Special issue on Tox-Hub: an internet-based platform facilitating access to toxicological information sources.

Editorial

Dear friends of HEROIC,

In this special issue we would like to introduce you the **Tox-Hub platform** through which HEROIC intends to close some of the gaps and remove impediments linked to the accessibility of scientific data.

A number of databases reporting toxicological or ecotoxicological data are currently available. However, to date, data are stored in heterogenous environments and formats so that the practitioner is faced with the time consuming task of separate access to each database. Overall, integration of ecotoxicological and toxicological information spread around the many existing databases is nowadays not possible. Therefore, there is a strong need to create an integrative tool to facilitate searching of all available toxicological and ecotoxicological information among different sources of information.

To this end, the Tox-Hub has been developed by HEROIC partner CSIC as part of the HEROIC project as an interactive internet platform enabling the simultaneous search and retrieval of information from already existing toxicological and ecotoxicological databases.

We invite you to use this tool, we hope that you find it useful, and we are looking forward to hearing your feedback.

Prof. Martin F. Wilks
Project Coordinator
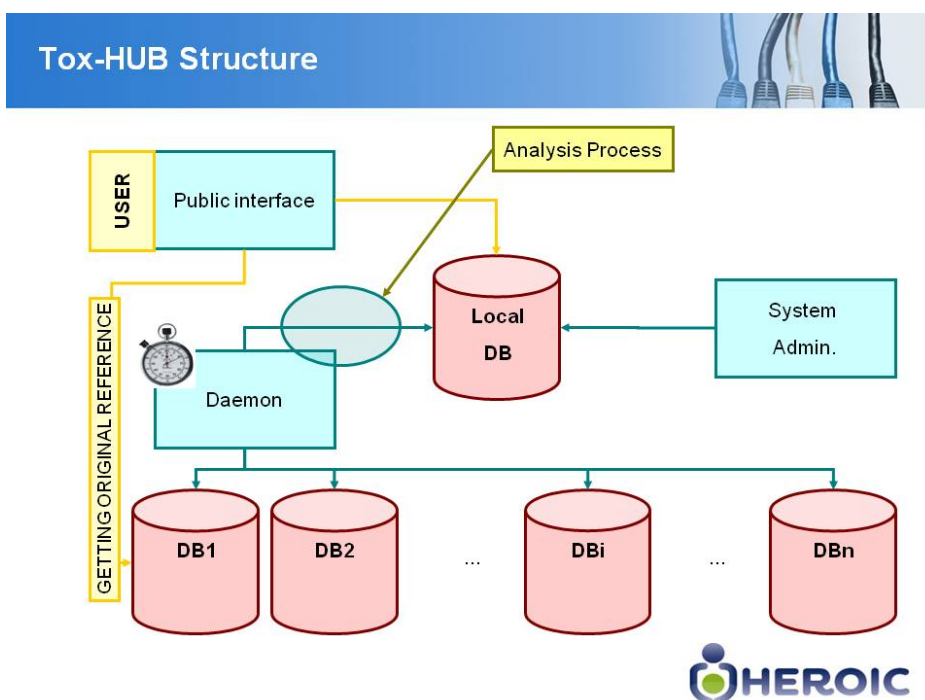
The Tox-Hub platform
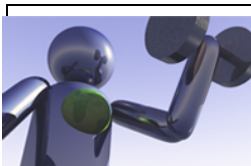
The concept of the Tox-Hub system relies on the idea that existing external information does not need to be duplicated.

We simply need to know where information is stored and how to access it. To that purpose, external databases are scanned, their contents analyzed and keywords extracted and carefully classified.

Every selected keyword is associated with a web-link of the external database where the original information related to it can be found. The Tox-Hub thus simply keeps "keyword/weblink" combinations (called "references") which are readily accessible to the user. This constitutes the so called "internal or built-in dictionary".

The process is programmed to be continuously running so that the references are revisited at a programmed schedule and thus permanently updated. This iterative process also ensures that any refinement introduced in the text processing, can be easily and quickly incorporated without any need of resetting the system.

The basic structure and functioning of Tox-Hub is depicted in **Figure 1**.

Figure 1: Basic structure and functioning of Tox Hub

Some 20 external databases are currently available to the user from the Tox-Hub platform, though not all of them have the same searching possibilities.

They were carefully selected based on expert judgment and access feasibility. External databases presently accessible from Tox-Hub are listed in **Table 1**.

