European digital content sharing services for health protection and occupational risk prevention

F. Rizzo[1], D. Ugolini[1,2], L. Maiorana[1], M. Gonzalez Rodriguez[3],
I. Laamanen[4], P. Boffetta[5], E. Mirkova[6], A. C. Wasilewska[7],
G. Viano[8] & D. Vecchio[1]
[1]*Department of Epidemiology and Prevention,
National Cancer Research Institute, Italy*
[2]*Department of Oncology, Biology and Genetics, Genoa University, Italy*
[3]*Fundacion Robotiker, Spain*
[4]*Finnish Institute for Occupational Health, Finland*
[5]*International Agency for Research on Cancer, France*
[6]*National Center of Public Health Protection, Bulgaria*
[7]*Jagellonian University Medical College, Polish*
[8]*Softeco Sismat S.p.A., Italy*

Abstract

The EuroWorsafe web aims at improving access to and use of large amounts of data related to the sector of health protection of workers and workplace safety. These data are presently available in many European public sector organisations, including research institutes, public health institutions and hospitals. The wide diffusion of this information would be of crucial interest for many organisations, citizens, workers, and companies to reduce risks from health hazards and to improve prevention and safety measures to contribute to the reduction of occupational cancer incidence. The access to and use of such data are often limited to scientific bodies and hampered by several barriers including language, data format, location at different organisations and non-accessibility on the internet. EuroWorksafe intends to overcome such problems by the establishment of a pan-European distributed database and the design and development of set based access services to allow the exploitation of available contents in the addressed domain and the evolving semantic web architecture based on meta-data modelling techniques, knowledge extraction methods and tools, web-based data access and semantic processing.
Keywords: occupational diseases, occupational exposure, risk factors, cancer.

1   Introduction

The most accepted estimates of cancers in the general population attributable to occupational exposures in developed countries are in the range of 4–5%. The proportion of cancer attributable to occupational exposure among those actually exposed to occupational carcinogens is estimated to be about 20% [1].

The high risk of activities dealing with the use of toxic and noxious substances has been studied for a long time and the EU and National Public Institutions have collected much information throughout Europe on this subject [2]. Currently, there is a large body of scientific, medical and legislative information available and scattered throughout the European Organisations. The most important problems in different workplaces (also in research Institutes) are arising from the number of agents used and their synergies [3], from lack of awareness of the chemical and toxicological characteristics of the substances, and continuous technological evolution [4]. Also in this domain the reproductive adverse effects are highlighted [5–7]. A large diffusion of this information can significantly improve cancer and chronic degenerative diseases risk awareness thus contributing to primary prevention and today's workplace safety and protection.

Web based information services can play a relevant role in making this information accessible and exploitable by the scientific community as well by the general public. Unfortunately, due to the structure and nature of the Internet itself, it is difficult to find and to access relevant information. Contents are published by several providers, uncoordinated and disorganized. No global indexing and classification services are available and adopted. Data are most often split up among several sources (agencies of governments, provinces, states, municipalities, entrepreneurs, associations, societies, interest groups) and it needs to navigate from databases to electronic libraries, technical reports, handbooks, guidelines, directories, bulletins, dictionaries, newspapers, toxic release inventories, and more.

In general, three approaches can be adopted to find relevant information on the web: search engines, free texts and subject directories. Current web search engines (like Google, AltaVista, Lycos, Excite, and so on) have a poor understanding of content semantics and typically return innumerable completely irrelevant "hits", requiring much manual weeding by the user and possibly missing important resources. They do not really distinguish officially approved guidelines from experts' opinions on the same topics, and cannot easily distinguish a personal home page from a major corporate web site [8,9]. A free text search is not always efficient and even when sophisticated statistical techniques are used for information retrieval, (e.g. PubMed's Related Articles function) the results are not relevant. Subject directories like "Yahoo!" or the Open Directory Project (http://dmoz.org), where resources are classified under different categories, are not much better. The web is very rapidly growing and changing, with so much information that people at these directory services cannot possibly keep up and there is frequently an inevitable overlap between the

categories in a web catalogue and a vagueness regarding the definition of their objectives, leading to confusion as to what to expect under a given category [10].

The EuroWorksafe project has instead developed a general, harmonised approach to content classification and provision in the occupational health area, based on a Europe-wide agreed meta-data description. The background of the project has taken into account the following elements: the absence of a comprehensive data collection in the prevention area, the use of different classification schemes, formats and languages in the main European research institutes and the lack of visibility and accessibility to the existing data.

