Toxicology in Vitro 28 (2014) 571–587

Contents lists available at ScienceDirect

Toxicology in Vitro

journal homepage: www.elsevier.com/locate/toxinvit

Automatic sorting of toxicological information into the IUCLID (International Uniform Chemical Information Database) endpoint-categories making use of the semantic search engine Go3R

Ursula G. Sauer [a,1], Thomas Wächter [b,c,1], Lars Hareng [d,1], Britta Wareing [d], Angelika Langsch [d], Matthias Zschunke [c], Michael R. Alvers [c], Robert Landsiedel [d,*]

[a] Scientific Consultancy – Animal Welfare, Neubiberg, Germany
[b] Biotechnology Center, Technische Universität Dresden, Dresden, Germany
[c] Transinsight GmbH, Dresden, Germany
[d] BASF SE, Product Safety – Experimental Toxicology and Ecology, Ludwigshafen, Germany

ARTICLE INFO

Article history:
Received 23 August 2013
Accepted 23 December 2013
Available online 2 January 2014

Keywords:
Semantic search engine
REACH
Registration dossiers
Regulatory toxicology
Data sharing
Alternative methods
3Rs principle

ABSTRACT

The knowledge-based search engine Go3R, www.Go3R.org, has been developed to assist scientists from industry and regulatory authorities in collecting comprehensive toxicological information with a special focus on identifying available alternatives to animal testing. The semantic search paradigm of Go3R makes use of expert knowledge on 3Rs methods and regulatory toxicology, laid down in the ontology, a network of concepts, terms, and synonyms, to recognize the contents of documents. Search results are automatically sorted into a dynamic table of contents presented alongside the list of documents retrieved. This table of contents allows the user to quickly filter the set of documents by topics of interest. Documents containing hazard information are automatically assigned to a user interface following the endpoint-specific IUCLID5 categorization scheme required, e.g. for REACH registration dossiers. For this purpose, complex endpoint-specific search queries were compiled and integrated into the search engine (based upon a gold standard of 310 references that had been assigned manually to the different endpoint categories). Go3R sorts 87% of the references concordantly into the respective IUCLID5 categories. Currently, Go3R searches in the 22 million documents available in the PubMed and TOXNET databases. However, it can be customized to search in other databases including in-house databanks.

1. Introduction

The revised European Union (EU) *Directive 2010/63/EU on the protection of animals used for scientific purposes* (Anon, 2010) has explicitly implemented the 3Rs principle to replace, reduce, and refine animal testing (Russell and Burch, 1959). Accordingly, animal testing may not be performed if "*a scientifically satisfactory method or testing strategy, not entailing the use of live animals*" can be used instead. Furthermore, it has to be ensured that the number of animals used in a procedure is reduced to the minimum, in the same way as any possible pain, suffering, distress, or lasting harm (Article 4 of Anon, 2010). Collecting comprehensive and up-to-date information on available 3Rs methods and on existing information and data related to the scientific topic in question is a prerequisite to fulfilling these legal provisions and to ensuring the indispensability of animal tests. In light of the continuously increasing abundance of information available in the World Wide Web, the gathering of relevant information is an increasingly complex and time-consuming challenge.

Since 2007, the semantic search engine Go3R (available free of charge at www.Go3R.org) has been developed to support scientists in searching for 3Rs relevant information in the Internet (Sauer et al., 2009). Semantic search tools 'understand' what the user is searching for and automatically select and sort relevant pieces of information (Schroeder, 2003). The semantic search paradigm of Go3R indexes documents with bibliometric metadata classes (i.e. authors, journals, cities, countries, year of publication) and with the classes of an ontology that has been specifically created for Go3R (see information box I for ontology-related definitions). In processing search queries, Go3R automatically compares the vocabulary used in the titles, abstracts, and keywords or Medical Subject Headings (MeSH; the indexing thesaurus of the PubMed

* Corresponding author. Address: Experimental Toxicology and Ecology, BASF SE, 67056 Ludwigshafen, Germany. Tel.: +49 (0)621 60 56 203.
E-mail address: robert.landsiedel@basf.com (R. Landsiedel).
[1] Equal first authorship.

http://dx.doi.org/10.1016/j.tiv.2013.12.011
0887-2333/

database; http://www.nlm.nih.gov/[2]) of the documents to the classes of the Go3R ontology and sorts the documents retrieved into the respective recognized classes. The outcome of this sorting is presented to the user alongside the search retrieval in the form of a dynamic 'table of contents' (see Fig. 1, presenting the exemplary search query 'eye irritation'). This table of contents can be used as a navigation tool: By filtering for its respective 'chapters' and 'sub-chapters' (i.e. by clicking onto them on the user interface), broad search results are quickly restricted to relevant information (see Fig. 2, showing subordinate levels of the ontology root concept '3Rs alternative methods' and restriction of the initial search query 'eye irritation' to documents in which the 'bovine corneal opacity and permeability assay' is mentioned).

Information Box I Ontology-related definitions
(Adapted from: Uschold and Gruninger (1996), Boyce and Pahl (2007); and the glossary of the US National Library of Medicine's Unified Medical Language System (UMLS®); available at: http://www.nlm.nih.gov/research/umls/new_users/glossary.html)

Ontology

A framework for representing concepts (things or ideas about things), the relationships that exist between the concepts, and the properties they might have. In the ontology, concepts are linked in a strictly hierarchical (tree-like) structure ensuring that subordinate ('parent–child') relationships not only hold between classes and their direct parents (e.g. 'mouse' and 'rat' as direct subordinate concepts to 'rodent'), but also to all further superordinate concepts (ancestors) in the given ontology branch (e.g. 'mammal', 'vertebrate', 'animal').

Concept (or class)

The fundamental unit of meaning in the knowledge source. A concept represents a single meaning and contains all textual labels from any source that express that meaning in any way, whether formal or casual, verbose or abbreviated (e.g. 'BCOP' and 'OECD TG (test guideline) 437' as textual labels for the 'bovine corneal opacity and permeability assay'). All of the textual labels within a concept are synonymous.

Root concept

The highest-level concept of a given thematic branch of the ontology.

Term

The textual label assigned to a concept.

Go3R is based upon prior work by the *Technical University Dresden* and *Transinsight GmbH Dresden* to develop GoPubMed (Doms and Schroeder, 2005; www.gopubmed.org), a semantic search tool to search general biomedical information provided in the PubMed database. Go3R has been designed to retrieve information on alternative methods in all areas of biomedical research. However, due to the main topic of the Go3R research project, the search engine currently has a special focus on retrieving information on 3Rs methods in the realm of regulatory toxicity testing. Accordingly, Go3R has been aligned to search the databases *PubMed* and *Toxicology Data Network* (TOXNET; http://toxnet.nlm.nih.gov/). Addressing the increasing economic importance of nanotechnological developments and the resulting amount of research investigating the safety of nanomaterials (Oomen et al., 2013), an additional 'nano'-ontology (Fig. 3) was created and linked to Go3R (Sauer et al., 2011). Thereby, the 3Rs relevant concepts of the Go3R ontology can be combined with concepts of the nanotechnology and nanotoxicology domains (such as 'endpoints methods for nanomaterial characterization and testing'; Fig. 3).

[2] *Note:* All websites were accessed in June 2013.

Depending on the vocabulary used in title and abstract, 3Rs-related documents of relevance for regulatory toxicity purposes might be sorted into a number of different Go3R ontology branches, i.e. '3Rs methods in the life sciences' (see Fig. 2 for subordinate levels of concepts to this root concept), 'cultured cells, tissues, etc.', '*in vitro* experimental design', 'cell culture technology, etc.', '3Rs method types' (listing different test systems), or '3Rs toxicity testing strategies' (see Fig. 1 for overview of root concepts). When *in vitro* studies are compared to *in vivo* tests or reduction and refinement methods are referred to, the ontology branches 'animal species', 'animal test method', and '*in vivo* experimental design' are also of relevance. Of note, the vocabulary used in different documents to describe one specific type of study is not uniform, but is incumbent upon the respective author's choice. Therefore a document containing information on a given toxicological endpoint (e.g. 'eye irritation') might be sorted into a combination of any of these (or even further) ontology branches. To allow searching for endpoint-specific information, documents from possibly many different branches need to be aggregated and structured by endpoint-specific topics to prevent users from manually checking multiple branches.

Against this background it was the aim of the present study to extend the Go3R search engine to automatically sort all relevant documents by toxicological endpoint *regardless of the vocabulary used by the respective authors*. Even if it did not explicitly mention the endpoint under consideration or the respective test method, a document should still be recognized as relevant for the given endpoint, e.g. due to a specific pattern of vocabulary mentioning specific cell lines, cellular endpoints, or endpoint detection methods relating to the respective endpoint, and it should be sorted accordingly.

The categorization scheme of the 5th version of the *International Uniform Chemical Information Database* (IUCLID5) was selected as a template for this sorting task. IUCLID5 (http://iuclid.eu) plays a central role in collecting, storing, submitting, and exchanging data in fulfilling the data submission requirements of, e.g. the EU *regulation on the Registration, Evaluation, Authorisation, and Restriction of Chemicals* (REACH; Anon, 2006), the *OECD Cooperative Chemicals Assessment Programme* (http://www.oecd.org/chemicalsafety/risk-assessment/oecdcooperativechemicalsassessmentprogramme.htm), or the EU *regulation concerning the making available on the market and use of biocidal products* (Anon, 2012). The IUCLID categorization scheme is closely linked to the OECD harmonized templates (http://www.oecd.org/ehs/templates/) for structuring data entry systems to report the results of tests determining human health and environmental effects of substances.

Chapter 7 of the IUCLID5 scheme provides the structure to submit toxicological information. This section encompasses 12 distinct endpoint-specific categories with sub-sorting into a total of 30 sub-categories (Table 1 and Fig. 4). Detailed guidance has been published on how to select appropriate information and sort it into the respective endpoint-specific categories (see: ECHA and OECD, 2007; ECHA, 2012a,b).

In regard to chemical substances, the REACH regulation prescribes that all substances manufactured or imported in quantities above 1 tonne per year have to be registered with the European Chemicals Agency (ECHA). Information requirements for the registration dossiers, making use of the IUCLID categorization scheme, increase with increasing tonnage. More than 100,000 registration dossiers for a total of 30,000 chemical substances are foreseen by the end of May 2018 (estimations of the *German Chemical Industry Association*, VCI – Verband der Chemischen Industrie; based upon the numbers of pre-registered substances). By this date, the so-called phase-in substances, i.e. substances that were already manufactured or imported before the implementation of the REACH regulation, are to be registered.