Name	Origin	Remarks
1. ATSDR-HazDat database	USA	
2. Chemical Sampling Information (CSI)	USA	
3. ESIS – European Chemical Substances Information System	EU	Only title searchable
4. EXTOXNET	USA	
5. GESTIS-database on hazardous substances	GER	Only title searchable
6. Haz-Map	USA	
7. High Production Volume Information System (HPVIS)	USA	
8. Hazardous Substances Data Bank (HSDB)	USA	No automatic Access
9. Hazardous Substances Information System (HSIS)	AUS	
10. IARC Monographs	UN	
11. IPCS INCHEM	UN	
12. ITER International Toxicity Estimates for Risk Database	USA	
13. IRIS database for risk assessment	USA	
14. MSDS Database – (Material Safety Data Sheet db)	CAN	pdf documents (only title + index searchable)
15. NIOSH Pocket Guide to Chemical Hazards	USA	Partially scanned
16. PAN Pesticides Database	USA	Partially scanned. Difficult to get references
17. Scorecard	USA	
18. Screening Information Data Set (SIDS) for High Volume Chemicals	UN	pdf documents (only title + index searchable)
19. SOLV-DB	USA	
20. The chemical database	USA	
21. TOXNET	USA	No automatic Access
22. OSHAS Occupational Chemical Data	USA	Difficult to get references

**Table 1: External databases currently linked to the Tox-Hub.**

The system is composed of the following parts:

a) External existing databases (referred to as DB_1 … DB_n) containing relevant toxicological/ecotoxicological information.

b) An internal or local database ("local DB") where terms ("words") are stored together with their corresponding web links ("references") to the external databases.

c) An analysis process that periodically scans each external database, analyzes its contents (text analysis and classification) and updates the local DB. Every pair word-reference is thus revisited at previously scheduled times.

d) An engine ("Daemon") that operates the above described process.

e) An interface allowing the interaction of the system with the user.

f) A system administration back-office.

The core of the system: Text Analysis and term classification process

The text analysis process is a complex operation and constitutes one of the key parts of the Tox-Hub (**Figure 2**) since it serves to build up the internal dictionary.

The word analysis and classification process is the most complex part of the Tox-Hub development, requiring skills and expertise that fall far beyond both toxicology and informatics. A too strict process (i.e., a too narrow "funnel" in **Figure 2**) would result in the loss of useful words, and conversely a too permissive process (broad funnel) would allow incorporating more words on the dictionary, but at the expense of some undesirable disturbing "noise".
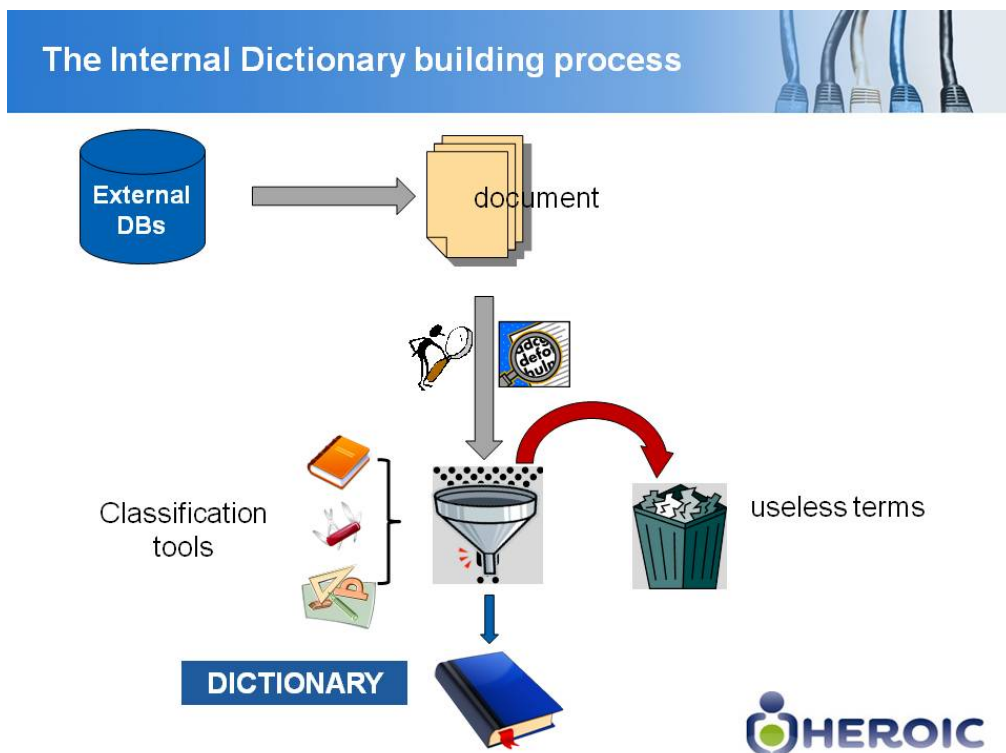
Figure 2. The word analysis and classification process

For that reason the analysis/classification process should be regarded as something subjected to permanent evolution and refinement by addition of new tools and procedures (or improvement of the already existing ones) that will be incorporated in forthcoming versions of Tox-Hub.