2   Materials and methods

The project has mapped Internet occupational health information resources in novel ways for enhanced retrieval and navigation. This vision fits into the evolving semantic web architecture, a web that provides intelligent access to heterogeneous and distributed information, enabling software products to mediate between user needs and available information sources where information semantic (or meaning) are represented in a form that can be understood by machines as well as by people. Metadata, ontologies and ontology representation languages are pivotal ingredients of this new intelligent knowledge-based approach [11].

Particularly the infrastructure of the EuroWorksafe project is based on fig. 1:

I) Web based distributed data access: Web technology is at the basis of the project, whose final objective is the development of a very large and complex web-based data access and distribution service. Data are distributed on the network, and the EuroWorksafe system manages a centralised index and classification service. Specifically, the project focuses on cancer and chronic degenerative disease risk evaluation, education and prevention. The addressed contents are multi-sectorial, combining scientific and medical contents, technical information, regulatory and legal, enriched by a glossary and a training area.

II) Semantic content modelling: constitute the core technology to allow a uniform and user-oriented access to the heterogeneous data collection. To this end a meta-data layer which will encapsulate the underlying data collections is implemented and the integration of the available up-to-date specific technologies is in progress, including: ontology analysis and management techniques/tools, for the creation of a structured vocabulary (i.e. entity and relationship name and relative description) for representing and communicating knowledge about the domain, and semantic annotation standards to model different types of contents, ranging from book-oriented document types to national law collocations. A specific metadata model has been designed following the particular requirements of the addressed domain. This model extends the well-known Dublin Core model [12] to include several domain specific elements. Contents are classified and described by exploiting a specialised ontology developed within the EuroWorksafe project. This ontology includes main concepts relevant in the occupational health field with special reference to cancer risks.
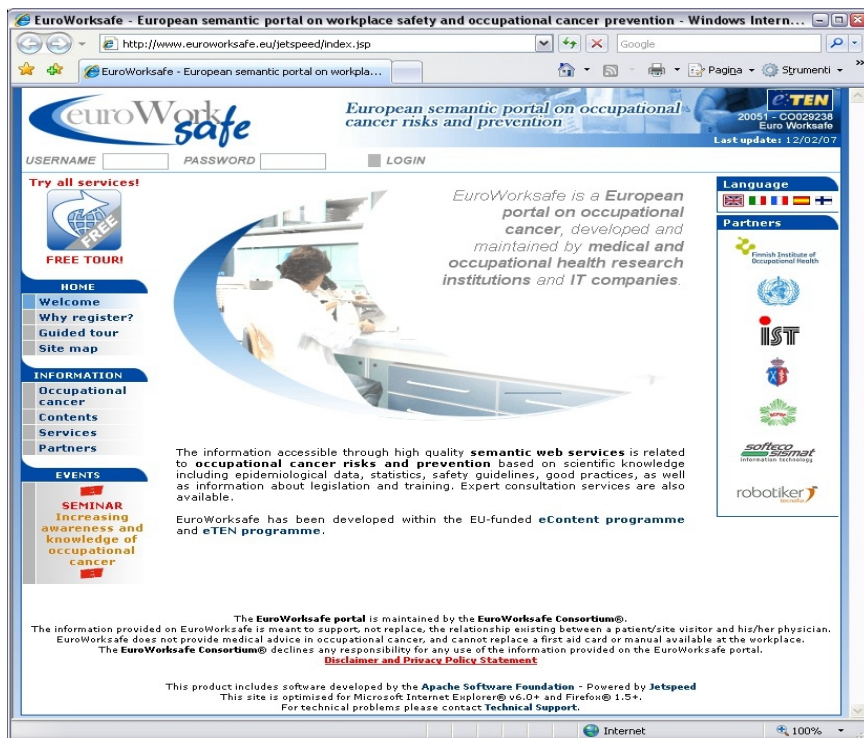
Figure 1.

III) Knowledge extraction: this technology supports the provision of rich information content through the larger possible base of potential users. The navigation through the complexity and the heterogeneity of the information available may be very difficult and the system needs to provide tools to support the retrieval of the required information, and the classification of all the available data collections to support data browsing. A specialised knowledge extraction layer is included in the system to provide specific search and filtering services to the user. These services exploit the content description and classification scheme based on the ontology and metadata in order to help users to find the information they are interested in through a simple and "intelligent" interface.

The tools we have structured are based on an architecture that helps to bridge the large gap between information needs of users on the one hand and available information sources on the other hand. It aggregates and processes statements (descriptions, annotations and ratings) made by a variety of participants, and integrates them with the individual preferences of the user, thereby harnessing the power of the Web as a decentralized medium.

Most Web content is posted without any form of review for accuracy or reliability and for general user it is very difficult to make sure that the online information find is credible.