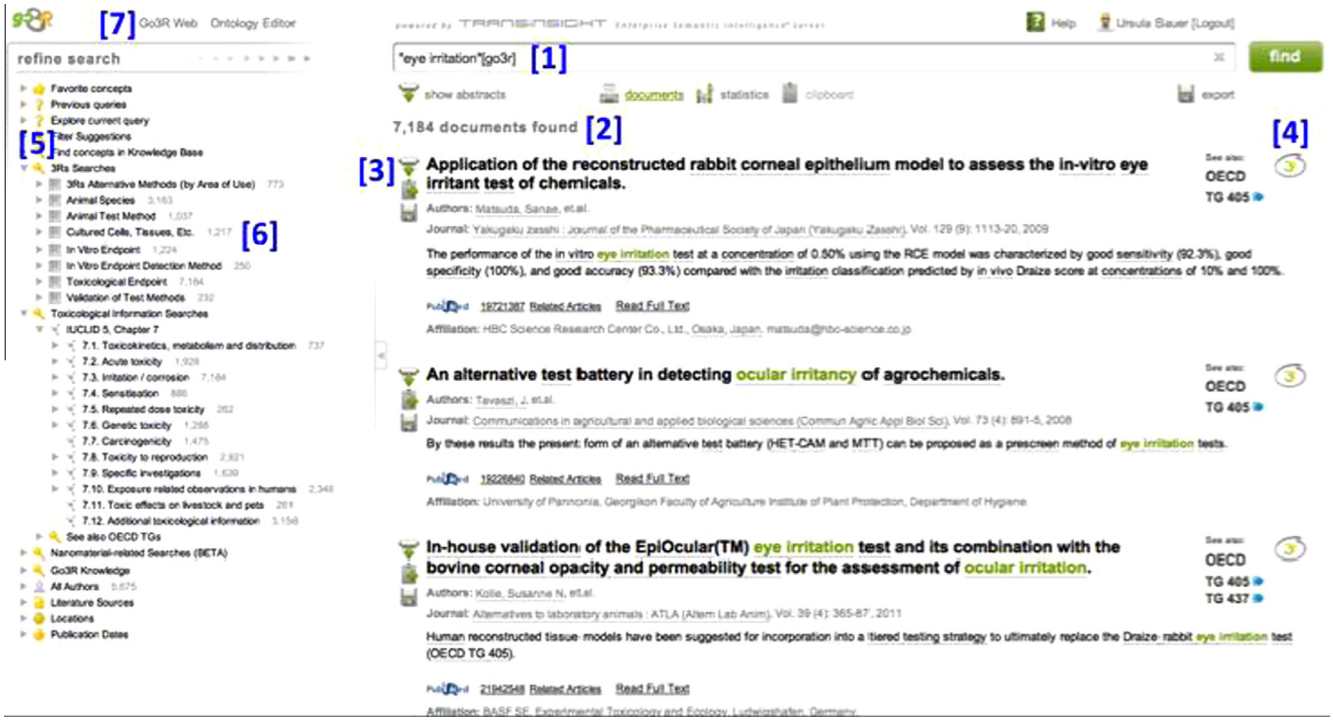

Fig. 1. Go3R user interface, available at www.Go3R.org, in this example presenting the results of the search query 'eye irritation' using the respective Go3R ontology concept (search date 19 November 2013). Footnote to Fig. 1: The Go3R user interface, as it is available free of charge at www.Go3R.org. (1) At the top of the right hand side of the screen, the input box for the search queries is located. In this example, the search query 'eye irritation' has been typed, making use of the respective Go3R ontology concept, as indicated by [go3r] behind the search query term. Matching ontology concepts are suggested to the user while he types his individual search query. If he decides to accept the proposed Go3R ontology concept, all textual labels (terms and synonyms, etc.) and subordinate concepts are automatically included in the search result. If he declines the proposed Go3R ontology concept, the search query is processed 'as written', i.e. in the case of 'eye irritation' without [go3r] icon, only documents mentioning this exact wording would be retrieved. (2) Upon clicking on the 'find' button, next to the search query input box, a total of 7184 documents are retrieved from PubMed and TOXNET (search date: 19 November 2013). (3) The information on these documents is presented to the user on the right hand side of the Go3R user interface below the search query input box. (4) The '3Rs relevant signet' assigned to individual documents is included to the right of the individual documents (see Information Box III for further information on the '3Rs relevant signet'). (5) A dynamic table of contents, built making use of the Go3R ontology, is presented to the user along-side the list of documents retrieved, on the left hand side of the screen: In this table of contents, all documents are automatically sorted into the 25 branches of the Go3R ontology. (6) Figures in square brackets behind each concept of the knowledge base indicate the number of documents assigned accordingly. Multiple assignments are possible, since individual documents can match several different concepts. (7) Link for access to the Go3R web tool (see Information Box III) allowing 'Google' searches with higher priority retrieval of 3Rs relevant information, thereby providing information e.g. on the status of regulatory acceptance of alternative test methods.

Due to long-time experience of use, an abundance of toxicological information is available for these 'existing' substances. Therefore collection of all available relevant data is an important pillar towards reducing the time and costs necessary to prepare registration dossiers and, most importantly, towards avoiding testing with sentient animals in line with the 3Rs principle. However, also when new testing has to be performed to obtain hazard information for the registration dossiers, in accordance with the provisions of REACH and EU Directive 2010/63/EU, testing on vertebrate animals may only be undertaken as a last resort, i.e. when there are no other means of obtaining the necessary information. Thus, the provisions of the REACH regulation not only require comprehensively gathering already existing data, but also continuously updating the information on newly developed replacement, reduction, or refinement methods.

Considering the plenitude of published and unpublished information that needs to be gathered, evaluated, and processed in preparing REACH registration dossiers, semantic search tools that aid the user by automatically recognizing and sorting relevant information are desirable not only for economic reasons but also for ensuring comprehensibility of the data collected, which in return directly serves the 3Rs principle. Annex XI (1) of the REACH regulation explicitly mentions the collection and use of existing data from *in vivo* studies as a first tier of an integrated approach to the testing and assessment of substances. The subsequent tiers involve application of weight-of-evidence tools, (quantitative) structure–activity relationships, and *in vitro* methods (Anon, 2006). As experience gained during the first registration phase highlighted, the collection of already existing *in vivo* data made a major contribution to avoiding new animal testing (Spielmann et al., 2011).

Accordingly, it was the aim of the present study to serve these goals by extending the Go3R search engine to automatically sort all relevant toxicological *in vivo*, *in silico*, and *in vitro* information for a given test substance (i.e. user search query 'test substance *xyz*') into the respective 12 IUCLID5 'toxicological information' categories and their sub-categories and present the categorized information on the Go3R user interface (Fig. 4).

2. Performance of the study

2.1. Go3R at the onset of the IUCLID search query-based annotation research

Since the beginning of the development of the Go3R ontology in 2007, it has been expanded to cover approximately 20,000 classes (see Information Box I) with nearly 90,000 textual labels (terms and synonyms). These classes were sorted hierarchically into 25 branches with up to 11 levels of increasingly subordinate classes. A total of 15,362 classes with 85,025 textual labels relating to 'diseases and symptoms', 'anatomical structures and organs', and 'chemical compounds' were taken over from the MeSH thesaurus branches. 14% of the classes were newly defined, corresponding to 2533 concepts with overall 5547 textual labels. The concepts

of the Go3R ontology include both classes that are specifically 3Rs relevant (e.g. 'Local Lymph Node Assay') and thematic-defining classes (e.g. 'dermatitis, allergic contact').

Go3R uses several direct and indirect assignment methods to automatically determine whether a document belongs to a specific class. For direct assignments, a number of independent methods decide whether a document can be assigned to a particular class. Prior to the IUCLID categorization annotation work, Go3R used five distinct technical tools to analyze and annotate documents: (1) lexical annotation, (2) word sense disambiguation, (3) topic classification, (4) integration of manual annotation of metadata provided by PubMed/TOXNET, and (5) user curations (see Information Box II for further details). The IUCLID search queries-based annotation presented in this article was incorporated into the Go3R search engine as a further, sixth, annotation tool.

Indirect assignment is provided by concept inclusion based on the taxonomic structure defined by higher levels of the respective ontology branch. For instance, a document containing the word 'mouse' (or 'mice', etc.) is directly annotated with 'mouse' and will be indirectly annotated with 'rodent' and all its further ancestors in the taxonomic structure ('mammal', 'vertebrate') up until the root concept ('animal species'). Searching for a parent class will retrieve all documents for all descendants of the class. Thereby generalized queries (e.g. 'rodent AND skin') retrieve documents relating to 'mice AND skin' as well as 'rat AND skin', and equally for all intersections between the rodent-related terms and the sub-classes of

Information Box II Go3R semantic document indexing and classification at the onset of the IUCLID-search query based annotation research.

Lexical annotation

Based upon the labels and synonyms of classes defined in the taxonomy, the documents are analyzed to find potential occurrences of associated lexical terms. For MeSH terms, which are intended for text indexing, most terms can be located with simple string matching algorithms. As determined by the domain experts, a number of ambiguous terms require further disambiguation to determine the true semantic meaning of the word in the analyzed text (e.g. 'monolayer' for monolayer cell cultures; or 'CASE' for 'Computer Automated Structure Evaluation' used e.g. for molecular modeling).

Word Sense Disambiguation (WSD)

WSD is a sub-task of semantic tagging and deals with relating the occurrence of a word in a text to a specific meaning, which is distinguishable from other meanings that can potentially be related to that same word (Schuemie et al., 2005). WSD is essentially a classification problem: Given an input text and a set of sense tags for the ambiguous words in the text, correct meanings are assigned to these words. The approach developed by Doms (2009) trains a machine learning classifier on text and metadata, such as journal, author, and date of publication. In a cross-validation evaluation on ambiguous Gene Ontology terms, this approach received an *F*-measure of 0.96 (see below for definitions). For the given examples 'CASE', the 5-fold cross-validation yielded an *F*-measure of 0.99 (±0.3) based upon 61 positive and negative examples, and ''monolayer'' achieved a *F*-measure of 0.95 (±0.2) based upon 186 positive and negative examples.

Cross-validation

In several rounds, subsets of the example documents are randomly selected as training and test sets. The training set is used to train the machine-learning model which is then tested to classify the test set into matching and non-matching (positive and negative) documents. For the performed 5-fold cross validation, the positive and negative document sets were divided into five parts. In five rounds, alternating four parts were used for training and one part for testing.

Precision and recall

'Precision' measures the proportion of documents correctly marked as relevant and 'recall' the proportion of documents marked as relevant from the total number of documents that should have been marked. In the *F*-measure, precision and recall are combined as a harmonic mean (Hripcsak and Rothschild, 2005).

Metadata mapping

MeSH headings originally assigned to documents from PubMed and keywords from TOXNET are syntactically mapped to the classes in the Go3R taxonomy. The corresponding documents are annotated with the mapped class. Thereby it is ensured that manual annotations from PubMed and TOXNET are considered in Go3R.

3Rs Relevance Classification

Similar to WSD, documents are classified using a machine learning technology called Maximum Entropy Method (Berger et al., 1996). The Maximum Entropy Method enables the classifier to learn the characteristics of documents belonging to a certain pre-defined category. The initial pool of positive examples for the 3Rs relevance classifier consisted of approximately 2000 PubMed entries that were indexed with the MeSH term 'Animal Testing Alternatives'. Of note, these documents frequently included articles from journals predominantly publishing 3Rs relevant information, such as 'Alternatives to Laboratory Animals', 'ALTEX', or 'Toxicology *in Vitro*' and were likely to be biased towards certain topics, e.g. toxicology. Based on these training examples, the algorithm automatically extracts a set of relationships inherent in the examples and then combines these rules into a model of the data that is both accurate and compact. Relationships can be found based on textual occurrences of terms as well as meta-information, and previously assigned annotations. The learned model is used to assign unknown documents to the pre-defined category.

Accordingly, Go3R assigns '3Rs relevance signets' (to be seen in the lists of search retrievals on the top right corner of the individual documents, Fig. 1) to documents if the Maximum Entropy Method assigns at least one of the five classes '3Rs relevant', '3Rs principle', 'Reduction', 'Refinement', 'Replacement' (of animal experiments). To obtain a stable and more reliable ranking, it was decided to increase the 3Rs relevance score if several classifiers coincide. In the list of documents retrieved, documents are ranked higher if they are expected to make a contribution towards replacing, reducing, or refining animal experiments.

To begin with, the Maximum Entropy Method was trained for the most general '3Rs relevant' class using 2346 PubMed abstracts hand-annotated as being 3Rs relevant and 2346 negative examples not belonging to that class. The first half of these documents was randomly selected from all PubMed documents, whereas the second half was randomly selected from the journals hand-annotated by the domain expert as being not 3Rs relevant. Thereby, the search engine was trained to distinguish between documents that were likely to be 3Rs relevant from those that were likely not to be 3Rs relevant. The 5-fold cross validation yielded an average *F*-measure of 0.91. Over five iterations, the mean precision (0.96) was higher than the mean recall (0.86). This evaluation suggests that most of the documents receiving the '3Rs relevance signet' will truly have some relation to animal testing alternatives.

Incorporation of user curation

All Go3R users can themselves curate the automatic annotations directly on the user interface. For this purpose, the curation tool can be opened for each individual displayed document allowing curation of each single annotation with (+) for correct, (o) for correct, but unimportant, or (−) for wrong annotations. Positive and negative curations will have an immediate impact on the documents displayed to the respective user.

