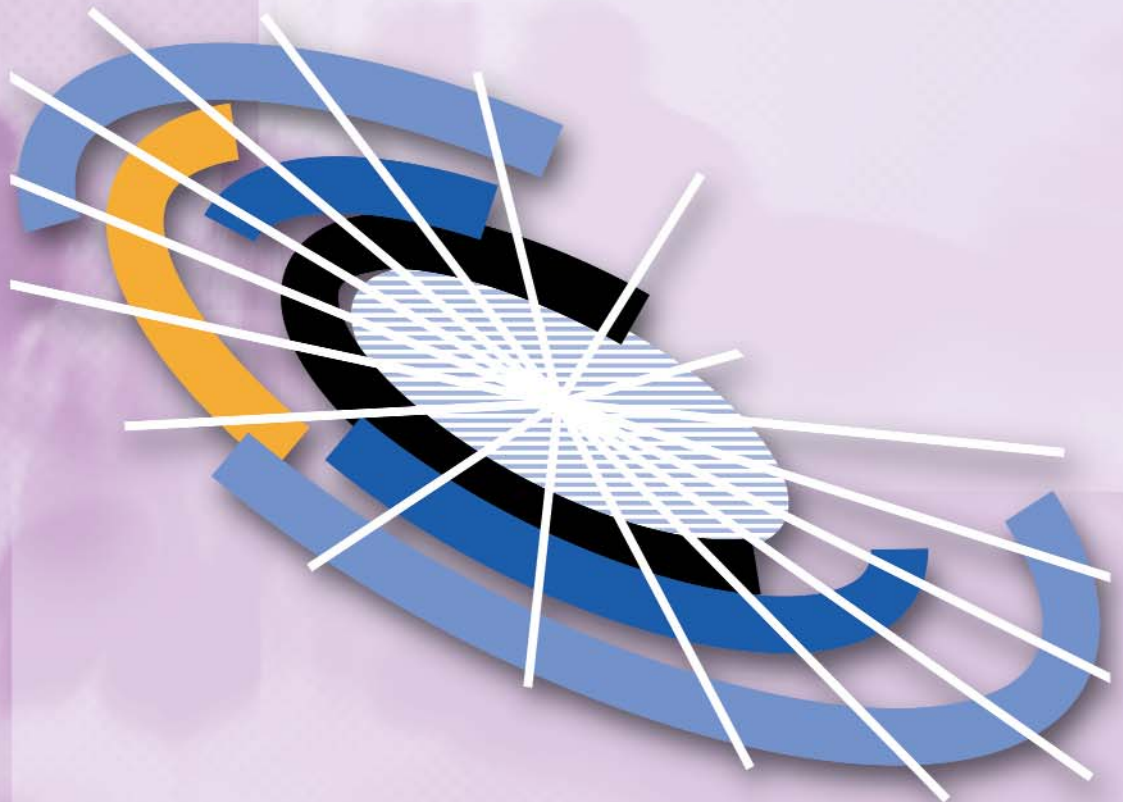


## Portals for Information Dissemination and Taxonomies for Classification



Report on CISTRANA Workshop  
Abingdon, 20 -21 February 2006



# Report on the CISTRANA Workshop:

## Portals for Information Dissemination and Taxonomies for Classification

**Abingdon, 20 - 21 February 2006**

### TABLE OF CONTENTS

<b>About CISTRANA</b> .....	<b>4</b>
<b>INTRODUCTION</b> .....	<b>5</b>
<b>SETTING THE SCENE</b> .....	<b>7</b>
<b>PORTALS</b> .....	<b>9</b>
The CISTRANA IST Research Portal .....	<b>9</b>
IST World: European RTD Information and Service Portal .....	<b>10</b>
The Research and Innovation Information Space, CRISs and Research Portals .....	<b>11</b>
HunCRIS: The Hungarian National Research Information System .....	<b>12</b>
Semantic portals .....	<b>13</b>
<b>TAXONOMIES</b> .....	<b>14</b>
Taxonomies, thesauri, classification schemes: structuring concepts and navigating structures .....	<b>14</b>
Issues in taxonomy development and application .....	<b>16</b>
Taxonomies and the Semantic Web .....	<b>17</b>
<b>THE MAIN MESSAGES</b> .....	<b>18</b>
<b>WORKSHOP PARTICIPANTS</b> .....	<b>21</b>

Rapporteur and Author of the Report: Simon Lambert, CCLRC



## Executive summary

This deliverable reports on the workshop held in February 2006 in Abingdon, UK, on the subject of portals and taxonomies in the context of provision of information on research activities, especially in the information technology area. The workshop was one of a series of five organised by CISTRANA on topics of special

importance to the project and its interested parties. The report summarises the presentations that were made, plus discussions and issues arising, and lists some key messages that arose about the stakeholders in portals, the information needs they satisfy, and perceived gaps in the current portal landscape.

## About CISTRANA

The CISTRANA project (<http://www.cistrana.org>) is a Coordination Action under the European Commission's Information Society Technologies (IST) programme. Its basic aim is to improve the coordination of national programmes in information technologies with each other and with European programmes, with a view to enhancing the impact of all research and development efforts in Europe and reinforcing European competitiveness in the area. Europe has a remarkably high reputation in the area of information technologies research, but fragmentation of efforts, limited cooperation between key players, and lack of information exchange about activities in other countries lead to loss of efficiency, duplication of effort and missed opportunities. Overcoming these barriers is the challenge for CISTRANA, and will help to build the vision of the European Research Area.