The quality of this information is critically important as it could potentially affect health outcomes for millions of people. Yet the quality of health information on the Internet is extremely variable and difficult to assess. To resolve these problems the EuroWorksafe web site has been designed with specific criteria for Evaluating Internet Health Information. These criteria are:

- Credibility: includes the source, currency, relevance/utility, and a review process for the information.
- Content: the source must be accurate and complete, and an appropriate disclaimer provided.
- Disclosure: includes informing the user of the purpose of the site, as well as any profiling or collection of information associated with using the site.
- Links: evaluated according to selection, architecture, content, and back linkages.
- Design: encompasses accessibility, logical organization (navigability), and internal search capability.

The EuroWorksafe project addresses the needs of a large number of users including: private individuals (employees) and professionals working at public or private institutions (a workers safety manager). Both can be interested in risks related to a specific working environment (for instance a laboratory for drug preparation) or occupation (for instance a hairdresser) and can ask to focus on a specific substance or group (for instance a petrol station attendant dealing with gasoline).

3   Conclusion

EuroWorksafe benefits to users by providing comprehensive and multidisciplinary information on the many aspects of occupational cancer risk and prevention, and by guiding users either experienced or not to analyse and discover the relations between such aspects. For example, people interested to know about a substance are guided to move from risk assessment information, to the legislation and provisions that apply to the substance, to the protective measure required by legislation and advisable from a technical point of view, and can access recent data on exposures at working places. All that without any specific background knowledge requirement with the aim of contributing to digital divide mitigation. The large and multidisciplinary content base accessible through the portal is provided by the project partner or is available on the Internet. The main added value of the EuroWorksafe portal is to provide a unified access point with a homogeneous classification supporting the correlation of contents from different sources and from different disciplines. This is supported by semantic web technology exploitation (fig. 2), based on the classification of information provided by content providers. In addiction while other portals focus on occupational health and safety in general, EuroWorksafe deals specifically with occupational cancer risks and prevention in Europe.

The possibility of accessing these data online from all European countries and the consideration for the multi language issue contributes to reduce disparities in occupational cancer prevention opportunities among EU citizens.

Figure 2.

Occupational diseases and accidents cost more than prevention: prevention is not only a tool to fight accidents, but it is also an economic investment for the whole community. It is necessary to create a new culture of safety and of prevention in the work environments, to stimulate enterprises to improve structures and establishments to invest in safety. It is important that this culture belongs to workers even before entering the work world, therefore raising the awareness of work risks. EuroWorksafe can contribute considerably to this objective. The EuroWorksafe web address is http://www.euroworksafe.org.

References

[1]   ILO Encyclopaedia of Occupational Health and Safety, 4th Ed., Stellman, J. M., Editor, ILO, Geneva, 1998.

[2]   Vecchio, D., Sasco, A.J., Cann, C.I. Occupational risk in health care and research. *Am J Ind Med,* **43(4)**, pp. 369-397, 2003.

[3]   Rachet, B., Partanen, T., Kauppinen, T., Sasco, A.J. Cancer risk in laboratory workers: an emphasis on biological research. *Am J Ind Med* **38(6)**, pp. 651–665, 2000.

[4] Andrup, L., Nielsen, B.H., Kolvraa, S. Biosafety considerations in industries with production methods based on the use of recombinant deoxyribonucleic acid. *Scand J Work Environ Health* **16(2)**, pp. 85–95, 1990.

[5] Ahlborg, G., Hemminki, K. Reproductive effects of chemical exposures in health professions. *J Occup Environ Med* **37(8)**, pp. 957–961, 1995.

[6] Boivin, J.F. Risk of spontaneous abortion in women occupationally exposed to anaesthetic gases: a meta-analysis. *Occup Environ Med* **54(8)**, pp. 541–548, 1997

[7] Little, J. Epidemiology of childhood cancer. IARC Scientific Publ No. 149. Lyon: International Agency for Research on Cancer 1999.

[8] Fensel, D. The semantic web and its languages. IEEE Intelligent Systems, **15**, pp. 67–73, 2000.

[9] Berners-Lee, T., Hendler, J. Publishing on the semantic web. *Nature* **410(6832)**, pp. 1023–1024, 2001.

[10] Risden, K. Toward usable browse hierarchies for the web. In H. Bullinger and J Ziegler (eds). Human Computer Interaction: Ergonomics and User Interfaces **1**, pp. 1098–1102, 1999.

[11] Berners-Lee, T., Hendler, J., Lassila, O. A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Sci Am* 279, 2001.

[12] Weibel, S. The state of the Dublin Core Metadata initiative. D-Lib Magazine, http://www.dlib.org/dlib/april99/04weibel.html 1999.