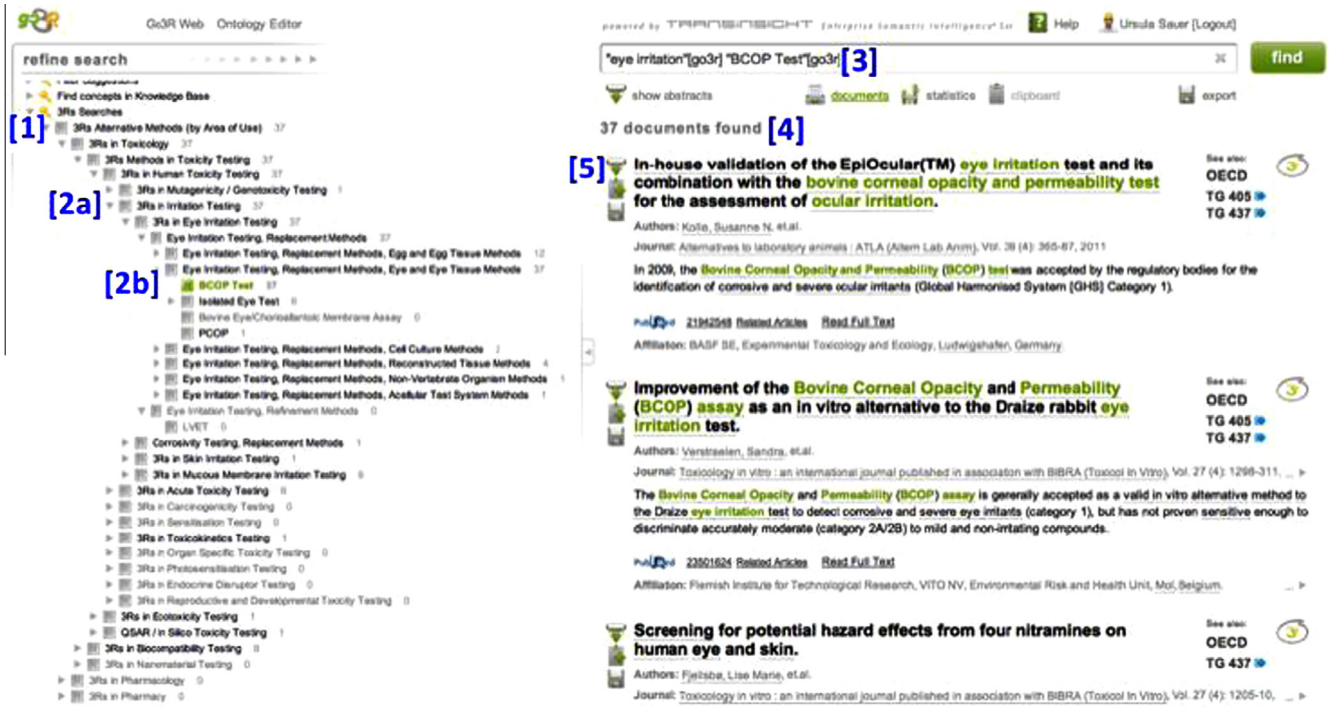

Fig. 2. Refinement of the results of the search query using the Go3R ontoloy concept 'eye irritation' (see Fig. 1) to filter documents referring to the 'Bovine Corneal Opacity and Permeability Test' (search date 19 November 2013). Footnote to Fig. 2: (1) The dynamic table of contents (left hand side) has been unfolded to reveal the different subordinate levels of the ontology concept '3Rs alternative methods'. Unfolding is performed by clicking onto the gray triangle to the left of the respective concept (which thereby points downwards). (2a and 2b) By filtering for the respective concepts of the table of contents, search results can be quickly restricted to relevant information. Filtering is performed by clicking onto the respective concept. This tool allows inclusion or exclusion of the concept (the latter application not shown). Included concepts are highlighted in green in the table of contents (2b); excluded concepts are crossed out and highlighted in red (again: not shown). (3) The filtering in accordance to the selected concept is reflected by the automatic search query expansion in the search query field. (4) The new list of documents consists of 37 of the original 7184 documents (search date: 19 November 2013). (5) This new list of documents is presented on the right hand side below the search query input box. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

skin presented in the taxonomic hierarchy ('epidermis', 'hair follicle', 'dermis', etc.).

2.2. IUCLID categorization and IUCLID search query-based annotation

The IUCLID categorization work aimed at developing so-called 'IUCLID search queries' (IUCLID SQ) to be used as annotation tool by the Go3R search engine. This categorization work is an interdisciplinary task involving experts in the knowledge domain of regulatory toxicology and 3Rs methods and computer scientists with special expertise with semantic technologies. In the following, the scientific research performed by these two expert groups is described separately, even though it was continuously interlinked throughout the entire study.

2.2.1. Domain expert work

The domain expert work covered the following steps:

(1) The project partner BASF, Ludwigshafen, Germany, made available the set of literature that had been collected, evaluated, and sorted manually by information retrieval experts for the compilation of the REACH registration dossiers for three chemical substances, i.e.
- Hydroxycitronellal (CAS No. 107-75-5).
- Pentanol (CAS No. 71-41-0).
- Toluene (CAS No. 108-88-3).

The spectrum of literature contained in the three registration dossiers (see below for further details) was laid down as the 'gold standard' for the formulation of the IUCLID SQs.

(2) One-by-one, the domain expert evaluated all documents that were referenced in the 3 registration dossiers. Firstly, this task served to discern whether the documents from the registration dossiers were available in PubMed or TOXNET at all. Secondly, based upon the assignment to the respective IUCLID5 categories and sub-categories undertaken by BASF, the titles, abstracts, and MeSH headings/TOXNET keywords of those documents available in PubMed or TOXNET were scrutinized. This served to identify relevant vocabulary and different patterns of vocabulary used in toxicological publications and reports relating to the 12 sections of the toxicological information chapter (section 7) of the IUCLID5 scheme and, more specifically, to the 30 respective sub-sections.

(3) Based upon these patterns of relevant vocabulary, relevant concepts of the Go3R ontology were selected that best reflect these patterns. These Go3R concepts were combined into one or more IUCLID SQs per toxicological endpoint. By exclusively making use of classes from the Go3R ontology, the endpoint-specific IUCLID SQs automatically took into account all textual labels and subordinate concepts linked to the given class of the ontology.

The IUCLID SQs were formulated to aim for data completeness (recall): For the test substance under investigation, all available *in vitro* and *in vivo* information should be retrieved that was relevant for risk assessment and classification and labeling, and it should be sorted into the toxicological endpoint categories as best possible. In the course of their development, the set of IUCLID SQs compiled for each endpoint was submitted to repeated test search trials. These test search trials were evaluated to determine, e.g. how inclusion of a given Go3R concept into the respective IUCLID

SQ affected overall search result precision and recall and to identify concepts missing in the given IUCLID SQ. Based upon the outcome of these repeated test search trials, the IUCLID SQs were adapted as necessary – and re-tested.

Of note, the work on the Go3R IUCLID SQs only addressed the issue of recognizing endpoint-specific patterns of vocabulary (e.g. 'genotoxicity', 'chromosome aberration', 'micronucleus assay'). It did not aim at recognizing documents that were relevant for a given test substance-specific search query (e.g. 'pentanol', 'formaldehyde', 'butadiene'), if the respective test substance was not specifically mentioned in the title, abstract or MeSH terms/TOXNET keywords of the document, but the author referred to the test substances in general terms (e.g. '50 compounds', '12 test substances').

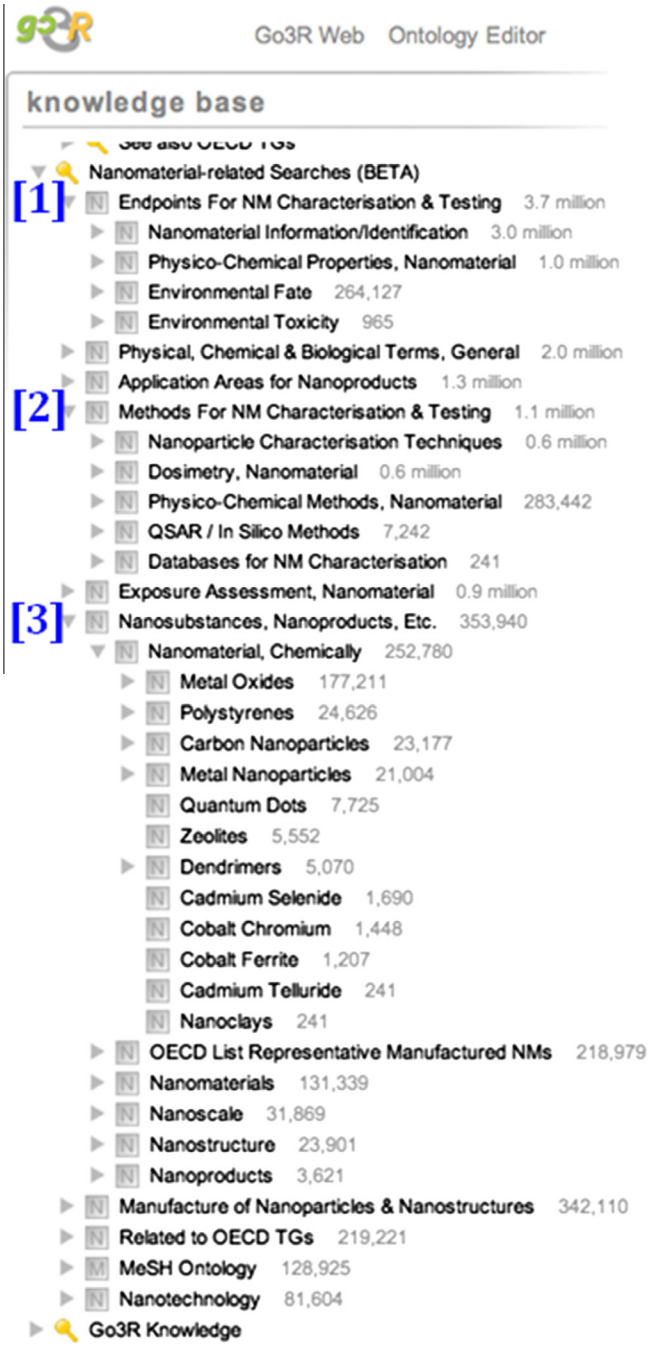

Fig. 3. The additional 'nano'-ontology that is linked to Go3R. Footnote to Fig. 3: The unfolded 'table of contents' reveals the first levels of subordinate concepts for 'endpoints for NM characterization and testing' (1), 'methods for NM characterization and testing' (2), and 'nanosubstances, nanoproducts, etc.' (3).

For this to become feasible, the Go3R technology would have to be expanded to allow searching in test substance-specific metadata (if available in the respective database) or to search in the full text documents (if permissible on copyright grounds).

2.2.2. Applied informatics work

The query-based classification and annotation of documents to IUCLID5 categories requires the development of descriptive models using various relationship types beyond the 'is a' parent–child relationship described above. For further processing of the formulated IUCLID search queries, they were stored as external database references (so-called 'dbxrefs') together with the corresponding class of the Go3R ontology. Documents found for a given IUCLID SQ will be annotated with the corresponding category. Opening the stored resource locator (URL) yields a Go3R search for the stored query with resulting automatic integration into the respective endpoint-specific sections of the IUCLID5 Chapter 7 user interface.

Making use of the concept of stored queries for document annotation, the description of a class is formulated directly in the query language of the Go3R application instead of using a standardized representation, such as OWL (Ontology Web Language; http://www.w3.org/TR/owl-features/). Thereby, the content of a class is defined via the documents indexed in Go3R. By storing the queries as external database references in form of URLs for each class, this additional information is part of the ontology file and leaves the option open to convert the queries into object or class properties at a later stage.

2.3. Evaluation of IUCLID search query-based annotation

Upon finalization of the work on the IUCLID SQ-based annotation tool, the usefulness of the Go3R IUCLID user interface was investigated. The correctness of the IUCLID sorting was evaluated for the set of references used in the three mentioned BASF registration dossiers. As an additional gold standard reference list, unrelated to the work on the IUCLID SQs, the references cited in an extensive literature review on non-animal test methods for predicting skin sensitization potentials (Mehling et al., 2012) was evaluated in respect to sorting of the literature references of the review into the endpoint-specific IUCLID user interface. This literature list was selected because of its comprehensiveness and up-to-dateness in regard to a given toxicological endpoint, and because of the importance of the endpoint skin sensitisation for chemical safety assessment: Skin sensitisation is the only toxicological endpoint for which information obtained by an *in vivo* method has been laid down as general information requirement for all chemical substances falling under the REACH system regardless of production volume (Anon, 2006).

3. Results

The BASF registration dossiers for pentanol (PT), toluene (TL), and hydroxycitronellal (HC) contain 284, 159, and 104 literature references, respectively. These figures include multiple citations of an individual document (allowed for in the registration dossiers by the so-called 'cross-referencing' of studies), both within an endpoint-specific category (e.g. to 7.6.1 if one document makes reference to several different *in vitro* genotoxicity studies), but also between different endpoints (e.g. to 7.3.1 and 7.3.2, if the document describes both skin and eye irritation studies). For the Go3R evaluation, all multiple citations within a given endpoint-specific category were excluded, whereas multiple citations (references) between endpoints were maintained since Go3R was expected to discern the different endpoints addressed. Upon vexclusion of the within-endpoint duplicates, the literature lists

Table 1
IUCLID5 categorization scheme (section 7, toxicological information) and Go3R IUCLID search queries assigned to each IUCLID5 category and sub-category.