The CISTRANA project has three objectives:

- ▶ To develop a map of the national research landscape

- ▶ To pinpoint areas where cooperation is essential
- ▶ To establish sustainable mechanisms including methodologies and procedures for transnational coordination initiatives

The project is conducting studies, writing reports and bringing stakeholders together in many ways. One of its outputs is a publicly accessible IST Research Portal giving access to information on national research activities in information technology in the 33 countries represented in CISTRANA. The portal can be accessed via the project's website, or directly at <http://www.portal.cistrana.org>. The development of the portal contributes directly to the first of the three objectives, by opening up to interested parties sources of information to help overcome the barriers identified.

## Introduction

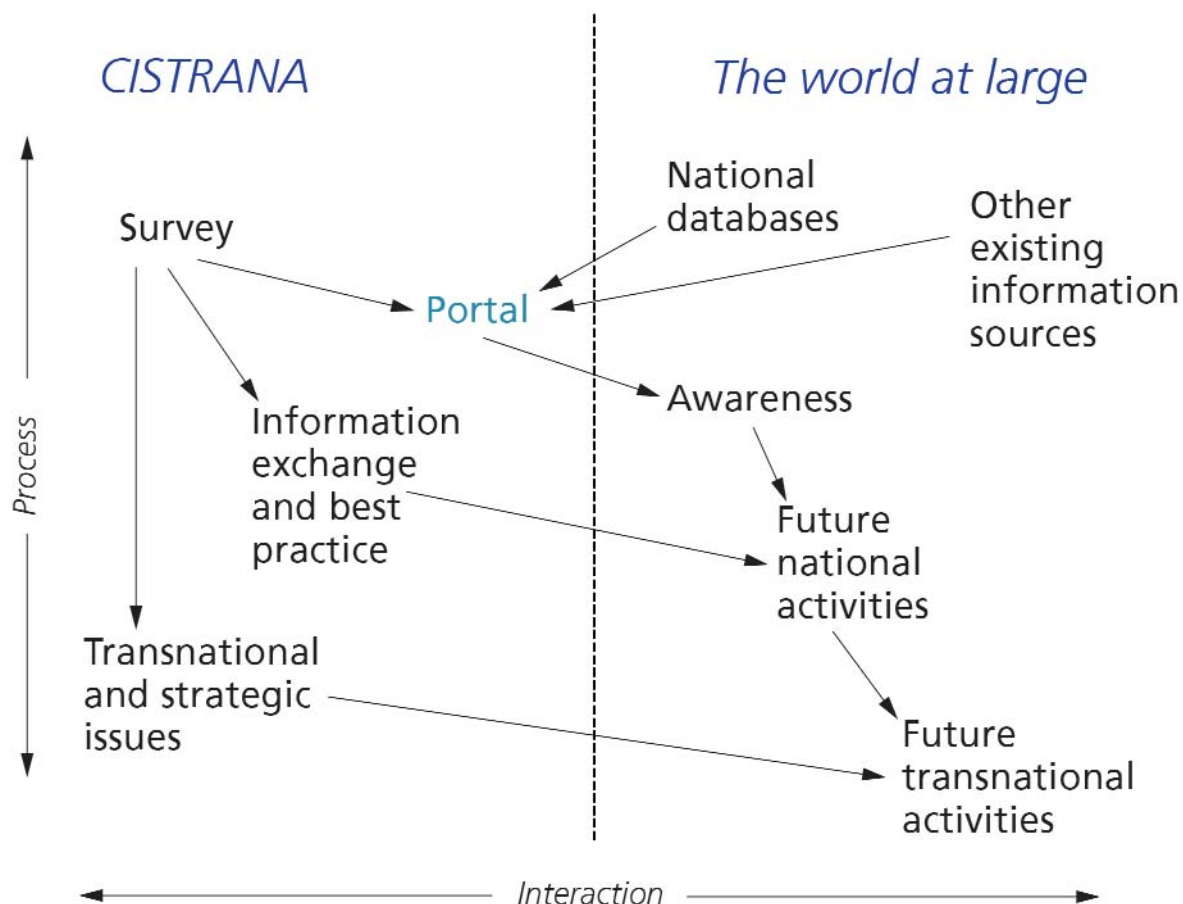
The CISTRANA project has organised a series of workshops that are part of the project's effort on analysis and exchange of information and best practices. Five topic areas were identified that are important in the context of CISTRANA's endeavour to improve coordination across national programmes:

- ▶ National policy priorities and RTD programmes in the field of ICT (November 2005, Brussels)
- ▶ Programme Impact Assessment in National IST Initiatives (December 2005, Paris)
- ▶ Best practice in multi-national programme collaboration (January 2006, Cologne)
- ▶ Portals for information dissemination and taxonomies for classification (February 2006, Abingdon)
- ▶ Design of national IST programmes in the context of ERA coordination (March 2006, Budapest)

In each case, the aim was to bring together a group of stakeholders to learn and exchange information and experience about the topic in question, which they will then be able to put into practice in their own national environments. Furthermore, the CISTRANA project itself is able to take advantage of the bringing together of these stakeholders and experts to gather and synthesize information, and to feed the results back to the community through the workshop reports and other actions.

So why a workshop devoted to portals and taxonomies? The CISTRANA portal is an important output of the project, providing visibility on a range of national research activities in IST. It contributes to the mapping of the IST landscape in Europe, and has a wide range of potential users, including policy makers, programme managers, researchers and people from industry.

