



ELSEVIER

Journal of Hazardous Materials 61 (1998) 291–297

**JOURNAL OF  
HAZARDOUS  
MATERIALS**

# An outline of the approach being used in developing the OECD Dictionary/Thesaurus of ‘risk assessment’ terminology

Isadore Rosenthal<sup>\*</sup>, Ben Ale, Lyse D. Helsing

*The Risk Management and Decision Processes Center, The Wharton School of the University of Pennsylvania, 1325 Steinberg Hall-Dietrich Hall, Philadelphia, PA 19104-6366, USA*

---

## Abstract

A system was developed to capture and make available the definitions of terms and the approaches and methodologies in the risk assessment process of the member-countries of the OECD. The system uses the principles of artificial intelligence. The system is running on the Internet. © 1998 Published by Elsevier Science B.V. All rights reserved.

*Keywords:* OECD; Communication; Internet; System

---

## 1. Introduction

The OECD Expert Group on Chemical Accidents set out the Project’s goals rather clearly:

“In summary, the workshop participants concluded that standardization of the risk assessment process, and approaches/methodologies used in each step of the process, is neither desirable nor feasible. Nonetheless, enhancing the mutual understanding of risk assessment in the context of chemical accidents, can be furthered by, e.g., efforts to map out the steps in the risk assessment process and the approaches/methodologies used therein, and an elaboration of the indicators influencing choices of particular approaches/methodologies. It must be emphasized that this is not intended to direct, still less to prescribe, a particular approach.

---

<sup>\*</sup> Corresponding author.

The objective of these efforts should be to help stakeholders see more clearly the range of possibilities, and to assist them in decisions which only they can make.’’ [1].

The project goals set out in this statement can be framed as follows:

- Illuminate the sense of the various approaches to risk assessment described in the literature.
- Promote understanding of the commonalities (and differences) among the various approaches to risk assessment.
- Facilitate communication of such commonalities (and differences) between different risk cultures and languages.

There is a need to constantly remind oneself that the project’s goal is to provide ‘products’ that promote understanding and communication among the project’s stakeholders, and not to set up a new terminology aimed at replacing or rationalizing present systems. Over time, stakeholders will likely take actions on their own that will lead to convergence, but that is not the project’s charge.

To say that achieving this project’s goals is an ambitious undertaking is somewhat of an understatement. The difficulty is compounded by the fact that some of our stakeholders who claim to do risk assessment do not, and others deny doing it, and in fact do.

The Thesaurus must therefore allow the ‘owners’ of an item they are entering to find terms in the locations they believe they belong, even if our generic representation or logic indicates they do not belong there. ‘Official’ and technical UK, French, Netherlands and American usages each have somewhat different logics. The Thesaurus makes an attempt to allow for the collection and illumination of terminology and concepts in all of these logics.

The measure of success will be whether or not stakeholders find the Thesaurus useful. To accomplish this, the Thesaurus must strike an appropriate balance between our stakeholders’ needs for completeness and uniqueness, and their need for a Thesaurus that is not too technically complex for practical use.

Clearly, there will have to be trade-offs and the project’s Steering Group will have to do a Pareto analysis and settle for meeting, say 95%, of our stakeholders needs for completeness and uniqueness in order to reduce technical complexity.

## **2. The conceptual approach used in developing the Thesaurus**

Each item entered into the Thesaurus will have an identified ‘owner’, the ‘client’. The client’s intended meaning of their entry will be elicited by a questionnaire containing a series of structured, operationally phrased, questions organized in a hierarchical system. These questions will be designed to elicit the greatest level of informational detail embedded in the owner’s meaning/understanding of the item being entered.

In some ways, the project’s approach to the Thesaurus can be visualized as conceptually similar to the taxonomic approach used for the description, identification, naming of organisms, and their classification into hierarchical groups. Others have

viewed the Thesaurus as a ‘translation Engine’ which captures the intended meaning of a risk assessment item and ‘translates’ it into defined operational language.

The questionnaire is available as a document on the Internet. Clients convey the meaning they attach to the risk assessment content of an entry (definition, regulation, code, risk assessment case) by their responses to a hierarchically organized series of questions. The client’s ‘response’ to each question is captured in a database linked to the client’s plain language description of their entry.

Users will be able to query or search the Thesaurus on-line. Standardized reports giving the content of an entered item or comparing two or more entries will be available. Some capacity for Boolean searches of the Thesaurus database will also be available in the future.

### 3. Construction and nature of the hierarchial system

Three major tasks needed to be addressed in order to reduce the Thesaurus concepts outlined above to practice: (1) A suitable hierarchical system (set of questions) to elicit the needed inputs on an item from the client, i.e. a suitable taxonomic system. (2) An effective and efficient physical system for collecting and storing the inputs. (3) A system that allows the user to obtain desired Thesaurus outputs.