Some specific tools have been developed and internally implemented to help in the analysis and classification process:

- Identification of chemical names
- Identification of CAS numbers
- Identification of numbers and units
- On-line English Dictionary (Word Net, Princeton University)
- etc

The final selected Words (Relevant + Synonyms) serve to build the internal Tox-Hub searching dictionary, which will be later accessible to users for searching purposes.

The Tox-Hub at work: overview of the platform functionalities

The main-page is organized in 5 areas (see **Figure 3**) which display different functionalities:

- Administrators Area
- Tabs Area
- Searching Facilities Area
- Query Sets Area
- Results Display Area

How is the Tox-Hub page organized?

a) Administrators login Area is specifically reserved for the system's manager login for internal maintenance and administration tasks. *It is NOT accesible to the USER*

b) **Tabs bar** contains **7 buttons** serving the following purposes:

- o *Heroic website*: directs to the HEROIC main webpage.

- o *Give us feedback:* opens an e-mail addressed to feedback@heroic-toxhub-platform.es to facilitate sending comments to the Tox-Hub system administrator.

- o *Clean up:* cleans up the page to start a new session.

- o *Retrieve session*: opens a dialog window allowing to retrieve a previous session.

- o *Help:* displays the help menu.

- o *DB List:* displays the list of external databases accessible from Tox Hub and provides a short information profile.

- o *IUPAC Glossary:* displays the contents of the **IUPAC GLOSSARY OF TERMS USED IN TOXICOLOGY, 2nd EDITION - IUPAC RECOMMENDATIONS 2007.**

c) **Searching facilities Area** contains **4** action windows:

- o *Select Search scope:* drop down menu allowing choice of the search domain ('*Title'* or '*Whole document'*).

- o *Select dictionary*: drop down menu allowing to choose the word dictionary ('*Internal built-in'* or '*IUPAC*').

- o *Words of the dictionary*: allows to insert a word for searching. On writing a string, a drop down menu appears showing the available words in the dictionary containing it. Once the word is selected, the query is performed by clicking the 'GO' button.

- o ***Operate on sets***: allows the operation of previously generated sets using boolean logical operators ('AND', 'OR', 'AND NOT'). After selection of sets and logical opertors from the corresponding drop down menus, the operation is executed by clicking the 'DO' button.

d) **Query Sets Area**:
- o Query sets generated either by direct search or by application of logical operations are listed sequentially.
- o Sets are identified by:
  - ▪ An order number (1,2,…n).
  - ▪ A figure indicating the number of hits (references).
  - ▪ The word or logical operation used to generate the query.
- o References contained in any set are diplayed by clicking See # button.
- o Any set can by deleted by clicking the rightmost button.

