

T-REKS[®]: a Contribution to the Environmental Information Management through a Computer- Supported Modular Knowledge Organisation System for the Environment

Paolo Plini¹, Sandra Lucke¹, Claudio Baffioni² and Bruno Felluga¹

Abstract

The utilisation of computer technology for the management of environmental issues has gained much importance in the last few years. The first reason is related to the increasing attention for environmental problems shown by governmental and non-governmental institutions, industries, opinion groups, professionals, and by the public. The concepts connected with the principle of sustained development have favoured this evolution through a complex route that led to extend the sustainability principle from global level to local level, down to the problems connected with the changes and growth of urban centres and to the daily choices of single citizens (Aalborg Chart). The second reason is related to the fast technological development of informatics and related issues. The increasing power of computers, the growing speed of information networks and the standardization of protocols, have favoured a wide circulation of data and information. This interaction between environmental issues and computer technology has produced good results in various fields: we will mention here only three. First, the contribution to the spreading of both environmental formation and information: the new computer instruments have rapidly changed the teaching methods that have been quickly introduced in the field of the environment, as well as of the occupational environment safety. A good example of this trend is offered by the projects financed by the European Program Leonardo da Vinci, of EC DG Education and Culture. Important is also the use of environmental information for implementing measures in land management and planning: administrative practices of management and control have been implemented thanks to these new instruments. An example of this is provided by the Italian GAIA (Environmental Management – Environmental Information) and Aquarium Projects, where innovative models for data banks (from data banks to data warehouses) and the GEMET thesaurus (General Multilingual Environmental Thesaurus, ETC/CDS - EEA, 1999) have been introduced for classifying

¹ CNR – Istituto di Tecnologie Biomediche, Via G.B. Morgagni 30/E, I-00161 Rome
E-mail: plini;lucke;felluga@itbm.rm.cnr.it, Internet: <http://www.t-reks.cnr.it>,
Phone: +39 06 44173704/8/3

² Indaco s.r.l., Rome

data and metadata. Eventually, we would like to mention the IST research programs of the EU, for the development of instruments to be used in transportation planning for the sustainability of metropolitan areas. In this program, the effort aimed not only at harmonising the information on mobility (traffic) characteristics with environmental problems, but also at achieving a dynamic exchange of information.

The common denominator of the aforementioned examples postulates the need and the possibility of using a general conceptual scheme for all these projects: a general reference scheme, developed through a common approach, tailored to the different applications in the different projects.

In the course of an international endeavour for the development of a multilingual thesaurus for the environment (GEMET, 1999), seven multilingual terminologies, totalling 50.000 terms, have been analysed and handled. These terminologies differed as regards document origin, source language, context, categorisation, level of control, semantic nuances, compliance to ISO norms, quality, and so on. Thus, a contextual frame was needed, to accommodate in a coherent system all the different expressions. This frame had to be controlled, flexible and ready to be applied to disparate environmental information systems. After a series of trials, T-REKS[®], the Thesaurus-based Reference Environmental Knowledge System[®], has been developed and applied.

Foreword

The T-REKS[®], Thesaurus-based Reference Environmental Knowledge System[®], is a computer-supported knowledge organization model developed by the Environmental Research and Documentation Unit of the Institute of Biomedical Technologies of CNR for handling the different terminologies of the Environmental Information Systems (EISs). The multilingual thesaurus that constitutes the core of the system contains terms representing reference environmental concepts, suitable for mastering data, metadata and classification systems. In the model, the thesaurus core is linked to other collections of terms needed by the EISs: standard lists of international usage; standard lists of national/local use; specific international, national, local terminologies. Basically, in T-REKS[®] two issues are envisaged, the activity on the terminological elements and their use within an application. A strict interrelation with each EIS and its feedback ensures the viability of the model.

On the thesaurus side, the activities deal with the basic environmental knowledge elements: selection and management of a controlled set of concepts/terms, internal accessory terminology and a catalogue to follow and keep track of the non-thesaurus terminology used in the various EISs. On the application side, the activities regard the specific semantic/formal control of any EIS and span from the data organisation graphs, data dictionaries and handling of international standard lists.

