Regulatory Toxicology and Pharmacology 59 (2011) 47–52

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

A knowledge-based search engine to navigate the information thicket of nanotoxicology

Ursula G. Sauer [a], Carsten Kneuer [b], Jutta Tentschert [b], Thomas Wächter [c,d], Michael Schroeder [c], Daniel Butzke [b], Andreas Luch [b], Manfred Liebsch [b], Barbara Grune [b], Mario E. Götz [b,*]

[a] Scientific Consultancy – Animal Welfare, Hallstattfeld 16, 85579 Neubiberg, Germany
[b] Federal Institute for Risk Assessment (BfR), Thielallee 88-92, 14195 Berlin, Germany
[c] Bioinformatics Group, Biotechnology Center, Technische Universität, Tatzberg 47/49, 01307 Dresden, Germany
[d] Transinsight GmbH, Tatzberg 47-51, 01307 Dresden, Germany

ARTICLE INFO

Article history:
Received 16 June 2010
Available online 17 September 2010

Keywords:
Nanotechnology
Nanotoxicology
3Rs principle
Animal testing alternatives
Information retrieval
Literature search
Knowledge-based search
Ontology
Semantic technology
Risk assessment

ABSTRACT

The risk assessment of nano-sized materials (NM) currently suffers from great uncertainties regarding their putative toxicity for humans and the environment. An extensive amount of the respective original research literature has to be evaluated before a targeted and hypothesis-driven Environmental and Health Safety research can be stipulated. Furthermore, to comply with the European animal protection legislation *in vitro* testing has to be preferred whenever possible. Against this background, there is the need for tools that enable producers of NM and risk assessors for a fast and comprehensive data retrieval, thereby linking the *3Rs* principle to the hazard identification of NM.

Here we report on the development of a knowledge-based search engine that is tailored to the particular needs of risk assessors in the area of NM. Comprehensive retrieval of data from studies utilising *in vitro* as well as *in vivo* methods relying on the *PubMed* database is presented exemplarily with a titanium dioxide case study. A fast, relevant and reliable information retrieval is of paramount importance for the scientific community dedicated to develop safe NM in various product areas, and for risk assessors obliged to identify data gaps, to define additional data requirements for approval of NM and to create strategies for integrated testing using alternative methods.

1. Introduction

1.1. A rise with risks

Nanotechnology is a rapidly growing technology, even considered by many to represent the key technology of the 21st century. Nano-sized materials (NM) have already made their way into consumer products, cosmetics, food, biomedical applications, etc. (Dekkers et al., 2008; Greßler et al., 2009; Hoffbauer, 2008; The Project on Emerging Nanotechnologies, 2010). Due to their unrivalled nature, however, current hazard and risk assessment procedures for NM safety testing may not be appropriate and thus are challenged (SCENIHR, 2006, 2007, 2009).

In response to uncertainties associated with the risk assessment of NM, the *OECD* Council established a *Working Party on Manufactured Nanomaterials* (*OECD WPMN*) with the purpose "to help member countries efficiently and effectively address the safety challenges of NM" (OECD, 2008). The working party is performing its work through steering groups, addressing amongst others the issue of safety testing and the role of alternatives to the use of animals in Nanotoxicology. This is in accordance with the *EU Directive on the Protection of Laboratory Animals* (Council of the European Communities, 1986), which requests the avoidance or limitation of animal testing whenever possible. The taken measures should follow the *3Rs principle – replace, reduce, refine animal experiments* – published by Russel and Burch (1959).

In Germany, the Higher Federal Authorities – i.e. the *Federal Institute for Risk Assessment* (BfR)[1], the *Federal Environment Agency* (UBA)[2] and the *Federal Institute for Occupational Safety and Health* (BAuA)[3] – are involved in assessing the possible risks of manufactured NM (i.e. nanoparticulate TiO_2, silver nanoparticles) for human health and the environment. These authorities concertedly pointed to data gaps that hamper a systematic risk assessment and thus have

* Corresponding author. Fax: +49 30 18412 4928.
E-mail address: Mario.Goetz@bfr.bund.de (M.E. Götz).

[1] BfR – Bundesinstitut für Risikobewertung
[2] UBA – Umweltbundesamt
[3] BAuA – Bundesanstalt für Arbeitsschutz und Arbeitsmedizin

0273-2300/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.yrtph.2010.09.005