The role of the portal in the project is illustrated in the figure below. It shows how the portal draws together disparate information sources to



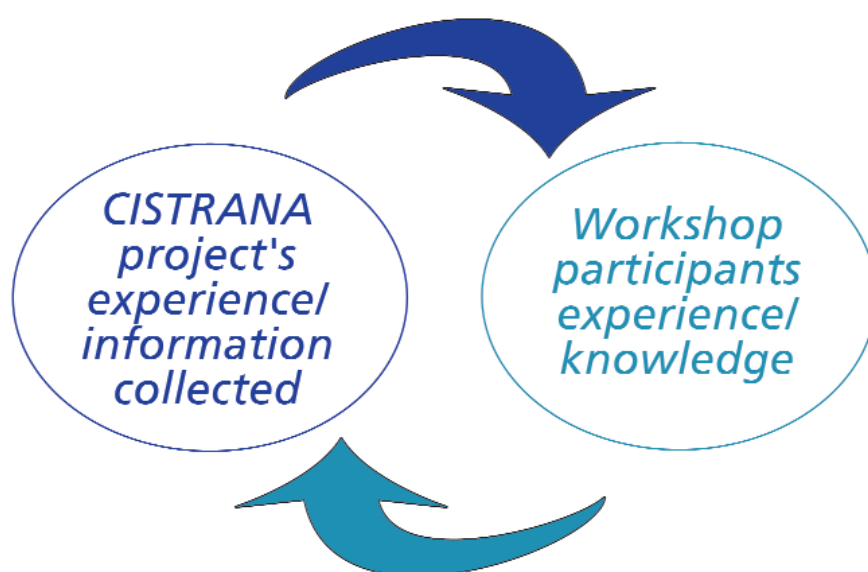
feed into awareness-building as a basis for other outcomes of the project with long-term significance.

The CISTRANA project began in September 2004 and a first beta version of the portal was produced in June 2005 for internal evaluation by a restricted set of representative end-users. This was followed by the first public release in September 2005. This release offered access to information about national research programmes in information technology, which had been specially collected by the CISTRANA project. A variety of ways of searching, browsing and viewing the information are available.

The CISTRANA IST Research Portal is itself part of a landscape of portals and information resources. It stands in relation to other European and national portal activities, including the well-established CORDIS, the early ground-breaker ERGO, and the innovative IST World. With the

first public release available, it was an opportune moment to draw together stakeholders to share their perspectives. The target group of participants included programme managers of national research and development programmes, national policy makers, developers and users of portals and Current Research Information Systems, and those working with taxonomies and classification schemes in the IST domain.

Specifically the aims of this workshop were exchange of information and increasing understanding of the needs, realities and future prospects of portals, through sharing experience, knowledge and good practice. This will lead to visions, goals and strategies for the future development of portals and taxonomies in European IST research, and it is hoped it will be influential at European level for those developing portals and using/maintaining taxonomies.





## Setting the scene

The first question to ask is 'What is a portal?' The term is widely used, and definitions abound. The online encyclopaedia of computer technology webopedia.com defines a portal as 'Web site or service that offers a broad array of resources and services, such as e-mail, forums, search engines, and on-line shopping malls', thereby emphasizing the broad functionality and commercial aspects of portals. For the purposes of the workshop, the following characteristics were considered as characterising a portal – not implying that all must be satisfied, but that their presence tends to suggest that the site in question is a portal:

- ▶ A portal is a single entry point
- ▶ It provides access to 'somewhere else'
- ▶ It is comprehensive
- ▶ It is suitably organised for its intended users

A portal may be contrasted with a gateway, which points at sites through selection, while a portal adds extra value through analysis and evaluation, and a vortal ('vertical portal') which is a portal that provides information and resources for a particular industry sector.

A key question about any portal is the motivation behind it. In some cases, as suggested above, the motivation might be commercial. In the case of CISTRANA, the motivation arose from a feasibility study undertaken by the European Commission in 2002-03, which found that only limited, incomplete and inconsistent information on national research activities is available, making identification of common and complementary activities virtually impossible. There is a need for a central access point to facilitate the exchange of information between states at the European level as a vital component of the European Research Area. The barriers to accessing relevant information include lack of consistent thematic search possibilities, lack of relevant information and lack of comparability between information sources.

In order to satisfy the above characteristics, a portal may utilise a taxonomy or other classification schemes. A taxonomy is a

hierarchical structure of terms used for information retrieval, but also for multilingual translation and other purposes. Its use helps to improve precision and recall in searching, and also to allow comparability between items. If different terms are used by different people, the use of a taxonomy can ensure that the expected results are obtained: for example, searching for all projects on 'neural networks' would also cover those labelled as 'connectionist systems', if the taxonomy establishes the equivalence of those terms. And if asking how many projects in formal software engineering exist in a certain country, the use of a taxonomy increases confidence that the retrieval is accurate, compared with a free text search.

The CISTRANA project has chosen a taxonomy for IST technology areas and implementation sectors, based on that used by IRC (Innovation Relay Centres) for classifying projects. The CISTRANA taxonomy may be seen at [http://www.cistrana.org/files/IST\\_ERA\\_taxonomy.pdf](http://www.cistrana.org/files/IST_ERA_taxonomy.pdf).

Some of the questions that arise in developing a portal include:

- ▶ How is it possible for a portal to accommodate a large number of users with diverse needs?
- ▶ What are the relations between national and international portals?
- ▶ How to obtain buy-in from information providers?
- ▶ Is there scope for a hierarchy of portals?
- ▶ How important are multilingual taxonomies in the European IST environment?
- ▶ How can taxonomies be kept up to date with the evolving field?
- ▶ Do different users have different needs for taxonomies? How to reconcile them?
- ▶ How to work with multiple established taxonomies?