T-REKS[®] is intended as a system of organisation of environmental concepts/terms. Its components and their interrelations may be summarised as follows:

1. The **CNR Environmental Thesaurus**, with its
 - 1.1. **Classification Scheme** and
 - 1.2. **Definitions (Glossary)**.
2. **Accessory Lists of Terms**.
3. The **Catalogue of Lists and Thesauri (CLiT[®])**, which provides the links with the EISs that have adopted the thesaurus.
4. The **T-REKS[®] Maintenance Group**;
5. **Support activities for the semantic control** in an Environmental Information System.

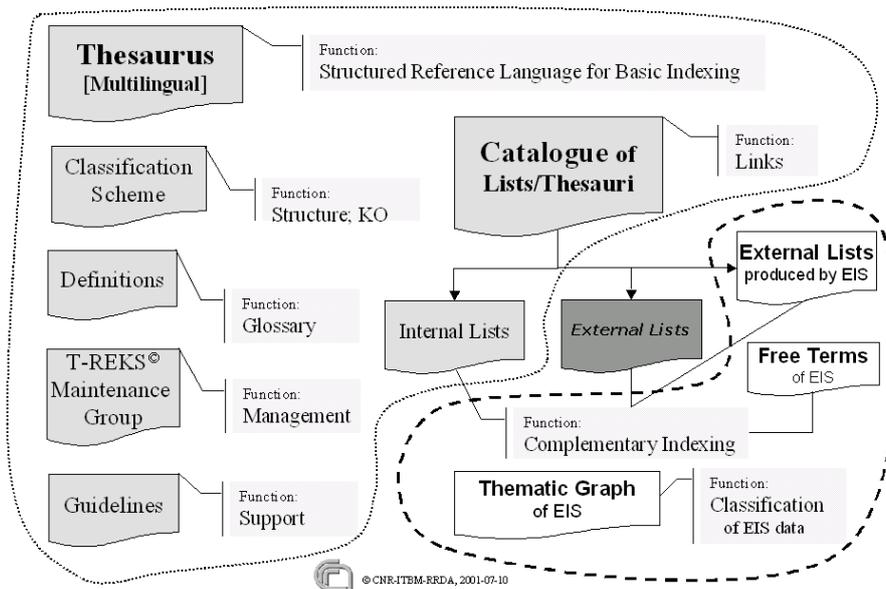


Fig. 1: The components of T-REKS[®]

1. The CNR Thesaurus for the Environment

Purpose: The CNR thesaurus has been developed and is maintained as a reference multilingual environmental controlled terminology, non-dedicated to a particular EIS.

Content: As an evolution of GEMET, the CNR Thesaurus contains about 12.000 terms arranged in 15 logical categories, corresponding to Top Terms. Its base of terms is largely shared with the most relevant environmental multilingual thesauri.

Structure: The Thesaurus terms are organised in 15 hierarchies with increasing levels of specificity. Terms belonging to different logical categories but conceptually close are clustered in themes. A series of extended horizontal relations ensures the exhaustive conceptual representation of terms in support of the system of faceted hierarchies and themes. The two-entry systematic presentation is completed by an alphabetic presentation. The thesaurus is compiled according to the ISO norms on thesauri.

Function: The Thesaurus is finalized to the conceptual and formal control of any information flow concerning the environment. It can be used as one of the link elements between different international, national and local environmental information systems. From the thesaurus it is possible to access the databases of any data repository that has adopted the thesaurus for the management of its environmental information. The thesaurus may also be used for work on ontologies; for the compilation of reports on the state of the environment; as an indexing and/or cataloguing tool in the organization of libraries and documentation centres; as a search engine in a Web site; as a dictionary for assisting in translation work.

In these respects, the thesaurus hierarchy is not only the traditional instrument for the semantic control of terms, but it also represents a way to assist indexing activities, facilitating the identification of the most appropriate term through the visualization of its semantic context.

The thesaurus terminology can act as an external checklist for the semi-automatic analysis of new corpora. Once identified, the new terms are checked and introduced either into the thesaurus or in specific lists of T-REKS[®].

1.1. The Classification Scheme

The Classification Scheme represents the framework for the knowledge organization of the thesaurus descriptors. It is based on:

- A systematic list of 15 Top Terms;
- A list of 40 Themes.

The hierarchical structure is facet-oriented. The Top Terms correspond to logical categories and reflect a systematic/taxonomic perspective. The descriptors are basically allocated in a mono-hierarchical order.