IUCLI5 No.	IUCLID5 category or sub-category	Go3R IUCLID search queries
7.1	Toxicokinetics, metabolism and distribution	
7.1.1	Basic toxicokinetics	"*in vivo* toxicokinetics study"[go3r] OR "Monte Carlo Method"[go3r] OR "Physiologically Based Toxicokinetic Modelling"[go3r] OR "Physiologically Based Pharmacokinetic Modelling"[go3r] "toxicokinetics"[go3r] NOT "skin absorption"[go3r]
7.1.2	Dermal absorption	"*in vivo* dermal absorption test"[go3r] OR "*in vitro* skin absorption test"[go3r] "Franz Cell Diffusion"[go3r] "skin absorption"[go3r] OR ("Human Skin Culture"[go3r] -"Local Toxicity"[go3r])
7.2	Acute toxicity	"Lethal Dose 50"[go3r] OR "Lethal Concentration 50"[go3r] "Acute Toxicity"[go3r] OR "Hygienic Assessment"[go3r] OR Poisoning[go3r]
7.2.1	Acute toxicity, oral	("Lethal Dose 50"[go3r] OR "No-Observed-Adverse-Effect Level"[go3r] OR "No Observed Effect Level"[go3r]) AND oral[go3r] "*in vivo* test methods, acute oral toxicity"[go3r] "Acute oral Toxicity"[go3r] OR "Oral Toxicity"[go3r]
7.2.2	Acute toxicity, inhalation	("Lethal Concentration 50"[go3r] OR "No-Observed-Adverse-Effect concentration"[go3r]) AND Mammal[go3r] AND inhalative[go3r] "Acute inhalation Toxicity"[go3r] OR "Pulmonary Toxicity"[go3r] OR "Inhalation Toxicity"[go3r] "*in vivo* test methods, acute inhalation toxicity"[go3r]
7.2.3	Acute toxicity, dermal	"Acute dermal Toxicity"[go3r] OR "Dermal Toxicity"[go3r] "*in vivo* test methods, acute dermal toxicity"[go3r]
7.2.4	Acute toxicity, other routes	("intraperitoneal administration"[go3r] OR "subcutaneous administration"[go3r] OR "Intravenous Injection"[go3r]) (Toxicity[go3r] -"Mammalian Toxicity"[go3r] -Cytotoxicity[go3r] -"Human Toxicity"[go3r] -Ecotoxicity[go3r])
7.3	Irritation/corrosion	"Irritation"[go3r] "*in vivo* test methods, local toxicity"[go3r] OR "3Rs in Irritation Testing"[go3r]
7.3.1	Skin irritation/corrosion	"Skin Irritation"[go3r] OR "Skin Corrosivity"[go3r] "*In Vivo* Test Methods, Skin Irritation"[go3r] OR "3Rs in Skin Irritation Testing"[go3r] OR "Skin Corrosivity Testing, Replacement Methods"[go3r]
7.3.2	Eye irritation	"*in vivo* test methods, eye irritation"[go3r] OR "3Rs in Eye Irritation Testing"[go3r] "Eye Irritation"[go3r] OR "Eye Corrosivity"[go3r]
7.4	Sensitization	(Sensitisation[go3r] AND "Guinea pig"[go3r]) NOT "Skin Sensitisation"[go3r]
7.4.1	Skin sensitization	"*in vivo* test methods, skin sensitisation"[go3r] "Skin Sensitisation"[go3r] AND (Rodent[go3r] OR "Guinea pig"[go3r] OR "*in vivo*"[go3r]) Mouse[go3r] AND Sensitisation[go3r]
7.4.2	Respiratory sensitization	"Respiratory Sensitisation"[go3r]
7.5	Repeated dose toxicity	"*in vivo* test methods, repeated-dose toxicity"[go3r] "Repeated-Dose Toxicity"[go3r]
7.5.1	Repeated dose toxicity, oral	"28 day oral toxicity study"[go3r] OR "90 day oral non-rodent toxicity study"[go3r] OR "90 day oral rodent toxicity study"[go3r] OR "90 day oral study"[go3r] ("Subacute Oral Toxicity"[go3r] OR "Subchronic Oral Toxicity"[go3r]) ((Subacute[go3r] OR Subchronic[go3r] OR Chronic[go3r]) AND oral[go3r]) NOT Toxicity[go3r]
7.5.2	Repeated dose toxicity, inhalation	"28 day inhalation toxicity study"[go3r] OR "90 day inhalation toxicity study"[go3r] "Subchronic Inhalation Toxicity"[go3r] OR "Subacute Inhalation Toxicity"[go3r] ((Subacute[go3r] OR Subchronic[go3r] OR Chronic[go3r]) AND Inhalation[go3r]) NOT Toxicity[go3r]
7.5.3	Repeated dose toxicity, dermal	"Subchronic Dermal Toxicity"[go3r] OR "Subacute Dermal Toxicity"[go3r] "28 day dermal toxicity study"[go3r] OR "90 day dermal toxicity study"[go3r] ((Subacute[go3r] OR Subchronic[go3r] OR Chronic[go3r]) AND Dermal[go3r]) NOT Toxicity[go3r]
7.5.4	Repeated dose toxicity, other routes	"Repeated-Dose Toxicity"[go3r] AND ("Intraperitoneal Injection"[go3r] OR "Subcutaneous Administration"[go3r] OR "Intravenous Injection"[go3r])
7.6	Genetic toxicity	"Gene-Related Endpoints"[go3r] OR Genotoxicant[go3r] Mutagenicity[go3r] OR Genotoxicity[go3r]
7.6.1	Genetic toxicity, *in vitro*	"3Rs in Mutagenicity/Genotoxicity Testing"[go3r]
7.6.2	Genetic toxicity, *in vivo*	"*in vivo* test methods, mutagenicity/genotoxicity"[go3r]
7.7	Carcinogenicity	Carcinogenicity[go3r] OR Neoplasms[go3r] "*in vivo* test methods, carcinogenicity"[go3r]
7.8	Toxicity to reproduction	
7.8.1	Toxicity to reproduction	"Multigeneration Study"[go3r] OR "Reproduction/Developmental Toxicity Screening Study"[go3r] OR "Repeated Dose Toxicity/Reproductive Screening Study"[go3r] "Reproductive Toxicity"[go3r] NOT "Developmental Toxicity"[go3r] NOT "Ovarian Toxicity"[go3r] NOT Embryotoxicity[go3r] NOT Teratogenicity[go3r] ("Fertility"[go3r] OR "Fertilization"[go3r] OR "Spermatogenesis"[go3r] OR "Genitalia, Male"[go3r] OR "Testis"[go3r]) NOT "Human Foreskin"[go3r]
7.8.2	Developmental toxicity/ teratogenicity	("Developmental Toxicity"[go3r] OR "Prenatal Developmental Toxicity"[go3r] OR Embryotoxicity[go3r] OR Teratogenicity[go3r] OR "Developmental Neurotoxicity"[go3r] OR "Ovarian Toxicity"[go3r]) NOT fish[go3r] NOT "Daphnia magna"[go3r] NOT amphibian[go3r] "neonatal exposure"[go3r] OR "prenatal exposure"[go3r] OR "juvenile exposure"[go3r] OR "maternal weight gain"[go3r] OR "weaning"[go3r] OR "Gestational Exposure"[go3r] ("Prenatal Developmental Toxicity Study"[go3r] OR "Developmental Neurotoxicity Study"[go3r]) NOT fish[go3r] NOT "Daphnia magna"[go3r] NOT amphibian[go3r] ("Congenital, Hereditary, and Neonatal Diseases and Abnormalities"[go3r] OR "Pregnancy Complications"[go3r] OR "Embryonic Structures"[go3r] OR "Developmental Biology"[go3r] OR "Embryo-Related Endpoints"[go3r] OR

(*continued on next page*)

Table 1 (*continued*)

IUCLI5 No.	IUCLID5 category or sub-category	Go3R IUCLID search queries
		"Embryonic and Fetal Development"[go3r]) NOT "Fetal Blood"[go3r]
7.8.3	Toxicity to reproduction, other studies	"*In vivo* test methods, endocrine disruption"[go3r] OR "Endocrine Disrupting Activity"[go3r]
7.9	Specific investigations	
7.9.1	Specific investigations: neurotoxicity	"behavioral studies"[go3r] OR "pain models"[go3r] Neurotoxicity[go3r] OR "Neuronal Endpoints"[go3r] OR "Nervous System Diseases"[go3r] OR "Nervous System"[go3r] "*In vivo* test methods, neurotoxicity"[go3r]
7.9.2	Specific investigations: immunotoxicity	(Immunotoxicity[go3r] -Immunosuppression[go3r]) AND "*in vivo*"[go3r] "*In vivo* test methods, immunotoxicity"[go3r] "Immune System"[go3r] NOT (Sensitisation[go3r] OR Genotoxicity[go3r] OR "3Rs Methods in Toxicity Testing"[go3r] OR "*In Vivo* Test Methods, Sensitisation"[go3r])
7.9.3	Specific investigations: other studies	"Cell Culture System"[go3r] OR "Tissue Culture System"[go3r] OR "Cell Lines, Specific Lines"[go3r] OR "Tissues, Cultured"[go3r] "Hearing Disorders"[go3r] "toxicology, replacement methods"[go3r] NOT ("*in vitro* skin absorption test"[go3r] OR ("BCOP Test"[go3r] AND "eye corrosivity"[go3r]) OR ("ICE Test"[go3r] AND "eye corrosivity"[go3r]) OR "skin corrosivity testing, replacement methods"[go3r] OR "skin irritation testing, replacement methods"[go3r] OR "3Rs in mutagenicity/genotoxicity testing"[go3r]) cytotoxicity[go3r] OR "3Rs in Acute Toxicity Testing"[go3r] OR Photosensitisation[go3r] "Organ Specific Toxicity"[go3r] -Immunotoxicity[go3r] -Neurotoxicity[go3r] "QSAR/*In Silico* Toxicity Testing"[go3r] OR "Structure–Activity Relationship"[go3r]
7.10	Exposure related observations in humans	Human[go3r] OR "Occupational Exposure"[go3r] OR Epidemiology[go3r] "Air Pollutants"[go3r] AND Human[go3r]
7.10.1	Health surveillance data	
7.10.2	Epidemiological data	
7.10.3	Direct observations, clinical cases, poisoning, etc.	
7.10.4	Sensitisation data, human	Sensitisation[go3r] AND Human[go3r] "Human Local Skin Tolerance Testing"[go3r] "Hypersensitivity, Delayed"[go3r] OR Dermatitis[go3r]
7.10.5	Exposure related observations in humans, other data	
7.11	Toxic effects in livestock and pets	(Dog[go3r] OR Cat[go3r] OR Cow[go3r] OR Pig[go3r]) AND (Toxicity[go3r] OR "*in vivo* test methods, mammalian toxicity"[go3r]) NOT "Serum Albumin, Bovine"[go3r] NOT "Fetal Calf Serum"[go3r] NOT cytotoxicity[go3r] NOT "BCOP Test"[go3r]
7.12	Additional toxicological information	Toxicity[go3r] NOT ("Mammalian Toxicity"[go3r] OR Cytotoxicity[go3r] OR Ecotoxicity[go3r]) -ADME[go3r] -Metabolism[go3r] -"Metabolic Processes"[go3r] (Mammal[go3r] AND (Exposure[go3r] OR Administration[go3r])) OR ((Exposure[go3r] OR Administration[go3r]) AND ("Dosage & Scoring, *In Vivo*"[go3r] OR "Study Duration"[go3r]) AND Mammal[go3r])

from the BASF registration dossiers amounted to 146, 150, and 82 references for PT, TL, and HC, respectively (total: 378 references). Of these, 116 (PT), 136 (TL), and 58 (HC) references (total: 310 references, i.e. 82%) were available in either PubMed or TOXNET. Amongst these 310 references, a mere 8 documents were available in TOXNET, but not in PubMed (Table 2).

3.1. IUCLID categorization – general observations

An overview of the IUCLID SQs assigned to each IUCLID5 category or sub-category is provided in Table 1. These IUCLID SQs reflect the vocabulary used in the documents to refer to the respective endpoints and hence do not reflect 'scientific definitions' of the endpoints. Compilation of relevant IUCLID SQs is dependent upon a sufficiently broad set of documents allowing recognition of the spectrum of relevant vocabulary and relevant patterns of vocabulary for the respective endpoints. None of the dossiers contained any documents for 4 of the 30 IUCLID5 (sub-)categories, i.e. (7.4.2) 'respiratory sensitisation'; (7.5.2) 'repeated dose toxicity: dermal'; (7.10) 'health surveillance data', or (7.11) 'toxic effects on livestock and pets' (Tables 3a–3c). For 6 IUCLID5 categories, only 1–5 references were listed in the three registration dossiers altogether (e.g. (7.2.2) 'acute toxicity: dermal'; (7.9.2) 'specific investigations: immunotoxicity'). Hence, the IUCLID SQs compiled for these endpoint categories can only be considered tentative.

However, also for those IUCLID5 endpoint categories, for which a broad set of references was available within the three registration dossiers, compilation of the IUCLID SQs was differently intricate for different endpoint categories, which is also reflected by the complexity of the assigned IUCLID SQs (Table 1). For once, the patterns of vocabulary used to describe the different endpoints turned out to be of differing complexity, e.g. overall unambiguous and straight-forward for (7.1.2) 'dermal absorption' or (7.6) 'genetic toxicity', but overall unspecific and intricate for (7.5) 'repeated dose toxicity' or (7.8) 'toxicity to reproduction'.

Furthermore, the IUCLID5 categories themselves are differently complex from a scientific point of view: A number of categories cover precise test methods, possibly even specified by one or more OECD test guidelines (TG; e.g. (7.8.1) 'reproduction toxicity, see below). Others, however, are defined by descriptions of relevant information unspecified by guideline-related test methods. Vagueness of the IUCLID5 categories further complicates the task of compiling high precision and high recall IUCLID SQs.

The following IUCLID5 categories can be related to a variety of toxicological endpoints or are not necessarily fully concordant with specific test methods:

- IUCLID5 categories referring to 'other routes' (i.e. (7.2.4) 'acute toxicity: other routes' and (7.5.4) 'repeated dose toxicity: other routes') or referring to 'other studies' (i.e. (7.8.3) 'toxicity to reproduction: other studies').
- (7.9.1.) 'Specific investigations: immunotoxicity' and (7.9.2) 'specific investigations: neurotoxicity': These endpoints can be addressed, e.g. in the course of 28-day or 90-day studies.