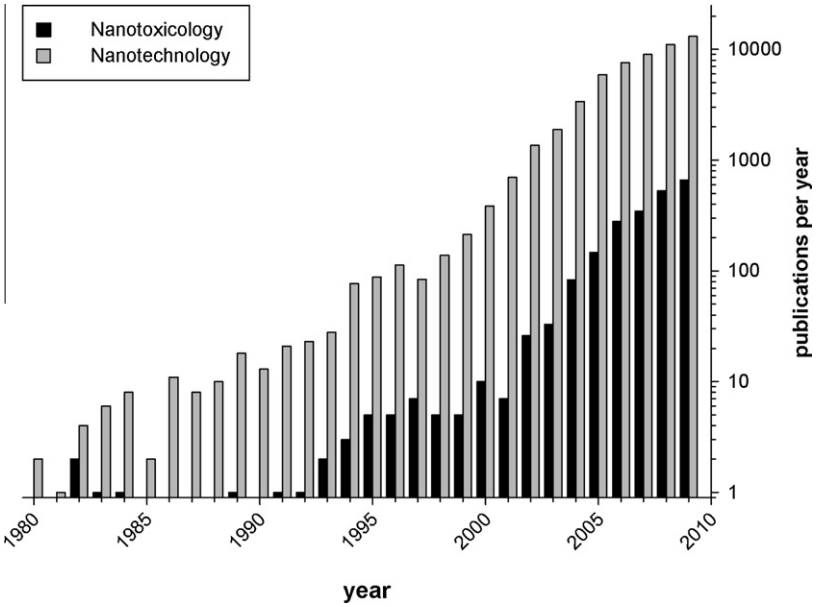

Fig. 1. The accumulation of publications in nanotechnology [retrieved from PubMed (using the query: "Nanotechnology" [Mesh] OR "Nanostructures"[Mesh] OR nanotub*OR nanopartic*OR nanocomposit*)] and nanotoxicology [retrieved from PubMed (using the query: "Nanotechnology" [Mesh] OR "Nanostructures"[Mesh] OR nanotub*OR nanopartic*OR nanocomposit*) AND ("Toxicity Tests" [Mesh] OR "adverse effects" [Sub heading])].

to be closed within the near future (BAuA, BfR, UBA, 2007). Other transgovernmental organizations, institutions and committees in charge of dealing with or commenting on the risks of the new technology, i.e. the *European Food Safety Authority* (EFSA, Europe), the *Scientific Committee on Emerging and Newly Identified Health Risks* (SCENIHR, Europe), the *National Academy of Sciences* (NAS, USA), the *New Energy and Industrial Technology Development Organization* (NEDO, Japan) and the *Food and Agriculture Organization* (FAO)/*World Health Organization* (WHO) confirmed this urgent necessity (SCENIHR, 2007; National Academy, 2008, EFSA, 2009; FAO/WHO, 2009; NEDO, 2009). To identify suitable methods and procedures to fill the data gaps, members of the mentioned authorities are participating in the *OECD WPMN*.

To achieve their aims, the Federal Authorities and the members of the OECD steering groups rely on literature searches to identify and compile a reliable body of scientific literature, which constitutes the knowledge base that subsequently will be considered.

1.2. The "thicket" of information

Since the importance of the Internet as a means of information dissemination is increasing steadily, the available data on any given topic is accumulating exponentially. The possibly relevant literature on Nanotechnology also follows this trend. Starting with the 1980s, an accelerating accumulation of patents and scientific papers reporting on Nanotechnology can be discerned with a steep increase since 2000 (Fig. 1). There is also a continuous increase in the number of publications addressing toxicological issues of NM, although staying behind the number of toxicologically unconcerned papers (Fig. 1).

At the same time, or even due to the tremendous increase in scientific literature, it is becoming increasingly difficult to extract relevant and reliable information. Especially to achieve a certain level of knowledge in a timely manner and to attain a high level of recall – i.e. to retrieve an exhaustive set of information – requires searcher skills that most people, including scientists, do not possess. As stated above, however, scientists need to collect reliable and complete information to evaluate research projects, scientific studies and both *in vivo* and *in vitro* data against the background of the current scientific state-of-the-art.

1.3. Semantic and knowledge-based technologies facilitate searching the Internet – Go3R, a semantic search engine for alternative methods

Search engines that integrate human expert knowledge are tools that can assist scientists in retrieving, sorting and evaluating extensive amounts of literature from the Internet. They are a subgroup of "semantic[4] search engines" that aim to gather the meaning of natural language documents from the occurrence (and co-occurrence) of certain terms (and their synonyms) within the text of the document. One example of such an engine is the knowledge-based *Go3R* tool that aids in retrieving *3Rs*-relevant literature from *PubMed* (Sauer et al., 2009). It is the worldwide first tool of its kind specially equipped with expert knowledge from the area of the *3Rs*. This unique knowledge is captured within a so-called "ontology", i.e. an extensive and detailed network of "concepts", terms that are unambiguous identifiers of the respective scientific area, such as *dendrimers* or *nanoclay* in the field of Nanotechnology, or *humane endpoints* in the field of the *3Rs*.

When a user performs a search query with *Go3R*, the search engine compares the terms and concepts of the ontology with the vocabulary used in the retrieved documents (Sauer et al., 2009). The technical process is called concept recognition. *Go3R* is based on *GoPubMed* (Doms and Schroeder, 2005; Dietze et al., 2009), which can find concepts in text as it is done in other commonly used tools like *ProMiner* (Hanisch et al., 2005) or *Textpresso* (Müller et al., 2004). *GoPubMed*'s word sense disambiguation, which is required to unambiguously identify ontology terms in text, performs well with an average *f*-measure of above 90% (Alexopoulou et al., 2009). A particularly difficult problem is the recognition of the mentioning of genes in text. Like *ProMiner*, *GoPubMed* participated in the gene normalisation task of the second *BioCreative challenge* (Morgan et al., 2008; Hakenberg et al., 2007) and achieved the best result.

[4] Semantics: the study of meaning.