- ▶ What basis is needed for portals: taxonomy, thesaurus, controlled vocabulary, ontology, ...?
- ▶ What motivates the choice of depth of a taxonomy?

The approach of the workshop was a series of talks by invited speakers, followed by free discussions run by facilitators. The speakers were chosen as representing a wide range of experience and perspectives, including well established portals, systems currently under

development, practical experience with taxonomies, and future technologies with an impact in the area.

The following sections on portals and taxonomies include summaries of the talks that were given, extracting the key points and where appropriate capturing some of the points made during the discussions. The original presentations may be seen on the Web at <http://www.cistrana.org/149.htm>. The graphic images associated with the individual talks were provided by the speakers themselves.

## Portals

### The CISTRANA IST Research Portal

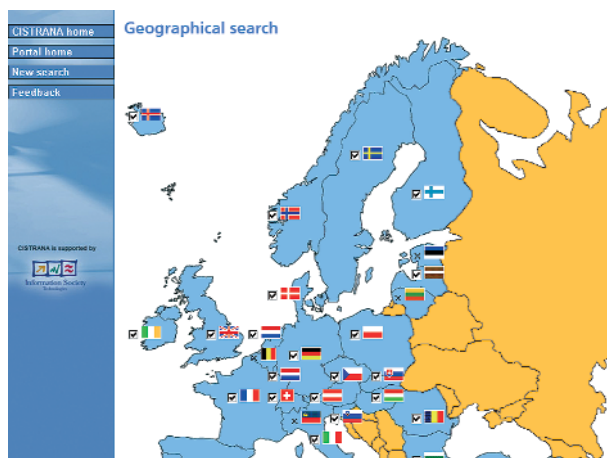
Simon Lambert, CCLRC (UK)

**A public portal on national research activities in the IST area. Developed within the CISTRANA project, a Coordination Action funded by the European IST programme. Intended users: policy makers, programme managers, project managers, industry.**

The CISTRANA IST Research Portal<sup>1</sup> was motivated by a feasibility study undertaken on behalf of the European Commission in 2002-03<sup>2</sup>, which found that only limited, incomplete and inconsistent information on national research activities in IST is available making identification of common and complementary activities virtually impossible. Barriers to accessing relevant information included:

- ▶ Lack of consistent thematic search possibilities
- ▶ Lack of relevant information
- ▶ Lack of comparability between information sources

A consequent need was identified for a central access point to facilitate the exchange of information between Member and Associated States at the European level as a vital component of the European Research Area.



The basic principle of the CISTRANA portal is to give access to information on national activities that is interesting, novel and not readily obtained elsewhere. The metaphor is that of mapping the landscape of IST research, presenting views on research policies, programmes, projects and related areas so as to allow users to apprehend and compare national activities.

Different methods of browse and search are available, including geographical search and keywords or free text. Some data, specially gathered by the CISTRANA project, is stored locally, but there is also the possibility of accessing remote databases (for example, of project information). The data model is based on CERIF<sup>3</sup>.

#### Discussion points

There is a trade-off between clearly targeted scope and comprehensiveness. Sometimes lack of absolute completeness may be justified, if what is available is valid and interesting. However, if a portal is to be used for benchmarking (and it might be tempting to do so), then completeness of content becomes critical. Consistency of presentation of information is required.

There are different strengths and weaknesses of local and remote data. Maintenance is a key issue. Third-party information providers need some commitment – what is the payback to them? How are the providers triggered to update information? In some cases peer pressure might be enough: keeping their information up to date makes the providers look good.

With regard to user experience, it is generally not desirable to embed pages from other sites into the portal. Rather a new window should be popped up for off-site pages, while leaving the original window open so that users will stay on the portal.

<sup>1</sup> <http://www.portal.cistrana.org>

<sup>2</sup> ERA Portal to National IST Research Information: Key Findings and Recommendations of the Feasibility Study, February 2003

<sup>3</sup> CERIF, the Common European Research Information Standard, is maintained and developed by euroCRIS. See <http://www.eurocris.org/en/taskgroups/cerif/>.



## IST World: European RTD Information and Service Portal

Brigitte Jörg, DFKI GmbH (Germany)

**A public portal<sup>4</sup> for IST competencies and their interrelationships in Europe.**

**Developed by the IST World project, a Special Support Action funded by the European IST programme.**

**Intended users: Organizations from all countries looking for specific RTD competencies; organizations and service providers from new member states and associated states wishing to promote their own competencies.**

The IST World portal offers innovative, automated analytic and visualisation services over information collected about RTD actors in Europe especially from the New Member States and accession countries. The aim is to promote RTD competencies in specific fields, thus supporting partner search for IST proposals and commercial projects. The portal concentrates on two out of four major thematic priorities within IST: knowledge and interface technologies and applied IST research addressing major societal and economic challenges.

The data model is CERIF-based, and the open Web directory DMOZ<sup>5</sup> was used for the taxonomy of topics. The repository contains data

imported from existing data sources, obtained from the community itself through self-registration, and harvested automatically from the Web. The portal allows flexible browse, search and navigation, and provides advanced analytic tools

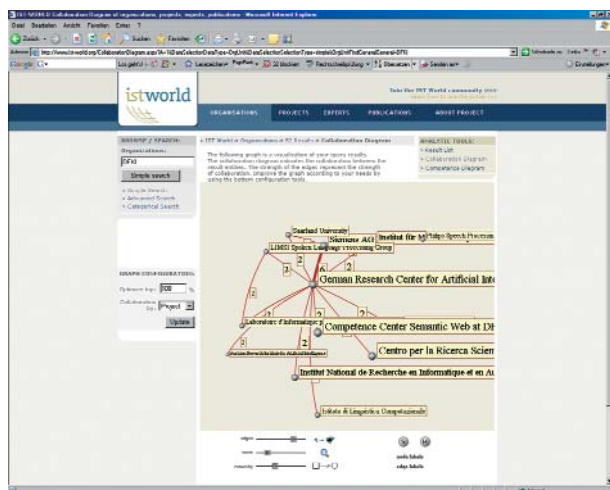
such as social network identification, partner finding, and expertise identification. A sample collaboration diagram is shown below.