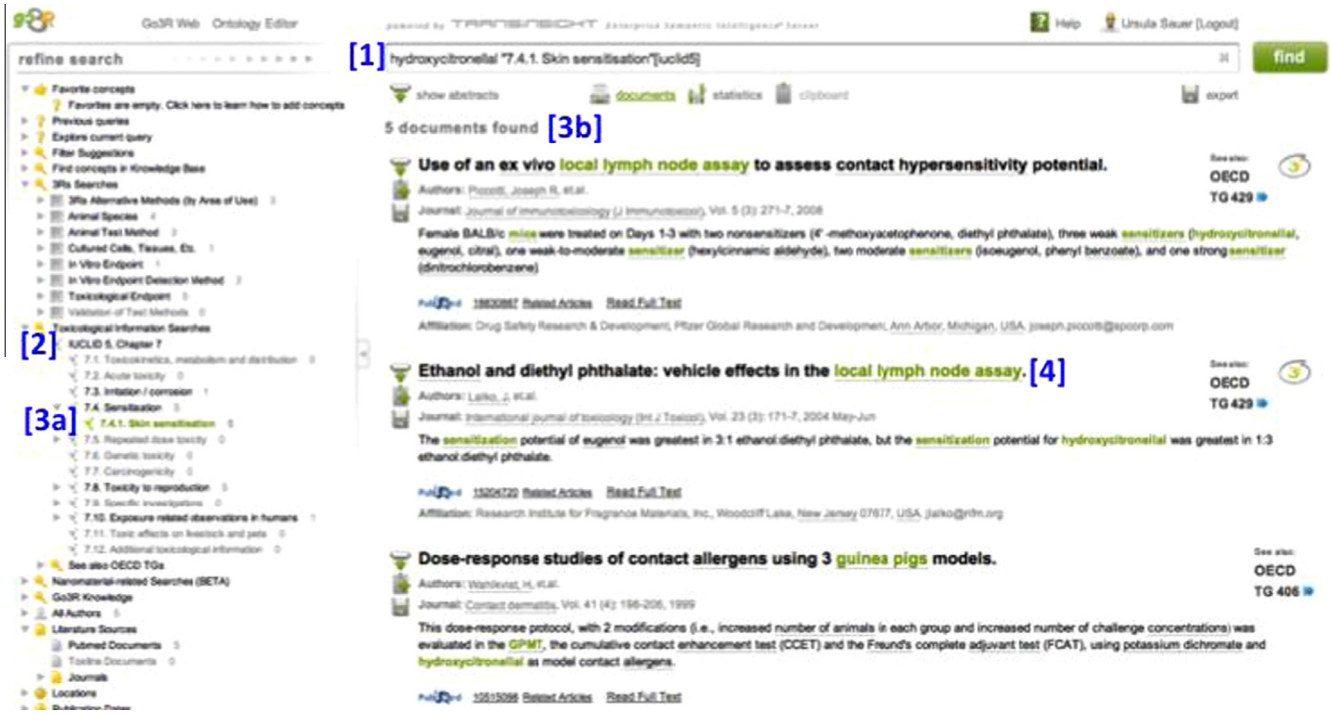

Fig. 4. Test substance-specific search query ('Hydroxycitronellal') making use of the IUCLID user interface to filter for endpoint-specific information (search date: 19 November 2013). Footnote to Fig. 4: (1) The test substance-specific search query 'hydroxycitronellal' has been typed into the search query input box, this time only searching for exactly this word, and not making use of Go3R ontology concepts. Therefore, unlike in the example presented in Fig. 1, there is no '[go3r]' extension behind the search query term. (2) The IUCLID user interface is part of the dynamic table of contents on the left hand side of the screen. (3a) Filtering for endpoint-specific information, again by unfolding the dynamic table of contents and clicking onto the selected concept (in the example: IUCLID category 7.4.1 'skin sensitisation'; see also Footnote to Fig. 2). (3b) Filtering for IUCLID category 7.4.1 'skin sensitisation' specifically filters 5 of originally 66 documents (search date: 19 November 2013). (4) In the list of documents, vocabulary that was used for sorting into the respective category is highlighted in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

vTherefore an overlap of the vocabulary used to describe such investigations with the one used for (7.5) 'repeated dose toxicity' categories is likely.

- (7.9.3) 'Specific investigations: other studies': In accordance with the IUCLID user manual (ECHA and OECD, 2007; page 1896), this category should be used to describe "*studies on behavioral effects*, *biochemical or cellular interactions*, *chemo-biokinetics general studies*, *cytotoxicity*, *endocrine system modulation*, *hematoxicity*, *hepatotoxicity*, *mechanistic studies*, *methaemoglobinaemia*, *nephrotoxicity*, *phototoxicity*, *respiratory irritation*, *splenic toxicity*, *or toxicogenomics*". (Of note, these study parameters listed are also addressed in, e.g. repeated dose toxicity studies. Therefore, an overlap in the respective endpoint-specific vocabularies is to be expected.)
- IUCLID5 categories referring to human data other than (7.10.4) 'human sensitisation data', i.e. (7.10.1) 'health surveillance data', (7.10.2) 'epidemiological data', (7.10.3) 'direct observations, clinical cases, poisoning incidents and other', (7.10.5) 'exposure related observations in humans: other data'.

Table 2
Overview of the Go3R IUCLID search query annotation work based upon evaluation of the references for the BASF SE registration dossiers for pentanol, toluene, and hydroxycitronellal.

Test substance	References in registration dossier	No. of references (one listing per endpoint)	No. of references in PubMed or TOXNET/only in TOXNET	No. of references sorted into correct IUCLID (sub-)category	No. of references sorted into 7.12 (lack of endpoint-specific vocabulary)	
Pentanol	284	146	116/4	69	16	Percent total available (n = 310)
Toluene	159	150	136/1	121	10	
Hydroxycitronellal	104	82	58/3	53	0	
Total	547	378	310/7	243	26	
Total number of correct or 7.12 sortings				269		87
Total number of incorrect sortings				41		13

Of which:		Percent total incorrect (n = 41)
Go3R sorting into 7.1.1 or 7.9.1 due to metabolism- or neurology-specific vocabulary in title and/or abstract	19	47
Go3R sorting into different categories than in registration dossier (but not 7.1.1 or 7.9.1)	14	34
Go3R sorting into different subcategory than in registration dossier (e.g. 7.8.2 instead of 7.8.3)	3	7
Incorrect Go3R sorting (or no sorting at all) of documents without abstract having inconclusive vocabulary in title	5	12

- (7.12) 'Additional toxicological information': In accordance with the IUCLID user manual (ECHA and OECD, 2007; page 1896), this category should be used to sort information "*that does not fit into any of the specific chapters*".

Initial work on the compilation of the IUCLID SQs revealed a number of further – linguistic – limitations in regard to sorting documents into exactly the identical IUCLID5 sub-category as the one selected manually by BASF, based exclusively on the vocabulary used in the titles, abstracts, and MeSH terms/keywords:

For most IUCLID5 categories (e.g. (7.2) 'acute toxicity'), the vocabulary used was frequently not sufficiently precise to allow distinguishing the respective different sub-categories with certainty (i.e. 7.2.1–7.2.4). Therefore, it was decided to compile general IUCLID SQs in addition to the sub-category-specific IUCLID SQs. Documents were sorted into the respective umbrella categories (e.g. (7.2) 'acute toxicity'; (7.6) 'genetic toxicity') whenever the vocabulary was not sufficiently precise to allow for a more specific sorting into the sub-categories (e.g. 7.2.1–7.2.4, distinguishing between different routes of exposure during acute toxicity studies; or 7.6.1–7.6.2, distinguishing between *in vitro* or *in vivo* genotoxicity studies). Documents with unambiguous vocabulary, however, were sorted into the more specific sub-categories (Table 1).

Based upon the 310 references evaluated, for the IUCLID5 categories 7.1 and 7.8, complete distinction into sub-categories appeared feasible. Therefore no general IUCLID SQs were assigned

Table 3a
Outcome of the IUCLID categorization for the references used in the registration dossier for pentanol.

Ch.	Chapter name	No. of assignments by BASF	References not available	Total Go3R available	Sorted correctly by IUCLID SQ	% Sorted correctly	References sorted incorrectly
7.1.1	Basic toxicokinetics	17	4	13	11	85	2
7.1.2	Dermal absorption	3	0	3	3	100	0
7.2.1	Acute toxicity: oral	11	3	8	5	63	3
7.2.2	Acute toxicity: inhalation	7	0	7	5	71	2
7.2.3	Acute toxicity: dermal	7	4	3	1	33	2
7.2.4	Acute toxicity: other routes	9	3	6	2	33	4
7.3.1	Skin irritation/corrosion	4	0	4	2	50	2
7.3.2	Eye irritation/corrosion	7	2	5	3	60	2
7.5.1	Repeated dose toxicity: oral	5	0	5	3	60	2
7.5.2	Repeated dose toxicity: inhalation	1	0	1	0	0	1
7.5.4	Repeated dose tox.: other routes	1	0	1	1	100	0
7.6.1	Genetic toxicity *in vitro*	9	4	5	5	100	0
7.6.2	Genetic toxicity *in vivo*	1	0	1	1	100	0
7.7	Carcinogenicity	2	0	2	2	100	0
7.8.1	Toxicity to reproduction	1	0	1	0	0	1
7.8.2	Developm. toxicity/teratogenicity	2	0	2	2	100	0
7.9.1	Spec. investigations: neurotoxicity	5	0	5	5	100	0
7.9.3	Spec. investigations: other studies	29	4	25	15	60	10
7.10.3	Direct observations: clin. cases, etc.	1	1	0	-		-
7.10.4	Sensitisation data (humans)	4	0	4	2	50	2
7.10.5	Exposure related observations in humans: other data	6	1	5	0	0	5
7.12	Additional toxicological information	14	4	10	1	10	9
Total		146	30	116	69		47

Table 3b
Outcome of the IUCLID categorization for the references used in the registration dossier for toluene.

Ch.	Chapter name	No. of assignments by BASF	References not available	Total Go3R available	Sorted correctly by IUCLID SQ	% Sorted correctly	References sorted incorrectly
7.1.1	Basic toxicokinetics	15	2	13	13	100	0
7.1.2	Dermal absorption	5	0	5	5	100	0
7.2.1	Acute toxicity: oral	3	0	3	2	67	1
7.2.2	Acute toxicity: inhalation	4	2	2	1	50	1
7.2.3	Acute toxicity: dermal	1	0	1	0	0	1
7.3.1	Skin irritation/corrosion	1	0	1	1	100	0
7.3.2	Eye irritation/corrosion	2	0	2	2	100	0
7.5.2	Repeated dose toxicity: inhalation	4	1	3	2	67	1
7.5.4	Repeated dose tox.: other routes	3	0	3	0	0	3
7.6.1	Genetic toxicity *in vitro*	6	3	3	3	100	0
7.6.2	Genetic toxicity *in vivo*	10	4	6	6	100	0
7.7	Carcinogenicity	5	0	5	5	100	0
7.8.1	Toxcity to reproduction	5	0	5	5	100	0
7.8.2	Developm. toxicity/teratogenicity	16	1	15	14	93	1
7.8.3	Toxicity to reprod.: other studies	1	0	1	0	0	1
7.9.1	Spec. investigations: neurotoxicity	22	0	22	22	100	0
7.9.2	Spec. investigations: immunotox.	1	0	1	1	100	0
7.9.3	Spec. investigations: other studies	11	0	11	7	55	4
7.10.2	Epidemiological data	23	0	23	23	100	0
7.10.3	Direct observations: clin. cases, etc.	3	1	2	2	100	0
7.10.5	Exposure related observations in humans: other data	9	0	9	7	78	2
Total		150	14	136	121		15

Table 3c
Outcome of the IUCLID categorization for the references used in the registration dossier for hydroxycitronellal.

Ch.	Chapter name	No. of assignments by BASF	References not available	Total Go3R available	Sorted correctly by IUCLID SQ	% Sorted correctly	References sorted incorrectly
7.1.1	Basic toxicokinetics	1	0	1	1	100	0
7.2.1	Acute toxicity: oral	1	1	0	–		–
7.3.1	Skin irritation/corrosion	9	2	7	6	86	1
7.4.1	Skin sensitisation	23	9	14	12	86	2
7.5.1	Repeated dose toxicity: oral	2	1	1	1	100	0
7.5.2	Repeated dose toxicity: inhalation	1	0	1	1	100	0
7.6.1	Genetic toxicity *in vitro*	1	0	1	1	100	0
7.6.2	Genetic toxicity *in vivo*	2	1	1	1	100	0
7.7	Carcinogenicity	1	0	1	1	100	0
7.8.2	Developm. toxicity/teratogenicity	2	0	2	2	100	0
7.9.3	Spec. investigations: other studies	3	0	3	3	100	0
7.10.4	Sensitisation data (humans)	28	5	23	22	96	1
7.10.5	Exposure related observations in humans: other data	7	5	2	2	100	0
7.12	Additional toxicological information	1	0	1	0	0	1
Total		82	24	58	53		5

to these categories, but only specific IUCLID SQs to the respective sub-categories. In regard to the IUCLID5 category (7.10) 'exposure related observations on humans', a sub-category-specific IUCLID SQ was only formulated for documents relating to the sub-category (7.10.4) 'human sensitisation data'. Due to vocabulary restraints, it was decided to sort documents for all other sub-categories only into the general 7.10 category (Table 1).