The themes reflect a more pragmatic approach to the environmental issues. They have been established according to practical considerations, corresponding to the information needs of an average environmental institution; thus, they can be used as checklists when dealing with environmental matters in general. The themes, being complementary to the Top-Terms, confer to the thesaurus a matrix structure. They allow a relational aggregation between terms belonging to different facets. They also provide useful information for the work on the horizontal relations (Related Terms).

1.2. The Glossary

An essential complementary element of the thesaurus is represented by the definitions provided for 5.700 preferred terms, in English and Italian. This Glossary is a necessary and important tool in checking the conceptual and linguistic correspondence of terms provided in different languages.

2. Accessory Lists of Terms.

Once the Thesaurus has been employed for one application, to perform an exhaustive indexing, additional systems of terms can be used, in the form of lists and specific thesauri, internal or/and external to the application. The thesaurus is intended to be a reference, non-dedicated thesaurus: its base of terms may be completed, if necessary, with a certain number of appendixes, like specific structured lists (e.g., relevant environmental organizations; national or regional geographic names; national environmental laws; national red lists of animals and plants; etc.) or specific thesauri. The purpose of these appendixes is to handle terms of specific sectors that, if allocated in the thesaurus categories, would make too big the base of terms. The CNR Thesaurus is predisposed for the link to such lists, whose titles are collected and displayed in the Catalogue of Lists and Thesauri.

It is important to note that the responsibility of the content, authority and updating of the individual lists lies with their compilers.

3. The Catalogue of Lists and Thesauri - CLiT[®]

The aforementioned complementary elements for indexing are listed in the CLiT[®], which contains the necessary references to access the content of these lists/thesauri from the thesaurus terms. Additional lists adopted by an EIS, identified during the evolution of the information system and notified to the Thesaurus Working Group, are examined in order to include them in the CLiT[®].

The access element to the CLiT[®] is a neutral identifier corresponding to a Thesaurus term.

CLiT[®] descriptors not present in the Thesaurus and exhaustively representing the meta-content of a list, can be candidate terms for the Thesaurus core.

4. The T-REKS[®] Maintenance Group

The T-REKS[®] Maintenance Group is permanently active at CNR, Rome. Thus far, its composition includes the CNR staff members and some co-workers, inside and outside CNR. Its activity coincides with the CNR research project on T-REKS[®] development and application, both in Italy and outside (e.g., 5th Framework Program). The Group is

responsible for selecting and handling the new terminologies proposed for insertion into T-REKS[®] and for its connection with the environmental ontologies.

5. Support activities for the semantic control in an Environmental Information System

5.1. The Specific Lists of an Application

These are lists containing various kinds of information, used exclusively by one application and cannot be considered as relevant for a shared use. Lists of bureaucratic-management-administrative jargons may be mentioned as an example.

5.2. The Free Terms

In the Thesaurus, all the relevant terms employed in the environmental field should be represented. Nevertheless, since no thesaurus system is expected to be exhaustive, any EIS, according to its indexing needs, can compile lists of free terms necessary for indexing and that are not included in the thesaurus or in the other reference documents (annexes). Some of these terms, identified by the user as “candidate terms”, can be submitted to the T-REKS[®] Maintenance Group for the inclusion in an updated version of the thesaurus. If not accepted, these terms will continue to reside in the list of free terms.

5.3. The Data Dictionary

This dictionary can be considered as an internal glossary of the whole information system. It identifies all the components of the data base domains. The data dictionary contains the list of object classes that have been identified during the analysis of the system, endowed with definitions. When relevant, it also contains the list of the attributes of an object.

5.4. The Thematic Graph

The thematic graph is a classifying tool organised with nodes (thematic areas; problems; entities; relevant environmental files; and so on) and arches (relations between nodes), which assists and gives the user direct access to the archives, starting from general conceptual areas to more specific areas. To each node one or more semantic qualifications are associated. Like in GAIA Project, it is advisable that these are built with controlled terms derived from the environmental Thesaurus, and struc-

tured in a way similar to the systems commonly used for the retrieval of the bibliographic databases. The metadata are linked to each graph node.

The definition of the catalogue structure implies among others, the following activities:

- The identification of the terminology sources to be used;
- The choice of the main thematic areas as starting points for navigating the catalogue;
- The identification of the semantic relations between themes even if different from the ones reported in the reference documents;
- The choice of a model that can be easily applied to different detail and extension level (worldwide, national, regional and local). The adoption of this model is the starting point for the definition of the transversal thematic organization.