e) **Results display Area:**
- o **References** (*hits*) obtained in the selected query set are displayed (paginated in blocks of 10). The following information is shown:
  - ▪ *Link to the External DB it is coming from.*
  - ▪ *Name of the External DB.*
  - ▪ *Indication if the word is contained in the title.*
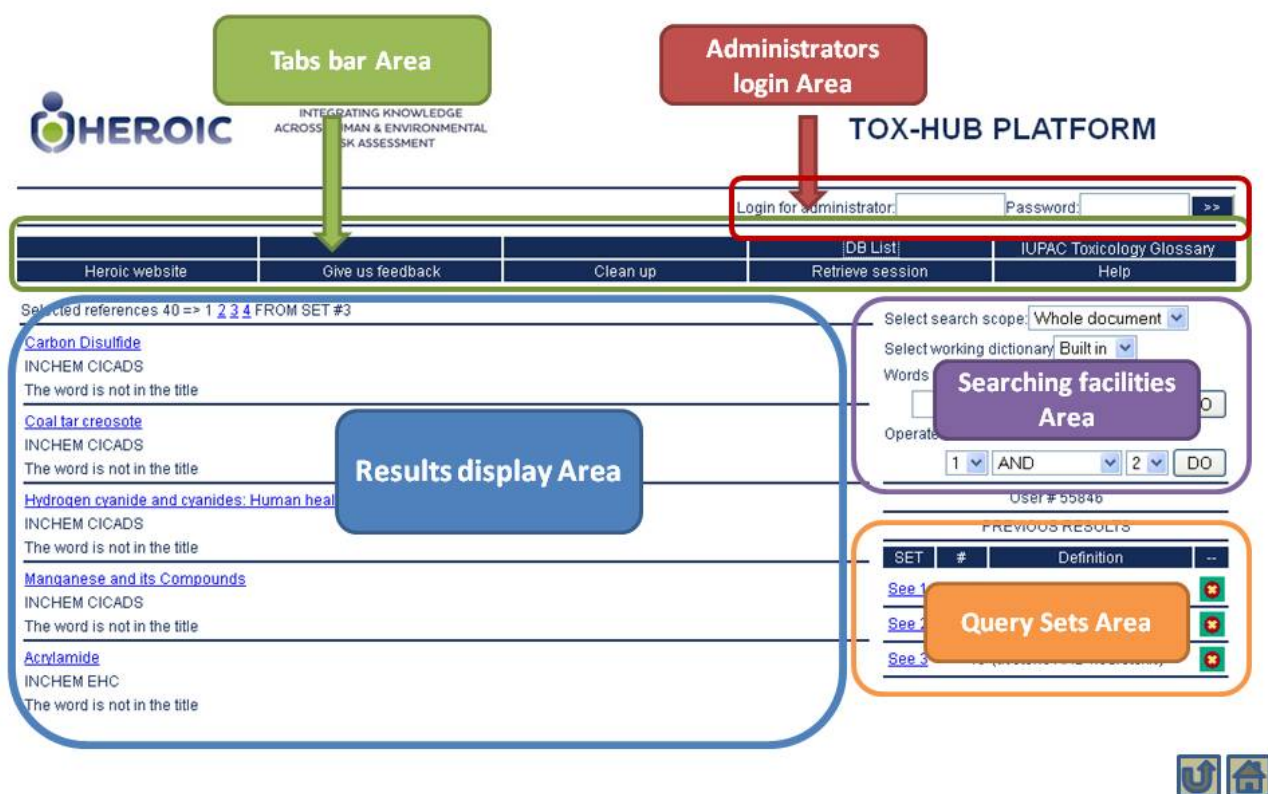
Figure 3: Tox-Hub platform, organization and functionalities.

Tox-Hub Example

When performing a search the user must specify three input fields:

- **Searching scope**: Title or Full document (referred to the original external databases).
- **Dictionary**: referred to the internally automatically generated (built in) dictionary, or alternatively, to the use of an external one (IUPAC).
- **Words**: a specific keyword related to the topic of interest should be selected from the drop-down menu.

Different primary searching sets are generated with each word. Further information processing is achieved using logical (Boolean) operators: AND, OR, AND NOT.

See the simple example below.

Figure 4. Example illustrating how words are searched for and refined using logical operators

Example: we are interested on the effects of insecticides on non-target organisms (i.e., other than insects), for instance birds or mammals: The relevant keywords thus are the terms 'insecticide', 'bird', 'mammal'.

First we proceed to build separately the corresponding sets with the three words selected:

Set #1: 'insecticide'   (168 references)

Set #2: 'bird'   (89 references)

Set #3: 'mammal'   (22 references)

Secondly, we obtain separately the intersections (AND) of #1 with #2 and #3. To do so, the set #1 is selected in the first dropdown, the operator AND is selected in the second and the set # 2 is selected in the second dropdown selector. We get:

Set #4: 'insecticide' AND 'bird'   (#1 AND #2)   (50 references)

Set #5: 'insecticide' AND 'mammal'   (#1 AND #3)   (12 references)

Then we combine sets #4 and #5 since possibly both are sharing some common references. This is done by an OR operation:

Set #6: (#4 OR #5)   (51 references)

The final 51 original references identified can be directly accessed by clicking the links shown in the Results Display Area.

Link

Tox-Hub is publicly accessible at: www.heroic-toxhub-platform.es.

Other useful informations

Users Manual is downloadable at www.heroic-toxhub-platform.es, or in the dedicated section of the HEROIC website at http://heroic-fp7.eu/en/content/TOX-HUB-platform.5/.