In evaluating the correctness of the IUCLID search query categorization, such general sorting (e.g. 7.2, 7.6, or 7.10) was rated as 'correct', even though the respective manual sorting of documents performed by BASF had resulted in sorting into the more refined sub-category (e.g. 7.2.1, 7.2.2, etc.).

For a number of documents, the vocabulary used in the titles and abstracts did indicate some form of toxicity study, but it was not sufficiently detailed to allow determining the specific toxicological endpoint(s) addressed (e.g. if the only relevant word used was 'toxicity'). Such documents frequently (but not always) did not have an abstract so that Go3R text mining was only possible for the document title and MeSH terms/TOXNET keywords. If the title mentioned the respective toxicological endpoint, the document could nevertheless be sorted correctly. However, in a number of documents without abstract, the title merely revealed that the document contained toxicologically relevant information and did not indicate any toxicological endpoint. In accordance with the IUCLID user manual instructions for (7.12) 'additional toxicological information' (see above), it was decided to sort such documents into 7.12 (Table 1). In the registration dossiers, these documents had been sorted into endpoint-specific categories, since manual document evaluation allowed use of the contents of the entire text of the document.

The initial set of IUCLID SQs (not shown) turned out to be very restrictive (e.g. by limiting the IUCLID SQ to the enumeration of relevant OECD TGs). As a result, recall was impaired and large numbers of relevant documents were not sorted into the respective IUCLID5 categories. Therefore the initial set of IUCLID SQs was revised to enable a more broad retrieval of documents. This revision took into account topic-relevant general vocabulary that was not necessarily only related to regulatory toxicity testing, but also to fundamental biomedical research (e.g. 'spindle disturbance' or 'chromosomal aberration' for 'genotoxicity', which is also used in the context of genetic research). This adaptation of the IUCLID SQs considerably increased the recall for endpoint-specific search queries, however at the expense of precision: Endpoint-specific searches (e.g. user search query 'genotoxicity', 'eye irritation', etc.) will also retrieve documents related to fundamental biomedical research that are not relevant for regulatory toxicity testing. However, the Go3R IUCLID user interface has been mainly developed for test substance-specific search queries (e.g. user search query 'butadiene', 'formaldehyde', etc.). Therefore the number (and type)

Table 4
Go3R IUCLID sorting of references from the sensitisation review article Mehling et al. (2012).

Types of references	Number of references	Number of references sorted correctly[a]	Number of references additionally mentioning endpoint irritation/corrosion	Number of references with correct additional sorting into (7.3) 'irritation/corrosion'
References mentioning a specific *in vitro* or *in vivo* skin sensitisation test method	66	66	5	5
References describing how test substances induce allergic immune reactions *in vitro* without indication of a specific test method	37	35	19	19
References describing molecular or cellular processes underlying allergic immune reactions without reference to test substance or test method	35	35	4	4
References describing cell culture technologies (e.g. cell isolation)	6	6	0	0
Total number of references related to topic sensitisation and available in PubMed/TOXNET	144	142	38	38
Number of references mentioning other toxicological endpoints (e.g. dermal absorption) or unavailable in PubMed/TOXNET (e.g. EU legislation, guidance documents)	22	–	–	–
Total number of references in sensitisation review (Mehling et al., 2012)	166	–	–	–

[a] 'Correct' sorting was defined as sorted into either 7.4 (or its sub-categories), .9.3, or 7.10.4. See text for explanations.

of documents retrieved is restricted by the search query itself, which by default mentions the respective test substances under investigation: All fundamental biomedical studies that do not mention chemical substances should not be retrieved in the first place.

3.2. Outcome of the IUCLID search query-based annotation

The numbers of correct and incorrect categorizations of the references of the three registration dossiers for pentanol, toluene, and hydroxycitronellal upon finalization of the compilation of the IUCLID SQs are presented in Tables 3a, 3b, and 3c, respectively. 100% Correct sorting (regardless of the total numbers of references) in at least 2 registration dossiers were obtained for a total of 7 IUCLID5 sub-categories. 100% Correct sorting in one of the registration dossiers was obtained for an additional 9 IUCLID5 sub-categories. Hence, for 16 of the 26 IUCLID5 sub-categories for which documents were available in the gold standards, Go3R achieved 100% correct sorting in at least one of the registration dossiers.

Overall, the Go3R IUCLID SQs allowed sorting 78.5% (i.e. 243) of the 310 references available in PubMed or TOXNET into the same sub-category or umbrella category as BASF had performed manually. For a further 8.5% (i.e. 26) of the references, the vocabulary used in title, abstract, and MeSH terms was unspecific and only allowed sorting into (7.12) 'additional toxicological information'. Hence, based upon the available vocabulary, 87% of the references were sorted into the 'best possible' IUCLID5 category (Table 2).

Consequently, 41 references (13% of all references) were neither sorted into the correct umbrella category, sub-category, nor into the general category 7.12. The following reasons were discerned for this (Table 2):

- For 19 of the 41 references (47% of the references sorted incorrectly), metabolism- or neurology-specific vocabulary used in the title and/or abstract resulted in Go3R IUCLID sorting into either (7.1.1) 'basic toxicokinetics' or (7.9.1) 'neurotoxicity' (see below, specific comments to the respective IUCLID SQ for 7.1.1 and 7.9.1). Manual BASF sorting, however, which took into account the concrete test methods applied in the respective studies (and not only the topic reported in title and abstract, i.e. metabolism, neurology), resulted in sorting into human data-specific categories (7.10) or the unspecific categories (7.12) 'additional toxicological endpoints' and (7.9.3) 'specific investigations: other studies'.
- Due to the vocabulary used, 3 of the 41 references (7%), were sorted into a different sub-category of the same umbrella category: Two references were sorted into (7.10.4) 'human sensitisation data' instead of (7.10.5) 'exposure related observations in humans: other data'. One of these documents did not have an abstract, but the 'human patch test' was mentioned in the MeSH terms. In the other document, the 'human patch test' was mentioned in the abstract, but its main topic encompassed presentation of a database on the skin irritation and sensitisation potential of chemical substances. One further document was sorted into (7.8.2) 'developmental toxicity/teratogenicity' instead of (7.8.3) 'toxicity to reproduction: other studies' because it referred to 'gestational exposure'. The described test system, however, was a test guideline-unrelated animal model using timed-pregnant rats to investigate the effects of solvent abuse.
- 14 References (34%) were sorted into a different IUCLID5 category than the one used in the registration dossiers. 10 of these references were sorted into the 'unspecific' IUCLID5 categories 7.9.3, 7.10, or 7.12 – either in the registration dossiers or in the Go3R sorting (e.g. one document was sorted into 7.12 instead of 7.10, because its only indication of 'human' data was the term 'automobile driver').
- Finally, for a further 5 references corresponding to 4 documents (1 document had been assigned to two different IUCLID5 sub-categories in the registration dossiers), Go3R did not recognize their relevance for regulatory toxicity purposes and did not sort them into the IUCLID user interface at all. These 4 references correspond to 12% of the references sorted incorrectly, and to 1.6% of all available references. None of these 4 documents had an abstract, and their titles were inconclusive as to toxicity testing: The authors of the 4 documents referred to the types of studies performed with the following formulations: 'influence of [test substance] on biological objects', 'odour properties', 'cardiovascular effects', and 'action of the blood pressure'.

3.3. Comments on specific IUCLID search queries

Documents that had been assigned to (*7.1.1*) *'basic toxicokinetics'* in the registration dossiers frequently only contained very general terms in their titles and abstracts to describe the endpoints addressed (e.g. 'elimination' or 'metabolism'). To improve recall of the search retrievals, IUCLID SQs were compiled for the 7.1.1 IUCLID sub-category containing such general terms (Table 1). As a result, however, the IUCLID-SQ-based annotation also sorts documents into this section in which studies other than toxicokinetics studies were described (see above, reasons for incorrect sorting). The gold standard further revealed that a number of authors use the term 'pharmacokinetics' even though they investigated the biokinetics of non-pharmacological chemical substances. For the time being, this term has not been included in the IUCLID SQs to avoid retrieving an abundance of irrelevant documents referring to pharmacological topics. The concepts 'Physiologically Based Toxicokinetic Modelling' and 'Physiologically Based Pharmacokinetic Modelling', on the other hand, were included in the 7.1.1 IUCLID SQs since application of such computer tools was described in a number of documents that had been assigned to 7.1.1 in the gold standard. These examples show how precision and recall have to be balanced against each other.

In the registration dossiers, documents that had been assigned to (*7.2*) *'acute toxicity'* or one of its sub-categories frequently did not contain an abstract, and the titles referred to the given endpoint in general terms (e.g. 'intoxication', 'poisoning', or 'hygienic assessment'). Therefore IUCLID SQs were compiled to cover these general terms. To avoid retrieving large numbers of irrelevant documents by making use of such general search queries, Go3R ontology terms relating to mammals were included into these IUCLID SQs. With this search query constraint, studies performed on species relevant for ecotoxicity testing (e.g. fish, daphnia, worms) were excluded from the search retrievals (Table 1).

Based upon the vocabulary used in documents assigned to (*7.3*) *'irritation/corrosion'*, full distinction between skin irritation studies performed in animal tests or *in vitro* tests, on the one hand, or in studies with human volunteers on the other hand is not possible. Therefore an overlap between (7.3.1) 'skin irritation' and (7.10) 'exposure related observations in humans', but also (7.9.3) 'specific investigations: other studies' (for studies describing the outcome of *in vitro* assays without regulatory acceptance) is currently inevitable. Test systems making use of human reconstituted skin equivalents have not been covered by the IUCLID SQs for the irritation–corrosion endpoints (but are listed in 7.9.3 instead), since such test systems are also used to evaluate a variety of other endpoints.

Documents referring to the endpoint sensitisation can be assigned to different IUCLID5 sub-categories depending on the types of studies performed. In the case of *in vivo* data from animal studies, they should be assigned to (*7.4*) *'sensitisation'*, further distinguishing between (7.4.1) 'skin' or (7.4.2) 'respiratory sensitisation'. If data were obtained *in vitro* making use of test methods

without regulatory acceptance, they should be assigned to (*7.9.3*) *'specific investigations: other studies'*. Finally, in the case of studies performed on humans, they should be assigned to *7.10.4 'human sensitisation data'* or in the case of clinical case reports, etc., to one of the other 7.10 sub-categories. However, the vocabulary used to describe any of these types of sensitisation studies is oftentimes very similar (some abstracts and titles, e.g. only contain the term 'allergic contact dermatitis'). Therefore it was not possible to compile the respective IUCLID SQs to fully distinguish between the 7.4.1, 7.4.2, 7.9.3, and 7.10.4 sub-categories (Table 1). In consequence, documents might be sorted into more than one of these categories or into a different sensitisation-related category than would be selected upon manual evaluation of the full-text document.

Especially in the context of (*7.5*) *repeated-dose toxicity*, authors oftentimes do not explicitly mention the study performed, but paraphrase it by using formulations such as "*...rats were exposed to... ppm for 28 days...in inhalation chamber*" or "*...mg/kg body weight of test substance... was administered to mice for 5 days/week for 13 weeks*". However, such indirect formulations were also recorded for documents assigned to the 'genetic toxicity' and 'reproduction toxicity' endpoint categories. Therefore, IUCLID SQs addressing indirect formulations were only included in the IUCLID category (*7.12*) *'additional toxicological information'* (Table 1). Further annotation work is required to aim to distinguish the underlying endpoint specific studies (possibly even distinguishing further e.g. between '28 days' as an indication for a sub-acute study, '13 weeks' for a sub-chronic study, and '2 years' for a chronic study).

For the (*7.6*) *'genetic toxicity'* category, one IUCLID SQ was compiled to cover gene-related endpoints (e.g. 'chromosomal aberration'). As mentioned above, such terms are also used in the context of fundamental genetic research. To prevent search result precision from being unduly diminished by retrieving documents that are unrelated to toxicity testing, this IUCLID SQ was restricted to documents available in TOXNET where the restriction to toxicity-related documents is inherent to the database. For documents only available in PubMed, a database which also covers medical and biological research, the IUCLID SQ on gene-related endpoints would have resulted in retrieval of too many irrelevant documents.

Further in regard to the (7.6) 'genetic toxicity' category, the terms 'comet assay' or 'micronucleus assay' repeatedly resulted in erroneous categorization into (7.6.1) '*in vitro* genetic toxicity' even though an *in vivo* study was recorded in the respective document: Since the terms 'comet assay' and 'micronucleus assay' are frequently being used to refer to an *in vitro* test method, they have been assigned accordingly in the Go3R ontology. However, strictly speaking, the comet or micronucleus assays are procedures for detecting DNA damage that can be performed in the course of either *in vivo* or *in vitro* testing. Therefore documents mentioning these assays might contain information on *in vivo* or *in vitro* test methods. This example underlines that imprecise linguistic use can neither be fully reflected, nor compensated in an ontology.