After examination and analysis of the manner in which the risk assessment process is described and categorized in the literature, a decision was taken on the broad categories that would be used to structure the hierarchical system for capturing inputs and furnishing outputs. The categories were chosen, to the extent practicable, to capture and mirror the present ways in which different risk cultures view the risk assessment process.

While the system was designed with terminology suited to describe accidental releases of chemicals from fixed facilities and the undesired outcomes from such releases, a reasonable effort was expended to develop categories that could be used in any other risk area.

Particular attention was paid to avoiding ‘terms of art’ such as dose, exposure, risk, etc. for obvious reasons. This, of course, led to increased verbosity which must be endured, though hopefully minimized. One cannot convey clear meaning using words whose meaning is equivocal.

The system is organized hierarchically in five levels. At the top level there are four broad Generic Elements. Each Generic Element has a varying number of sub-Elements, and each Sub-Element has a varying number of Terms. Each Term has is broken into categories containing a set of Descriptors which make up the lowest level of the system.

The terminology used to describe these hierarchical levels is as follows:

*Element*: A group of related, operationally defined, general concepts.

*Sub-Element*: One or more closely related, operationally defined, concepts embodied in the Element to which it relates.

*Term*: A single concept in the particular Sub-Element to which it relates.

*Category*: Examples that give general operational meaning to a concept and contain a set of related Descriptors.

*Descriptor*: Examples that give a specific operational meaning to a concept.

The Categories and Descriptors in the Thesaurus are only a set of examples selected because they were most likely to cover common usage. Clients can create their own Categories and Descriptors as required to capture their intended meanings or understanding of their entry, and the system will capture and display such entries. Over time, multiple entries of a new Category or Descriptor will be incorporated into revised versions of the Thesaurus, and unused ones will be dropped.

The number of sub-Elements and their Terms, Categories of Descriptors and Descriptors reflect both logic and system needs. They are set to promote ease and effectiveness in collecting the information needed to capture the intended meaning of the person submitting an item and to convey this meaning to users of the Thesaurus. While the Sub-Elements of Element I contain only one Term each, the sub-Elements of Element IV contain as many as four Terms.

The number of Elements, Sub-Elements, Terms, Categories and Descriptors is as follows: *Elements*: 4, *Sub-Elements*: 14, *Terms*: 19 *Categories*: 70, *Descriptors*: 368.

In order to give an impression of the terminology used in the system, the definition of the four terms is given here.

*Generic Element I*: Identification of sources with the potential to cause undesired outcomes to subjects of concern.

*Generic Element II*: Identification of possible sequences of events leading to loss of containment of the potential to cause undesired outcomes to a subject of concern resulting in its entry into a domain of the ecosystem. Estimation of possible distributions of both the released potential and the subjects of concern over time periods within compartments delimited by specified boundaries or end-points.

*Generic Element III*: Identification and description of how the specified undesired outcome is related to the intensity, time and mode of contact of a specified potential to cause the undesired outcome to the subject(s) of concern.

*Element IV*: Identification of the basis for estimating and expressing the likelihood that a specified undesired effect\* will occur and description of the quality/uncertainty of such estimates; comparison of the estimates with relevant standards and guidelines and evaluation of the impact of specified alternative assumptions on the estimates.

As will be discussed later, the client has the ability to generate new 'Categories' and 'Descriptors', and insert them into the report on his entry. However, the number of Elements, Sub Elements and Terms are fixed.

The four Risk assessment Generic Elements are designed to capture only a segment of a continuum of activities that start with scoping the risk assessment, run through the risk assessment process and continue onto activities related to risk management, risk reduction, emergency response, risk communication, etc. Two additional Elements, O and V, are presented as simple text boxes. The purpose of these text boxes are to capture information that a client submitting an item feels is part of the risk assessment process not captured within the somewhat arbitrary boundaries used in the Thesaurus.

Element 0 is meant to capture aspects of the item being entered that the client believes should be included as part of the risk assessment process, and which generally precede Element I, e.g. scope and purpose of the Risk Assessment, stakeholder interactions re assumptions, etc.

Similarly, Element V covers aspects that generally follow Element IV, e.g. the Risk Assessment–Risk Management interface.

#### **4. The computer system for collecting inputs and providing outputs**

A large amount of complex information needs to be collected from and subsequently be made available to stakeholders who are widely distributed geographically in order to meet the project's objectives. Moreover, the Thesaurus needs to be capable of easy, quick and inexpensive updating, since its contents can be expected to change fairly rapidly over time. Ideally, it should also be available in a number of languages (commitment for a French version already exists).