### Discussion points

The IST World portal illustrates a balance between revealing what is actual and what is potential: for example, a map of potential collaborators (a hypothetical social network) might be valuable in building project consortia.

The analytic tools of the IST World portal are highly general and could be transferred to other CERIF-based systems, opening up many possibilities for visualising and manipulating data in novel ways.

Automatic harvesting, though efficient, leads to problems of data quality, and inevitably data cleaning is required.



<sup>4</sup> <http://www.ist-world.org/>

<sup>5</sup> DMOZ (<http://www.dmoz.org>) is 'the largest, most comprehensive human-edited directory of the Web. It is constructed and maintained by a vast, global community of volunteer editors.'

## The Research and Innovation Information Space, CRISs and Research Portals

Geert Van Grootel, Ministry of Flanders (Science Division) (Belgium)

### Reflections on the requirements for information portals in a research environment, based on the experience in Flanders.

The Flemish research database IWETO can be taken as a case study of a long-established portal to research information, allowing some lessons to be learnt for the future. There are a number of problems concerning acquisition of content. The update frequency is low, and data is often out of date. The completeness of coverage is variable, and there is often redundancy and ambiguity. These are all typical problems arising from a data collection process that is not well aligned with organisational processes.

Understanding the research and innovation added-value chain allows an alternative vision of a Current Research Management System. This idea has some basic principles:

- ▶ Distributed information architecture that is scalable
- ▶ Authentic sources with identifiers, and data quality produced in the processes

- ▶ Crossroads databases such as researcher CV environment and a central metadata repository
- ▶ An agreed information framework with metadata schemes and protocols

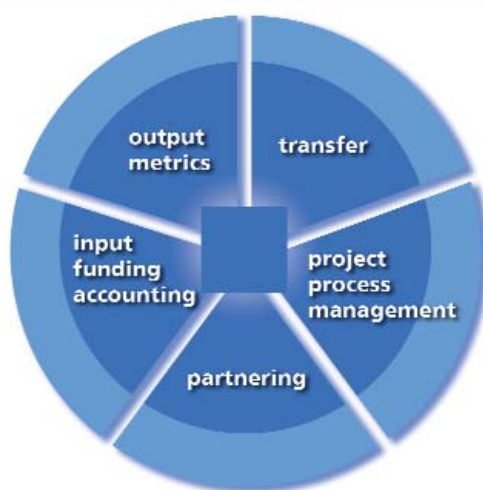
Information should be captured directly from within the electronic business processes, and quality and timeliness should be improved by collecting data at the time of first use in the processes.

### Discussion points

It is generally recognised that the problems of timeliness, completeness and accuracy of information are much less IT problems than organisational ones.

A different set of issues concerning input and maintenance of data arises in this context. Making data available might mean a loss of power; the benefit must go to supplier of information, and for each group of stakeholders there has to be an argument of the benefits. The cost argument is key: the electronic process is simply quicker.

## Output requirements for a CRIS



The different output requirements for a research information management system (Copyright: uniCRIS AG)

## HunCRIS: The Hungarian National Research Information System

*Adam Tichy-Racs, National Technical Information Centre and Library, Budapest University of Technology and Economics (Hungary)*

**The Hungarian National Research and Development Registry and Information Service, established in 2001.**

**Research projects receiving public money have to be registered.**

**Intended users: experts of different disciplines, research managers of organizations, research managers of funding agencies, researchers, mediators (for partner search and for take-up of new technologies), science and technology representatives of Hungary, industrial managers (for expert search).**

The HunCRIS<sup>6</sup> portal has a large amount of content and is growing. It uses several classification schemes: the Hungarian national classification (8 fields, 60 branches); also Ortelius (1300 expressions) and EPSS (2300 expressions, used in FP6). The taxonomy evolves; project managers supply free text which is transformed into standard format, and inserted into the best-matching branches, or intermediate level expressions may be created. Interestingly, the end result resembles EPSS.

The HunCRIS system is CERIF-based and bilingual (Hungarian and English). It contains data on 4500 projects, 1050 organisations and 11000 researchers. It is used to produce thematic compilations, such as current projects in physics or in the automotive industry. It is also used to generate organisational compilations such as university 'yearbooks'.

### Discussion points

There may be problems of usability for classification schemes when there are a very large number of items to be indexed. Having an estimation of the number of national and multinational IST projects throughout Europe, a two-level taxonomy such as CISTRANA's would lead to the number of projects recalled using a single low level item of the taxonomy being above 100. It is therefore desirable to use a classification system which is more selective even when using a single key expression for searching. That more selective system should consist of more than two levels.

<sup>6</sup> [http://www.info.omikk.bme.hu/nkr1/HunCRIS\\_eng.htm](http://www.info.omikk.bme.hu/nkr1/HunCRIS_eng.htm)

## Semantic portals

Alistair Miles, CCLRC (UK)

### A view of what Semantic Web technologies can offer to portal developers, with reference to developments in the SWAD-Europe project.