In the registration dossiers, as a rule, documents were assigned to the sub-categories of (*7.8*) *'toxicity to reproduction'* if they presented results from the following studies:

- (7.8.1) 'Toxicity to reproduction': OECD TG 415 (one-generation reproduction toxicity study), OECD TG 416 (two-generation reproduction toxicity), OECD TG 421 (reproduction/developmental toxicity screening test), OECD TG 422 (combined repeated dose toxicity study with the reproduction/developmental toxicity screening test), OECD TG 443 (extended one-generation reproduction toxicity study);
- (7.8.2) 'Developmental toxicity/teratogenicity': OECD 414 TG (prenatal development toxicity study), OECD TG 426 (developmental neurotoxicity study);
- (7.8.3) 'Toxicity to reproduction: other studies': specific tests, e.g. OECD TG 440 (uterotrophic bioassay in rodents) or OECD TG 441 (Hershberger bioassay).

An overlap of the outcome of such studies between the sub-categories 7.8.1 and 7.8.2 was possible and also in regard to (7.9.1) neurotoxicity (in the case of developmental neurotoxicity studies). In the registration dossiers, such an overlap is accounted for by 'cross-referencing' of documents between endpoints. In the Go3R IUCLID SQ-based annotation, however, the documents were categorized in accordance to the pattern of vocabulary used in the respective document. Therefore full accordance with the manual sorting is unlikely. In regard to (7.8.3), further work is required to adapt the IUCLID SQs to exclude ecotoxicological studies performed in the context of endocrine disruption testing.

Due to the spectrum of vocabulary used in documents assigned to the IUCLID5 category (*7.9.1*) *'neurotoxicity'*, the corresponding IUCLID SQs were designed to include documents containing terms relating to neuronal endpoints, neuronal body systems and body structures, neurological diseases, as well as pain models or behavioral studies (Table 1). Again, further work is required to determine how best to balance precision and recall of search retrievals when taking into account this broad spectrum of vocabulary.

In accordance with the IUCLID user manual, the IUCLID SQ relating to (*7.9.3*) *'specific investigations: other studies'* cover organ specific toxicity other than neurotoxicity (7.9.1) or immunotoxicity (7.9.2); *in vitro* test methods that have not yet achieved regulatory acceptance (covering not only concrete test methods, but also *in vitro* studies making use of, e.g. specific cell lines or tissue culture systems); and *in silico* methods. IUCLID SQ were not compiled to cover studies in which primary cells are used, since the vocabulary depicting *in vitro* studies with primary cells was not found to be sufficiently different from that describing cellular effects observed during *in vivo* investigations.

3.4. Go3R IUCLID sorting of references from the 'sensitisation review' Mehling et al. (2012)

The extensive literature review on non-animal test methods for predicting skin sensitisation potentials (Mehling et al., 2012) contains a total of 166 references. Of these, 144 were discerned to be directly related to the toxicological endpoint 'sensitisation' and to be available in PubMed or Toxnet. In 66 references, a specific *in vitro* or *in vivo* method was indicated in title or abstract. All of these references were sorted into the IUCLID5 sub-categories (7.4.1) 'skin sensitisation', (7.9.3) 'specific investigations: other studies' (in the case of *in vitro* studies), or (7.4.10) 'sensitisation: human data' (Table 4).

A further 37 references presented fundamental research describing how test substances induce allergic immune reactions *in vitro* without making reference to a specific test method. 35 of these references were sorted into one of the IUCLID5 sub-categories 7.4.1, 7.9.3, or 7.10.4. Two references were not sorted into any of these sub-categories.

A further 35 references described molecular or cellular processes underlying allergic immune reactions and 6 references specific cell culture technologies (e.g. cell isolation). Due to the pattern of vocabulary used in the titles and abstracts of these documents, related to the endpoint sensitisation, all of these documents were also sorted into the IUCLID5 sub-categories 7.4.1, 7.9.3, or 7.10.4.

A total of 38 documents, in which skin sensitisation was compared to skin irritation were additionally correctly sorted into the IUCLID5 category (7.3) 'irritation/corrosion' or its sub-category 7.3.1. 63 documents were sorted into (7.9.2) 'specific investigations: immunotoxicity' in addition to one of the sensitisation endpoints 7.4.1, 7.9.3, 7.10.4 pointing to an overlap of the pattern of

vocabulary used to describe the endpoints immunotoxicity and skin sensitisation (data not shown). Moreover, in a number of documents, the term 'immunotoxicity' was used to describe 'skin sensitisation'.

In vitro endpoints mentioned in the 166 reference documents were predominantly 'immune cell-related endpoints', 'protein-related endpoints', and 'cell-related endpoints' but also 'gene related-endpoints' (e.g. 'gene expression'). The latter documents were additionally sorted into the IUCLID category (7.6) 'genetic toxicity'. Further ontological work is required to determine patterns of vocabulary allowing distinguishing types of gene expression pointing to genotoxicity studies in comparison to studies related to other toxicological endpoints, such as skin sensitisation.

4. Discussion

It was the goal of the present study to explore the feasibility of automatically categorizing and sorting scientific literature according to a given scientific topic. Specifically, Go3R, the knowledge-based search engine for alternatives to animal testing, has been extended to automatically sort documents of relevance for regulatory purposes by toxicological endpoint *regardless of the vocabulary used by the respective authors*. Both the collection of information from non-animal test methods and from *in vivo* methods can serve to prevent new animal testing thereby directly furthering the 3Rs principle as it has been implemented in Directive 2010/63/EU.

The underlying challenge of the automatic sorting implies recognizing the spectrum of relevant patterns of vocabulary that authors use to describe studies performed for the different toxicological endpoints. Evidently, a sufficiently large set of documents is required as gold standard to allow discerning such patterns. Considering that the gold standard used consisted of 310 references from a mere three REACH registration dossiers, the outcome of the present study is entirely satisfactory (i.e. 87% of the references were sorted into the 'best possible' IUCLID5 category, and for 16 of the 26 IUCLID5 toxicological endpoint categories that were covered by the gold standard, Go3R achieved 100% correct sorting for the respective set of references of at least one of the registration dossiers; see above and Tables 2, 3a, 3b and 3c).

Limitations in regard to the gold standard are not only related to the relatively small number of references evaluated, but also to the circumstance that individual substances have different patterns of toxicity, resulting in different foci of endpoints addressed in the respective registration dossiers: All 12 IUCLID5 toxicological information categories were not equally addressed in three registration dossiers or the 310 references. As a result, the observations (in regard to patterns of vocabulary) underlying the annotation work are specific to the spectrum of toxic effects observed for the three substances pentanol, toluene, and hydroxycitronellal. The task of formulating IUCLID SQs to be used for annotation in Go3R cannot be considered finalized. Further work taking into account additional registration dossiers (for substances with different patterns of toxicity) would further improve precision and recall of the IUCLID SQ-based annotation work. Furthermore, the IUCLID5 sections relating to effects on biotic systems, i.e. ecotoxicological information, have not yet been included in the Go3R IUCLID user interface.

Inherent features of the PubMed and TOXNET databases also produce unsurpassable limitations of the automatic IUCLID sorting. The PubMed and TOXNET databases allow searching in the titles, abstracts, and MeSH terms/keywords of the documents, but not in their full-text bodies. Important information on the studies recorded will only be available, e.g. in the 'Materials and Methods' sections of publications. The vocabulary used in document abstracts, on the other hand, oftentimes is not sufficiently precise to allow determination of the concrete toxicological endpoint addressed in the study.

This was also shown by evaluation of the references of the sensitisation review article, Mehling et al. (2012): Whereas Go3R was able to identify the topic 'sensitisation' for all but 2 of the 144 references from this article, exact sub-sorting into the precise subcategories 7.4.1, 7.4.2, 7.9.3, or 7.10.4 was not possible based upon the vocabulary of titles and abstracts alone. In this context, it should further be noted that review articles are unlikely to cover all toxicologically relevant information in their abstracts. Nevertheless only singular documents (5 references out of 310, i.e. 1.6%) were not recognized as relevant for regulatory toxicity purposes and hence were not sorted into the Go3R IUCLID user interface at all. None of these 5 references (corresponding to 4 different documents) had an abstract so that text mining was only possible in their titles and MeSH terms/keywords.

Furthermore, a large number of publications in PubMed are medical publications that are (widely) irrelevant for regulatory purposes in regard to chemical substances. Search results might include documents, which mention the test substance under investigation and, e.g. the endpoint under investigation, but in the context the two are not related. Without full-text text mining it is frequently not possible to recognize such correlations or to determine whether the substance under investigation has been used as a control or reference substance (or as test substance in a pharmacological context).

Considering that the PubMed and TOXNET data pools contain 82% of the references recorded in the 3 registration dossiers, it is further evident that these databases are not sufficient for comprehensive searching for toxicological information to be used for regulatory purposes. Special purpose databases, such as the *GESTIS information system on hazardous substances* of the *German Social Accident Insurance Institutions* (Deutsche Gesetzliche Unfallversicherung; http://www.dguv.de/ifa/gestis) or the *Toxicity Reference Database* (ToxRefDB; Martin et al., 2009) run by the USA *Environmental Protection Agency* (EPA), specifically collect toxicologically relevant information for chemical substances. The EPA additionally provides the ACToR database (*Aggregated Computational Toxicology Resource*; http://actor.epa.gov/; Judson et al., 2008), a repository combining 500 public sources with information on over 500,000 environmental chemicals that are searchable by chemical name, identifier, or structure. Processed toxicological information is also available in the ECHA's inventory of registered substances (http://echa.europa.eu) and the OECD SIDS (Screening Information Data Sets) database (http://www.inchem.org/pages/sids.html).

It remains a matter of additional annotation work, exceeding the work on the IUCLID SQs, to allow retrieving documents in which the test substance in question is not mentioned in title, abstract, or MeSH terms/keywords. Linking the Go3R search engine to chemical substances databases would allow expanding search queries to retrieve documents in which synonyms, superordinate, or subordinate terms for the chemical substance under consideration might be used or which indicate specific nomenclature (e.g. CAS registry numbers, EC numbers). However, if the document titles and abstracts only use imprecise, collective formulations to refer to the test substances (e.g. 10 compounds, 25 chemicals), full-text text mining continues to be indispensible to obtain information on the specific test substances.

As a result, the restrictions of the vocabulary used in the documents further limit the precision of the IUCLID SQ-based sorting. Upon automatic sorting into the Go3R IUCLID toxicological endpoint categories, some form of manual processing of the literature will remain necessary. In this regard, it is promising that a mere 1.6% of the gold standard references were not recognized as relevant for regulatory purposes.

Manual processing of automatically sorted literature continues to be inevitable to evaluate the validity of the data. Thereby, the key studies for a given endpoint are determined or a weight-of-evidence regarding a given effect is established. In the context of REACH, this will form the basis of the overall toxicological evaluation of the respective substance including a proposal for its classification and labeling. Manual processing further continues to be required for comprehensive cross-referencing of documents between different endpoints. Such cross-referencing indicates which toxicological endpoint is the main endpoint addressed in the respective study and for which additional endpoints the study results provide supportive information. Further manual processing of the data obtained for different test substances is also necessary for the 'grouping of substances'. During grouping, a set of chemical substances is assigned to a specific group based upon common properties. During read-across the overall toxicological information available for the substances of the given group is used to predict unknown effects of further substances within the group (Anon, 2006; ECHA, 2008, 2012c).

All comparisons of automatic data sorting to manual sorting, however, have to take into account that also manual sorting can differ considerably between different information retrieval experts: The inter-annotator agreement is usually as low as 80–90% (Morgan et al., 2008; Verberne et al., 2010). Against this background, the search tools of the Go3R search engine that not only collect and highlight information on alternative methods to animal experiments, but also allow for automatic retrieval and sorting of relevant toxicological information obtained *in vivo* can make an important contribution to preventing animal testing. Both of these aspects directly serve the 3Rs principle.

Information box III Additional tools and applications for Go3R providing added value to the IUCLID-SQ-based annotation tool.

The **data pool that Go3R searches** in is not fixed or unchangeable. Expanding the initial data pool (only PubMed) to the TOXNET database did not serve to significantly extend the proportion of obtainable literature relevant for regulatory purposes (only 8 of the 310 references were only available in TOXNET, but not in PubMed). Nevertheless, this expansion underlines that the Go3R search paradigm can be adapted to search in different (and differently coded and structured) databases. Most likely, vast amounts of toxicological information of relevance for regulatory purposes are not available in the open literature at all, but stored in in-house databases. The example of expansion of the Go3R data pool to TOXNET shows that Go3R can provide an integrated, unified view on documents and their content throughout different databases, including in-house databases.

