accordingly it is allowed to put a second, third record etc. to a certain chapter.

This makes sense thinking of some plausible reasons:

- The submission reports on more than one fact on one and the same issue, like for example two different tests.
- The submission reports on several opinions.
- The submission combines data from various sources.

Separate data handling functions have been implemented for that particular purpose: insertion and deletion of a record, navigation between

chapters and records, and the display of counters pointing at the currently displayed record.

4 INFORMATION MANAGEMENT

4.1 Self-Reproduction

With any IUCLID installation one can create new datasets ("external reproduction") and create new records within existing datasets ("internal reproduction"). Both cases make the local database grow in size thus decisively augmenting IUCLID's functionalities.

An external reproduction asks for insertion of those three identifiers (as described in chapter 2.1), and displays all chapters with empty input fields. Once the user has filled in the minimum required, such a new dataset can be saved. Likewise during the loading of an incoming submission the system checks whether none of the locally residing datasets has identical identifiers.

Similarly, an existing dataset, once opened for editing, can receive a new record for a certain chapter (see also 3.3) when the foreseen "add" function has been activated. This function will provide the user with a chapter with empty fields which, when saved, will be added to the overall number of records for that particular chapter.

4.2 Data Fusion

IUCLID-type data fusion means to merge information coming from different datasets (i.e. sources). In order to avoid semantic mix-up only information referencing the same (i.e. pointing at the same substance) can be considered for merge. The only exception in this context is to merge also with "neutral" information as it is kept in template datasets (see 2.1).

In functional terms data fusion copies parts of one or more sources onto a destination dataset thus augmenting the number of records if the destination dataset is not empty at the time of merging. The destination dataset is the only one in this context that receives an internal reproduction.

4.3 Flagging

Users can manage datasets under different conditions and in various contexts:

- Data has been collected or is used for different programmes
- Data has been generated by a different organisation and has to be re-used somehow
- Parts of a dataset are confidential and cannot be published
- Work on filling in data is in an on-going status and cannot be considered to be final.

In order to better deal with those conditions and to give the user mechanisms to distinguish relevant input from non-relevant, a flag and reliability (attaching a "degree of trust") mechanism had been implemented. Both mechanisms work on record level so that even for different records of a single chapter a distinction can be made.

Both flagging mechanisms act as filters when the user invokes one of the data exchange (import / export) or the data publication (print) functions. In this way, for example, records marked as "confidential" could be excluded from publication.

5 INFORMATION SECURITY

The availability of functions like merging, referencing, editing, exchanging of datasets requires a detailed plan of distribution of rights if security measures should be applied on all kinds of data. Security issues relate to alteration and re-use of data from other organisations, and are mainly driven by business-associated concerns.

Currently the following security policy is applied:

- Create a submission and point at dataset details created by another organisation according to Type 2.1 (see 2.1). Consequently the details that have been pointed at can be viewed but not edited.
- Create a submission and take over information from a template dataset according to Type 2.2 (see 2.1). Similarly to the first case those records which originally come from the template dataset, are not editable and are flagged as such.

Augment a dataset with information from another by merging parts of the other dataset into the destination dataset. In this particular case the merged parts become editable; during the merging, however, a "SO"-type ("source") freetext is created and attached to all those records which are from the source dataset. These additional freetexts indicate the coordinates of the organisation that is responsible for the source dataset.

Basic security principles are indirectly defined by incorporating a pin code during installation of IUCLID; consequently write mechanisms are limited to one's own or to partner organisation. As an aim of such a pin code, a system of mutual acknowledgement is set up.

6 INFORMATION DEPLOYMENT

6.1 Data Exchange

IUCLID installations can mutually exchange data sets using the export/import functionality.

Technically, the exportation of a IUCLID dataset means to encrypt parts of its contents on an ASCII file. The user can customise this if applicable.

The importation of a IUCLID dataset means to launch an internal (customised) function that reads the contents of an encrypted file and stores parts of it as an add-on to the local database.

It can thus easily be seen that through the bi-lateral acceptance of datasets and their re-use the IUCLID community is continuously expanding. Consequently the ever augmenting knowledge on chemical substances and in particular the conclusions drawn from toxicological properties and results are deployed to an ever increasing community of interested users.

6.2 Data Publication and Use

Beyond the transferring of information among partners, IUCLID also stands for publication of its (consolidated) internals. The merge function allows the generation of summaries of contributions concerning a certain substance. Such harmonized versions are of particular use when external output is required; this could implicitly be done by generating a safety data sheet (which contains all relevant and acknowledged information concerning the safe use of a chemical substance), or explicitly by searching the entire database.

Being a relational database, IUCLID allows to search for every single detail of its contents. A number of pre-fabricated searches are offered by IUCLID; the user might add own search queries if necessary, and might also define the layout of search results. Searching the database is limited to structured information within entry fields, while freetext-type information, obviously, cannot be searched.

Coming back to its original purpose the IUCLID collection database is also serving as basis for doing further risk assessment for EU legislation. As a result a couple of risk assessment reports have already been published (see, for example, Hansen et al.). Such reports contain a final conclusion on a chemical substance and determine whether there is a "need for limiting the risk" or whether there is a "need for further information and/or testing".

The use of IUCLID within chemical industry has been extended towards internal use, too; companies can use IUCLID and its contents by integrating it into internal business structures.

7 PRÉCIS AND OUTLOOK

Much experience with IUCLID has been made over the past 10 years. Although designed for a particular purpose (see also Heidorn et al.), the software shows some particularities and assets that made it accepted and acknowledged by a worldwide community.

What is currently modernized is the way the data exchange file looks like. Originally only data exchange between IUCLID installations had been foreseen. Modern business principles, however, demand a much higher integration of IUCLID with other business applications. For this purpose one has to understand which type of information is kept in the export file in order to re-use this information differently. As state-of-the-art solution the use of a XML Schema is proposed.

A XML layer as the main data exchange format will also foster a higher degree of customisations of the user interface. Allowing only extensions to the existing data capturing facilities, it will become quite easy to add user or business driven information to the worldwide exchange of data. Purpose of this exercise is to get IUCLID even more accepted and better integrated and thus further contributing to the success of IUCLID.

REFERENCES

Heidorn, C.J.A., Hansen, B.G., Nørager, O., 1996. *IUCLID: A Database on Chemical Substances Information as a Tool for the EU Risk Assessment Programme*, J. Chem. Inf. Comput. Sci., 1996, 36m 949 – 954.

Allanou, R., Hansen, B.G., Bilt, Y.v.d., 1999. *Public Availability of Data on EU High Production Volume Chemicals*, EUR Report 18996 EN, Joint Research Centre. Ispra.

Hansen, B.G., et al. (editors), 2001. *European Union Risk Assessment Report on Hydrogen Fluoride*, EUR Report EUR 19729, Luxembourg.

Heidorn C.J.A. et al., IUCLID: An Information Management Tool for Existing Chemicals and Biocides, J. Chem. Inf. Comput. Sci. 2003, 43, 779 – 786.