Semantic Web technologies<sup>7</sup> are not something in the future; they are present technologies available for use now. An exemplar of their application is the 'Semantic Web Environmental Directory'<sup>8</sup> from the SWAD-Europe project, which supported W3C's Semantic Web initiative in Europe. This is an online directory of environmental organisations and projects in the UK, and serves as a proof of concept demonstrator.

The principle is that information is decentralised. An organisation creates a machine-understandable description of itself, and publishes it on its own website. It then registers with SWED, submitting only the URL of its self-

description. SWED then periodically harvests descriptions from registered locations and updates the directory.

Organisations use shared ontologies and taxonomies to describe themselves. Ontologies, taxonomies and descriptions are published using (emerging) Web standards (RDF, OWL, SKOS). The portal allows faceted browse: that is, browsing by particular characteristics such as 'topic of interest' or 'organisation type'. It also allows combinations of browse and search, taking advantage of the semantics that have been assigned to the content.

### Discussion points

There is a difference of approach between, on the one hand, designing a system and later drawing in content from heterogeneous databases, and on the other hand offering a standard form for what the distributed data sources are doing anyway. The latter can be seen as growing best practice.

Semantic portals offer the possibility for add-ons such as quality check by rules.

There is a cost-benefit trade-off: 'semantically rich' metadata may be expensive to create, and expensive to maintain, but a little bit of 'meaning' goes a long way.



<sup>7</sup> See <http://www.w3.org/Consortium/activities#SemanticWebActivity> for an overview of the World Wide Web Consortium's activities in the semantic Web arena.

<sup>8</sup> <http://www.swed.org.uk>

## Taxonomies

### ***Taxonomies, thesauri, classification schemes: structuring concepts and navigating structures***

*Leonard Will, Willpower Information (UK)*

#### **An overview of structuring concepts and navigating structures, clarifying the sometimes inconsistent uses of terminology.**

The term 'taxonomy' is used with a wide variety of loosely defined meanings. It would be better to restrict it to meaning a monohierarchical classification of concepts, as used, for example, in the classification of biological organisms. A better expression for the general concept is knowledge organization system (KOS). These schemes deal with and relate concepts rather than words; concepts and relationships are their building blocks. One term is chosen as a descriptor to label the concept, but there may be alternatives – for example 'cars USE automobiles'. There is no implication that the chosen term is 'better' or 'more correct' than the others. A 'scope note' may be provided defining the concept.

There are two kinds of relationships between concepts: paradigmatic and syntagmatic. The former are a priori: they apply generally, independently of any specific document; the latter are a posteriori: concepts that are related only in the context of a specific document (for example 'shoes: history').

There are several ways of providing access to concepts

- ▶ Free text searching
- ▶ Uncontrolled tagging
- ▶ Controlled vocabularies

Uncontrolled tagging leads to 'folksonomies': users tagging items with words they find familiar. This has disadvantages that there is no control of consistency.

Thesauri and classification schemes are alternative ways of showing concepts and their relationships. They are complementary and both approaches are needed. They can both be built on the principles of facet analysis. Facets are groups of concepts of the same kind, e.g. materials, actions, disciplines, people and organisations, place, time.

Given structures such as these, it is possible to implement highly effective modes of browse and search, for example, finding related terms and synonyms. Most current interfaces do not provide the interaction, guidance and unobtrusive training that would help users to make full use of the underlying knowledge organisation systems.

#### **Discussion points**

Maintenance of taxonomies will normally be required, raising issues such as how to update previously indexed documents to new taxonomy terms. Maintaining a taxonomy must take place throughout its lifetime – it is not just a creation cost. Automated categorisation is lower cost and may be the only cost-effective solution in some areas, as when dealing with large numbers of documents in a restricted subject field.



## The same concepts viewed in different ways

### Classification view

- Good for browsing or surveying a topic
- Like a map
- Like a book's contents page
- Shows related concepts together
- Usually arranged by discipline
- Shows syntagmatic and paradigmatic relationship
- Shows compound topics as pre-combined subjects strings

### Thesaurus view

- Good for searching if you know what you want
- Like a gazetteer
- Like a book's index
- Gets quickly to individual concepts
- Usually arranged by facet
- Shows paradigmatic relationships
- Lets you combine concepts when searching

## Issues in taxonomy

*Stella G Dextre Clarke, Information consultant (UK)*

**An outline of what is involved in using taxonomies, raising the questions: What is a taxonomy for? Is it worth the bother? Multilingual versus monolingual; coping with different user needs; what's involved in building and maintenance?**

Some taxonomies are designed to enable browsing/navigation, while others primarily support the search process. One very useful option is search filtering – for example, selecting which of the meanings of the English word 'bat' is of interest (sports bat or flying mammal).

At its inception many years ago Yahoo! achieved immense popularity with its hierarchical taxonomy designed for drill-down navigation. Subsequently Google has tipped the balance of user preference in favour of searching, without the use of a taxonomy. Another search engine called Northern Light applies a very extensive and sophisticated taxonomy, but hides it from users. The benefit in this case is perceived as disambiguation of search queries. For example, a search for 'chips' presents the results in taxonomy-driven clusters, enabling the user to choose between the electronics context, snackfood, blue chip companies, etc.

The success of Google on the World Wide Web is rarely matched by any search engine in a smaller scale application, or within an enterprise; hence a continuing demand for taxonomies.