But also when searching in PubMed and TOXNET, Go3R provides a number of unique features, both in regard to general literature searching and, specifically, in searching for 3Rs relevant information: Generally, literature searches with Go3R provide **automatic indexing of literature** ahead of PubMed and TOXNET and, additionally, with terms that are not existent in MeSH. The Go3R search interface allows users to see the full index at hand and provides a means of sorting literature rather than searching for it, which is especially important when scanning large quantities of literature, such as when evaluating existing toxicological literature. Whenever queries lead to large result sets, one can browse the dynamic table of contents, constituted by the taxonomic backbone of the ontology, to refine the search by relevant categories. (Not addressed in the present article, Go3R also allows sorting search retrievals by the names of authors, dates of publication, specific journals, and affiliations, including countries and cities. Additionally, an automatic statistical evaluation of the search retrieval presents e.g. the numbers of publications over time, highlights the most frequently used journals, or helps discern major research institutions for a given topic.) During filtering, the user can immediately see how the number of documents assigned to the respective categories of the table of contents changes, a valuable feedback for query formulation. Searching for alternative methods benefits from this methodology: During each Go3R search, the user will receive direct feedback, whether or not there are any alternative methods explicitly mentioned in the documents retrieved. Already before filtering for a given 3Rs relevant branch of the table of contents, the respective number in square brackets provided next to it indicates the number of assigned documents.

The system is designed to **update information fully automated**. Documents that have been newly taken up into the underlying PubMed database are added to Go3R within 1 h; TOXNET updates as soon as they are published on the servers of the US National Library of Medicine. Individual user curations will take effect instantly with the next search for the documents displayed for this user and will be propagated to the search index within minutes after confirmation. Changes in the ontology will trigger automatic re-indexing of affected documents for the whole or parts of the annotation pipeline introduced in the following sections.

A further feature provided in Go3R is the assignment of the **3Rs signet** displayed next to documents that are likely to be 3Rs relevant. This signet, however, can only provide a first overview on 3Rs relevant documents available in a given search result: Automatic assignment of the 3Rs signet relies on learning from examples, and in the selected examples the topic toxicology was overrepresented. But also within the domain of toxicology, the most frequently mentioned endpoints are irritation, corrosion, or sensitisation, embryotoxicity, and carcinogenicity. This can lead to a bias in the underlying automatic classification scheme. In terms of the type of alternative method, replacement is clearly overrepresented compared to refinement or reduction methods. Whereas the purely technical outcome of the 3Rs relevance signet assignment is very good (86% recall, while maintaining 95% precision; see Information Box II), this is not necessarily evidence that documents are valuable and informative with respect to alternative methods for a given scientific topic.

Distinctions between *in vivo and in vitro* data can further be made manually by **selecting or excluding ontology branches** for '3Rs methods in the life sciences' versus 'animal experiments' and by selecting or excluding the different subordinate terms from 'method specification' (i.e. *'in vitro'*, *'in vivo'*, *'ex vivo'*, *'in silico'*). To obtain an overview on the spectrum of 3Rs test methods available for a given endpoint, users can search for that endpoint making use of the IUCLID user interface and further sort the documents retrieved making use of the '3Rs methods in the life sciences' ontology branch. This allows them to filter the search result by different animal testing alternatives.

To select the test method that is best suited in meeting the 3Rs principle (to replace tests on live animals altogether or, if that is not yet possible, to reduce the number of animals used and the suffering of the animals), **information on the status of the test method** is required:

Such information distinguishing whether a given test method is still in the stage of development, whether it is currently under (pre-)validation, or has already reached the status of regulatory acceptance, however, is not (or not regularly) provided in peer review publications indexed in PubMed or in toxicological reports covered by TOXNET. To allow users to retrieve such additional information, a further tool has been introduced into the Go3R search engine, enabling Google searches in the entire web (not shown). This **'Go3R web' tool**, however, does not make use of the Go3R semantic search paradigm or ontology-based annotation tool, but merely allows for a higher ranking of 3Rs relevant websites, such as the ones from the OECD test guidelines programme (http://www.oecd.org/env/ehs/testing/), the European Centre for the Validation of Alternative Methods (ECVAM; http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam), or the US Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM; http://iccvam.niehs.nih.gov/), where information, e.g. on the status of test methods or standard operating protocols, is made publicly available.

Similarly, linking documents sorted into the IUCLID user interface with the **Go3R 'nano'-ontology** allows restricting search retrievals to documents addressing nanotoxicological issues. In this context, further work is required to expand the IUCLID SQs to cover early cellular effects that might be caused by nanomaterials, such as oxidative stress or pro-inflammatory reactions (Nel et al., 2013).

5. Conclusion

The search engine Go3R is accessible free of charge at www.Go3R.org. In addition to allowing retrieving information on alternatives to animal testing, especially in the area of regulatory toxicology, for the first time it provides an endpoint-centered literature search, based upon the IUCLID5 'toxicological information' categorization scheme. Search results are automatically linked to a dynamic 'table of contents' which enables the user to sort the literature listed in the PubMed and TOXNET databases. Linking of the Go3R IUCLID user interface with 3Rs-relevant branches of the Go3R knowledge base further allows recognizing available *in vitro* assays for the respective endpoints. Thereby, Go3R can contribute to preventing animal testing both by allowing comprehensive retrieval of all existing information and also by identifying available 3Rs methods.

Whereas the work on the IUCLID SQ-based annotation tool cannot be considered accomplished at this point in time (since it is based upon the documents listed in a mere 3 registration dossiers), the current Go3R IUCLID sorting can already be used as a starting point for subsequent manual processing of the data retrieved. Comprehensiveness and correctness of the automatic data retrieval and data sorting will improve over time, the more the search engine is being used: Users can curate individual search results, thereby reporting missing (or incorrectly assigned) concepts, terms, and synonyms. The more such knowledge from individual users will be taken up into the search engine, the more precision and recall of the IUCLID-based annotation work will improve. Finally, Go3R can be adapted to search in other databases, and also in-house databases, just as the annotation tools can be extended to cover other topics apart from regulatory toxicity testing.

Conflict of interest

The authors declare no conflict of interests.

Acknowledgements

The Go3R research project was funded by the *German Federal Institute for Risk Assessment* (BfR; Bundesinstitut für Risikobewertung; Grant No. FK 3-1328-201) and by the *German Federal Ministry for Education and Research* (BMBF; Bundesministerium für Bildung und Forschung; Grant No. 614 40003 0315489A) taking into account prior work by the project partners *Technical University Dresden* (TUD) and *Transinsight GmbH*, *Dresden*. The development of the 'nano'-ontology was funded by the BfR. Throughout, the TUD, Transinsight GmbH, and *BASF SE*, *Ludwigshafen*, *Germany*, further provided technical and/or financial support to ensure the continued Internet presence of the Go3R search engine.

Appendix A. Supplementary material

Transparency documents associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.tiv.2013.12.011.

References

Anon, 2006. Regulation (EC) No. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), Establishing a European Chemicals Agency, Amending Directive 1999/45/EC and Repealing Council Regulation (EEC) No. 793/93 and Commission Regulation (EC) No. 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. O.J. L 396/1, 30 December 2006, pp. 1–849.

Anon, 2010. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the Protection of Animals Used for Scientific Purposes. O.J. L 276/33, 20 October 2010.

Anon, 2012. Regulation (EU) No. 528/2012 of the European Parliament and of the Council of 22 May 2012 Concerning the Making Available on the Market and Use of Biocidal Products. OJ L 167, 27 June 2012, pp. 1–123.

Berger, A.L., Della Pietra, S.D., Della Pietra, V.J.D., 1996. A maximum entropy approach to natural language processing. Comput. Linguist. 22 (1), 39–71.

Boyce, S., Pahl, C., 2007. Developing domain ontologies for course content. Educ. Technol. Soc. 10 (3), 275–288.

Doms, A., 2009. GoPubMed: Ontology-based Literature Search for the Life Science. PhD Thesis, Technical University Dresden.

Doms, A., Schroeder, M., 2005. GoPubMed: exploring PubMed with the gene ontology. Nucleic Acids Res. 33 (Web Server Issue), W783–W786.

ECHA, 2008. Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.6. QSARs and grouping of chemicals. European Chemicals Agency, 134pp. <http://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf>.

ECHA, 2012a. European Chemicals Agency. Guidance for the Implementation of REACH. Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.7a. Endpoint Specific Guidance. Version 2.0. ECHA-12-G-11-EN. November 2012, 381pages. <http://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf>.

ECHA, 2012b. European Chemicals Agency. Guidance for the Implementation of REACH. Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.7c. Endpoint Specific Guidance. Version 1.1. ECHA-12-G-23-EN. November 2012, 239pages. <http://echa.europa.eu/documents/10162/13632/information_requirements_r7c_en.pdf>.

ECHA, 2012c. Background Paper. An Introduction to the Assessment of Read Across. European Chemicals Agency, 8pp. <http://echa.europa.eu/documents/10162/5649897/ws_raa_20121003_background_paper_an_introduction_to_the_assessment_of_read-across_in_echa_en.pdf>.

ECHA and OECD, 2007. European Chemicals Agency on Association with the Organisation for Economic Cooperation and Development. IUCLID 5 Guidance and Support – End User Manual. Version 1. July 2007, 2044pp. <http://echa.europa.eu/documents/10162/13632/iuclid_en.pdf>.

Hripcsak, G., Rothschild, A.S., 2005. Agreement, the *f*-measure, and reliability in information retrieval. J. Am. Med. Inform. Assoc. 12 (3), 296–298.

Judson, R., Richard, A., Dix, D., Houck, K., Elloumi, F., Martin, M., Cathey, T., Transue, T.R., Spencer, R., Wolf, M., 2008. Actor – aggregated computational toxicology resource. Toxicol. Appl. Pharmacol. 233 (1), 7–13.

Martin, M.T., Judson, R.S., Reif, D.M., Kavlock, R.J., Dix, D.J., 2009. Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef database. Environ. Health Perspect. 117 (3), 392–399.

Mehling, A., Eriksson, T., Eltze, T., Kolle, S.N., Ramirez, T., Teubner, W., van Ravenzwaay, B., Landsiedel, R., 2012. Non-animal test methods for predicting skin sensitization potentials. Arch. Toxicol. 86 (8), 1273–1295.

Morgan, A.A., Lu, Z., Wang, X., Cohen, A.M., Fluck, J., Ruch, P., Divoli, A., Fundel, K., Leaman, R., Hakenberg, J., Sun, C., Liu, H.-H., Torres, R., Krauthammer, M., Lau,

W.W., Liu, H., Hsu, C.-N., Schuemie, M., Cohen, K.B., Hirschman, L., . Overview of BioCreative II gene normalization. Genome Biol. 9 (Suppl. 2), S3.
Nel, A.E., Xia, T., Meng, H., Wang, X., Lin, S., Ji, Z., Zhang, H., 2013. Nanomaterial toxicity testing in the 21st century: use of a predictive toxicological approach and high-throughput screening. Acc. Chem. Res. 46 (3), 607–621.
Oomen, A.G., Bos, P.M.J., Fernandes, T.F., Hund-Rinke, K., Boraschi, D., Byrne, H.J., Aschberger, K., Gottardo, S., van der Kammer, F., Kühnel, D., Hristozov, D., Marcomini, A., Migliore, L., Scott-Fordsmand, J., Wick, P., Landsiedel, R., 2013. Concern-driven integrated approaches to nanomaterial testing and assessment – report of the NanoSafety Cluster Working Group 10. Nanotoxicology, 28 May 2013, epub ahead of print.
Russell, W., Burch, R., 1959. The Principles of Humane Experimental Technique. Methuen, London.
Sauer, U.G., Wächter, T., Grune, B., Doms, A., Alvers, M.R., Spielmann, H., Schroeder, M., 2009. Go3R – semantic Internet search engine for alternative methods to animal testing. ALTEX 26 (1), 17–31.
Sauer, U.G., Kneuer, C., Tentschert, J., Wächter, T., Schroeder, M., Butzke, D., Luch, A., Liebsch, M., Grune, B., Götz, M.E., 2011. A knowledge-based search engine to navigate the information thicket of nanotoxicology. Regul. Toxicol. Pharmacol. 59 (1), 47–52.
Schroeder, M., 2003. Intelligent information integration: from information integration infrastructure through consistency management to information visualisation. In: Dykes, J., Kraak, M.-J., McEachren, A. (Eds.), Exploring geoVisualization. International Cartographic Association, <http://www.biotec.tu-dresden.de/research/schroeder/publications.html>.
Schuemie, M.J., Kors, J.A., Mons, B., 2005. Word sense disambiguation in the biomedical domain: an overview. J. Comput. Biol. 12 (5), 554–565.
Spielmann, H., Sauer, U.G., Mekenyan, O., 2011. A critical evaluation of the 2011 ECHA reports on compliance with the REACH and CLP regulations and on the use of alternatives to testing on animals for compliance with the REACH regulation. ATLA 39, 481–493.
Uschold, M., Gruninger, M., 1996. Ontologies: principles, methods, and applications. Knowl. Eng. Rev. 11 (2), 1–63.
Verberne, S., D'Hondt, E.K.L., Oostdijk, N.H.J., Koster, C., 2010. Quantifying the challenges in parsing patent claims. In: Proceedings of the 1st International Workshop on Advances in Patent Information Retrieval (AsPIRe 2010), pp. 14–21.