Given these needs, the Thesaurus is being built as a computer document, easily accessible on the Internet. The 'input' questionnaire can be filled in by selection of appropriate 'boxes' on a HTML document. Instructions and clarification of terminology are available at each data entry point.

#### **5. Entering information into the Thesaurus**

The Thesaurus can handle the risk assessment content of four classes of items: Definitions, Laws and regulations, Specific risk assessment studies and Guidelines, policies, or codes.

The system presents the client with the opportunity of serially selecting applicable Elements, sub-Elements and ultimately any appropriate descriptors related to the item being entered. Provisions exist to collect reference information on each selected Descriptor, related 'Criteria' determining membership in the Descriptor class and 'Tools', (i.e. methodology, models, procedures etc.) associated with the selected descriptor. The client entry process has been summarized elsewhere [3].

##### *5.1. Thesaurus outputs*

Outputs will be furnished by essentially reversing the hierarchical system used for collecting inputs. Users will be able to ask for the intended meaning of a definition or the risk assessment provisions of a regulation, guideline, or an actual risk assessment case entered by a specified agency, society or individual.

Three types of Thesaurus outputs are visualized:

A complete report on an entered item.

Comparisons of entries on the same entered item by different clients, comparisons of different entry items by the same client. These comparisons will be able to be made at a chosen level of depth, i.e. Elements, sub-Elements, terms, Category of Descriptor and Descriptors.

Reports identifying all entries that contain or refer to Elements, Sub-Elements, Terms, Category of descriptor or descriptors specified by the person making a search.

## **6. Discussion and conclusions**

The initial test of the Thesaurus involved four countries. Each country had two participants enter the risk assessment provisions of the old 'Seveso' Directive and the country's laws implementing or corresponding to the Directive. This test revealed several deficiencies in both the ability of the system to capture a client's intended meaning (content) and the operation of the computer entry system [2].

The principle deficiency in the area of content arose in regard to interpreting and capturing explicit vs. implicit provisions of the document being entered. For example, a provision in a regulation that 'regulated facilities must notify all residents potentially affected by an accidental release' was interpreted by some clients as a risk assessment requirement for quantitative consequence analysis, although the particular regulation did not explicitly mandate such an analysis *per se*. Very many examples of this type of ambiguity were uncovered. This difficulty has been resolved by making system provisions allowing the client to differentiate between descriptors selected because they believed that the item being entered contained them explicitly or implicitly. Reports on entries will also allow for this type of differentiation. Limitations on the length of submitted papers does not permit a more detailed analysis of this trial in this paper.

A second trial of the Thesaurus based on the improved system started on July 1, 1997. This trial will involve entry of risk assessment guidance for facilities that store and process specified quantities of Chlorine and or LPG. In addition, each participant will enter their definitions for ten specified risk assessment terms, e.g. hazard, risk assessment, risk analysis, risk, exposure, etc. The results of this test will be available from the system via a greatly enhanced reporting/search system at the time the paper is being presented at this conference.

At this point in time, it is not possible to evaluate the utility of the OECD risk assessment Dictionary/Thesaurus to a wide audience of involved but non-technical specialists. However, it is already clear that the Thesaurus provides the technical community new insights into the structure and communication of the risk assessment process and the preparation and comparison of risk assessment studies and regulations.

## **Acknowledgements**

This work was sponsored and supported in large part by the OECD Expert Group on Chemical Accidents. The authors also wish to acknowledge the considerable contributions made by members of the project's OECD Steering group, and significant contributions by members of an Ad Hoc industry group, composed of D. Ferguson of ICI, M. Jayjock and D. Hendershot of Rohm & Haas and A. Ignatowski, S. Schechter, J. Lynch and P. McNulty. Special acknowledgement must be made for C. Ross at the University of Pennsylvania for developing the Internet presentation with help from S. Kimbrough. Work supported in part under the U.S. Environmental Protection Agency's Cooperative Agreement 821947.

## References

- [1] Draft conclusions and recommendations concerning the sessions on risk assessment, OECD Workshop on Risk Assessment and Risk Communication in the Context of Accident Prevention, Preparedness and Response, Item 15, Paris, July, 1995.
- [2] Second meeting of the OECD Steering Group on Risk Assessment in the Context of Accident Prevention, Emergency Preparedness and Response, Document 2, 14–15 February 1996, Washington, DC.
- [3] A.J. Ignatowski, I. Rosenthal, L.D. Helsing, An Internet Thesaurus/Dictionary for Analyzing Risk Assessment Processes, Laws and Regulations, International Conference and Workshop on Risk Analysis in Process Safety, October 21–24, 1997, Atlanta, GA, Workshop D: Methodology for Comparing Risk Assessment, American Institute of Chemical Engineers, New York, NY, 1997, p. 16.