When planning a new taxonomy, it is helpful to explore those available on the Internet, as they cast light on the pitfalls, for example:

- ▶ The limitations of the precoordinated approach implicit in most hierarchies
- ▶ The frustration of users who cannot locate complex concepts in the 'tree'

- ▶ The problem of choosing between a large number of headings on one screen

'There is some research – and many anecdotes – that shows that the best retrieval situation is a mix of human indexing using a controlled vocabulary and the ability to search full text. But there appears to be no work that will assist the index designer in deciding how much expensive human effort is really warranted.'<sup>9</sup>

Multilingual taxonomies can be much more useful than monolingual, but are also much harder to build and maintain.

Key steps for building and maintenance of a taxonomy include:

- ▶ Decide what you want it for
- ▶ Involve all stakeholders to get their buy-in, and acceptance of inevitable compromises
- ▶ Assign responsibilities
- ▶ Organise a good budget
- ▶ Plan for sustainability
- ▶ Choose software and understand operational environment
- ▶ Get on with it! And keep it up!

### Discussion points

A taxonomy needs to be complemented by other search tools, and planned within an overall information architecture. For some purposes, free text searching will continue to be desirable to allow the discovery of unexpected relationships.

<sup>9</sup> Jessica L. Milstead, 'Needs for research in indexing', *Journal of the American Society for Information Science*, Volume 45 Issue 8, 1994

## Taxonomies and the Semantic Web

Alistair Miles, CCLRC (UK)

### An overview of the technologies in the Semantic Web arena that support taxonomies and other knowledge organisation systems.

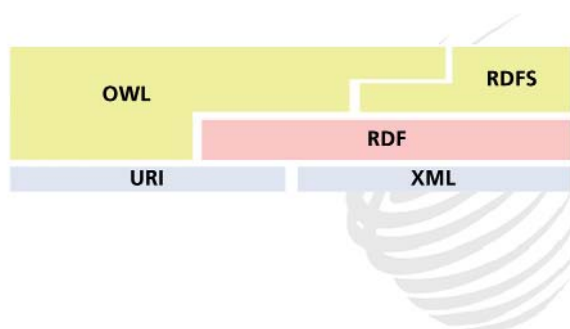
The Semantic Web allows individuals to present machine understandable data (e.g. about themselves, their projects, their discipline) which can be harvested and aggregated. It can be repurposed, so if the information is presented for a purpose, then it can be used for another purpose. A semantic web is a web of machine-understandable content. Computers can do a lot more with machine-understandable content:

- ▶ Merge
- ▶ Analyse
- ▶ Repurpose

The key technologies include XML, RDF, OWL and SKOS. Modelling frameworks that use these technologies are ontologies, taxonomies, thesauri, and classification schemes.



### Technology stack



Formal semantics with RDF is what makes content machine-understandable. An ontology is a way of declaring what types of things exist, and what types of relationships they have with each other, and provides common meaning to use in machine-understandable content. OWL is the Web Ontology Language, an extension of RDF, and is a W3C Recommendation as of February 2004.

SKOS is an application of RDF, and is concerned with declaring and publishing taxonomies, thesauri or classification schemes, for use in a distributed, decentralised information system (i.e. a semantic web). SKOS and OWL allow the exploration of the cost/benefit tradeoffs involved in investing in semantics. The SWED portal (as described in the preceding talk by the same speaker) uses a combination of OWL and SKOS.

Basic application toolkits for storing, querying and manipulating RDF are good and stable, but currently collaborative development environments for ontologies, taxonomies, thesauri and classification schemes are missing.

### Discussion points

Though there is some scepticism about the readiness of semantic web technology, there are good, scalable, stable systems such as Jena from HP Labs, an RDF storage and retrieval system that has been tested on tens of millions of triples.

There are several models for collaborative creation of taxonomies and thesauri. UDC (Universal Decimal Classification) is at one end of the spectrum, with strict editorial control, with delicious at the opposite end. A good solution is probably a hybrid approach, in which users play a role but there is also an editorial team.

Change management in distributed ontologies is urgently required: what are the available change models, what are the risks and tradeoffs?

## The main messages

Portals can be an effective way of delivering pertinent information to groups of users. However, it is not only the users who should be identified and whose needs should be understood; there are other stakeholders as well,

principally those involved in supplying content. For portals to research information in general, these stakeholders can be classed in the following way<sup>10</sup>:

<b>Information service providers (providers of the data set)</b>	They aim to increase the awareness, usage of and interest in portals as a tool to communicate RTD information. In some case, portals are made available as a policy-supporting tool.
<b>Information providers (providers of the content)</b>	Data provision is often mandatory for participants in publicly funded programmes. However, information providers also supply data on their own initiative. Providers use portals as a communication tool, even as a promotional instrument for their activities and achievements. Information providers also might be end-users seeking to benefit from published information.
<b>Institutions and policy makers</b>	They use portals for policy making background information, research trends analysis, etc. They might want to set up priorities, objectives, policy definition, budgets and programmes, or to evaluate research funding allocation, avoid duplication of research activity, analyse research trends and get concrete research results examples for citizens. Policy makers also commonly need statistical analysis on research activities. Furthermore, policy makers can use CRIS to support information services in meeting government policy requirements, as a formal log of research in progress, or to assist project planning.
<b>RTD community</b>	They are users of portals as a source of research-related information. They may be seeking a kind of interactive channel (for example to seek for project partners). In general, portals can be a key tool to use in supporting their RTD activities: <ul style="list-style-type: none"> <li>▶ To avoid duplication of research</li> <li>▶ To identify experts for exchange of ideas or for collaboration</li> <li>▶ To identify information about research results</li> <li>▶ To locate equipment and services</li> <li>▶ To find funding opportunities for research activities</li> <li>▶ To make their research activities known</li> <li>▶ To communicate their own research results to the CRIS user community</li> </ul>