A catalogue should be constructed as a classification graph with the nodes (the environmental themes) arranged in a hierarchical structure, allowing the user to move top-down, from more general to specific terms.

This system for handling and organising environmental information, based on a rigid hierarchical structure, provides the user with a tool for visualizing the internal organisation of the catalogue but it must be integrated with thematic transversal (horizontal) associations.

In some cases, in order to overcome the rigidity of the scheme and to enlighten different aspects of the same environmental topic, it could be necessary to duplicate some nodes of the graph, thus providing different pathways to access the same environmental topics.

In case of problems related to the choice of the most appropriate terms, it is possible to refer to indications coming from reference classifications, to environmental legislation and to the common use of a term in respect to its scientific use.

6. Technical Aspects

The Thesaurus is continuously maintained using a set of interconnected software tools. It is handled in order to be displayed on intranet/Internet. Specific exporting functions are developed using the SGML-family formats.

The Communication Network Service of CNR has activated the specific domain <http://www.t-reks.cnr.it>, where the user component of the system is going to be available. Specific Web pages are being developed for a close interaction with users through the Web.

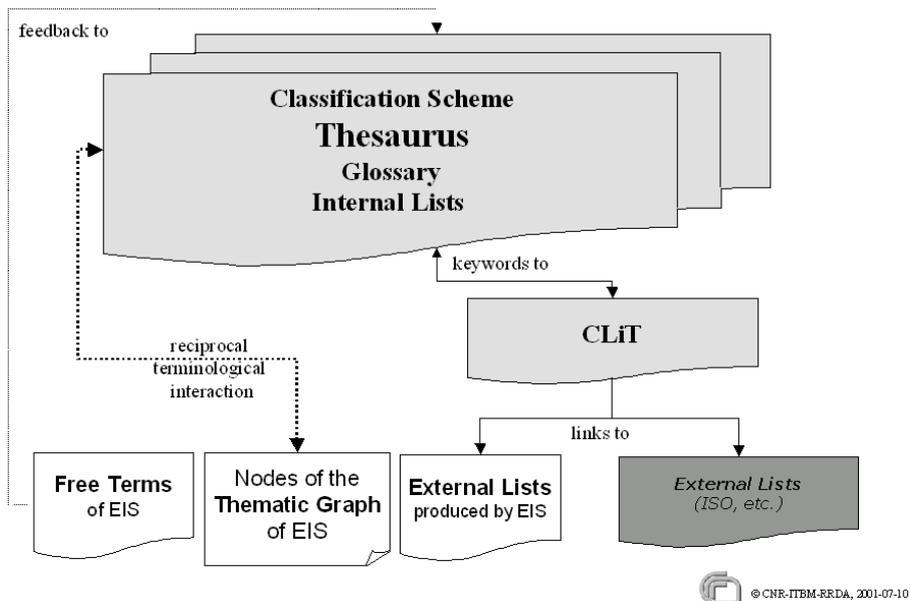


Fig. 2: The T-REKS[®] Relationships

Bibliography

Felluga B., Lucke S., Pàlmera M., Plini P., de Lavieter L., Deschamps J., Eds.: *Thesaurus per l'ambiente - Versione quadrilingue / Thesaurus for the Environment - Quadrilingual Version / Milieu-thesaurus - Vierteralige vertaling / Thesaurus für die Umwelt*, CNR-SIAM & CNR-UPIS edition on CD-ROM, Milan, 1994.

EEA, European Environment Agency. *GEMET, GEneral Multilingual Environmental Thesaurus*. pp. ix + Volume 1: Systematic List of Descriptors, pp. 44; Volume 2: Thematic List of Descriptors, pp. 78; Volume 3: Alphabetical List of Terms, pp. 550; Volume 4: Concordance List, pp. 127; Volume 5: Multilingual List of Descriptors, pp. 536. EEA, Copenhagen, August 1999.

For GAIA Project, 1999, see: <http://www.regione.piemonte.it/ambiente/sina/gaia/adwd.htm>.

For Aquarium Project, 2000, see: <http://www.aquarium.infotn.it/>.

Plini P., Felluga B., 2000 - T-REKS, Thesaurus-based Reference Environmental Knowledge System. TDCNet, Conférence pour une infrastructure terminologique en Europe. 13-15 mars 2000, Maison de l'UNESCO, Paris, France.

<http://www.unilat.org/dtil/etis/actasTDCnet/feluga.htm>.