<sup>10</sup> Adapted from CERIF 2000 Guidelines: Final Report of the CERIF Revision Working Group Co-ordinated by DG XIII-D.4, European Commission, September 1999

<b>Intermediary organisations (any body offering assistance and support in the field of innovation and/or research and technological development)</b>	<p>Intermediaries are often frequent and pragmatic users of portals to assist their clients. They might use portals to get overviews, with the purpose to produce synthesis reports for their clients. CRIS facilitates day-to-day contact with innovative enterprises and finding relevant practical information (on on-going and completed research, exploitable research results, potential partners and experts).</p>
<b>Enterprises</b>	<p>CRIS can help businesses:</p> <ul style="list-style-type: none"> <li>▶ To know what is going on in their sector and in the business world at large</li> <li>▶ To promote and locate transferable technologies</li> <li>▶ To identify funding sources</li> <li>▶ To identify potential partners and experts all over Europe and to build the right relationships</li> <li>▶ To turn research efforts into products, but also to encourage them to participate in and exploit research activities</li> <li>▶ To compile and create publications</li> </ul>
<b>Non-profit sector</b>	<p>Non-profit users can use portals for different purposes:</p> <ul style="list-style-type: none"> <li>▶ as a useful tool in their work to locate innovative technologies or results related to their 'societal' objectives</li> <li>▶ CRIS is an easy tool to get relevant information and contacts with experts</li> <li>▶ CRIS is a tool that can provide them with help and information with policy options and guidelines</li> </ul> <p>These users are particularly looking for research projects leading to specific social applications in a wide variety of domains such as health, education, culture, social services, needs of disabled, environment, transportation and leisure, etc.</p>
<b>Media</b>	<p>The media are potential users of portals as their information source to find accurate and up-to-date information on on-going research, to detect relevant experts for interviews etc. Portals can also generate inspiration about the use of research information in their daily working process.</p>



In the domain of IST research, some particular information needs are of interest and not well met currently.

When dealing with information about individual people, data protection issues might arise. The important distinction to bear in mind is between individuals and their roles.

There has to be an alignment of the effort to input and maintain information with the benefit of putting it into the system. This can be achieved by taking the information, not from the source, but from a funding agency whom the end user was strongly motivated to pass it to. There may be privacy or data protection issues with such an approach that require an extension of the clearance given when the data is submitted – otherwise such a route cannot be used, and a direct query needs to be made for which there is no motivation. Data protection needs to be resolved urgently.

If data is made available through the portal in a harvestable format, that would be a motivation to contribute content, since it would be accessible to more people.

In the World Wide Web at large, a portal for public use (such as Yahoo) should be perceived as something trustworthy, so that users consider that it is 'on their side' and its pointers are reliable. Therefore there is a need to clearly market a portal with a slogan or brand image to convey its scope and function.

The question of multilingualism cannot be ignored. In the IST World portal, there are multilingual interfaces but no content translation. Human translation would cost CORDIS six times its budget, and this would be a recurrent cost. Machine translation such as that offered by Google gives the gist of content but is not reliable for the detail.

### The drivers of research

This research was done in response to...

### The outputs and impacts of research

### Trend detection

### Highly specific questions

How much does each country spend in a certain area? Number of patents? Number of citations etc.

## Workshop participants

### Speakers

Name	Organisation	Country
Dextre-Clarke, Stella	Information Consultant	UK
Joerg, Brigitte	DFKI	Germany
Lambert, Simon	CCLRC	UK
Miles, Alistair	CCLRC	UK
Rogers, Michael	European Commission - OPOCE - CORDIS	Luxembourg
Tichy-Rács, Adám	Budapest University of Technology and Economics	Hungary
Van Grootel, Geert	Ministry of Flanders - Science Division	Belgium
Will, Leonard	Willpower Information	UK

## 7.2 Rapporteurs

Name	Organisation	Country
Jeffery, Keith	CCLRC	UK
Lambert, Simon	CCLRC	UK
Wilson, Michael	CCLRC	UK

## 7.3 Participants

Name	Organisation	Country
Asserson, Anne	University of Bergen	Norway
Barkham, Mary	Environment Research Funders	UK
Bartholomew, Marion	Environment Research Funders	UK
Baus, Erika	Royal Belgian Institute of Natural Sciences	Belgium
Chik, Ken	i-exhibitions Ltd.	UK
Grabczewski, Eddy	CCLRC	UK
Hogenaar, Arjan	KNAW	Netherlands
Jindra, Petr	CD-Telematika	Czech Republic

Name	Organisation	Country
Kay, Zoe	European Commission	Belgium
Lawson, Gerry	Natural Environment Research Council	UK
Matthews, Brian	CCLRC	UK
Minter, Kate	Natural Environment Research Council	UK
Morris, Tara	IST Results	Belgium
Németh, Edina	NKTH	Hungary
Palmer, Claire-Louise	Environment Agency	UK
Riccalton, Carol	OPOCE, European Commission	Luxembourg
Sarzi Amadè, Fabio	Regione Piemonte	Italy
Steinmann, Rolf	uniCRIS AG	Switzerland
Vass, Gábor	NKTH	Hungary



### **For more information**

CISTRANA web-site: **<http://www.cistrana.org/>**

Council for the Central Laboratory of the  
Research Councils  
Chilton, Didcot,  
OX 110 QX  
UK

Tel.: +44 1235 44 5716  
Fax: +44 1235 44 5831  
Email: **[S.C.Lambert@rl.ac.uk](mailto:S.C.Lambert@rl.ac.uk)**